BMJ

BMJ 2011;343:d7163 doi: 10.1136/bmj.d7163 (Published 28 November 2011)

RESEARCH

Risk models and scores for type 2 diabetes: systematic review


OPEN ACCESS

Douglas Noble *lecturer*[1], Rohini Mathur *research fellow*[1], Tom Dent *consultant*[2], Catherine Meads *senior lecturer*[1], Trisha Greenhalgh *professor*[1]

[1]Centre for Primary Care and Public Health, Barts and the London School of Medicine and Dentistry, London E1 2AT, UK; [2]PHG Foundation, Cambridge, UK

Abstract

Objective To evaluate current risk models and scores for type 2 diabetes and inform selection and implementation of these in practice.

Design Systematic review using standard (quantitative) and realist (mainly qualitative) methodology.

Inclusion criteria Papers in any language describing the development or external validation, or both, of models and scores to predict the risk of an adult developing type 2 diabetes.

Data sources Medline, PreMedline, Embase, and Cochrane databases were searched. Included studies were citation tracked in Google Scholar to identify follow-on studies of usability or impact.

Data extraction Data were extracted on statistical properties of models, details of internal or external validation, and use of risk scores beyond the studies that developed them. Quantitative data were tabulated to compare model components and statistical properties. Qualitative data were analysed thematically to identify mechanisms by which use of the risk model or score might improve patient outcomes.

Results 8864 titles were scanned, 115 full text papers considered, and 43 papers included in the final sample. These described the prospective development or validation, or both, of 145 risk prediction models and scores, 94 of which were studied in detail here. They had been tested on 6.88 million participants followed for up to 28 years. Heterogeneity of primary studies precluded meta-analysis. Some but not all risk models or scores had robust statistical properties (for example, good discrimination and calibration) and had been externally validated on a different population. Genetic markers added nothing to models over clinical and sociodemographic factors. Most authors described their score as "simple" or "easily implemented," although few were specific about the intended users and under what circumstances. Ten mechanisms were identified by which measuring diabetes risk might improve outcomes. Follow-on studies that applied a risk score as part of an intervention aimed at reducing actual risk in people were sparse.

Conclusion Much work has been done to develop diabetes risk models and scores, but most are rarely used because they require tests not routinely available or they were developed without a specific user or clear use in mind. Encouragingly, recent research has begun to tackle usability and the impact of diabetes risk scores. Two promising areas for further research are interventions that prompt lay people to check their own diabetes risk and use of risk scores on population datasets to identify high risk "hotspots" for targeted public health interventions.

Introduction

The prevalence of diabetes is rising rapidly throughout the world.[1] By 2010 its prevalence in the adult populations of the United Kingdom, the United States, mainland China, and the United Arab Emirates had exceeded 7%,[2] 11%,[3] 15%,[4] and 17%,[5] respectively. Americans born in 2000 or later have a lifetime risk of more than one in three of developing diabetes.[6] Type 2 diabetes (which accounts for over 95% of diabetes worldwide) results from a complex gene-environment interaction for which several risk factors, such as age, sex, ethnicity, family history, obesity, and hypertension, are well documented. The precise interaction of these and other risk factors with one another is, however, a complex process that varies both within and across populations.[7-11] Epidemiologists and statisticians are striving to produce weighted models that can be presented as scores to reflect this complexity but which at the same time are perceived as sufficiently simple, plausible, affordable, and widely implementable in clinical practice.[12 13]

Cohort studies have shown that early detection of established diabetes improves outcome, although the evidence base for screening the entire population is weak.[14 15] The proportion of cases of incident type 2 diabetes in people with impaired glucose tolerance or impaired fasting glucose levels was reduced in landmark trials from China,[16] Finland,[17] and the United States[18] by up to 33%, 50%, and 58%, respectively, through lifestyle

Correspondence to: D Noble d.noble@qmul.ac.uk

Extra material supplied by the author (see http://www.bmj.com/content/343/bmj.d7163?tab=related#webextra)
Details of search strategy

changes (increased exercise, weight loss) or pharmacotherapy, or both, although changes may be more modest in a non-trial population. Some have argued that because combining known risk factors predicts incident diabetes at least as effectively as impaired glucose metabolism, a diabetes risk score may be a better and more practical means of identifying people for preventive interventions than either a glucose tolerance test or a fasting blood glucose level.[19] Others favour targeting the assessment of diabetes risk in those with established impaired glucose metabolism on the basis that interventions in this group are particularly effective.[20]

Risk models and scores first emerged for cardiovascular disease, and these are widely used in clinical and public health practice. In the United Kingdom, for example, all electronic patient record systems in general practice offer the facility to calculate the Framingham score, a patient's risk of a cardiovascular event within 10 years. This risk score features in many guidelines and decision pathways (such as the cut-off for statin therapy[21]), and general practitioners receive financial rewards for calculating it.[22] In contrast, although numerous models and scores have been developed for diabetes risk, we found limited evidence for use of these as part of a formal health policy, guideline, or incentive scheme for practitioners in any country (one Australian scheme incentivises general practitioners' measurement of the risk of diabetes in adults aged 40-49[23]). This is perhaps surprising, given that morbidity and mortality from cardiovascular disease has been decreasing in many countries since the 1970s,[24] whereas those from diabetes continue to increase.[3]

A diabetes risk score is an example of a prognostic model.[25] Such scores should ideally be developed by taking a large, age defined population cohort of people without diabetes, measuring baseline risk factors, and following the cohort for a sufficiently long time to see who develops diabetes.[26] Although prospective longitudinal designs in specially assembled cohorts are expensive, difficult, and time consuming to execute, cross sectional designs in which risk factors are measured in a population including people both with and without diabetes are methodologically inferior. They use prevalence as a proxy for incidence and conflate characteristics of people with diabetes with risk factors in those without diabetes, and thus are incapable of showing that a putative risk factor predated the development of diabetes. In practice, researchers tend to take one of two approaches: they either study a cohort of people without diabetes, which was assembled some years previously with relevant baseline metrics for some other purpose (for example, the British Regional Heart Study[27]), or they analyse routinely available data, such as electronic patient records.[8] Both approaches are potentially susceptible to bias.

Some diabetes risk scores are intended to be self administered using questions such as "have you ever been told you have high blood pressure?" Scores that rely entirely on such questions may be hosted on the internet (see for example www.diabetes.org.uk/riskscore). Some researchers have used self completion postal questionnaires as the first part of a stepwise detection programme.[28] To the extent that these instruments are valid, they can identify two types of people: those who already have diabetes whether or not they know it (hence the questionnaire may serve as a self administered screening tool for undiagnosed diabetes) and those at high risk of developing diabetes—that is, it may also serve as a prediction tool for future diabetes. Prevalence rates for diabetes derived from self assessment studies thus cannot be compared directly with the rate of incident diabetes in a prospective longitudinal sample from which those testing positive for diabetes at baseline have been excluded.

A good risk score is usually defined as one that accurately estimates individuals' risk—that is, predictions based on the score closely match what is observed (calibration); the score distinguishes reliably between high risk people, who are likely to go on to develop the condition, and low risk people, who are less likely to develop the condition (discrimination); and it performs well in new populations (generalisability). Validating a risk model or score means testing its calibration and discrimination either internally (by splitting the original sample, developing the score on one part and testing it on another), temporally (re-running the score on the same or a similar sample after a time period), or, preferably, externally (running the score on a new population with similar but not identical characteristics from the one on which it was developed).[26 29] Caution is needed when extrapolating a risk model or score developed in one population or setting to a different one—for example, secondary to primary care, adults to children, or one ethnic group to another.[30]

Risk scores and other prognostic models should be subject to "impact studies"—that is, studies of the extent to which the score is actually used and leads to improved outcomes. Although most authors emphasise quantitative evaluation of impact such as through cluster randomised controlled trials,[30] much might also be learnt from qualitative studies of the process of using the score, either alone or as an adjunct to experimental trials. One such methodology is realist evaluation, which considers the interplay between context, mechanism (how the intervention is perceived and taken up by practitioners), and outcome.[31] In practice, however, neither quantitative nor qualitative studies of impact are common in the assessment of risk scores.[30]

We sought to identify, classify, and evaluate risk models and scores for diabetes and inform their selection and implementation in practice. We wanted to determine the key statistical properties of published scores for predicting type 2 diabetes in adults and how they perform in practice. Hence we were particularly interested in highlighting those characteristics of a risk score that would make it fit for purpose in different situations and settings. To that end we reviewed the literature on development, validation, and use of such scores, using both quantitative data on demographics of populations and statistical properties of models and qualitative data on how risk scores were perceived and used by practitioners, policy makers, and others in a range of contexts and systems.

Methods

Theoretical and methodological approach

We followed standard methodology for systematic reviews, summarised in guidance from a previous study and the York Centre for Reviews and Dissemination.[32 33] The process was later extended by drawing on the principles of realist review, an established form of systematic literature review that uses mainly qualitative methods to produce insights into the interaction between context, mechanism, and outcome, hence explaining instances of both success and failure.[34] Our team is leading an international collaborative study, the Realist and Meta-narrative Evidence Synthesis: Evolving Standards (RAMESES) to develop methodological guidance and publication standards for realist review.[35]

Search strategy

We identified all peer reviewed cohort studies in adults over age 18 that had derived or validated, or both, a statistically weighted risk model for type 2 diabetes in a population not preselected for known risk factors or disease, and which could

be applied to another population. Studies were included that had developed a new risk model based on risk factors and that used regression techniques to weight risk factors appropriately, or validated an existing model on a new population, or did both. Exclusion criteria were cross sectional designs, studies that had not finished recruiting, studies on populations preselected for risk factors or disease, studies that did not link multiple risk factors to form a scoring system or weighted model, screening or early detection studies, genetic studies, case series, studies on under 18s, animal studies, and studies that applied a known risk model or score to a population but did not evaluate its statistical potential.

In January 2011 we undertook a scoping search, beginning with sources known to the research team and those recommended by colleagues. We used the 29 papers from this search to develop the definitive protocol, including search terms and inclusion and exclusion criteria. In February 2011 a specialist librarian designed a search strategy (see web extra) with assistance from the research team. Key words were predict, screen, risk, score, [type two] diabetes, model, regression, risk assessment, risk factor, calculator, analysis, sensitivity and specificity, ROC and odds ratio. Both MESH terms and text words were used. Titles and abstracts were searched in Medline, PreMedline, Embase, and relevant databases in the Cochrane Library from inception to February 2011, with no language restrictions.

Search results from the different databases were combined in an endnote file and duplicates removed electronically and manually. In February and March 2011 two researchers independently scanned titles and abstracts and flagged potentially relevant papers for full text analysis.

Two researchers independently read the interim dataset of full text papers and reduced this to a final dataset of studies, resolving disagreements by discussion. Bilingual academic colleagues translated non-English papers and extracted data in collaboration with one of the research team. To identify recently published papers two researchers independently citation tracked the final dataset of studies in Google Scholar. Reference lists of the final dataset and other key references were also scanned.

Quantitative data extraction and analysis

Properties of included studies were tabulated on an Excel spreadsheet. A second researcher independently double checked the extraction of primary data from every study. Discrepancies were resolved by discussion. Where studies trialled multiple models with minimal difference in the number of risk factors, a judgment was made to extract data from the authors' preferred models or (if no preferences were stated in the paper) the ones judged by two researchers to be the most complete in presentation of data or statistical robustness. Data extraction covered characteristics of the population (age, sex, ethnicity, etc), size and duration of study, completeness of follow-up, method of diagnosing diabetes, details of internal or external validation, or both, and the components and metrics used by authors of these studies to express the properties of the score, especially their calibration and discrimination—for example, observed to predicted ratios, sensitivity and specificity, area under the receiver operating characteristic curve. We aimed to use statistical meta-analysis where appropriate and presented heterogeneous data in disaggregated form.

Qualitative data extraction and analysis

For the realist component of the review we extracted data and entered these on a spreadsheet under seven headings (box 1). One researcher extracted these data from our final sample of papers and another checked a one third sample of these. Our research team discussed context-mechanism-outcome interactions hypothesised or implied by authors and reread the full sample of papers with all emerging mechanisms in mind to explore these further.

Impact analysis

We assessed the impact of each risk score in our final sample using three criteria: any description in the paper of use of the score beyond the population for whom it was developed and validated; number of citations of the paper in Google Scholar and number of these that described use of the score in an impact study; and critical appraisal of any impact studies identified on this citation track. In this phase we were guided by the question: what is the evidence that this risk score has been used in an intervention which improved (or sought to improve) outcomes for individuals at high risk of diabetes?

Prioritising papers for reporting

Given the large number of papers, statistical models, and risk scores in our final sample, we decided for clarity to highlight a small number of scores that might be useful to practising clinicians, public health specialists, or lay people. Adapting previous quality criteria for risk scores,[26] we favoured those that had external validation by a separate research team on a different population (generalisability), statistically significant calibration, a discrimination greater than 0.70, and 10 or fewer components (usability).

Results

Figure 1⇓ shows the flow of studies through the review. One hundred and fifteen papers were analysed in detail to produce a final sample of 43. Of these 43 papers, 18 described the development of one or more risk models or scores,[8 27 36-51] 17 described external validation of one or more models or scores on new populations,[9 10 19 52-65] and eight did both.[7 66-72] In all, the 43 papers described 145 risk models and scores, of which 94 were selected for extraction of full data (the other 51 were minimally different, were not the authors' preferred model, or lacked detail or statistical robustness). Of the final sample of 94 risk models, 55 were derivations of risk models on a base population and 39 were external validations (of 14 different models) on new populations. Studies were published between 1993 and 2011, but most appeared in 2008-11 (fig 2⇓). Indeed, even given that weaker cross sectional designs had been excluded, the findings suggest that new risk models and scores for diabetes are currently being published at a rate of about one every three weeks.

Table 1⇓ gives full details of the studies in the sample, including the origin of the study, setting, population, methodological approach, duration, and how diabetes was diagnosed. The studies were highly heterogeneous. Models were developed and validated in 17 countries representing six continents (30 in Europe, 25 in North America, 21 in Asia, 8 in Australasia, 8 in the Middle East, 1 in South America, and 1 in Africa).

Comparisons across studies were problematic owing to heterogeneity of data and highly variable methodology, presentation techniques, and missing data. Cohorts ranged in size from 399 to 2.54 million. The same data and participants were often included in several different models in the same paper. Ten research populations were used more than once in different papers.[9 10 27 37 41 42 44 46-49 51-56 63-66 70 71] In total, risk models

Box 1: Categories for data entry

Intended users

Authors' assumptions (if any) about who would use the risk score, on which subgroups or populations

Proposed action based on the score result

Authors' assumptions (if any) on what would be offered to people who score above the designated cut-off for high risk

Mechanism

Authors' hypothesised (or implied) mechanism by which use of the score might improve outcomes for patients

Descriptor

Authors' adjectives to describe their risk model or score

Relative advantage

Authors' claims for how and in what circumstances their model or score outperforms previous ones

Concerns

Authors' stated concerns about their model or score

Real world use, including citation tracking

Actual data in this paper or papers citing it on use of the score in the real world

were tested on 6.88 million participants, although this figure includes duplicate tests on the same dataset. Participants aged 18 to 98 were studied for periods ranging from 3.15 to 28 years. Completeness of follow-up ranged from 54% to 99% and incidence of diabetes across the time periods studied ranged from 1.3% to 20.9%.

None of the models in the sample was developed on a cohort recruited prospectively for the express purpose of devising it. Rather, all authors used the more pragmatic approach of retrospectively studying a research dataset that had been assembled some years previously for a different purpose. Forty two studies excluded known diabetes in the inception cohort. Diagnosis of diabetes in a cohort at inception and completion of the study was done in different ways, including self report, patient questionnaires, clinician diagnosis, electronic code, codes from the *International Classification of Diseases*, disease or drug registers, diabetes drugs, dietary treatment, fasting plasma glucose levels, oral glucose tolerance test, and measurement of haemoglobin A_{1c}. In some studies the method was not stated. Half the studies used different diagnostic tests at inception and completion of the study.

One third of the papers focused almost exclusively on the statistical properties of the models. Many of the remainder had a clinician (diabetologist or general practitioner) as coauthor and included an (often short and speculative) discussion on how the findings might be applied in clinical practice. Three described their score as a "clinical prediction rule."[45 51 59]

Quantitative findings

Table 2⇓ gives details of the components of the 94 risk models included in the final sample and their statistical properties—including (where reported) their discrimination, calibration, sensitivity, specificity, positive and negative predictive value, and area under the receiver operating characteristic curve. Many papers offered additional sophisticated statistical analysis, although there was no consistency in the approach used or statistical tests. Heterogeneity of data (especially demographic and ethnic diversity of validation cohorts and different score components) in the primary studies precluded formal meta-analysis.

All 94 models presented a combination of risk factors as significant in the final model, and different models weighted different components differently. The number of components in a single risk score varied from 3 to 14 (n=84, mean 7.8, SD 2.6). The seven risk scores that were classified as having high potential for use in practice offered broadly similar components and had similar discriminatory properties (area under receiver operating characteristic curve 0.74-0.85, table 4). Overall, the areas under the receiver operating characteristic curve ranged from 0.60 to 0.91. Certain components used in some models (for example, biomarkers) are rarely available in some pathology laboratories and potentially too expensive for routine use. Some models that exhibited good calibration and discrimination on the internal validation cohort performed much less well when tested on an external cohort,[62 67] suggesting that the initial model may have been over-fitted by inclusion of too many variables that had only minor contributions to the total risk.[73] Although this study did not seek out genetic components, those studies that had included genetic markers alongside sociodemographic and clinical data all found that the genetic markers added little or nothing to the overall model.[9 10 36 50]

Reporting of statistical data in some studies was incomplete—for example, only 40 of the 94 models quantified any form of calibration statistic. Forty three presented sensitivity and specificity, 27 justified the rationale for cut-off points, 22 presented a positive predictive value, 19 presented a negative predictive value, and 26 made some attempt to indicate the percentage of the population that would need clinical follow-up or testing if they scored as "high risk." Some models performed poorly—for example, there was a substantial gap between expected and observed numbers of participants who developed diabetes over the follow-up period. The false positive and false negative rates in many risk scores raised questions about their utility in clinical practice (for example, positive predictive values ranged from 5% to 42%, negative predictive values from 88% to 99%). However, some scores were designed as non-invasive preliminary instruments, with a recommended second phase involving a blood test.[7 43 52 53 55 58 65]

Risk models and scores tended to "morph" when they were externally validated because research teams dropped components from the original (for example, if data on these were not available), added additional components (for example, to compensate for missing categories), or modified what counted in a particular category (for example, changing how ethnicity was classified); in some cases these modifications were not clarified. A key dimension of implementation is appropriate

adaptation to a new context. It was considered that this did not negate the external validation.

Qualitative findings

Table 3⇓ provides the qualitative findings from the risk scores. Of the 43 papers in the full sample, three did not recommend use of the model tested because the authors believed it had no advantage over existing ones.[50 56 60] Authors of the other 40 papers considered that at least one of their scores should be adopted and used, and to justify this made various claims. The commonest adjective used by authors to describe their score was "simple" (26 of 43); others mentioned "low cost," "easily implemented," "feasible," and "convenient."

Sixteen of the 43 studies that recommended use of a particular risk model or score did not designate an intended user for it. Some authors assigned agency to a risk score—that is, they stated, perhaps inadvertently, that the score itself had the potential to prevent diabetes, change behaviour, or reduce health inequalities. Although most authors did state an intended target group, this was usually given in vague terms, such as "the general population" or "individuals who are likely to develop diabetes." Eleven of the 43 papers gave a clear statement of what intervention might be offered, by whom, to people who scored above the cut-off for high risk; the other papers made no comment on this or used vague terms such as "preventive measures," without specifying by whom these would be delivered.

In all, authors of the papers in the full sample either explicitly identified or appeared to presume 10 mechanisms (box 2) by which, singly or in combination, use of the diabetes risk score might lead to improved patient outcomes (see table 3).

Risk models and scores had been developed in a range of health systems. Differences in components could be explained partly in terms of their intended context of use. For example, the QDScore, intended for use by general practitioners, was developed using a database of electronic records of a nationally representative sample of the UK general practice population comprising 2.5 million people. The QDScore is composed entirely of data items that are routinely recorded on general practice electronic records (including self assigned ethnicity, a deprivation score derived from the patient's postcode, and clinical and laboratory values).[8] Another score, also intended to be derived from electronic records but in a US health maintenance organisation (covering people of working age who are in work), has similar components to the QDScore except that ethnicity and socioeconomic deprivation are not included. In contrast, the FINDRISC score was developed as a population screening tool intended for use directly by lay people; it consists of questions on sociodemographic factors and personal history along with waist circumference but does not include clinical or laboratory data; high scorers are prompted to seek further advice from a clinician.[52] Such a score makes sense in many parts of Finland and also in the Netherlands where health and information literacy rates are high, but would be less fit for purpose in a setting where these were low.

Prioritising scores for practising clinicians

Table 4⇓ summarises the properties of seven validated diabetes risk scores which we judged to be the most promising for use in clinical or public health practice. The judgments on which this selection was based were pragmatic; other scores not listed in table 4 (also see tables 1 and 2) will prove more fit for purpose in certain situations and settings. One score that has not yet been externally validated was included in table 4 because it is the only score already being incentivised in a national diabetes prevention policy.[23]

Studies of impact of risk scores on patient outcomes

None of the 43 papers that validated one or more risk scores described the actual use of that score in an intervention phase. Furthermore, although these papers had been cited by a total of 1883 (range 0-343, median 12) subsequent papers, only nine of those 1883 papers (table 5⇓) described application and use of the risk score as part of an impact study aimed at changing patient outcomes. These covered seven studies, of which (to date) three have reported definitive results. All three reported positive changes in individual risk factors, but surprisingly none recalculated participants' risk scores after the intervention period to see if they had changed. While one report on the ongoing FIN-D2D study suggests that incident diabetes has been reduced in "real world" (non-trial) participants who were picked up using a diabetes risk score and offered a package of preventive care,[74] this is a preliminary and indirect finding based on drug reimbursement claims, and no actual data are given in the paper. With that exception, no published impact study on a diabetes risk score has yet shown a reduction in incident diabetes.

Discussion

Numerous diabetes risk scores now exist based on readily available data and provide a good but not perfect estimate of the chance of an adult developing diabetes in the medium term future. A few research teams have undertaken exemplary development and validation of a robust model, reported its statistical properties thoroughly, and followed through with studies of impact in the real world.

Limitations of included studies

We excluded less robust designs (especially cross sectional studies). Nevertheless, included studies were not entirely free from bias and confounding. This is because the "pragmatic" use of a previously assembled database or cohort brings an inherent selection bias (for example, the British Regional Heart Study cohort was selected to meet the inclusion criteria for age and comorbidity defined by its original research team and oriented to research questions around cardiovascular disease; the population for the QDScore is drawn from general practice records and hence excludes those not registered with a general practitioner).

Most papers in our sample had one or more additional limitations. They reported models or scores that required collection of data not routinely available in the relevant health system; omitted key statistical properties such as calibration and positive and negative predictive values that would allow a clinician or public health commissioner to judge the practical value of the score; or omitted to consider who would use the score, on whom, and in what circumstances. We identified a mismatch between the common assumption of authors who develop a risk model (that their "simple" model can now be taken up and used) and the actual uptake and use of such models (which seems to happen very rarely). However, there has recently been an encouraging—if limited—shift in emphasis from the exclusive pursuit of statistical elegance (for example, maximising area under the receiver operating curve) to undertaking applied research on the practicalities and outcomes of using diabetes risk scores in real world prevention programmes.

Box 2: 10 suggested mechanisms by which diabetes risk scores could help improve patient outcomes

Clinical

Direct impact—clinicians will pick up high risk patients during consultations and offer advice that leads to change in patients' behaviour and lifestyle

Indirect impact—routine use of the score increases clinicians' awareness of risk for diabetes and motivation to manage it

Self assessment

Direct impact—people are alerted by assessing their own risk (for example, using an online tool), directly leading to change in lifestyle

Indirect impact—people, having assessed their own risk, are prompted to consult a clinician to seek further tests or advice on prevention

Technological

Individual impact—a risk model programmed into the electronic patient record generates a point of care prompt in the clinical encounter

Population impact—a risk model programmed into the electronic patient record generates aggregated data on risk groups, which will inform a public health intervention

Public health

Planners and commissioners use patterns of risk to direct resources into preventive healthcare for certain subgroups

Administrative

An administrator or healthcare assistant collects data on risk and enters these onto the patients' records, which subsequently triggers the technological, clinical, or public health mechanisms

Research into practice

Use of the risk score leads to improved understanding of risk for diabetes or its management by academics, leading indirectly to changes in clinical practice and hence to benefits for patients

Future research

Use of the risk score identifies focused subpopulations for further research (with the possibility of benefit to patients in later years)

Strengths and limitations of the review

The strengths of this review are our use of mixed methodology, orientation to patient relevant outcomes, extraction and double checking of data by five researchers, and inclusion of a citation track to identify recently published studies and studies of impact. We applied both standard systematic review methods (to undertake a systematic and comprehensive search, translate all non-English texts, and extract and analyse quantitative data) and realist methods (to consider the relation between the components of the risk score, the context in which it was intended to be used, and the mechanism by which it might improve outcomes for patients).

The main limitation of this review is that data techniques and presentation in the primary studies varied so much that it was problematic to determine reasonable numerators and denominators for many of the calculations. This required us to make pragmatic decisions to collate and present data as fairly and robustly as possible while also seeking to make sense of the vast array of available risk scores to the general medical reader. We recognise that the final judgment on which risk scores are, in reality, easy to use will lie with the end user in any particular setting. Secondly, authors of some of the primary studies included in this review were developing a local tool for local use and made few or no claims that their score should be generalised elsewhere. Yet, the pioneers of early well known risk scores[49 68] have occasionally found their score being applied to other populations (perhaps ethnically and demographically different from the original validation cohort), their selection of risk factors being altered to fit the available categories in other datasets, and their models being recalibrated to provide better goodness of fit. All this revision and recalibration to produce "new" scores makes the systematic review of such scores at best an inexact science.

Why did we not recommend a "best" risk score?

We have deliberately not selected a single, preferred diabetes risk score. There is no universal ideal risk score, as the utility of any score depends not merely on its statistical properties but also on its context of use, which will also determine which types of data are available to be included.[75 76] Even when a risk model has excellent discrimination (and especially when it does not) the trade-off between sensitivity and specificity plays out differently depending on context. Box 3 provides some questions to ask when selecting a diabetes risk score.

Risk scores as complex interventions

Our finding that diabetes risk scores seem to be used rarely can be considered in the light of the theoretical literature on diffusion of innovation. As well as being a statistical model, a risk score can be thought of as a complex, technology based innovation, the incorporation of which into business as usual (or not) is influenced by multiple contextual factors including the attributes of the risk score in the eyes of potential adopters (relative advantage, simplicity, and ease of use); adopters' concerns (including implications for personal workload and how to manage a positive score); their skills (ability to use and interpret the technology); communication and influence (for example, whether key opinion leaders endorse it); system antecedents (including a healthcare organisation's capacity to embrace new technologies, workflows, and ways of working); and external influences (including policy drivers, incentive structures, and competing priorities).[77 78]

Challenges associated with risk scores in use

While the developers of most diabetes risk scores are in little doubt about their score's positive attributes, this confidence seems not to be shared by practitioners, who may doubt the accuracy of the score or the efficacy of risk modification strategies, or both. Measuring diabetes risk competes for practitioners' attention with a host of other tasks, some of which

Box 3: Questions to ask when selecting a diabetes risk score, and examples of intended use

What is the intended use case for the score?

If intended for use:
In clinical consultations, score should be based on data on the medical record

For self assessment by lay people, score should be based on things a layperson would know or be able to measure

In prevention planning, score should be based on public health data

What is the target population?

If intended for use in high ethnic and social diversity, a score that includes these variables may be more discriminatory

What is expected of the user of the score?

If for opportunistic use in clinical encounters, the score must align with the structure and timeframe of such encounters and competencies of the clinician, and (ideally) be linked to an appropriate point of care prompt. Work expected from the intended user of the score may need to be incentivised or remunerated, or both

What is expected of the participants?

If to be completed by laypeople, the score must reflect the functional health literacy of the target population

What are the consequences of false positive and false negative classifications?

In self completion scores, low sensitivity may falsely reassure large numbers of people at risk and deter them from seeking further advice

What is the completeness and accuracy of the data from which the score will be derived?

A score based on automated analysis of electronic patient records may include multiple components but must be composed entirely of data that are routinely and reliably entered on the record in coded form, and readily searchable (thus, such scores are only likely to be useful in areas where data quality in general practice records is high)

What resource implications are there?

If the budget for implementing the score and analysing data is fixed, the cost of use must fall within this budget

Given the above, what would be the ideal statistical and other properties of the score in this context of use?

What trade-offs should be made (sensitivity *v* specificity, brevity *v* comprehensiveness, one stage *v* two stage process)?

bring financial and other rewards. At the time of writing, few opinion leaders in diabetes seem to be promoting particular scores or the estimation of diabetes risk generally—perhaps because, cognisant of the limited impacts shown to date (summarised in table 5), they are waiting for further evidence of whether and how use of the risk score improves outcomes. Indeed, the utility of measuring diabetes risk in addition to cardiovascular risk is contested within the diabetes research community.[79] In the United Kingdom, the imminent inclusion of an application for calculating QDScore on EMIS, the country's most widely used general practice computer system, may encourage its use in the clinical encounter. But unless the assessment of diabetes risk becomes part of the UK Quality and Outcomes Framework, this task may continue to be perceived as low priority by most general practitioners. Given current evidence, perhaps this judgment is correct. Furthermore, the low positive predictive values may spell trouble for commissioners. Identifying someone as "[possibly] high risk" will inevitably entail a significant cost in clinical review, blood tests, and (possibly) intervention and follow-up. Pending the results of ongoing impact studies, this may not be the best use of scarce resources.

Delivering diabetes prevention in people without any disease requires skills that traditionally trained clinicians may not possess.[80] We know almost nothing about the reach, uptake, practical challenges, acceptability, and cost of preventive interventions in high risk groups in different settings.[12] The relative benefit of detecting and targeting high risk people rather than implementing population-wide diabetes prevention strategies is unknown.[13] Effective prevention and early detection of diabetes are likely to require strengthening of health systems and development of new partnerships among the clinicians, community based lifestyle programmes, and healthcare funders.[81]

Mechanisms by which risk scores might have impact

Although most authors of papers describing diabetes risk scores have hypothesised (or seem to have assumed) a clinical mechanism of action (that the score would be used by the individual's clinician to target individual assessment and advice), the limited data available on impact studies (see table 5) suggest that a particularly promising area for further research is interventions that prompt self assessment—that is, laypeople measuring their own risk of diabetes. The preliminary findings from the impact studies covered in this review also suggest that not everyone at high risk is interested in coming forward for individual preventive input, nor will they necessarily stay the course of such input. It follows that in areas where aggregated data from electronic patient records are available, the diabetes risk scores may be used as a population prediction tool—for example, to produce small area statistics (perhaps as pictorial maps) of diabetes risk across a population, thereby allowing targeted design and implementation of community level public health interventions.[82] Small area mapping of diabetes risk may be a way of operationalising the recently published guidance on diabetes prevention from the National Institute for Health and Clinical Excellence, which recommends the use of "local and national tools . . . to identify local communities at high risk of developing diabetes to assess their specific needs."[83]

Towards an impact oriented research agenda for risk scores

We recommend that funding bodies and journal editors help take this agenda forward by viewing the risk score in use as a complex intervention and encouraging more applied research studies in which real people identified as at "high risk" using a particular risk score are offered real interventions; success in risk score development is measured in terms of patient relevant intermediate outcomes (for example, change in risk score) and

final outcomes (incident diabetes and related morbidity) rather than in terms of the statistical properties of the tool; a qualitative component (for example, process evaluation, organisational case study, patient's experience of lifestyle modification) explores both facilitators and barriers of using the score in a real world setting; and an economic component evaluates cost and cost effectiveness.

Conclusion

Millions of participants across the world have already participated in epidemiological studies aimed at developing a diabetes risk score. An extensive menu of possible scores are now available to those who seek to use them clinically or to validate them in new populations, none of which is perfect but all of which have strengths. Nevertheless, despite the growing public health importance of type 2 diabetes and the enticing possibility of prevention for those at high risk of developing it, questions remain about how best to undertake risk prediction and what to do with the results. Appropriately, the balance of research effort is now shifting from devising new risk scores to exploring how best to use those we already have.

We thank Helen Elwell, librarian at the British Medical Association Library, for help with the literature search; Samuel Rigby for manually removing duplicates; and Sietse Wieringa, Kaveh Memarzadeh, and Nicholas Swetenham for help with translation of non-English papers. *BMJ* reviewers Wendy Hu and John Furler provided helpful comments on an earlier draft.

Contributors: DN conceptualised the study, managed the project, briefed and supported all researchers, assisted with developing the search strategy and ran the search, scanned all titles and abstracts, extracted quantitative data on half the papers, citation tracked all papers, checked a one third sample of the qualitative data extraction, and cowrote the paper. TG conceptualised the qualitative component of the study, extracted qualitative data on all papers, independently citation tracked all papers, and led on writing the paper. RM independently scanned all titles and abstracts of the electronic search, extracted quantitative data from some papers, assisted with other double checking, and helped revise drafts of the paper. TD helped revise and refine the study aims, independently double checked quantitative data extraction from all papers, and helped revise drafts of the paper. CM advised on systematic review methodology, helped develop the search strategy, extracted quantitative data from some papers, and helped revise drafts of the paper. TG acts as guarantor.

Funding: This study was funded by grants from Tower Hamlets, Newham, and City and Hackney primary care trusts, by a National Institute of Health Research senior investigator award for TG, and by internal funding for staff time from Barts and the London School of Medicine and Dentistry. 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.

1 Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Res Clin Pract* 2010;87:4-14.
2 Holman N, Forouhi NG, Goyder E, Wild SH. The Association of Public Health Observatories (APHO) Diabetes Prevalence Model: estimates of total diabetes prevalence for England, 2010-2030. *Diabetic Med* 2011;28:575-82.
3 National Diabetes Clearinghouse. National Diabetes Statistics 2011. National Institute of Health, USA. http://diabetes.niddk.nih.gov/dm/pubs/statistics/#Diagnosed20.
4 Yang SH, Dou KF, Song WJ. Prevalence of diabetes among men and women in China. *N Engl J Med* 2010;362:2425-6.
5 Saadi H, Carruthers SG, Nagelkerke N, Al-Maskari F, Afandi B, Reed R, et al. Prevalence of diabetes mellitus and its complications in a population-based sample in Al Ain, United Arab Emirates. *Diabetes Res Clin Pract* 2007;78:369-77.
6 Williamson DF, Vinicor F, Bowman BA. Primary prevention of type 2 diabetes mellitus by lifestyle intervention: implications for health policy. *Ann Intern Med* 2004;140:951-7.
7 Aekplakorn W, Bunnag P, Woodward M, Sritara P, Cheepudomwit S, Yamwong S, et al. A risk score for predicting incident diabetes in the Thai population. *Diabetes Care* 2006;29:1872-7.
8 Hippisley-Cox J, Coupland C, Robson J, Sheikh A, Brindle P. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. *BMJ* 2009;338:b880.
9 Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 2008;359:2208-9.
10 Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M, et al. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *BMJ* 2010;340:b4838.
11 Park KS. The search for genetic risk factors of type 2 diabetes mellitus. *Diabetes Metab J* 2011;35:12-22.
12 Simmons RK, Unwin N, Griffin SJ. International Diabetes Federation: an update of the evidence concerning the prevention of type 2 diabetes. *Diabetes Res Clin Pract* 2010;87:143-9.
13 Yates T, Davies M, Khunti K. Preventing type 2 diabetes: can we make the evidence work? *Postgrad Med J* 2009;85:475-80.
14 Simmons RK, Echouffo-Tcheugui JB, Griffin SJ. Screening for type 2 diabetes: an update of the evidence. *Diabetes Obes Metab* 2010;12:838-44.
15 Gillies CL, Lambert PC, Abrams KR, Sutton AJ, Cooper NJ, Hsu RT, et al. Different strategies for screening and prevention of type 2 diabetes in adults: cost effectiveness analysis. *BMJ* 2008;336:1180-5.
16 Pan XR, Li GW, Hu YH, Wang JX, Yang WY, An ZX, et al. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. *Diabetes Care* 1997;20:537-44.
17 Tuomilehto J, Lindstrom J. The major diabetes prevention trials. *Curr Diab Rep* 2003;3:115-22.
18 Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 2002;346:393-403.
19 Mann DM, Bertoni AG, Shimbo D, Carnethon MR, Chen H, Jenny NS, et al. Comparative validity of 3 diabetes mellitus risk prediction scoring models in a multiethnic US cohort: the Multi-Ethnic Study of Atherosclerosis. *Am J Epidemiol* 2010;171:980-8.
20 Tuomilehto J, Lindstrom J, Hellmich M, Lehmacher W, Westermeier T, Evers T, et al. Development and validation of a risk-score model for subjects with impaired glucose tolerance for the assessment of the risk of type 2 diabetes mellitus-The STOP-NIDDM risk-score. *Diabetes Res Clin Pract* 2010;87:267-74.
21 National Institute for Health and Clinical Excellence. Lipid modification. 2011. www.nice.org.uk/nicemedia/live/11982/40675/40675.pdf.
22 McElduff P, Lyratzopoulos G, Edwards R, Heller RF, Shekelle P, Roland M. Will changes in primary care improve health outcomes? Modelling the impact of financial incentives introduced to improve quality of care in the UK. *Qual Saf Health Care* 2004;13:191-7.
23 Anon. Type 2 diabetes risk evaluation for men and women aged 40-49 years at high risk of type 2 diabetes. MBS item 713. 2011. www.agpn.com.au/__data/assets/pdf_file/0015/15009/20081126_brf_practice-detailing-card-item-713.pdf.
24 Sytkowski PA, Kannel WB, D'Agostino RB. Changes in risk factors and the decline in mortality from cardiovascular disease. The Framingham Heart Study. *N Engl J Med* 1990;322:1635-41.
25 Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009;338:b375.
26 Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.
27 Wannamethee SG, Papacosta O, Whincup PH, Thomas MC, Carson C, Lawlor DA, et al. The potential for a two-stage diabetes risk algorithm combining non-laboratory-based scores with subsequent routine non-fasting blood tests: results from prospective studies in older men and women. *Diabet Med* 2011;28:23-30.
28 Wareham NJ, Griffin SJ. Risk scores for predicting type 2 diabetes: comparing axes and spades. *Diabetologia* 2011;54:994-5.
29 Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009;338:b604.
30 Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009;338:b606.
31 Pawson R, Tilley N. *Realistic evaluation* . Sage, 1997.
32 Kahn KS, Kunz R, Kleijnen J, Antes G. *Systematic reviews to support evidence-based medicine* , 2nd ed. Royal Society of Medicine, 2011.
33 Centre for Reviews and Dissemination. Systematic Reviews: centre for reviews and dissemination guidance for undertaking reviews in health care. York Publishing Services. 2011. www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf.
34 Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review—a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005;10(suppl 1):21-34.
35 Greenhalgh T, Wong G, Westhorp G, Pawson R, on behalf of the RAMESES study group. Protocol-realist and meta-narrative evidence synthesis: evolving standards (RAMESES). *BMC Med Res Methodol* 2011;11:115.
36 Balkau B, Lange C, Fezeu L, Tichet J, de Lauzon-Guillain B, Czernichow S, et al. Predicting diabetes: clinical, biological, and genetic approaches: data from the Epidemiological Study on the Insulin Resistance Syndrome (DESIR). *Diabetes Care* 2008;31:2056-61.
37 Chen L, Magliano DJ, Balkau B, Colagiuri S, Zimmet PZ, Tonkin AM, et al. AUSDRISK: an Australian type 2 diabetes risk assessment tool based on demographic, lifestyle and simple anthropometric measures. *Med J Aust* 2010;192:197-202.
38 Chuang SY, Yeh WT, Wu YL, Chang HY, Pan WH, Tsao CK. Prediction equations and point system derived from large-scale health check-up data for estimating diabetic risk in the Chinese population of Taiwan. *Diabetes Res Clin Pract* 2011;92:128-36.

What is already known on this topic

The many known risk factors for type 2 diabetes can be combined in statistical models to produce risk scores

What this study adds

Dozens of risk models and scores for diabetes have been developed and validated in different settings

Sociodemographic and clinical data were much better predictors of diabetes risk than genetic markers

Research on this topic is beginning to shift from developing new statistical risk models to considering the use and impact of risk scores in the real world

39 Gao WG, Qiao Q, Pitkaniemi J, Wild S, Magliano D, Shaw J, et al. Risk prediction models for the development of diabetes in Mauritian Indians. *Diabet Med* 2009;26:996-1002.
40 Joseph J, Svartberg J, Njolstad I, Schirmer H. Incidence of and risk factors for type-2 diabetes in a general population: the Tromso Study. *Scand J Public Health* 2010;38:768-75.
41 Kahn HS, Cheng YJ, Thompson TJ, Imperatore G, Gregg EW. Two risk-scoring systems for predicting incident diabetes mellitus in US adults age 45 to 64 years. *Ann Intern Med* 2009;150:741-51.
42 Kolberg JA, Jorgensen T, Gerwien RW, Hamren S, McKenna MP, Moler E, et al. Development of a type 2 diabetes risk model from a panel of serum biomarkers from the Inter99 cohort. *Diabetes Care* 2009;32:1207-12.
43 Liu M, Pan C, Jin M. A Chinese diabetes risk score for screening of undiagnosed diabetes and abnormal glucose tolerance. *Diabetes Technol Ther* 2011;13:501-7.
44 Mehrabi Y, Sarbakhsh P, Hadaegh F, Khadem-Maboudi A. Prediction of diabetes using logic regression. *Iran J Endocrinol Metab* 2010;12:16-24.
45 Rathmann W, Kowall B, Heier M, Herder C, Holle R, Thorand B, et al. Prediction models for incident type 2 diabetes mellitus in the older population: KORA S4/F4 cohort study. *Diabet Med* 2010;27:1116-23.
46 Schmidt MI, Duncan BB, Bang H, Pankow JS, Ballantyne CM, Golden SH, et al. Identifying individuals at high risk for diabetes: the Atherosclerosis Risk in Communities study. *Diabetes Care* 2005;28:2013-8.
47 Schulze MB, Weikert C, Pischon T, Bergmann MM, Al-Hasani H, Schleicher E, et al. Use of multiple metabolic and genetic markers to improve the prediction of type 2 diabetes: the EPIC-Potsdam Study. *Diabetes Care* 2009;32:2116-9.
48 Stern MP, Morales PA, Valdez RA, Monterrosa A, Haffner SM, Mitchell BD, et al. Predicting diabetes. Moving beyond impaired glucose tolerance. *Diabetes* 1993;42:706-14.
49 Stern MP, Williams K, Haffner SM. Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test? *Ann Intern Med* 2002;136:575-81.
50 von Eckardstein A, Schulte H, Assmann G. Risk for diabetes mellitus in middle-aged Caucasian male participants of the PROCAM study: implications for the definition of impaired fasting glucose by the American Diabetes Association. Prospective Cardiovascular Munster. *J Clin Endocrinol Metab* 2000;85:3101-8.
51 Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med* 2007;167:1068-74.
52 Alssema M, Feskens EJ, Bakker SJ, Gansevoort RT, Boer JM, Heine RJ, et al. [Finnish questionnaire reasonably good predictor of the incidence of diabetes in The Netherlands]. *Ned Tijdschr Geneeskd* 2008;152:2418-24.
53 Alssema M, Vistisen D, Heymans MW, Nijpels G, Glumer C, Zimmet PZ, et al. The Evaluation of Screening and Early Detection Strategies for Type 2 Diabetes and Impaired Glucose Tolerance (DETECT-2) update of the Finnish diabetes risk score for prediction of incident type 2 diabetes. *Diabetologia* 2011;54:1004-12.
54 Bozorgmanesh M, Hadaegh F, Azizi F. Transportability of the updated diabetes prediction model from Atherosclerosis Risk in Communities Study to a Middle Eastern adult population: community-based cohort study. *Acta Diabetol* 2010; doi: 10.1007/s00592-010-0241-1.
55 Bozorgmanesh M, Hadaegh F, Zabetian A, Azizi F. San Antonio heart study diabetes prediction model applicable to a Middle Eastern population? Tehran glucose and lipid study. *Int J Public Health* 2010;55:315-23.
56 Cameron AJ, Magliano DJ, Zimmet PZ, Welborn TA, Colagiuri S, Tonkin AM, et al. The metabolic syndrome as a tool for predicting future diabetes: the AusDiab study. *J Intern Med* 2008;264:177-86.
57 Collins GS, Altman DG. External validation of QDSCORE((R)) for predicting the 10-year risk of developing type 2 diabetes. *Diabet Med* 2011;28:599-607.
58 Guerrero-Romero F, Rodriguez-Moran M. [Validation of an instrument for screening cases of type 2 diabetes and monitoring at-risk individuals in Mexico]. *Rev Panam Salud Publica* 2010;27:181-6.
59 Kanaya AM, Wassel Fyr CL, de Rekeneire N, Shorr RI, Schwartz AV, Goodpaster BH, et al. Predicting the development of diabetes in older adults: the derivation and validation of a prediction rule. *Diabetes Care* 2005;28:404-8.
60 Mainous AG, III, Diaz VA, Everett CJ. Assessing risk for development of diabetes in young adults. *Ann Fam Med* 2007;5:425-9.
61 McNeely MJ, Boyko EJ, Leonetti DL, Kahn SE, Fujimoto WY. Comparison of a clinical model, the oral glucose tolerance test, and fasting glucose for prediction of type 2 diabetes risk in Japanese Americans. *Diabetes Care* 2003;26:758-63.
62 Nichols GA, Brown JB. Validating the Framingham Offspring Study equations for predicting incident diabetes mellitus. *Am J Manag Care* 2008;14:574-80.
63 Rahman M, Simmons RK, Harding AH, Wareham NJ, Griffin SJ. A simple risk score identifies individuals at high risk of developing type 2 diabetes: a prospective cohort study. *Fam Pract* 2008;25:191-6.
64 Urdea M, Kolberg J, Wilber J, Gerwien R, Moler E, Rowe M, et al. Validation of a multimarker model for assessing risk of type 2 diabetes from a five-year prospective study of 6784 Danish people (Inter99). *J Diabetes Sci Technol* 2009;3:748-55.
65 Wannamethee SG, Shaper AG, Lennon L, Morris RW. Metabolic syndrome vs Framingham Risk Score for prediction of coronary heart disease, stroke, and type 2 diabetes mellitus. *Arch Intern Med* 2005;165:2644-50.
66 Bozorgmanesh M, Hadaegh F, Ghaffari S, Harati H, Azizi F. A simple risk score effectively predicted type 2 diabetes in Iranian adult population: population-based cohort study. *Eur J Public Health* 2011;21:554-9.
67 Chien K, Cai T, Hsu H, Su T, Chang W, Chen M, et al. A prediction model for type 2 diabetes risk among Chinese people. *Diabetologia* 2009;52:443-50.
68 Lindstrom J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 2003;26:725-31.
69 Rosella LC, Manuel DG, Burchill C, Stukel TA. A population-based risk algorithm for the development of diabetes: development and validation of the Diabetes Population Risk Tool (DPoRT). *J Epidemiol Community Health* 2011;65:613-20.
70 Schulze MB, Hoffmann K, Boeing H, Linseisen J, Rohrmann S, Mohlig M, et al. An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. *Diabetes Care* 2007;30:510-5.
71 Simmons RK, Harding AH, Wareham NJ, Griffin SJ. Do simple questions about diet and physical activity help to identify those at risk of type 2 diabetes? *Diabet Med* 2007;24:830-5.
72 Sun F, Tao Q, Zhan S. An accurate risk score for estimation 5-year risk of type 2 diabetes based on a health screening population in Taiwan. *Diabetes Res Clin Pract* 2009;85:228-34.
73 Harrell F. *Regression modelling strategies with applications to linear models, logistic regression and survival analysis* . Springer-Verlag, 2001.
74 Lindstrom J, Absetz P, Hemio K, Peltomaki P, Peltonen M. Reducing the risk of type 2 diabetes with nutrition and physical activity—efficacy and implementation of lifestyle interventions in Finland. *Public Health Nutr* 2010;13(6A):993-9.
75 Herman WH. Predicting risk for diabetes: choosing (or building) the right model. *Ann Intern Med* 2009;150:812-4.
76 Wright C, Dent T. Quality standards in risk prediction: summary of an expert meeting. PHG Foundation, 2011.www.phgfoundation.org.
77 Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, Peacock R. Diffusion of innovations in service organisations: systematic literature review and recommendations for future research. *Milbank Q* 2004;82:581-629.
78 Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* 1989;13:319-40.
79 Chamnan P, Simmons RK, Jackson R, Khaw KT, Wareham NJ, Griffin SJ. Non-diabetic hyperglycaemia and cardiovascular risk: moving beyond categorisation to individual interpretation of absolute risk. *Diabetologia* 2011;54:291-9.
80 Saaristo T, Peltonen M, Keinanen-Kiukaanniemi S, Vanhala M, Saltevo J, Niskanen L, et al. National type 2 diabetes prevention programme in Finland: FIN-D2D. *Int J Circumpolar Health* 2007;66:101-12.
81 Narayan KM, Williamson DF. Prevention of type 2 diabetes: risk status, clinic, and community. *J Gen Intern Med* 2010;25:154-7.
82 Stewart JE, Battersby SE, Lopez-De FA, Remington KC, Hardin JW, Mayfield-Smith K. Diabetes and the socioeconomic and built environment: geovisualization of disease prevalence and potential contextual associations using ring maps. *Int J Health Geogr* 2011;10:18.
83 National Institute for Health and Clinical Excellence. Preventing type 2 diabetes—population and community interventions. 2011. www.nice.org.uk/guidance/PH35.
84 Stone MA, Camosso-Stefinovic J, Wilkinson J, de Lusignan S, Hattersley AT, Khunti K. Incorrect and incomplete coding and classification of diabetes: a systematic review. *Diabet Med* 2010;27:491-7.
85 Rathmann W, Kowall B, Schulze MB. Development of a type 2 diabetes risk model from a panel of serum biomarkers from the Inter99 cohort: response to Kolberg et al. *Diabetes Care* 2010;33:e28.
86 Stern M, Williams K, Eddy D, Kahn R. Validation of prediction of diabetes by the Archimedes model and comparison with other predicting models. *Diabetes Care* 2008;31:1670-1.
87 Colagiuri S, Vita P, Cardona-Morrell M, Singh MF, Farrell L, Milat A, et al. The Sydney Diabetes Prevention Program: a community-based translational study. *BMC Public Health* 2010;10:328.
88 Absetz P, Oldenburg B, Hankonen N, Valve R, Heinonen H, Nissinen A, et al. Type 2 diabetes prevention in the real world: three-year results of the GOAL lifestyle implementation trial. *Diabetes Care* 2009;32:1418-20.
89 Jallinoja P, Pajari P, Absetz P. Repertoires of lifestyle change and self-responsibility among participants in an intervention to prevent type 2 diabetes. *Scand J Caring Sci* 2008;22:455-62.
90 Kulzer B, Hermanns N, Gorges D, Schwarz P, Haak T. Prevention of diabetes self-management program (PREDIAS): effects on weight, metabolic risk factors, and behavioral outcomes. *Diabetes Care* 2009;32:1143-6.
91 Laatikainen T, Dunbar JA, Chapman A, Kilkkinen A, Vartiainen E, Heistaro S, et al. Prevention of type 2 diabetes by lifestyle intervention in an Australian primary health care setting: Greater Green Triangle (GGT) Diabetes Prevention Project. *BMC Public Health* 2007;7:249.
92 Vermunt PW, Milder IE, Wielaard F, van Oers JA, Westert GP. An active strategy to identify individuals eligible for type 2 diabetes prevention by lifestyle intervention in Dutch primary care: the APHRODITE study. *Fam Pract* 2010;27:312-9.

Accepted: 5 October 2011

Cite this as: *BMJ* 2011;343:d7163

Tables

Table 1| Summary of 43 papers from which 94 diabetes risk models or scores were identified for systematic review

Study*	Country	Name of study	Name of risk score	Study design and sampling frame	Why inception cohort was assembled	Sample size	Duration: mean (SD), range (years), or as reported	Age: mean (SD) or range	How diabetes was excluded at inception	How incident diabetes was diagnosed
Aekplakorn 2006[7] (two of six models reported)	Thailand	Electric Generating Authority of Thailand Study	NS	Power plant workers: cohort derivation study; and cohort external validation study	Study of vascular risk; implicitly, study of diabetes risk	3254; 2420	12, 1985-97; 5, 1998-2003	35-54	History of diabetes, fasting plasma glucose, oral glucose tolerance test; and not stated	Diagnosis of diabetes, fasting plasma glucose, oral glucose tolerance test, diabetes drugs; and fasting plasma glucose
Alssema 2008[52] (two of three models reported)	Netherlands	Hoorn study, PREVEND study	Modified FINDRISC for Dutch population	Cohort external validation study, sample NS	Studies of glucose tolerance; cardiovascular disease and renal disease	2439; 3345	6.4 (0.5), 1989-98; 4.2 (0.4), 1997-2003	≥45; 28-75	Oral glucose tolerance test; fasting plasma glucose	NS
Alssema 2011[53] (two of three models reported)	Netherlands, Denmark, Sweden, UK, Australia, Mauritius	DETECT-2 (includes Ausdiab, Hoorn, Inter99, MONICA, Whitehall-II)	Based on FINDRISC	Cohort external validation study of FINDRISC in combined samples from five studies	NS	18 301	4.8-5, 1986-2001	Ranged from 46.3 (7.8) to 60.3 (6.9) in five studies	Oral glucose tolerance test	Oral glucose tolerance test
Balkau 2008[36] (both models reported)	France	DESIR	NS	Cohort derivation study in volunteers for free health examinations	Study of insulin resistance syndrome	1863 and 1954	9 (<1996)	47 (10)	NS	Fasting plasma glucose, diabetes drugs
Bozorgmanesh 2010[54]	Iran	Tehran Lipid and Glucose Study	Modified ARIC (Atherosclerosis Risk In Communities)	Cohort external validation study in general population	Study of lipid and glucose risk factors	5018	6, 1999-2008	Men 42.8 (14.8); women 40.7 (12.5)	Oral glucose tolerance test, fasting plasma glucose, diabetes drugs	Oral glucose tolerance test, fasting plasma glucose, diabetes drugs
Bozorgmanesh 2011[66] (all five models reported)	Iran	Tehran Lipid and Glucose Study	NS	Cohort derivation study, and cohort external validation study, in general population	Study of lipid and glucose risk factors	5018	6, 1999-2008	41.6 (13.2)	Oral glucose tolerance test, fasting plasma glucose, diabetes drugs	Oral glucose tolerance test, fasting plasma glucose, diabetes drugs
Bozorgmanesh 2010[55] (one of six models reported)	Iran	Tehran Lipid and Glucose Study	San Antonio diabetes prediction model	Cohort external validation study in general population	Study of lipid and glucose risk factors	5018	6.3, 1999-2008	Men 42.8 (14.8); women 40.7 (12.5)	Oral glucose tolerance test, fasting plasma glucose, diabetes drugs	Oral glucose tolerance test, fasting plasma glucose, diabetes drugs
Cameron 2008[56] (both models reported)	Australia	AusDiab	Diabetes prediction model; and Finnish diabetes risk score	Cohort external validation study in general population	Diabetes incidence/prevalence study	11 247	5, 2000	50.9 (50.6-51.2)	WHO criteria	WHO criteria

Table 1 (continued)

Study*	Country	Name of study	Name of risk score	Study design and sampling frame	Why inception cohort was assembled	Sample size	Duration: mean (SD), range (years), or as reported	Age: mean (SD) or range	How diabetes was excluded at inception	How incident diabetes was diagnosed
Chen 2010[37] (all six models reported)	Australia	Ausdiab	Ausdrisk	Cohort derivation study in general population	Diabetes incidence/prevalence study	11 247	5, 1999-2005	≥25	NS	Fasting plasma glucose, oral glucose tolerance test, diabetes drugs
Chien 2009[67] (seven of eight models reported)	Taiwan	Chin-Shan Community Cardiovascular Cohort	Cambridge risk score as well as several unnamed	Cohort derivation study in general population	NS	2960	10, 1990	54	Fasting plasma glucose, diabetes drugs	Fasting plasma glucose, diabetes drugs
Chuang 2011[38] (all six models reported)	Taiwan	MJ Health Screen	NS	Cohort derivation study in private health clinic patients	Data from routine health checks	19 919 (3 scores), 6111 (3 scores)	5.61 (3.33), 1994-2006	49.2 (10.4)	Fasting plasma glucose, diabetes drugs	Fasting plasma glucose, diabetes drugs
Collins 2011[57]	UK	THIN database	QDScore	Cohort external validation study in UK general practice population	Data from primary care database	2 396 392	15, 1993-2008	Median (interquartile range) men 44 (34-57), women 43 (34-56)	Read code C10 (diagnosis of diabetes)	Read code C10 (diagnosis of diabetes)
Gao 2009[39] (one of three models reported)	Mauritius	NS	NS	Cohort derivation study in random sample of entire island population	Study of non-communicable diseases	1544	11, 1987-98	<65	History of diabetes, fasting plasma glucose, oral glucose tolerance test	Diagnosis of diabetes, fasting plasma glucose, oral glucose tolerance test, diabetes drugs
Guerrero-Romero 2010[58] (one of two models reported)	Mexico	NS	ITD (Instrumento Para El Tamizaje de la diabetes tipo 2)	Cohort external validation study, sample NS	NS	525	7 (range 4.5-10), 1996-2006	20-65	NS	NS
Hippisley-Cox 2009[8] (two of four models reported)	UK	QResearch database	QDScore	Cohort derivation study in general practice electronic record database	Data from primary care database	2 samples 2 540 753 and 1 232 832	15, 1993-2008	25-79 (median 41)	Read code C10 (diagnosis of diabetes) less those receiving insulin <age 35	Read code C10 (diagnosis of diabetes) less those receiving insulin <age 35
Joseph 2010[40]	Norway	Tromsø Study	NS	Cohort derivation study in single academic health centre (Tromsø)	NS	26 168	10.8 (median), 1994-2005	25-98	Self report, haemoglobin A_{1c}, ICD-10, plasma glucose, diabetes drugs	"T2DM event"
Kahn 2009[41] (all three models reported)	USA	ARIC (Atherosclerosis Risk in Communities)	NS	Cohort derivation study in four US communities	Study of atherosclerosis risk	9587; 3142; 3142	14.9, 1987-2003	45-64	NS	Varied over study period. Fasting plasma glucose, oral glucose tolerance test, self report,

Table 1 (continued)

Study*	Country	Name of study	Name of risk score	Study design and sampling frame	Why inception cohort was assembled	Sample size	Duration: mean (SD), range (years), or as reported	Age: mean (SD) or range	How diabetes was excluded at inception	How incident diabetes was diagnosed
										record, survey
Kanaya 2005[59]	USA	Health, Aging, and Body Composition Study (Validation)	NS	Cohort external validation study in two clinics (Memphis and Pittsburgh)	NS	2503	6, 1997-2003	70-79	Self report, diabetes drugs, fasting plasma glucose	Fasting plasma glucose
Kolberg 2009[42]	USA	Inter99	NS	Cohort derivation study, sample from Danish civil register	Lifestyle intervention trial for cardiovascular disease	632	5, NS	30-60	Fasting plasma glucose, oral glucose tolerance test	Fasting plasma glucose, oral glucose tolerance test
Lindstrom 2003[68] (both models reported)	Finland	FINRISK Studies	Diabetes risk score	Cohort derivation study, national population register; and cohort external validation study, FINRISK	NS	4746; 4615	10, 1987-97; 5, 1992-7	45-64	Fasting plasma glucose, oral glucose tolerance test, diabetes drugs	Fasting plasma glucose, oral glucose tolerance test, diabetes drugs
Liu 2011[43](all three models reported)	China	NS	Chinese diabetes risk score	Cohort derivation study in hospital screening centre for military officers	Analysis of routine data from health checks	1457	10, 1996-2006	48-87	Fasting plasma glucose, oral glucose tolerance test	Self report, fasting plasma glucose, oral glucose tolerance test, diabetes drugs
Mainous 2007[60]	USA	Coronary Artery Risk Development in Young Adults (CARDIA)	NS	Cohort external validation study in young adults recruited to CARDIA study	Study of coronary heart disease risk	2543	10, 1985-95	18-30	Self report, fasting plasma glucose	Self report, fasting plasma glucose
Mann 2010[19] (all three models reported)	USA	Multi-ethnic Study of Atherosclerosis (MESA)	NS	Cohort external validation study in adults without cardiovascular disease in six diverse US communities	Study of atherosclerosis risk	5329	4.75, 2000-6	61.6 (45-84)	Fasting plasma glucose, diabetes drugs	Fasting plasma glucose, diabetes drugs
McNeely 2003[61] (one of two models reported)	USA	Japanese American Community Diabetes Study	NS	Cohort external validation study, sample NS	Community diabetes study	518	5-10, NS	52.1 (34-75)	Fasting plasma glucose, oral glucose tolerance test, diabetes drugs	Oral glucose tolerance test
Mehrabi 2010[44] (one of four models reported)	Iran	Tehran Lipid and Glucose Study	NS	Cohort derivation study, sample NS	Study of lipid and glucose risk factors	5114	9, 1998-2007	Men 43.4 (14.1), women 40.4 (12.6)	Fasting plasma glucose, oral glucose tolerance test, diabetes drugs	NS

Table 1 (continued)

Study*	Country	Name of study	Name of risk score	Study design and sampling frame	Why inception cohort was assembled	Sample size	Duration: mean (SD), range (years), or as reported	Age: mean (SD) or range	How diabetes was excluded at inception	How incident diabetes was diagnosed
Meigs 2008[9]	USA	Framingham Offspring Study	Genotype score	Cohort external validation study, sample NS	Study of children of Framingham Heart Study participants	2377	28, 1971-2001	28-62	Fasting plasma glucose, diabetes drugs	Fasting plasma glucose, diabetes drugs
Nichols 2008[62] (all three models reported)	USA	Kaiser Permanente Northwest electronic records	Framingham Offspring Study score	Cohort external validation study in health maintenance organisation registered population	Analysis of health maintenance organisation electronic records	20, 644	7, 1999-2007	57.4	NS	Diagnosis of diabetes (ICD-9 codes), fasting plasma glucose, diabetes drugs
Rahman 2008[63]	UK	European Prospective Investigation of Cancer (EPIC)-Norfolk	Cambridge risk score	Cohort external validation study in UK general practice	Study of causes of cancer	24, 495	4.8 (1.3), 1993-2000	58.9 (40-79)	Self report, diabetes drugs, clinic registers, death certificates	As inception
Rathmann 2010[85] (all three models reported)	Germany	KORA S4/F4 study	NS	Cohort derivation study, sample NS	NS	1202	Implicitly, 7, 1999-2008	55-74	Oral glucose tolerance test	Diagnosis of diabetes, oral glucose tolerance test
Rosella 2010[69] (all three models reported)	Canada	National Population Health Survey—Ontario	Dport (Diabetes population at risk tool)	Cohort derivation study, sample NS	Health survey	19 795; 9899; 26 465	9, 1996-7; 9, 1996-2005; 5, 2000-5	Men 44, women 46; men 44, women 47; men 44, women 46	NS	Hospital diagnosis of diabetes (ICD code), physician claims
Schmidt 2005[46] (all three models reported)	USA	ARIC (Atherosclerosis Risk in Communities)	NS	Cohort derivation study in four US communities	Study of atherosclerosis risk	7915	9, 1987-98	Median 54 (45-64)	Diagnosis of diabetes (including self report), fasting plasma glucose, diabetes drugs	Diagnosis of diabetes, fasting plasma glucose, oral glucose tolerance test, diabetes drugs
Schulze 2007[70] (both models reported)	Germany	EPIC-Potsdam; and EPIC-Heidelberg	German diabetes risk score	Cohort derivation study (Potsdam); cohort external validation study (Heidelberg)	Study of causes of cancer	27 548; 25 540	7, NS; 5, NS	Men 40-65, women 35-65; NS	NS	Self report, verified by ICD-10; self report, record, death certificate
Schulze 2009[47]	Germany	EPIC-Potsdam	Adaptation of German diabetes risk score	Cohort derivation study in general population (Potsdam)	Study of causes of cancer	1962	7.1, 1994	35-65	Self report verified by physician	Self report verified by physician
Simmons 2007[71] (both models reported)	UK	EPIC-Norfolk	NS; Cambridge risk score	Cohort derivation study; cohort external validation study, sample NS	Study of causes of cancer	12 591	4.6, 1993-2000	40-79	Self report	Health check, clinic registers, diabetes drugs, haemoglobin A_{1c}

Table 1 (continued)

Study*	Country	Name of study	Name of risk score	Study design and sampling frame	Why inception cohort was assembled	Sample size	Duration: mean (SD), range (years), or as reported	Age: mean (SD) or range	How diabetes was excluded at inception	How incident diabetes was diagnosed
Stern 1993[48] (two of six models reported)	USA	San Antonio Heart Study	NS	Cohort derivation study, sample NS	Population based study of diabetes and cardiovascular disease	2217	8, 1979-87	25-64	Fasting plasma glucose, oral glucose tolerance test, diabetes drugs	Fasting plasma glucose, oral glucose tolerance test, diabetes drugs
Stern 2002[86] (both models reported)	USA	San Antonio Heart Study	NS	Cohort derivation study, sample NS	Population based study of diabetes and cardiovascular disease	5158	7-8, 1979-88	25-64	Fasting plasma glucose, oral glucose tolerance test, diabetes drugs	Fasting plasma glucose, oral glucose tolerance test, diabetes drugs
Sun 2009[72] (three of six models reported)	Taiwan	Taiwan health-check-up database (MJLPD)	Atherosclerosis Risk in Communities (ARIC) score	Cohort derivation study in private patient sample	NS	10 294	Median 3.15, 1997-2006	47.5 (35-74)	Fasting plasma glucose, diabetes drugs	NS
Talmud 2010[10] (two of three models reported)	UK	Whitehall II	Cambridge Risk Score; and Framingham Offspring Study score	Cohort external validation study in civil servant sample	Study of health in civil servants	8713	11.7 (median), NS	49 (35-55)	Oral glucose tolerance test	Oral glucose tolerance test, diabetes drugs, self report of doctor diagnosis
Urdea 2009[64] (one score, two studies, both reported)	Denmark	Inter99	PreDx diabetes risk score training set; PreDx diabetes risk score validation set	Cohort external validation study, sample not stated	Primary prevention study of cardiovascular disease	399; 400	5, NS	40-55	NS	NS
Von Eckardstein 2000[50]	Germany	PROCAM (Prospective Cardiovascular Münster Study)	Multiple logistic function model	Cohort derivation study in employees of 52 companies and authorities in Münster	To examine cardiovascular risk factors, events, and mortality	3737	4-10, 1979-95	30-60	Self report, fasting plasma glucose, diabetes drugs	Self report, diabetes drugs, fasting plasma glucose
Wannamethee 2011[27] (all three models reported)	UK	British Regional Heart Study and British Women's Heart and Health Study	NS	Cohort derivation study, sample not stated	Study of cardiovascular risk	6927	7, 1998-2007	60-79	Doctor diagnosis of diabetes, fasting plasma glucose	Record review, self report
Wannamethee 2005[65]	UK	British Regional Heart Study	Framingham risk score	Cohort external validation study in sample of mostly manual social class	Heart study	5128	21.3, 1978-2000	50.3 (5.7), 40-59	Recall of doctor diagnosis, high blood glucose	NS
Wilson 2007[51] (one of seven models reported)	USA	Framingham Offspring Study	NS	Cohort derivation study, sample not stated	Population based study of health outcomes	3140	7, mid-1990-2001	54	History of diabetes, oral glucose tolerance test, fasting plasma glucose, diabetes drugs	Fasting plasma glucose, diabetes drugs

NS=not stated; WHO=World Health Organization; ICD-10=*International Classification of Disease*, 10th revision; ICD-9=*International Classification of Diseases*, ninth revision.

Table 1 (continued)

Study*	Country	Name of study	Name of risk score	Study design and sampling frame	Why inception cohort was assembled	Sample size	Duration: mean (SD), range (years), or as reported	Age: mean (SD) or range	How diabetes was excluded at inception	How incident diabetes was diagnosed

Some studies tested multiple models, with minimal difference in number of risk factors; in such cases authors' preferred models were selected or, if no preference stated, we made our own judgment.

*Bracketed information shows how many scores tested by the original authors were included in this systematic review.

Table 2| **Key characteristics of 94 diabetes risk models or scores included in systematic review**

Study	Diabetes incidence (%)*	Components of score	Sensitivity/specificity† %	AUROC (95% CI)	Positive/negative predictive value (%)	Calibration	% needing further tests
Aekplakorn 2006[7]	11.1	Age, BMI, waist circumference, hypertension, family history of diabetes in first degree relative	77/60	0.74 (0.71 to 0.78)	NS/NS	Hosmer-Lemeshow P=0.8	NS
Aekplakorn 2006[7]	5.2	Age, BMI, waist circumference, hypertension, family history of diabetes in first degree relative	84.4/52.5	0.75 (0.71 to 0.80)	NS/NS	NS	NS
Alssema 2008[52]	22.3 per 1000 person years	Age, BMI, waist circumference, use of antihypertensive drugs, parental history of diabetes, family history of diabetes in first degree relative	84/42 (cut-off ≥7); 52/76 (cut-off ≥10)	0.71 (0.68 to 0.75)	19/94 (cut-off ≥7); 26/91 (cut-off ≥10)	NS	28
Alssema 2008[52]	10.7 per 1000 person years	Age, BMI, waist circumference, use of antihypertensive drugs, parental history of diabetes, family history of diabetes in first degree relative	78/64 (cut-off ≥7); 43/85 (cut-off ≥10)	0.77 (0.73 to 0.80)	9/98 (cut-off ≥7); 12/97 (cut-off ≥10)	NS	16
Alssema 2011[53]	Range 2.3-9.9 across five substudies	Age, BMI, waist circumference, use of antihypertensive drugs, history of gestational diabetes	NS/NS	0.77 (0.75 to 0.78)	NS/NS	NS	NS
Alssema 2011[53]	Range 2.3-9.9 across five substudies	Age, BMI, waist circumference, use of antihypertensive drugs, history of gestational diabetes, sex, smoking, family history of diabetes	76/63	0.76 (0.75 to 0.78)	11/NS	Hosmer-Lemeshow P=0.27	40
Balkau 2008[36]	7.5	Waist circumference, smoking, hypertension	NS/NS	0.71 (NS)	NS/NS	Hosmer-Lemeshow P=0.8	NS
Balkau 2008[36]	3.2	Waist circumference, family history of diabetes, hypertension	NS/NS	0.83	NS/NS	Hosmer-Lemeshow P=0.9	NS
Bozorgmanesh 2011[54]	4.6	Age, family history of diabetes, hypertension, waist circumference, fasting plasma glucose level, height, pulse, triglyceride-high density lipoprotein ratio	Men 71.6/75.3, women 67.1/85.0	Men 0.79, women 0.829	NS/NS	Hosmer-Lemeshow P=0.129	NS
Bozorgmanesh 2011[66]	4.6	Age, family history of diabetes, systolic blood pressure, waist-hip ratio, waist-height ratio	NS/NS	0.75 (0.72 to 0.78)	NS/NS	NS	NS
Bozorgmanesh 2011[66]	4.6	Family history of diabetes, systolic blood pressure, waist-height ratio, triglyceride-high density lipoprotein ratio, fasting plasma glucose level	NS/NS	0.85 (0.82 to 0.87)	NS/NS	NS	NS
Bozorgmanesh 2011[66]	4.6	Family history of diabetes, systolic blood pressure, waist-height ratio, triglyceride-high density lipoprotein ratio, fasting plasma glucose level, two hour postprandial plasma glucose level	NS/NS	0.86 (0.83 to 0.89)	NS/NS	NS	NS
Bozorgmanesh 2011[66]	4.6	Systolic blood pressure, waist-height ratio, fasting plasma glucose level, triglyceride-high density lipoprotein ratio, family history of diabetes	75/77	0.83 (0.80 to 0.86)	NS/NS	Hosmer-Lemeshow P=0.631	NS
Bozorgmanesh 2011[66]	4.6	NS	NS/NS	0.78 (0.75 to 0.81)	NS/NS	Hosmer-Lemeshow P=0.264	NS
Bozorgmanesh 2010[55]	4.6	"San Antonio diabetes prediction model"	NS/NS	0.83 (0.80 to 0.86)	NS/NS	Hosmer-Lemeshow P<0.001, when recalibrated P=0.131	NS
Cameron 2008[56]	2.0	Age, sex, ethnicity, fasting plasma glucose level, systolic blood pressure, high density lipoprotein cholesterol level, BMI, family history of diabetes	62.4/82.3	NS	11.9/98.3	NS	19.3
Cameron 2008[56]	2.0	NS	62.3/70.5	NS	6.8/98.2	NS	30.6

Table 2 (continued)

Study	Diabetes incidence (%)*	Components of score	Sensitivity/specificity† %	AUROC (95% CI)	Positive/negative predictive value (%)	Calibration	% needing further tests
Chen 2010[37]	3.2	Age, sex, ethnicity, parental history of diabetes, history of high blood glucose levels, use of antihypertensive drugs, lipid lowering drugs, smoking, physical inactivity, waist circumference, BMI, education, occupation	NS/NS	0.79 (0.76 to 0.81)	NS/NS	Hosmer-Lemeshow P=0.06	NS
Chen 2010[37]	3.2	Age, sex, ethnicity, parental history of diabetes, history of high blood glucose levels, use of antihypertensive drugs, lipid lowering drugs, smoking, physical inactivity, waist circumference, BMI, education	NS/NS	0.79 (0.76 to 0.81)	NS/NS	Hosmer-Lemeshow P=0.02	NS
Chen 2010[37]	3.2	Age, sex, ethnicity, parental history of diabetes, history of high blood glucose levels, use of antihypertensive drugs, lipid lowering drugs, smoking, physical inactivity, waist circumference, BMI	NS/NS	0.79 (0.76 to 0.81)	NS/NS	Hosmer-Lemeshow P=0.06	NS
Chen 2010[37]	3.2	Age, sex, ethnicity, parental history of diabetes, history of high blood glucose levels, antihypertensive drugs, smoking, physical inactivity, waist circumference, BMI	NS/NS	0.79 (0.76 to 0.81)	NS/NS	Hosmer-Lemeshow P=0.02	NS
Chen 2010[37]	3.2	Age, sex, ethnicity, parental history of diabetes, history of high blood glucose levels, use of antihypertensive drugs, smoking, physical inactivity, waist circumference	NS/NS	0.78 (0.76 to 0.81)	NS/NS	Hosmer-Lemeshow P=0.85	NS
Chen 2010[37]	3.2	Age, sex, ethnicity, parental history of diabetes, history of high blood glucose levels, use of antihypertensive drugs, smoking, physical inactivity, BMI	NS/NS	0.78 (0.75 to 0.80)	NS/NS	Hosmer-Lemeshow P=0.66	NS
Chien 2009[67]	18.5	Age, BMI, white blood cell count, triglyceride level, high density lipoprotein cholesterol level, fasting plasma glucose level	52/78	0.70 (0.68 to 0.73)	NS/NS	Hosmer-Lemeshow P=0.874	NS
Chien 2009[67]	18.5	Age, BMI, white blood cell count, triglyceride level, high density lipoprotein cholesterol level, fasting plasma glucose level, family history of diabetes, systolic blood pressure	69/62	0.70 (0.68 to 0.73)	NS/NS	NS	NS
Chien 2009[67]	18.5	Age, sex, BMI, family history of diabetes, use of antihypertensive drugs	NS/NS	0.65 (0.62 to 0.67)	NS/NS	NS	NS
Chien 2009[67]	18.5	NS	66/56	NS	NS/NS	Hosmer-Lemeshow P=0.008	NS
Chien 2009[67]	18.5	NS	72/40	NS	NS/NS	Hosmer-Lemeshow P=0.001	NS
Chien 2009[67]	18.5	NS	55/72	NS	NS/NS	Hosmer-Lemeshow P=0.002	NS
Chien 2009[67]	18.5	NS	48/78	NS	NS/NS	Hosmer-Lemeshow P=0.032	NS
Chuang 2011[38]	6.4	Age, sex, education, alcohol, BMI, waist circumference	NS/NS	0.71 (0.70 to 0.73)	NS/NS	NS	NS
Chuang 2011[38]	6.4	Age, sex, education, alcohol, BMI, waist circumference, blood pressure, hypertension	NS/NS	0.720 (0.71 to 0.74)	NS/NS	NS	NS

Table 2 (continued)

Study	Diabetes incidence (%)*	Components of score	Sensitivity/specificity† %	AUROC (95% CI)	Positive/negative predictive value (%)	Calibration	% needing further tests
Chuang 2011[38]	6.4	Age, sex, education, alcohol, BMI, waist circumference, triglyceride level, blood pressure, hypertension, fasting plasma glucose level	NS/NS	0.82 (0.81 to 0.83)	NS/NS	NS	NS
Chuang 2011[38]	6.4	Age, sex, education, alcohol, BMI, waist circumference, family history of diabetes	NS/NS	0.75 (0.73 - 0.78)	NS/NS	NS	NS
Chuang 2011[38]	6.4	Age, sex, education, family history of diabetes, alcohol, BMI, waist circumference, blood pressure, hypertension	NS/NS	0.76 (0.73 to 0.79)	NS/NS	NS	NS
Chuang 2011[38]	6.4	Age, sex, education, alcohol consumption, BMI, waist circumference, blood pressure, hypertension, fasting plasma glucose level, triglyceride level, family history of diabetes	NS/NS	0.84 (0.81 to 0.86)	NS/NS	NS	NS
Collins 2011[57]	3.0	Age, sex, ethnicity, BMI, smoking, family history of diabetes, cardiovascular disease, Townsend score, treated high blood pressure, current use of corticosteroids	NS/NS	Women 0.81, men 0.80	NS/NS	Brier score: men 0.053 (0.051-0.054), women 0.041 (0.040-0.043)	NS
Gao 2009[39]	16.5	BMI, waist circumference, family history of diabetes	Men 72 (71-74)/0.47 (0.45-0.49), women 77 (75-78)/0.50 (0.48-0.52)	Men 0.62 (0.56 to 0.68), women 0.64 (0.59 to 0.69)	NS/NS	NS	NS
Guerrero-Romero 2010[58]	11.8	Age, sex, family history of diabetes, family history of hypertension, family history of obesity, history of gestational diabetes or macrosomia, fasting plasma glucose level, physical inactivity, triglyceride level, systolic or diastolic blood pressure, BMI	92/71	0.91	35/97.5	NS	NS
Hippisley-Cox 2009[8]	3.1	Age, sex, ethnicity, BMI, smoking, family history of diabetes, Townsend score, treated hypertension, cardiovascular disease, current use of corticosteroids	NS/NS	NS	NS/NS	NS	NS
Hippisley-Cox 2009[8]	3.0	Age, sex, ethnicity, BMI, smoking, family history of diabetes, Townsend score, treated hypertension, cardiovascular disease, current use of corticosteroids	NS/NS	Women 0.85 (0.85 to 0.86), men 0.83 (0.83 to 0.84)	NS/NS	Brier score: men 0.078 (0.075-0.080), women 0.058 (0.055-0.060)	NS
Joseph 2010[40]	Men 2.5, women 1.5	Age, BMI, total cholesterol, triglyceride level, high density lipoprotein cholesterol level, hypertension, family history of diabetes, education, physical inactivity, smoking	NS/NS	Men 0.87, women 0.88	NS/NS	NS	NS
Kahn 2009[41]	Men 19.4, women 18.6	See next two rows for description of both models	NS/NS	NS	NS/NS	NS	NS
Kahn 2009[41]	17.7 at 10 years	Waist circumference, parental history of diabetes, hypertension, short stature, black race, age >55, weight, pulse, smoking	69/64	0.71 (0.69 to 0.73)	NS/NS	NS	NS

Table 2 (continued)

Study	Diabetes incidence (%)*	Components of score	Sensitivity/specificity† %	AUROC (95% CI)	Positive/negative predictive value (%)	Calibration	% needing further tests
Kahn 2009[41]	17.7 at 10 years	Glucose, waist circumference, parental history of diabetes, hypertension, triglyceride level, black race, high density lipoprotein cholesterol level, short stature, high uric acid level, age >55, pulse, alcohol consumption	74/71	0.79 (0.77 to 0.81)	NS/NS	NS	NS
Kanaya 2005[59]	5.7	Age, sex, triglyceride level, fasting plasma glucose level	NS/NS	0.71 (NS)	NS/NS	NS	NS
Kolberg 2009[42]	2.7	Six biomarkers: adiponectin, C reactive protein, ferritin, glucose, interleukin 2 receptor A, insulin	NS/NS	0.78 (NS)	NS/NS	NS	10% classified as high risk
Lindstrom 2003[68]	4.1	Age, BMI, waist circumference, use of antihypertensive drugs, history of hypertension, physical inactivity, diet (vegetables, fruits or berries)	78 (71-84)/77 (76-79)	0.85 (NS)	0.13 (0.11-0.15)/0.99 (0.98-0.99)	NS	25% in two highest risk categories
Lindstrom 2003[68]	1.5	Age, BMI, waist circumference, use of antihypertensive drugs, history of hypertension, physical inactivity, diet (vegetables, fruit or berries)	81 (69-89)/76 (74-77)	0.87 (NS)	0.05 (0.04-0.06)/0.996 (0.993-0.998)	NS	26% of men and 24% of women in two highest risk categories
Liu 2011[43]	20.9	Age, hypertension, history of high blood glucose level, BMI	NS/NS	0.68 (0.65 to 0.72)	NS/NS	NS	NS
Liu 2011[43]	20.9	Age, hypertension, history of high blood glucose level, BMI, fasting plasma glucose level	NS/NS	0.71 (0.68 to 0.74)	NS/NS	NS	NS
Liu 2011[43]	20.9	Age, hypertension, history of high blood glucose level, BMI, fasting plasma glucose level, triglyceride level, high density lipoprotein cholesterol level	64.5/71.6	0.72 (0.69 to 0.76)	37.70/88.60	NS	NS
Mainous 2007[60]	3.9	Waist circumference, hypertension or use of antihypertensive drugs, low density lipoprotein cholesterol level, triglyceride level, BMI, hyperglycaemia	15/98	0.70	NS/NS	NS	NS
Mann 2010[19]	8.4	Overweight or obese, impaired fasting glucose, high density lipoprotein cholesterol level, triglyceride level, hypertension, parental history of diabetes	NS/NS	0.78 (0.74 to 0.82)	NS/NS	Hosmer-Lemeshow P<0.001 before calibration, P>0.10 after recalibration	27.7 in highest risk fifth
Mann 2010[19]	8.4	Height, waist circumference, black ethnicity, systolic blood pressure, fasting plasma glucose level, high density lipoprotein cholesterol level, triglyceride level, parental history of diabetes, age	NS/NS	0.84 (0.82 to 0.86)	NS/NS	Hosmer-Lemeshow P<0.001 before calibration, P>0.10 after recalibration	27.6 in highest risk fifth
Mann 2010[19]	8.4	Age, sex, Mexican-American ethnicity, fasting plasma glucose level, systolic blood pressure, high density lipoprotein cholesterol level, BMI, family history of diabetes	NS/NS	0.83 (0.81 to 0.85)	NS/NS	Hosmer-Lemeshow P<0.001 before calibration, P>0.10 after recalibration	27.6 in highest risk fifth
McNeely 2003[61]	9.7 at 5 years 14.3 at 10 years	Age, sex, ethnicity, BMI, systolic blood pressure, fasting plasma glucose level, high density lipoprotein cholesterol level, family history of diabetes in first degree relative	60 and 73.3 at 5-6 years/64.9 and 78.4 at 10 years	0.76 (0.70 to 0.81) at 5-6 years, 0.79 (0.74 to 0.85) at 10 years	NS/NS	NS	NS

Table 2 (continued)

Study	Diabetes incidence (%)*	Components of score	Sensitivity/specificity† %	AUROC (95% CI)	Positive/negative predictive value (%)	Calibration	% needing further tests
Mehrabi 2010[44]	4.2	Impaired fasting glucose, family history of diabetes, impaired glucose tolerance, waist circumference, triglyceride level	NS/NS	0.843 (0.813 to 0.874)	NS/NS	NS	NS
Meigs[9]	9.2	Age, sex, family history of diabetes, BMI, triglyceride level, fasting plasma glucose level, systolic blood pressure, high density lipoprotein cholesterol level (Framingham simple clinical model)	NS/NS	0.90 (0.88 to 0.92)	NS/NS	NS	NS
Nichols 2008[62]	16.5	Age, sex, parental history of diabetes, BMI	NS/NS	0.68 (NS)	NS/NS	NS	NS
Nichols 2008[62]	16.5	Age, sex, parental history of diabetes, BMI, hypertension or antihypertensive drugs, high density lipoprotein cholesterol level, triglyceride level, fasting plasma glucose level	NS/NS	0.82 (NS)	NS/NS	Hosmer-Lemeshow P<0.001	NS
Nichols 2008[62]	16.5	Age, sex, parental history of diabetes, BMI, systolic blood pressure, high density lipoprotein cholesterol level, triglyceride level, fasting plasma glucose level, waist circumference	NS/NS	0.84 (NS)	NS/NS	NS	NS
Rahman 2008[63]	1.3	Age, sex, current use of corticosteroids, use of antihypertensive drugs, family history of diabetes, BMI, smoking	54.5/80	0.74 (NS)	NS/NS	NS	20
Rathmann 2010[85]	7.6	Age, sex, BMI, parental history of diabetes, smoking, hypertension	69.2/74	0.76 (0.71 to 0.81)	23.7/95.4	Hosmer-Lemeshow P=0.66, Brier score 0.0848	NS
Rathmann 2010[85]	7.6	Age, sex, BMI, parental history of diabetes, smoking, hypertension, fasting plasma glucose level, haemoglobin A_{1c} concentration, uric acid level	82.4/72.9	0.84 (0.80 to 0.89)	26.1/97.3	Hosmer-Lemeshow P=0.45, Brier score 0.0716	NS
Rathmann 2010[85]	7.6	Age, sex, BMI, parental history of diabetes, smoking, hypertension, fasting plasma glucose level, haemoglobin A_{1c} concentration, uric acid level, oral glucose tolerance test	81.3/84.1	0.89 (0.85 to 0.92)	37.4/97.5	Hosmer-Lemeshow P=0.70, Brier score 0.0652	NS
Rosella 2010[69]	7.1	Age, ethnicity, BMI, hypertension, immigrant status, smoking, education, cardiovascular disease	NS/NS	Men 0.77 (0.76 to 0.79), women 0.78 (0.76 to 0.79)	NS/NS	Hosmer-Lemeshow	NS
Rosella 2010[69]	5.3	Age, ethnicity, BMI, hypertension, immigrant status, smoking, education, cardiovascular disease	NS/NS	Men 0.77 (0.76 to 0.79), women 0.76 (0.74 to 0.77)	NS/NS	Hosmer-Lemeshow	NS
Rosella 2010[69]	4.2	Age, ethnicity, BMI, hypertension, immigrant status, smoking, education, cardiovascular disease	NS/NS	Men 0.79 (0.77 to 0.82), women 0.80 (0.77 to 0.82)	NS/NS	Hosmer-Lemeshow	NS
Schmidt 2005[46]	16.3	Age, waist circumference, height, systolic blood pressure, family history of diabetes, ethnicity	Range 40-77/55-84 (at different cut-offs)	0.71	Range 25-32/range 88-93 (at different cut-offs)	NS	50
Schmidt 2005[46]	16.3	Age, waist circumference, height, systolic blood pressure, family history of diabetes, ethnicity, fasting plasma glucose level	Range 51-83/56-86 (at different cut-offs)	0.78	Range 27-41/90-94 (at different cut-offs)	NS	50

Table 2 (continued)

Study	Diabetes incidence (%)*	Components of score	Sensitivity/specificity† %	AUROC (95% CI)	Positive/negative predictive value (%)	Calibration	% needing further tests
Schmidt 2005[46]	16.3	Age, ethnicity, waist circumference, height, systolic blood pressure, family history of diabetes, fasting plasma glucose level, triglyceride level, high density lipoprotein cholesterol level	Range 52-85/57-86 (at different cut-offs)	0.80	Range 27-42/range 90-95 (at different cut-offs)	NS	50
Schulze 2007 [70]	3.1	Age, waist circumference, height, history of hypertension, physical inactivity, smoking, consumption of red meat, whole grain bread, coffee, and alcohol	83.1, 67.5, 50.3/68.3, 80.6, 89.9 (at different cut-offs)	0.84	5.9, 7.7, 10.7 at different cut-offs/NS	Observed to predicted incidence	23.20
Schulze 2007[70]	2.6	Age, waist circumference, height, history of hypertension, physical inactivity, smoking, consumption of red meat, whole grain bread, coffee, and alcohol	94.4 ≥500 points, 79.7 ≥550 points/66.7 ≥500 points, 79.3 ≥550 points	0.82	NS/NS	Observed to predicted incidence	NS
Schulze 2009[47]	3	Diabetes risk score plus haemoglobin A_{1c} concentration, glucose level, triglyceride level, high density lipoprotein cholesterol level, γ-glutamyltransferase level, alanine aminotransferase level	NS/NS	0.90 (0.89 to 0.91)	NS/NS	Hosmer-Lemeshow tests showed better calibration with haemoglobin A_{1c} or glucose included	NS
Simmons 2007[71]	1.7	Age, sex, use of antihypertensive drugs, BMI, family history of diabetes, physical inactivity, diet (green leafy vegetables, fresh fruit, wholemeal bread)	NS/NS	0.76 (0.73 to 0.79)	NS/NS	NS	NS
Simmons 2007[71]	1.7	Age, sex, current use of corticosteroids, use of antihypertensive drugs, family history of diabetes, BMI, smoking	NS/NS	0.76 (0.73 to 0.79)	NS/NS	NS	NS
Stern 1993[48]	3.7	Fasting plasma glucose level, two hour postprandial plasma glucose level, BMI, high density lipoprotein cholesterol level, pulse pressure	75/88.5	NS	26.80/98.40	NS	12.8
Stern 1993[48]	3.7	Sex, fasting plasma glucose level, BMI, high density lipoprotein cholesterol level, pulse pressure	69.6/88.1	NS	25.20/98.10	NS	14.7
Stern 2002[86]	6.0	Age, sex, ethnicity, triglyceride level, total cholesterol level, low and high density lipoprotein cholesterol levels, fasting plasma glucose level, family history of diabetes in first degree relative, two hour postprandial plasma glucose level, systolic and diastolic blood pressure, BMI	NS/NS	0.86 (0.84 to 0.88)	NS/NS	Hosmer-Lemeshow P>0.2	NS
Stern 2002[86]	6/0	Age, sex, ethnicity, fasting plasma glucose level, systolic blood pressure, high density lipoprotein cholesterol level, BMI, family history of diabetes in first degree relative	NS/NS	0.84 (0.82 to 0.87)	NS/NS	Hosmer-Lemeshow P>0.2	NS
Sun 2009[72]	4.7	Age, sex, education, family history of diabetes, smoker, sport time, high blood pressure, BMI, waist circumference, fasting plasma glucose level	72.3/82.8	0.85 (0.83 to 0.87)	17.18/98.38	Observed to predicted incidence P=0.410	31.2
Sun 2009[72]	4.7	Age, ethnicity, waist circumference, height, systolic blood pressure, family history of diabetes, fasting plasma glucose level	75.2/79.0	0.84	13.54/98.47	NS	23.5

Table 2 (continued)

Study	Diabetes incidence (%)*	Components of score	Sensitivity/specificity† %	AUROC (95% CI)	Positive/negative predictive value (%)	Calibration	% needing further tests
Sun 2009[72]	4.7	Age, ethnicity, waist circumference, height, systolic blood pressure, family history of diabetes, fasting plasma glucose level, triglyceride level, high density lipoprotein cholesterol level	75.0/79.7	0.84	15.39/98.47	NS	22.7
Talmud 2010 [10]	3.5	NS	NS/NS	0.72 (0.69 to 0.76)	NS/NS	Hosmer-Lemeshow P=0.77	19.2
Talmud 2010[10]	3.5	NS	NS/NS	0.78 (0.75 to 0.82)	NS/NS	Hosmer-Lemeshow P=0.42	26.6
Urdea 2009[64]	3.2	Levels of adiponectin, C reactive protein, ferritin, glucose, haemoglobin A_{1c}, interleukin 2, insulin	NS/NS	0.84 (NS)	NS/NS	Observed to predicted risk	NS
Urdea 2009[64]	3.2	Levels of adiponectin, C reactive protein, ferritin, glucose, haemoglobin A_{1c}, interleukin 2, insulin	NS/NS	0.84 (NS)	NS/NS	Observed to predicted risk	NS
Von Eckardstein 2000[50]	5.4	Age, BMI, hypertension, glucose, family history of diabetes, high density lipoprotein cholesterol level	69.5 (62.6-73.9) at 80% specificity, 57.0 (49.8-64.0) at 90% specificity/set at 80% and 90%	0.79 (0.78 to 0.81)	16.7 at 80% specificity, 24.6 at 90% specificity/NS	NS	NS
Wannamethee 2011[27]	4.3	Age, sex, family history of diabetes, smoking status, BMI, waist circumference, hypertension, recall of doctor diagnosed coronary heart disease	79.2 (top 40%) 50.3 (top 20%)/61.8 (top 40%) 81.4 (top 20%)	0.77 (0.74 to 0.79)	NS/NS	Hosmer-Lemeshow P=0.006	47
Wannamethee 2011[27]	4.3	Age, sex, family history of diabetes, fasting plasma glucose level, smoking status, BMI, waist circumference, hypertension, recall of doctor diagnosed coronary heart disease, high density lipoprotein cholesterol level, triglyceride level	84.2 (top 40%), 63.8 (top 20%)/62% (top 40%) 82 (top 20%)	0.82 (0.79 to 0.84)	NS/NS	Hosmer-Lemeshow P=0.43	NS
Wannamethee 2011[27]	4.3	Age, sex, family history of diabetes, smoking, BMI, waist circumference, hypertension, recall of doctor diagnosed coronary heart disease, high density lipoprotein cholesterol level, γ-glutamyltransferase level,, haemoglobin A_{1c} concentration	85.1 (top 40%), 62% (top 20%)/62.1 (top 40%), 82% (top 20%)	0.81 (0.79 to 0.83)	NS/NS	Hosmer-Lemeshow P=0.61	NS
Wannamethee 2005[65]	5.8	NS	35.6/75.7 (both at 20 years)	0.60 (0.56 to 0.64) at 20 years	NS/NS	NS	10.8
Wilson 2007[51]	5.1	Fasting plasma glucose level, BMI, high density lipoprotein cholesterol level, parental history of diabetes, triglyceride level, blood pressure	NS/NS	0.85 (NS)	NS/NS	NS	15.6

NS=not stated; BMI=body mass index.

*Incidence of diabetes was measured differently by different authors, such as annually, every five years, every 10 years, or per 1000 patient years.

†Sensitivity and specificity are based on authors' preferred cut-off score.

Table 3| Summary of authors' assumptions and claims about their diabetes risk models or scores

Study	Authors' assumptions			Authors' adjectives to describe their risk score	Authors' claims for risk score over others	Authors' stated concerns about their risk score	Data in paper on use of risk score in real world	Citation tracking (Google Scholar) for studies of real world use
	Who will use risk score, on which subgroups or populations	What will be offered to people who score above cut-off for "caseness"	Mechanism by which use of risk score may improve outcome					
Aekplakorn 2006[7]	"Primary health care" will use score on "individuals who are likely to develop diabetes"	Fasting plasma glucose test, "health education and the opportunity to engage in healthy lifestyles"	Clinical	Simple, "a practical tool," low tech, no lab tests, non-invasive	"Almost as good as" and less expensive than models that rely on blood tests	Generalisability has not been shown beyond Thai population	Validated on another cohort in same factory	64 citations, not relevant
Alssema 2008[52]	General practitioners, for use on high risk patients. Public health clinicians, for use on high risk populations	Blood test, preventive management according to protocol	Clinical, public health	"Pretty good"	NS	Only predicts getting diabetes, does not predict complications	None	0
Alssema 2011[53]	Intended users not stated. Refined previous risk score	Blood test, "integrated strategies" (addressing risk of cardiovascular disease as well)	Clinical, public health	Updated, refined, simple	Better discrimination	Some missing data in dataset	None	1 citation, not relevant
Balkau 2008[36]	Implicit target audience epidemiologists and population geneticists	Focuses on population level, not clinical care of high risk people	None specifically hypothesised	Simple	Better area under receiver operating characteristic curve, simple (requires 3 variables for men, 4 for women)	2 hour glucose level rarely used in practice	None	34 citations, not relevant
Bozorgmanesh 2010[54]	Clinical ("targeted interventions") and public health ("efficient allocation of resources")	"Intensive diabetes prevention interventions"	Clinical	Simple, parsimonious	Better discrimination capacity, developed on large cohort	Sample may not be representative (too "urban")	None	1 citation, not relevant
Bozorgmanesh 2011[66]	Clinicians in Iran and other Middle Eastern countries; unselected Middle Eastern population	NS	Clinical	Simple, superior, pragmatic, parsimonious, comprehensive	Better discrimination capacity, developed on large cohort	NS	None	2 citations, not relevant
Bozorgmanesh 2010[55]	Clinical practice ("to be ordinarily available in a routine clinical setting"), Middle Eastern countries	Formal test for diabetes, for example, oral glucose tolerance test, plus "Individualised primary prevention"	Clinical	Simple, clinical, parsimonious	Likely to be acceptable to patients and doctors	Response 65%; short follow-up, predictive value reduces with time	NA	0
Cameron 2008[56]	Intended users not stated. Does not consider how scores will be used	Implicitly, general population (Australians). "Lifestyle measures"	Clinical	No better at predicting diabetes than random blood glucose level	NA	Authors unconvinced that it adds value	NA	22 citations, not relevant
Chen 2010[37]	Not stated but score has been converted to an online tool for self assessment of risk by lay people	"Interventions to prevent or delay [diabetes] onset"	Lay people	Simple, non-invasive	Better discrimination, easier to measure (for example, waist circumference more practicable than BMI for lay people)	Developed on narrow age band hence age not very significant in final model	Validated on second population as part of this study	6 citations, of which one was an impact study
Chien 2009[67]	"Clinical practice" (Chinese population)	"Preventive and treatment strategies"	Clinical	Simple	First to be validated in Chinese (but others claim this too)	AUROC only 70%, diabetes not excluded at baseline	None	24 citations, not relevant

Table 3 (continued)

Study	Authors' assumptions: Who will use risk score, on which subgroups or populations	Authors' assumptions: What will be offered to people who score above cut-off for "caseness"	Authors' assumptions: Mechanism by which use of risk score may improve outcome	Authors' adjectives to describe their risk score	Authors' claims for risk score over others	Authors' stated concerns about their risk score	Data in paper on use of risk score in real world	Citation tracking (Google Scholar) for studies of real world use
Chuang 2011[38]	"Clinical professionals and general subjects," for use in "middle aged Chinese adults living in Taiwan"	NS	Clinical	Simple	Menu of scores (some simple, some more complex with better discrimination); large validation cohort	None	None	0
Collins 2011[57]	Implicitly, epidemiologists and public health clinicians, for use in UK population	NS	Public health	Useful	Validated by an independent team on an independent cohort (unlike most others)	None	NA (not their risk score)	0
Gao 2009[39]	"To be used by laypersons" to detect diabetes and raise awareness, "particularly in low-income countries"	NS	Lay people	Simple	Simple, uses absolute risk, based on prospective cohort	Only moderately good predictive power (AUROC 71%)	None	0
Guerrero-Romero 2010[58]	Intended users not stated. For use on unselected Latin American population	Blood test, monitoring of risk, preventive intervention targeting particular risk factors	Implicitly, clinical	Quick and easy to use, few laboratory investigations, cheap	Statistically better than other scores for use on a Latin American population	Not shown to be cost effective or to improve quality of life, needs external validation	None	0
Hippisley-Cox 2009[8]	General practice and public health in areas of high socioeconomic and ethnic diversity; use in "clinical settings" and by lay public through a "simple web calculator"	"To identify and proactively intervene"	Clinical	Simple, good discrimination, well calibrated, readily implementable in primary care, cost effective	Includes deprivation and ethnicity, based on data from general practice record, good statistical properties, well validated, "likely to reduce . . . health inequalities"	Missing values (for example, smoking, ethnicity); internal validation on EMIS only; better design would be a prospective study of inception cohort	None, but authors emphasise that it could be used easily	46 citations, not relevant
Joseph 2010[40]	Implicitly, epidemiologists (focus of paper is identification and refinement of risk factors in a population)	"Lifestyle advice advocating physical activity, healthy low fat diet, and weight reduction"	None specifically hypothesised	NS	More comprehensive, AUROC 0.85, longer follow-up, less bias (for example, in how incident diabetes was diagnosed)	None mentioned	None	0
Kahn 2009[41]	"Insurers or public health agencies . . . to optimise allocation of preventive medicine resources"	"Preventive interventions"	Clinical, public health	Low cost, clinical, simple	Prospectively validated, may illuminate cause of diabetes by demonstrating new associations	Limited to age 45-65 and to white or black ethnic groups	None	29 citations, not relevant
Kanaya 2005[59]	To identify "older persons who should receive intensive lifestyle intervention"	"Lifestyle modification"	Clinical	Simple	Very simple, validated in several samples	Needs validating in a longitudinal study	None	0

Table 3 (continued)

Study	Authors' assumptions		Mechanism by which use of risk score may improve outcome	Authors' adjectives to describe their risk score	Authors' claims for risk score over others	Authors' stated concerns about their risk score	Data in paper on use of risk score in real world	Citation tracking (Google Scholar) for studies of real world use
	Who will use risk score, on which subgroups or populations	What will be offered to people who score above cut-off for "caseness"						
Kolberg 2009[42]	For use on "individuals at highest risk of developing type 2 diabetes"	"for whom the most comprehensive prevention strategies should be considered"	None specifically hypothesised	Objective, quantitative	Biologically plausible ("multi-biomarker"), convenient, fewer logistical challenges to implementation, better discrimination	Developed in overweight middle aged white people, hence transferability may be limited	None	29 citations, not relevant
Lindstrom 2003[68]	Intended users not stated. Implicitly, those who (like the authors) seek to undertake intervention studies of diabetes prevention. For use with "the general public"	"Direct attention to modifiable risk factors." Also, doing one's own risk score might prompt people to modify their lifestyle and prompt them to get their blood glucose level checked	Clinical, lay people	Simple, practical, informative, fast, non-invasive, inexpensive, reliable, safe	Prospective, large cohort. "The public health implications of the Diabetes Risk Score are considerable"	Possible circular argument—identifying people based on same risk factors that would have prompted their clinician to measure random blood glucose level in the first place	Not in this paper, but see citation track	343 citations, of which eight described impact studies
Liu 2011[43]	Clinicians. "initial instrument for opportunistic screening in general population", "could enhance people's awareness"	Oral glucose tolerance test, education, "opportunity to engage in healthy lifestyles at an early stage"	Clinical	Practical, effective, simple, easily used in clinical practice	Validated on a mainland Chinese population, large cohort, prospective, stable prediction model	Validated in middle aged to older cohort so unproved benefit in younger people. Did not include family history of diabetes, as not on database	None	0
Mainous 2007[60]	Implicitly, clinicians. Paper describes validation of a previous risk score in a younger cohort	"Early recognition and treatment"	Clinical	NA (they don't recommend it in this group)	NA	Poor discriminatory ability	None	8 citations, not relevant
Mann 2010[19]	"Clinicians . . . to stratify their patient populations"	NS	None specifically hypothesised	High discriminative ability	Recalibration and revalidation of Framingham based score in large ethnically diverse population	Inability to isolate Mexicans	None	3 citations, not relevant
McNeely 2003[61]	"Clinical practice." To predict diabetes risk in Japanese Americans	NS	None specifically hypothesised	None, all data expressed in numbers	Better in short term than fasting blood glucose test but not in long term (younger people). Not as good as oral glucose tolerance test (older people)	"Further refinements that take into account the differential effects of age are needed"	None	29 citations, not relevant
Mehrabi 2010[44]	NS	NS	Not specifically hypothesised	Useful, novel	Higher predictability rate than use of single risk factors alone	New and relatively untested, some missing data	None	0
Meigs 2008[9]	NA—negative study showing that genetic factors add nothing to clinical scores	NA	NA (authors suggest further research on key subgroups)	Less useful than data collected at a routine clinical examination	NA	Did not help to refine the prediction of diabetes risk	NA	163 citations, but not relevant as paper cited for its negative findings

Table 3 (continued)

Study	Authors' assumptions						Data in paper on use of risk score in real world	Citation tracking (Google Scholar) for studies of real world use
	Who will use risk score, on which subgroups or populations	What will be offered to people who score above cut-off for "caseness"	Mechanism by which use of risk score may improve outcome	Authors' adjectives to describe their risk score	Authors' claims for risk score over others	Authors' stated concerns about their risk score		
Nichols 2008[62]	Health maintenance organisations. Based on analysis of electronic record data, to identify members at high risk of developing diabetes	"Interventions" and targeting of healthcare resources	Clinical, public health, technology	"Extremely accurate," simple	Better AUROC	If health maintenance organisation population has different incidence of type 2 diabetes from validation cohort, score will be inaccurate	None	1 citation, not relevant
Rahman 2008[63]	Primary care and public health clinicians. Use for "defining individuals and populations for testing, treatment and prevention"	Not explicitly stated but authors suggest potential avenues for impact studies	Clinical, public health	Simple, effective	Based on data routinely available on general practice records	Will need to be validated in other prospective cohorts	None	29 citations, not relevant
Rathmann 2010[85]	Intended users not stated. Use "to identify high-risk populations for preventive strategies"	"Preventive strategies"	Public health	Simple	Validated in older population	No external validation yet	None	1 citation, not relevant
Rosella 2010[69]	Public health clinicians and health planners "to estimate diabetes incidence, to stratify the population by risk, and quantify the effect of interventions"	"New intervention strategies"	Public health, clinical	Simple	Uses data available on population registries	Could be further tested on other populations. Family history and poor diet not collected, relies on self reports	None	1 citation, not relevant
Schmidt 2005[46]	Use "in clinical encounters," "by managed care organizations . . . to identify high-risk individuals," and to enrol to clinical trials	"Preventive actions of appropriate intensity"	Clinical, public health, research	Simple, based on readily available clinical information and simple laboratory tests	Good predictor for white and African-American men and women; may apply also to other ethnic groups in United States	High losses to follow-up, oral glucose tolerance test not done at baseline	None	111 citations, not relevant
Schulze 2007[70]	Intended users not stated. "Identifying individuals at high risk of developing T2D [type 2 diabetes] in the general population"	Not explicitly stated	The public	Precise, non-invasive, accurate, useful	Good AUROC (0.84), used absolute values for age rather than broad categories	Self reports may have been biased	None	114 citations, not relevant
Schulze 2009[47]	NS	NS	None specifically hypothesised	Improved discrimination	"A comprehensive basic model," significantly improved by routine blood tests but not chemical or genetic biomarkers	Predictive for onset of diabetes in middle age but not from birth, since diabetes was excluded from inception cohort	None	17 citations, not relevant
Simmons 2007[71]	Primary care: "could inform . . . health behaviour information . . . routinely collected in GP consultations or by administrative staff," identify groups for targeted prevention	"Could be incorporated into new patient health checks and may provide a more feasible means of identifying those at risk than OGTT [oral glucose tolerance test], or select those	Clinical, administrative	Simple, feasible	Relies only on simple questions about lifestyle, which would be asked in a routine health check. AUROC (0.76) is as good as many complex risk scores	No better than standard clinical dataset routinely collected in UK general practice (but may be feasible in other health settings)	Feasible to collect	21 citations, not relevant

Table 3 (continued)

Study	Authors' assumptions: Who will use risk score, on which subgroups or populations	Authors' assumptions: What will be offered to people who score above cut-off for "caseness"	Mechanism by which use of risk score may improve outcome	Authors' adjectives to describe their risk score	Authors' claims for risk score over others	Authors' stated concerns about their risk score	Data in paper on use of risk score in real world	Citation tracking (Google Scholar) for studies of real world use
		suitable for OGTT"						
Stern 1993[48]	Implicitly, epidemiological researchers	"Identifying high-risk cohorts for prevention trials"	Research, clinical	Predictive, multivariate	Uses commonly measured clinical variables	NS	None	45 citations, not relevant
Stern 2002[86]	"Could be incorporated as it stands into clinical practice and public health practice with the aid of a calculator or personal computer"	Clinical: "patient counselling." Public health: "to identify target populations for preventive interventions"	Clinical, public health, technological, research	Simple	Less expensive and more convenient than oral glucose tolerance testing	Possible missing data	None	245 citations, not relevant
Sun 2009[72]	Use in clinical encounter, by managed care organisations to identify high risk people, and to enrol to clinical trials	Further research	Clinical, technological, research	Simple, effective, accurate	Simple, uses readily available clinical information	Losses to follow-up, oral glucose tolerance test not done at baseline so some cases detected, especially early on, may be prevalent ones	None	3 citations, not relevant
Talmud 2010[10]	Intended users not stated (but study used an existing risk score as a "control" for testing a genetic profile)	NS	Not specifically hypothesised	NA (revalidation)	Simple clinical risk scores performed much better than assessment of genetic risk from 40 polymorphisms	NA	None	21 citations, not relevant
Urdea 2009[64]	"Current clinical practice"; for "identifying individuals at highest risk of developing T2DM [type 2 diabetes mellitus]"	"so that clinicians can implement an effective diabetes prevention program"	Clinical	Simple, accurate, convenient	"Better than any other clinical measure", not over-fit, based on multiple biomarkers hence highly plausible	NS	None	6 citations, not relevant
Von Eckardstein 2000[50]	NA (negative study, no better than fasting blood glucose test alone in this cohort)	NA (negative study)	NA	NA (negative study)	NA	Negative study	NA	56 citations, not relevant
Wannamethee 2011[27]	Intended users not stated	Not stated, unit of analysis is the population	Not specifically hypothesised	NA (less effective than Framingham risk score)	"Useful predictor" (but not as good as Framingham score)	NS	None	273 citations, not relevant
Wannamethee 2005[65]	Intended users not stated	Blood tests	Not specifically hypothesised	Simple, routine	Stepwise	Diabetes diagnosed by self reports	None	0
Wilson 2007[51]	Implicitly, clinicians	Implicitly, lifestyle advice and metformin	Clinical	Simple, effective, easy	Very good AUROC (85%)	NS	None	143 citations, not relevant

NS=not stated; NA=not applicable; BMI=body mass index; AUROC=area under receiver operating characteristic curve.

Table 4| Components of seven diabetes risk models or scores with potential for adaptation for use in routine clinical practice

Score/study name, country, reference	Risk factors included in score	AUROC	Calibration	External validation		
				Year, country	AUROC	Calibration
ARIC (Atherosclerosis Risk in Communities), Germany, Schmidt 2005[46]	Age, ethnicity, waist circumference, height, systolic blood pressure, family history of diabetes, fasting plasma glucose levels, triglyceride levels, high density lipoprotein cholesterol levels	0.80	NS	2010,[19] USA	0.84	Hosmer-Lemeshow P<0.001, after recalibration P>0.10
Ausdrisk, Australia, Chen 2010[37]	Age, sex, ethnicity, parental history of diabetes, history of high blood glucose, use of antihypertensive drugs, smoking, physical inactivity, waist circumference	0.78	Hosmer-Lemeshow P=0.85	Not externally validated but has been studied as part of an intervention to improve outcomes[87]		
Cambridge risk score, UK, Rahman 2008[63]	Age, sex, use of current corticosteroids, use of antihypertensive drugs, family history of diabetes, body mass index, smoking	0.74 with threshold of 0.38	NS	2010,[10] UK*	0.72	Hosmer-Lemeshow P=0.77
FINDRISC, Finland, Lindstrom 2003[68]	Age, body mass index, waist circumference, use of antihypertensive drugs, history of high blood glucose, physical inactivity, daily consumption of vegetables, fruits, and berries	0.85	NS	2010,[53] Holland, Denmark, Sweden, UK, Australia*	0.76	Hosmer-Lemeshow P=0.27
Framingham Offspring Study, USA, Wilson 2007[51]	Fasting plasma glucose levels, body mass index, high density lipoprotein cholesterol levels, parental history of diabetes, triglyceride levels, blood pressure	0.85	NS	2010,[19] USA	0.78	Hosmer-Lemeshow P<0.001, after recalibration P>0.10
San Antonio risk score, clinical model, USA, Stern 2002[49]	Age, sex, ethnicity, fasting plasma glucose levels, systolic blood pressure, high density lipoprotein cholesterol levels, body mass index, family history of diabetes in first degree relative	0.84	Hosmer-Lemeshow P>0.2	2010,[19] USA; 2010,[55] Iran*; 2010,[10] UK*; 2010,[66] Iran*	0.83; 0.83; 0.78; 0.78	Hosmer-Lemeshow P<0.001, after recalibration P>0.10; Hosmer-Lemeshow P≤0.001, after recalibration P=0.131; Hosmer-Lemeshow P=0.42; Hosmer-Lemeshow P=0.264
QDScore, UK, Hippisley-Cox 2009[8]	Age, sex, ethnicity, body mass index, smoking, family history of diabetes, Townsend deprivation score, treated hypertension, cardiovascular disease, current use of corticosteroids	0.83 men, 0.85 women	Brier score: 0.078 men, 0.058 women	2011,[57] UK	0.80 men, 0.81 women	Brier score: 0.053 men, 0.041 women

AUROC=area under receiver operating characteristic curve; NS=not stated.

*Validation used more, less, or substituted risk factors from original risk score or did not state the exact factors it used. See table 2 for further details.

Table 5| Results of impact citation search (studies using diabetes risk models or scores as part of an intervention to improve outcomes)

Study (acronym)	Score used	Research question	Setting and sample	Study design, intervention	Main findings or expected reporting date	Comment
Absetz 2009 (GOAL study)[88]	FINDRISC[68]	Can diabetes risk be reduced by lifestyle counselling?	Australia, 352 high risk adults	Real world feasibility study: eight lifestyle counselling sessions	271/352 completed study. Showed statistically significant reduction in weight, body mass index, and total cholesterol level, maintained at 36 months	Changes only reported on "completers"; those lost to follow-up were not included in analysis. Absolute changes were small and probably not clinically significant—for example, mean 1 kg weight loss. Change in FINDRISC score was not reported
Jallinoja 2008 (GOAL study)[89]	FINDRISC[68]	What is the experience of lifestyle change in people recruited into diabetes prevention studies?	Australia, 30 weight losers and 30 weight gainers from GOAL study	Focus groups with weight losers and weight gainers studied separately	Many found dietary change difficult and stressful; some who did not achieve weight loss felt despondent	Some but not all people encouraged to change lifestyle will achieve it, but most will struggle
Colaguiri 2010 (Sydney DPP)[87]	AUSD-RISK[37]	Can diabetes risk be reduced by a programme of intensive behaviour change?	Australia, 1550 high risk adults (100 indigenous people)	Real world feasibility study: individual assessment followed by group sessions	Results expected 2013. Main outcomes will be change in weight, physical activity, diet, fasting glucose levels, blood pressure, lipid levels, quality of life, and health service utilisation	Participants will be recruited in primary care, but intervention will be delivered as a public health/community based programme
Kulzer 2009 (PREDIAS)[90]	FINDRISC[68]	Can diabetes risk be reduced by lessons in lifestyle modification?	Germany, 182 high risk adults	Randomised trial. Intervention group received 12 group lessons in lifestyle modification, controls had leaflet	Statistically significant changes in weight, physical activity, diet, and fasting glucose levels at 12 months compared with controls	Weight loss in intervention group was clinically significant (3.8 kg); fasting glucose in the control group increased, whereas that in the intervention group decreased. However, follow-up was short
Laatikainen 2007 (GGTDPP)[91]	FINDRISC[68]	Can risk factor reduction be achieved in a high risk non-trial population?	Australia, 237 high risk adults	Real world feasibility study: six sessions of nurse led group education	Statistically significant improvements in weight, fasting and two hour glucose levels, and lipid levels at 12 months	Mean weight loss 2.52 kg. Authors view findings as "convincing evidence that a type 2 diabetes prevention programme using lifestyle intervention is feasible in Australian primary health care with reductions in risk factors approaching those observed in randomised controlled trials"
Saaristo 2007 (FIN-D2D)[80] and Lindstrom 2010 (FIN-D2D)[74]	FINDRISC[68]	Can a population approach detect high risk people, modify their risk through educational intervention, and thereby reduce the incidence of new diabetes?	Finland, high risk adults (part of a national diabetes prevention programme that also included population component)	High scorers on FINDRISC had oral glucose tolerance and lipid levels tested; those without diabetes were offered nurse led community based individual or group sessions, or both, based on stages of change and tailored to individual profile	Preliminary results only. Numbers and detailed findings not given. "Desirable changes" at 12 months in risk factors and glucose tolerance in high risk cohort. Incident diabetes reduced (as measured by drug reimbursement registration data). Full results expected 2012-13	Authors report that "certain problems and challenges were encountered, especially in relation to the limited resources allotted to preventive health-care."[74] A smaller ongoing prevention programme using FINDRISC along with occupational health screening on an occupational cohort in an airline company (FINNAIR diabetes prevention study) is also briefly outlined in Lindstrom paper[74]
Schwarz 2007 (TUMANI)[59]	FINDRISC[68]	Can an intensive, multifaceted public health intervention prevent incident diabetes in high risk people?	Germany, high risk adults (part of a national prevention programme)	High scorers on FINDRISC had oral glucose tolerance test before being assigned a "prevention manager" for education, support, and telephone counselling	Results expected 2012-13	Authors recognise that prevention on a large scale sits oddly within the existing treatment oriented health system. Key features of TUMANI are prevention managers working within the existing infrastructure, a structured quality control programme, and a population component—for example, website and links to mass media
Vermunt 2010 (APHRODITE)[92]	FINDRISC[68]	Can a mailed questionnaire from general practice identify high risk people to participate in a preventive intervention?	Netherlands, 48 general practices	General practitioners mailed questionnaires to their adult patients. High scorers were offered oral glucose tolerance test	16 032 people were mailed; response rate to questionnaire 54.6%, of which 17.5% were classified as high risk. Of these, 73.1% booked a consultation with their general practitioner. Full results expected 2014	Findings to date suggest that half of high risk patients were willing to fill out the FINDRISC questionnaire and follow-up with their general practitioner. Response rates to questionnaire varied significantly among practices

Figures

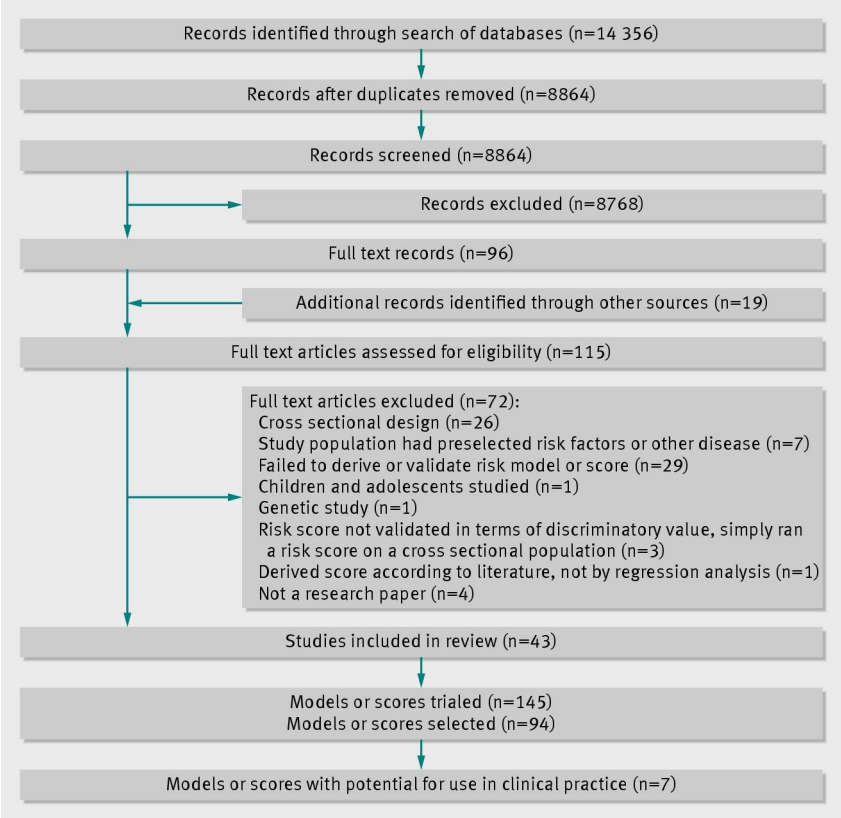

Fig 1 Flow of studies through review

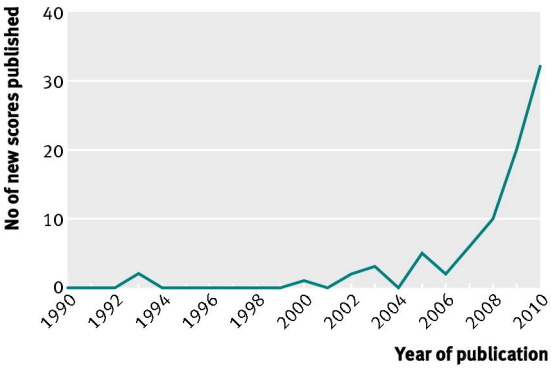

Fig 2 Publication of diabetes risk models and scores 1990-2010. Eleven new risk models and scores had been published in the first five months of 2011