

Systematic Review or Meta-analysis

Risk scores for type 2 diabetes mellitus in Latin America: a systematic review of population-based studies

R. M. Carrillo-Larco^{1,2,4} , D. J. Aparcana-Granda^{2,*}, J. R. Mejia^{5,*}, N. C. Barengo^{6,7,8} and A. Bernabe-Ortiz^{2,3} 

¹Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK, ²CRONICAS Centre of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia, ³Universidad Científica del Sur, Lima, ⁴Centro de Estudios de Poblacion, Universidad Catolica los Ángeles de Chimbote (ULADECHCatolica), Chimbote, Perú, ⁵Facultad de Medicina Humana, Universidad Nacional del Centro del Perú, Huancayo, Perú, ⁶Department of Medical and Population Health Sciences Research, Herbert Wertheim College of Medicine, Florida International University, Miami, FL, USA, ⁷Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland and ⁸Faculty of Medicine, Riga Stradins University, Riga, Latvia

Accepted 20 August 2019

Abstract

Aim To summarize the evidence on diabetes risk scores for Latin American populations.

Methods A systematic review was conducted (CRD42019122306) looking for diagnostic and prognostic models for type 2 diabetes mellitus among randomly selected adults in Latin America. Five databases (LILACS, Scopus, MEDLINE, Embase and Global Health) were searched. type 2 diabetes mellitus was defined using at least one blood biomarker and the reports needed to include information on the development and/or validation of a multivariable regression model. Risk of bias was assessed using the PROBAST guidelines.

Results Of the 1500 reports identified, 11 were studied in detail and five were included in the qualitative analysis. Two reports were from Mexico, two from Peru and one from Brazil. The number of diabetes cases varied from 48 to 207 in the derivations models, and between 29 and 582 in the validation models. The most common predictors were age, waist circumference and family history of diabetes, and only one study used oral glucose tolerance test as the outcome. The discrimination performance across studies was ~ 70% (range: 66–72%) as per the area under the receiving-operator curve, the highest metric was always the negative predictive value. Sensitivity was always higher than specificity.

Conclusion There is no evidence to support the use of one risk score throughout Latin America. The development, validation and implementation of risk scores should be a research and public health priority in Latin America to improve type 2 diabetes mellitus screening and prevention.

Diabet. Med. 36, 1573–1584 (2019)

Introduction

Type 2 diabetes mellitus is a leading cause of morbidity, disability and mortality worldwide [1–3], disproportionately affecting low- and middle-income countries in Latin America [4]. In addition, type 2 diabetes mellitus imposes a heavy financial burden on local healthcare systems [5]. Therefore, the increasing number of newly detected type 2 diabetes mellitus cases creates challenges for low- and middle-income countries [6]. The United Nations/World Health Organization have set several goals to reduce the burden of non-communicable diseases, including a 0% increase in diabetes

[7]. For Latin America, in particular, the Pan American Health Organization has issued policies and guidelines for the control and prevention of diabetes [8]. Epidemiological evidence along with the active participation of international health organizations, support the relevance of identifying pragmatic strategies to reduce type 2 diabetes mellitus burden at the population level.

A pragmatic, although still challenging solution is the early identification of people with type 2 diabetes or those at high risk of developing type 2 diabetes so that non-pharmacological and pharmacological prevention strategies can be initiated. Diagnostic and prognostic models such as risk scores are convenient for this purpose and yet their use is limited to the population for which they were developed, hence internal and external validation before application in new populations are recommended. Although there have been previous efforts to synthesize available risk scores

Correspondence to: Rodrigo M. Carrillo-Larco. E-mail: rcarrill@ic.ac.uk

*These authors contributed equally to this work

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

What's new?

- Risk scores are tools that could support screening, diagnosis and prognosis decisions in clinical medicine and public health.
- Risk scores for undiagnosed diabetes or to predict diabetes are available worldwide with a few in Latin America. However, the characteristics of risk scores available for Latin America, their performance, pitfalls and other attributes have not been summarized or appraised.
- A lack of synthesized information makes it difficult to understand the strengths and limitations of the available tools, hampering their implementation in clinical and screening guidelines.
- We conducted a thorough search for risk scores for type 2 diabetes developed in Latin America, providing the clinical and public health communities with evidence to inform their decisions regarding these risk scores.
- Local and regional health organizations could recommend one risk score or foster the development of a stronger tool to overcome the limitations signalled herein.

globally [9–12], even focusing on Latin American populations in the USA [13], scientific information on type 2 diabetes risk scores in Latin American countries has been limited. Therefore, whether there are scientifically validated type 2 diabetes mellitus risk scores for populations in Latin America remains unknown.

We aimed to critically review the current scientific evidence on developed diabetes risk scores for Latin American populations. In so doing, we provide a list of risk scores that could be further studied in different Latin American countries, used by practitioners in countries where the models were developed, or integrated by guideline/policy-makers in the current standard of practice for diabetes screening at the population level. Emphasis is placed on tools developed for the general population because of their ability to be used in different communities, thus benefiting populations beyond those accessing the health system [14–16].

Methods**Protocol and registration**

This review is a systematic and critical appraisal of the scientific literature following PRISMA guidelines and registered at PROSPERO (CRD42019122306) [17]. The review framework adheres to international recommendations for systematic reviews of prediction models and followed the CHARMS strategy [18,19].

Eligibility criteria

Eligibility criteria for studies following the CHARMS checklist are given in Table 1. In brief, we searched both diagnostic and prognostic models aiming to inform general practitioners (GPs), clinicians, researchers and the general population about their current type 2 diabetes status (i.e. diagnostic) or future risk (i.e. prognostic). The studies could present results for models with or without external validation. The target population was adults in Latin America with no restrictions on age.

Information sources

Five search engines were used systematically: LILACS, Scopus, MEDLINE, Embase and Global Health; the last three through Ovid. The search was conducted on 15 January 2019 with no time or language restrictions. The search terms used are given in Appendices S1–S3.

Study selection

Reports were selected if the study population included men and women who were both from and living in any Latin American country. Thus, studies including Latin American populations outside Latin America or those including only foreigners living in Latin America were excluded. To be included in this review, the study participants had to be a randomly selected sample of the general population. Studies of convenience samples were excluded. Furthermore, studies including a specific subsample of people (e.g. studies in obese or hypertensive people) as well as hospital-based samples were excluded. The outcome of interest was previously undiagnosed type 2 diabetes mellitus, defined using at least one biomarker such as fasting glucose, random glucose, oral glucose tolerance test or HbA_{1c}. Studies in which the outcome was defined solely based on self-reported diagnosis were excluded. Reports needed to present the development and/or validation procedures of a multivariable model. Thus, studies assessing the diagnostic or predictive power of one variable or biomarker alone were excluded. In addition, both cross-sectional and cohort studies were included.

Data collection process

Results from the literature search were downloaded into EndNote and duplicates were removed. All unique results were uploaded to Rayyan [20], an online systematic review tool, whereby titles and abstracts were independently screened by two reviewers (pairwise combinations between RMC-L, DJA-G, JRM) and disagreements were solved by consensus or by a third party (AB-O). Before screening, all reviewers underwent a standardization process. Reports selected from the screening phase were studied in detail by

Table 1 Criteria to guide the literature search and selection criteria

Concept	Criteria
Prognostic or diagnostic? Scope	Both, this review focuses on diagnostic and prognostic risk scores for type 2 diabetes mellitus Diagnostic/prognostic models to inform physicians, researchers and general population about their current type 2 diabetes mellitus status (i.e. diagnostic) or risk of type 2 diabetes mellitus in the future (i.e. prognostic)
Type of prediction modelling studies	Focus on the three types: (i) diagnostic/prognostic models with external validation, (ii) diagnostic/prognostic models without external validation, and (iii) diagnostic/prognostic models validation
Target population to whom the prediction model applies	General adult population in Latin America and the Caribbean; no age or gender restrictions
Outcome to be predicted	type 2 diabetes mellitus (diagnostic or prognostic)
Time span of prediction	Any; prognostic models will not be included/excluded based on the prediction time span
Intended moment of using the model	Diagnostic/prediction models to be used in asymptomatic adults in Latin America to ascertain current type 2 diabetes mellitus status (i.e. diagnostic) or future risk of type 2 diabetes mellitus (i.e. prognostic); these models could be used for research purposes, screening and treatment allocation in primary prevention

Based on the CHARMS checklist. [19]

two reviewers independently (RMC-L, DJA-G, JRM), and disagreements were solved by consensus or by a third party (AB-O). These processes led to the selection of reports for inclusion in the qualitative summary, from which key information was extracted onto a data collection form developed by the authors based on international guidelines for systematic reviews of prognosis models (CHARMS checklist) [18,19]. The data collection form was not modified afterwards. Information was extracted by one reviewer (RMC-L) and independently verified by another (AB-O); disagreements were solved by consensus.

Risk of bias of individual studies

Risk of bias was assessed following the PROBAST recommendations, a risk of bias assessment tool developed exclusively for prognosis models [21,22]. This process was conducted by two reviewers independently (DJA-G and JRM) and verified by a third (AB-O).

Synthesis of results

Only a qualitative synthesis was conducted. A quantitative synthesis was not possible because of the small number of reports using the same variables in the prediction models.

This study was classified as low risk because no human participants were studied. This is a systematic review of the scientific literature, which is public and can be accessed freely.

Results

The initial search yielded 1546 results; 1500 titles and abstracts were screened and 11 reports were studied in full. Five reports were included in the qualitative synthesis (Fig. 1) [23–27].

What has been done?

In 2018, Bernabe-Ortiz and colleagues derived a simplified version of the FINDRISC, and validated the original FINDRISC, the Latin America-FINDRISC and the Peruvian Risk Score, i.e. they provided estimates for four models (one derivation and three validations) [23]. In 2016, Bernabe-Ortiz *et al.* derived a diagnostic model, which was validated externally, using both cross-sectional and prospective data, i.e. they provided estimates for three models (one derivation and two validations) [24]. In 2018, Félix-Martínez and Godínez-Fernández derived and validated two models using cross-sectional data collected in 2006 and 2012, i.e. they provided estimates for four models (two derivations and two validations) [25]. In 2010, Guerrero-Romero and Rodríguez-Morán derived a model using cross-sectional data and validated it in prospective data, i.e. provided estimates for two models (one derivation and one validation) [26]. Finally, in 2009, Pires de Sousa and colleagues derived and validated a diagnostic model, i.e. they provided estimates for two models (one derivation and one validation) [27]. Overall, six models were derived and nine underwent validation analysis.

General characteristics

Two of the five reports studied people in Peru [23,24], two studied Mexicans [25,26], and one was conducted in Brazil [27]. The oldest analysed data was collected in 1996 [26], and the remainder of the studies used data collected after 2000 [23–25,27]. The mean age of the participants in the derivation models ranged from 42 to 50 years, and the proportion of men varied from 38% to 51% [23–27]. The mean age of people analysed in the validation models ranged from 40 to 55 years, and the proportion of men ranged from 25% to 49% (Appendix S4) [23–27].

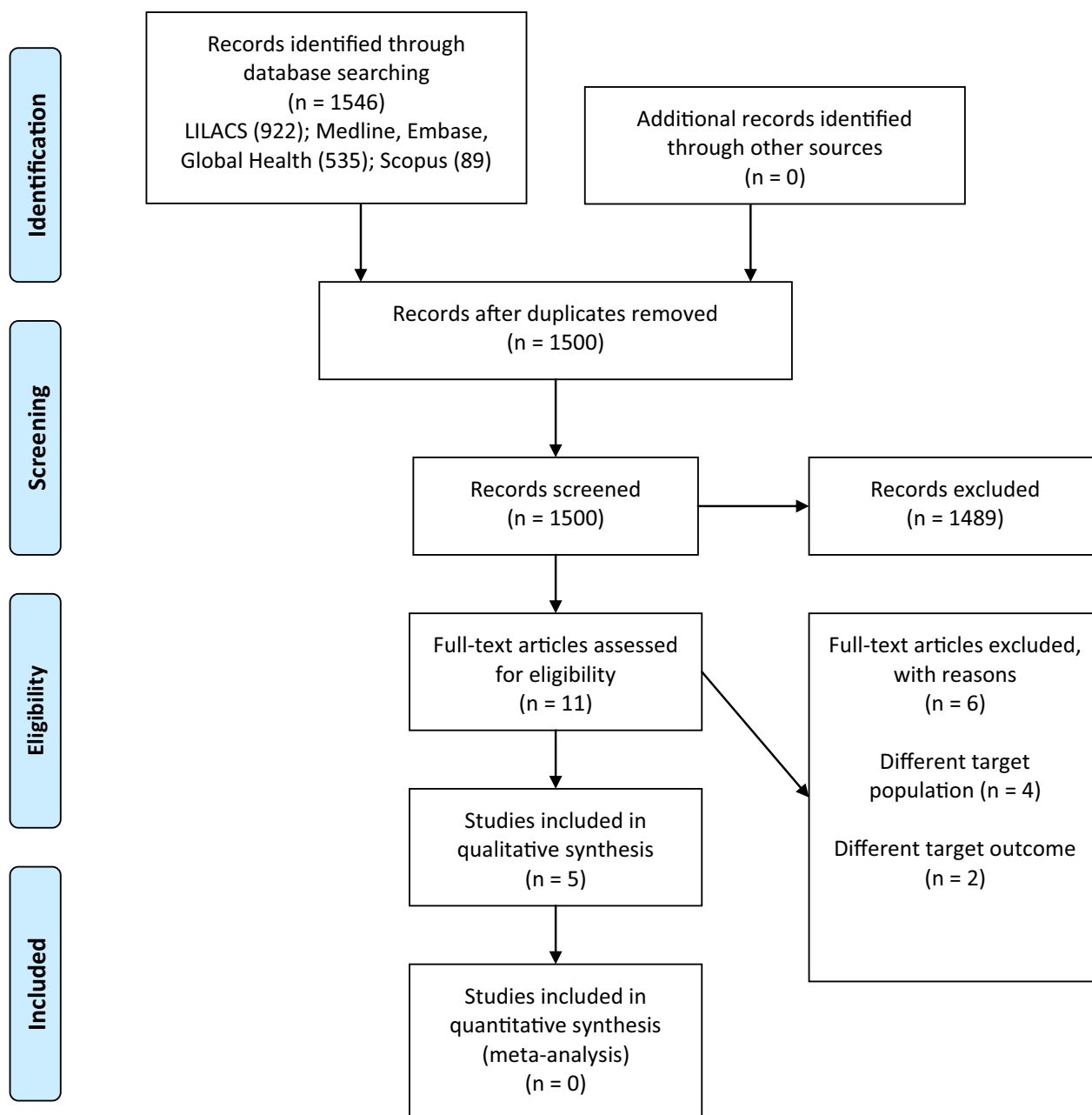


FIGURE 1 Flowchart of the study selection process.

The sample size analysed to derive the diagnostic models ranged from 711 [26] to 6995 people [25], and from 438 [26] to 28 913 [25] for the validation models. The number of diabetes cases varied greatly in the derivation models, from 48 [24] to 207 [25], with only two derivations models having ≥ 100 events; the equivalent numbers in the validation models were 29 [27] and 582 [25], with two derivation models having ≥ 100 events. Of note, this information (number of outcome events or diabetes cases) could not be extracted from Guerrero-Romero and Rodríguez-Morán [26]. The ratio of outcome events per number of candidate predictors in the derivation analyses ranged from 3.43 [24] to

15.92 [25]. Across all reports, missing data were handled by conducting a complete-case analysis [23–25,27], although this information was not available in Guerrero-Romero and Rodríguez-Morán's study [26] (Appendix S4)

All derivations models used a logistic regression analysis [23–27]. In all but one report [26], it was clear that preselection of predictors was conducted (i.e. choosing the final list of predictors based on statistical significance), mostly following a stepwise backward selection approach (Appendix S4) [23–25,27]. In Félix-Martínez and Godínez-Fernández's work, numerical variables were not categorized [25], but this approach was followed in the other studies

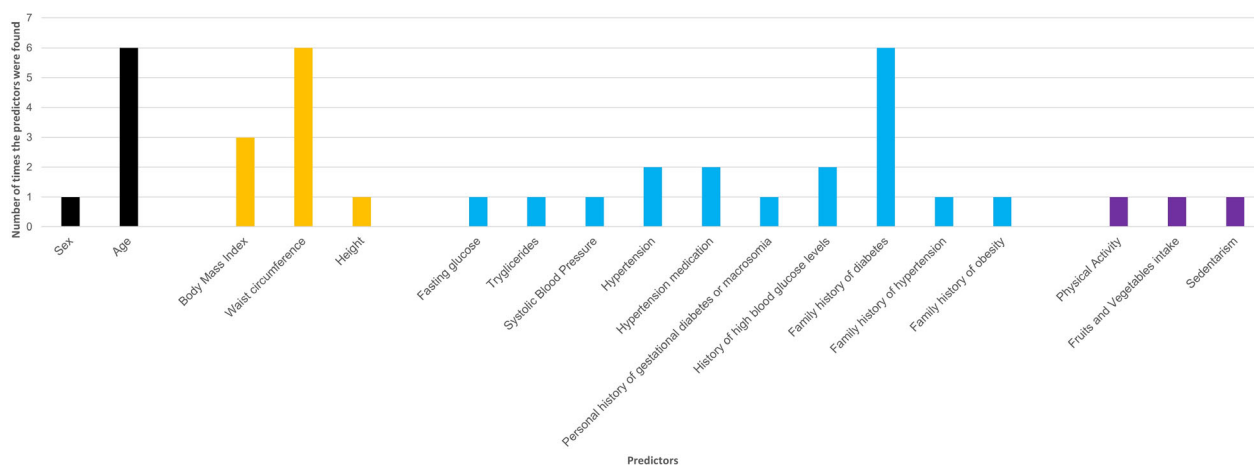


FIGURE 2 Predictors included in the final diagnosis and prognosis models. The colours of the bars identify the underlying characteristic of predictors inherent to: the subject (black), anthropometrics (orange), clinical assessment and history (blue) and lifestyle-related behaviours (purple).

[23,24,26,27]. As shown in Fig. 2, the most common predictors used in the models were: age, waist circumference and family history of diabetes (Appendix S4) [23–27].

Outcome ascertainment

Across reports, the outcome was new type 2 diabetes cases based on standard laboratory procedures. It was not possible to extract the definition used to diagnose new diabetes cases from Guerrero-Romero and Rodríguez-Morán's work [26]. The other reports relied mostly on fasting plasma glucose ≥ 7.0 mmol/l (126 mg/dl) [23–25,27]. In addition to fasting plasma glucose, Bernabe-Ortiz *et al.* also used 2-h plasma glucose ≥ 11.1 mmol/l (≥ 200 mg/dl) [23]. Félix-Martínez and Godínez-Fernández [25] also defined new diabetes cases according to random glucose ≥ 11.1 mmol/l (≥ 200 mg/dl) (Appendix S4).

Model performance

Figure 3 shows the performance metrics for each derivation model as presented in the original reports (Appendix S4) [23–27]. Discrimination performance across studies was $\sim 70\%$ as per the area under the receiving-operator curve, ranging from 66% [25] to 72% [24,27]. Where reported, the negative predictive value was the best metric, achieving almost 100% [23,24,27]. Sensitivity was always larger than specificity [23–25,27] and the largest absolute difference was 39.2% (sensitivity, 85.9%; specificity, 46.7%) [23].

Figure 4 shows the performance metrics for the validation models as reported originally (Appendix S4) [23–27]. Discrimination performance ranged from 64.0% [25] to 91.0% [26]. The best performance metric was the negative predictive value, and sensitivity was always larger than specificity (Appendix S4) [23–27]. In both prediction and validation analyses, calibration metrics or plots were not

presented, though Bernabe-Ortiz *et al.* reported the Hosmer–Lemeshow *P*-value (0.21) [24].

Risk of bias

Table 2 summarizes the risk of bias ascertainment, and full details are provided in Appendix S5. Across all reports, the participants' criterion in the risk of bias assessment revealed a low risk of bias; by contrast, the analysis criterion showed high risk of bias mostly because of few numbers of outcome events (i.e. diabetes cases) and conducting complete-cases analysis instead of performing imputation methods. Across summarized studies, there was a low concern regarding applicability as models were created from population-based studies.

Discussion

Main findings

This systematic review of the literature synthesized the available risk scores for type 2 diabetes mellitus that can be used in Latin American general populations and countries, providing evidence and tools for practitioners as well as guideline/policy-makers across Latin America. Five reports from three countries (Brazil, Mexico and Peru) were summarized [23–27], which developed five diagnostic tools; two of them also conducted a longitudinal assessment [24,26]. In addition, these five reports provided results for the validation of nine models [23–27]. Although discrimination estimates were largely acceptable, calibration metrics were not reported. The negative predictive value was the highest metric across risk scores [23–27]. Even though several type 2 diabetes mellitus risk scores have been developed for Latin American populations, few have followed optimal analytical approaches regarding internal and

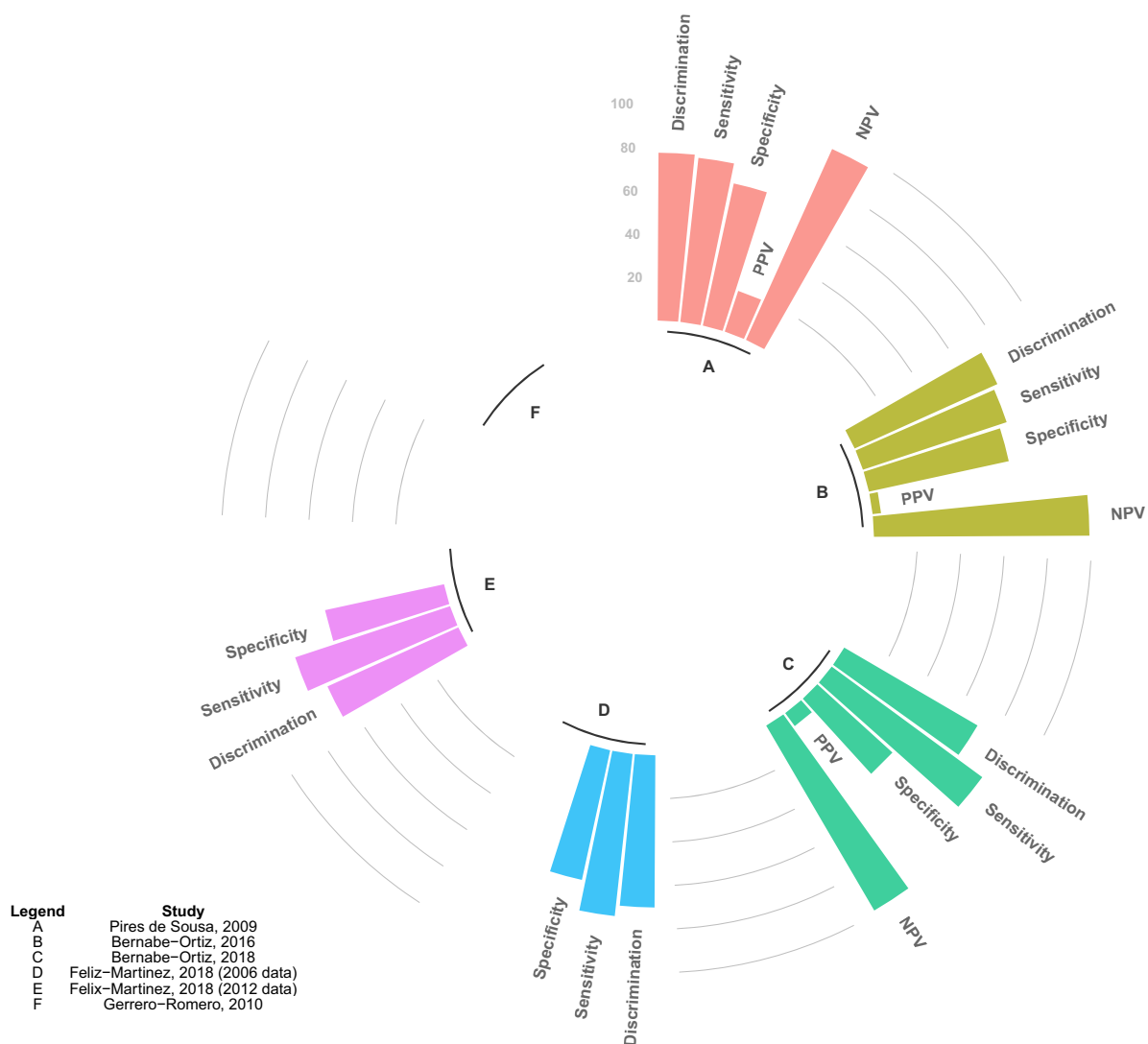


FIGURE 3 Performance metrics of the derivation models. Guerrero-Romero and Rodríguez-Morán [26] did not provide these details for the derivation models, thus it was left blank. Pires de Sousa *et al.* [27] presented these metrics for several thresholds (Appendix S4), the ones shown here correspond to the cut-off point with sensitivity closest to the pre-specified value. Bernabe-Ortiz *et al.* [24] presented these metrics for several thresholds (Appendix S4), the ones shown here correspond to those with the best Youden Index. A, Pires de Sousa, 2009; B, Bernabe-Ortiz, 2016; C, Bernabe-Ortiz, 2018; D, Félix-Martínez, 2018 (2006 data); E, Félix-Martínez, 2018 (2012 data); F, Guerrero-Romero, 2010. NPV, negative predictive value; PPV, positive predictive value.

external validation. For countries (Brazil, Mexico and Peru) where risk scores were generated and validated both cross-sectionally and prospectively, there is enough scientific evidence to implement them as part of the standard of care for type 2 diabetes mellitus screening at the population level.

Limitations of the review

This is a sound methodological review following international guidelines for the systematic reviews of prognosis models [18,19,21,22]. In addition, we used several search engines including ones based in Latin America, hence most, if not all, available evidence should have been retrieved. However, we did not systematically search grey literature,

e.g. dissertations. We argue that this potential limitation would not change our overall findings and conclusions, because these sources would usually not retrieve population-based studies and would have the same or more methodological issues.

Limitations of the selected reports

Most of the reports ascertained the outcome based on fasting glucose, yet one effort in Peru also used oral glucose tolerance test [23].

It could be argued that results based on fasting glucose, or any single biomarker, could lead to underestimation, i.e. some cases might have not been detected. Nonetheless, we



FIGURE 4 Performance metrics of the validation models. Pires de Sousa *et al.* [27] presented these metrics for several thresholds (Appendix S4), the ones shown here correspond to the cut-off point with the sensitivity closest to the pre-specified value. Bernabe-Ortiz *et al.* [24] presented these metrics for several thresholds (Appendix S4), the ones shown here correspond to those with the best Youden Index. A, Pires de Sousa, 2009; B, Bernabe-Ortiz, 2016 (cross-sectional); C, Bernabe-Ortiz, 2016 (longitudinal); D, Bernabe-Ortiz, 2018 (FINDRISC); E, Bernabe-Ortiz, 2018 (LA-FINDRISC); F Bernabe-Ortiz, 2018 (Peruvian); G, Félix-Martínez, 2018 (2006 data); H, Félix-Martínez, 2018 (2012 data); I, Guerrero-Romero, 2010. NPV, negative predictive value; PPV, positive predictive value.

need to acknowledge that these studies were conducted in low- and middle-income countries, sometimes in rural areas, where laboratory facilities to analyse a wider range of biomarkers is limited. In any case, this limitation does not invalidate the results, but rather invites additional investigations to further confirm them using more/other biomarkers. Furthermore, most of the selected reports followed a cross-sectional design, which is not suitable for assessing prognostic models (i.e. long-term outcomes). In addition, the study populations were rather young, which further limits the implementation of the available tools in very young individuals (e.g. adolescent or early adulthood) as well as among the elderly.

There were three main methodological limitations: (i) continuous predictors were categorized, (ii) there was preselection of the predictors, and (iii) some studies included a

limited number of diabetes cases. The first two limitations have been identified as common but suboptimal approaches that hamper the prediction accuracy of the models [28]. Some authors may argue that categorizing continuous predictors helps to make the risk score friendlier thus fostering their use. Whether this argument supersedes the statistical limitations remains unknown. Nevertheless, there are other ways to make the risk scores more accessible such as the use of mobile apps that could include a ‘complex’ algorithm without compromising statistical power. Alternatively, a spreadsheet could accompany the main report as supplementary material, also containing a ‘complex’ algorithm ready to be used. Preselection of predictors was a common practice, following a backward elimination technique [23–25,27]. This could lead to the omission of important predictors that by chance, are not statistically associated with the outcome in the training data

Table 2 Risk of bias assessment of individual diagnostic/prognostic models (PROBAST)

Study; first author, date		Risk of bias				Analysis	Applicability			Overall	
		Objective	Participants	Predictors	Outcome		Participants	Predictors	Outcome	Risk of bias	applicability
Pires de Sousa, 2009 [27]	Derivation	+	+	+	–	+	+	+	?	+	
Pires de Sousa, 2009 [27]	Validation	+	+	+	–	+	+	+	?	+	
Bernabe-Ortiz, 2016 [27]	Derivation	+	+	+	–	+	+	+	?	+	
Bernabe-Ortiz, 2016 [27]	Validation	+	+	+	–	+	+	+	?	+	
Bernabe-Ortiz, 2018 (FINDRISC) [23]	Validation	+	+	+	–	+	+	+	–	+	
Bernabe-Ortiz, 2018 (Latin America-FINDRISC) [23]	Validation	+	+	+	–	+	+	+	–	+	
Bernabe-Ortiz, 2018 (Peruvian) [23]	Validation	+	+	+	–	+	+	+	–	+	
Bernabe-Ortiz, 2018 (Simplified FINDRISC) [23]	Derivation	+	+	+	–	+	+	+	–	+	
Félix-Martínez, 2018 (NHNS-2006) [25]	Derivation	+	?	?	–	+	+	+	–	+	
Félix-Martínez, 2018 (NHNS-2012) [25]	Derivation	+	?	?	–	+	+	+	–	+	
Félix-Martínez, 2018 (NHNS-2006) [25]	Validation	+	?	?	–	+	+	+	–	+	
Félix-Martínez, 2018 (NHNS-2012) [25]	Validation	+	?	?	–	+	+	+	–	+	
Guerrero-Romero, 2010 [26]	Derivation	?	–	–	–	?	–	–	–	–	
Guerrero-Romero, 2010 [26]	Validation	+	–	–	–	+	–	–	–	–	

PROBAST, Prediction model Risk of Bias ASsessment Tool [21,22]; +, low risk of bias/low concern regarding applicability; –, high risk of bias/high concern regarding applicability; ?, unclear risk of bias/unclear concern regarding applicability.

set; moreover, this could lead to over-fitting the risk model [29]. A general recommendation could be to conduct a systematic review of available models in the field to identify the most common and relevant predictors; alternatively, expert knowledge should be included rather than statistical significance alone. Our work could help to overcome this limitation for future studies. We have summarized the most common predictors, so that future efforts could select these instead of ‘sampling’ within a pool of variables available in the data. The number of predictors was small in some derivation models; most importantly, this was also the case in the validation models. It has been suggested that for external validation, at least 100 events should be available [30]. An additional methodological limitation, although one that has little impact in the selected studies, was analysing a complete-case data set, i.e. not conducting methods to account for the missing observations. Multiple imputation techniques still seem to be conducted poorly or not be particularly popular among Latin American health data analysis.

Several metrics for the performance of the risk scores were reported, although calibration estimates were not

available. Calibration is important because it tells us whether the prediction computed by the model agrees with what is actually observed; in other words, a poor calibration could result in overestimation (when the model predicts higher risk than the actual observed risk) or underestimation (when the model predicts lower risk than the actual observed risk) [29,31]. Although Bernabe-Ortiz and colleagues reported the Hosmer–Lemeshow *P*-value [24], further details such as a calibration plot comparing observed vs. predicted cases were missing. The absence of this performance metric but the presentation of other clinically relevant metrics such as sensitivity, specificity and negative/positive predictive values, highlight a need for further training in diagnostic/prediction models analysis. Regarding negative/predictive values, it is relevant to signal that these depend on the underlying prevalence in the population; therefore, these metrics should be interpreted in line with the prevalence estimates and would not be useful to compare prediction models across countries with very different prevalence rates. Given the relevance that risk scores may have in clinical medicine and public health,

strengthening the analytical skills in this field appears to be necessary. Even though friendly technical literature is available [29,32–36], the equivalent in Spanish, the language mostly spoken throughout Latin America, is limited.

Because the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement was published in 2015 [37,38], studies published before that date could not have adhered to this reporting checklist. Studies published after, by contrast, could have adhered to TRIPOD but probably did not because they were unaware of it, suggesting a lack of experience in the field and poor penetration of this statement across professions and regions. Another limitation regarding presentation of the results was the fact that only one study reported the baseline risk, i.e. the intercept of the logistic regression [25]. This parameter is not generally reported and can make it difficult for other researchers to recalibrate these tools for other populations or countries.

Overall, the synthesized prediction models exhibited some methodological limitations. Although these do not invalidate the results, they further support the need to improve this research area in Latin America, for diabetes and other conditions, including several non-communicable diseases. Conducting sound and methodologically robust analyses is key to taking advantage of all the available data and produce better tools that could be easily scaled to clinical medicine and supported by guidelines or policies.

Additional evidence

This review focused on population-based studies with random sampling, although this does not mean that studies following different sampling methods are of little relevance. To develop risk scores, random samples of the general population are not essential. Reports with different sampling criteria have provided valuable information and pragmatic tools for Latin American countries, and thus deserve to be acknowledged as well.

In Mexico, Rojas-Martinez *et al.* using data of a cohort of public and private servants developed a risk score for undiagnosed diabetes; although the external validation was conducted on a population-based sample, this endeavour was not selected for the main synthesis because the model was generated in a closed population [39]. Their score yielded a discrimination of 60% in men and 63% in women, specificity was larger than sensitivity, and the negative predictive value was the largest metric [39]. In addition, the authors compared this new tool with the one currently recommended in Mexico, concluding that the new one performed better [39]. This work signalled that additional research on prediction models benefiting of new and larger data could be useful to improve and update current guidelines.

In Colombia, Barengo and colleagues analysed data of an insurance company to develop a risk score for undiagnosed

type 2 diabetes mellitus [40]. Their model had a discrimination of 74%, slightly higher than the internationally known FINDRISC (73%) [40]. However, this study was not tested externally [40], leaving room for further validation in the general population. Other Colombian researchers have also tested the accuracy of the FINDRISC score, this time using data for people at a primary care facility [41]. They reported a discrimination of 72% in women and 75% in men for undiagnosed type 2 diabetes mellitus [41]; these numbers for incident diabetes were 68% in women and 72% in men [41]. Although these two Colombian experiences should undergo further validation, they signal that available risk scores have a relevant prediction accuracy that could provide valuable tools to improve the early diagnosis of type 2 diabetes mellitus in Colombia.

A recent study in Venezuela also tested the Latin American version of the FINDRISC score, concluding that people above the proposed threshold must have an additional diagnostic test, e.g. oral glucose tolerance test [42]. Because Munoz-Gonzalez *et al.* studied volunteers attending cardiometabolic screening campaigns, further validation is warranted with a larger and more heterogeneous study population.

Central America has contributed poorly to this systematic review, which calls researchers and health officers from this region to conduct studies to develop efficient approaches to early identify people with type 2 diabetes mellitus. However, Milton and collaborators developed a prognostic model benefiting from data of a primary care clinic; their model yielded a discrimination of 89% [43]. Despite the limitations of this work, it is worth acknowledging that the model was intended for rural populations in Honduras, who have been underrepresented to date in the selected and discussed type 2 diabetes mellitus risk scores.

Clinical and public health relevance

Clinical guidelines provide recommendations for type 2 diabetes mellitus screening. The Latin American guidelines, issued by the Latin America Diabetes Association (Asociación Latinoamericana de Diabetes), recommends screening with fasting glucose if a person has one of more risk factors (e.g. overweight, abdominal obesity, family history of type 2 diabetes mellitus) [44]. In addition, if a person is ≥ 45 years old, they should be screened with a fasting glucose test at least once every 5 years, although this could be more often depending on the co-existence of other risk factors [44]. Of note, this guideline also recommends the use of a validated risk score such as the FINDRISC, which could guide the decision on whether or not someone should be screened using fasting glucose [44]. However, about one third of people with undetected type 2 diabetes have normal fasting glucose levels but 2-h postprandial glucose values of > 200 mg/dl. Our work provides evidence on additional type 2 diabetes mellitus risk scores locally developed and validated in Latin

America, thereby this and other guidelines can update their recommendations with strong regional evidence to secure better and more reliable diabetes screening in Latin American populations.

American guidelines also propose screening individuals with risk factors, and suggest using the American Diabetes Association risk test to inform the decision on who should undergo further diagnostic tests (e.g. fasting glucose) [45]. Similarly, the Canadian guidelines propose screening people based on risk factors, or using the Canadian Diabetes Risk Assessment Questionnaire (CANRISK) [46]. Importantly, both, the American and Canadian guidelines, include locally developed and validated risk scores. Furthermore, the Canadian guidelines offer a brief but solid preface on the relevance of using validated risk scores [46]. By contrast, the Latin American guidelines simply suggest use of a ‘validated risk score’. Apparently, this general recommendation was made in the absence of a comprehensive list of available type 2 diabetes mellitus risk scores for Latin American populations. Thus, our systematic review fills this knowledge gap providing scientific evidence to improve regional- and country-based guidelines for the detection of type 2 diabetes mellitus.

It may seem bold to seek one risk score for Latin America as a region, but still worth trying because it could bring great benefits in population screening and disease prevention. With relevant methods, as shown in cardiovascular medicine [47,48], along with support from stakeholders and professional bodies, one or a series of country-specific risk scores could be acceptable and strongly recommended throughout Latin America.

Conclusions

This systematic review of risk scores for the diagnosis and prognosis of type 2 diabetes mellitus could not find compelling evidence to strongly support the use of one single diabetes risk score throughout Latin America. Conversely, there was good evidence to support the use of validated risk scores in Peru and Mexico, whereas further studies need to be conducted with a multi-country or regional scope. Because risk scores could provide additional options to identify type 2 diabetes mellitus cases early, hence decreasing the burden of this disease, the development, validation and implementation of accurate risk scores should be a research and public health priority in Latin America and other low- and middle-income regions.

Funding sources

Strategic Award, Wellcome Trust-Imperial College Centre for Global Health Research (100693/Z/12/Z). Imperial College London Wellcome Trust Institutional Strategic Support Fund [Global Health Clinical Research Training Fellowship] (294834/Z/16/Z ISSF ICL). Rodrigo M Carrillo-Larco is supported by a Wellcome Trust International

Training Fellowship (214185/Z/18/Z). The funder had no role in the conception of this work, neither in the preparation or presentation of results. The authors are responsible for the results and opinions in this work.

Competing interests

None declared.

References

- 1 GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; **392**: 1736–1788.
- 2 GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; **392**: 1859–1922.
- 3 GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; **392**: 1789–1858.
- 4 Carrillo-Larco RM, Barengo NC, Albitres-Flores L, Bernabe-Ortiz A. The risk of mortality among people with type 2 diabetes in Latin America: a systematic review and meta-analysis of population-based cohort studies. *Diabetes Metab Res Rev* 2019; **35**: e3139.
- 5 Bommer C, Sagalova V, Heesemann E, Manne-Goehler J, Atun R, Barnighausen T *et al.* Global economic burden of diabetes in adults: projections from 2015 to 2030. *Diabetes Care* 2018; **41**: 963–970.
- 6 NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet* 2016; **387**: 1513–1530.
- 7 World Health Organization. Noncommunicable Diseases and Mental Health. *NCD Global Monitoring Framework*. Available at https://www.who.int/nmh/global_monitoring_framework/en/ Last accessed 26 August 2019.
- 8 Pan American Health Organization/World Health Organization. Health Policies: Regulatory Measures for Diabetes Prevention and Control. Available at https://www.paho.org/hq/index.php?option=com_docman&view=list&slug=health-policies-regulatory-measures-for-diabetes-prevention-control-8623&Itemid=270&xml:lang=en Last accessed 26 August 2019.
- 9 Abbasi A, Peelen LM, Corpeleijn E, van der Schouw YT, Stolck RP, Spijkerman AM *et al.* Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. *BMJ* 2012; **345**: e5900.
- 10 Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med* 2011; **9**: 103.
- 11 Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. *BMJ* 2011; **343**: d7163.
- 12 Kengne AP, Beulens JW, Peelen LM, Moons KG, van der Schouw YT, Schulze MB *et al.* Non-invasive risk scores for prediction of type 2 diabetes (EPIC-InterAct): a validation of existing models. *Lancet Diabetes Endocrinol* 2014; **2**: 19–29.

- 13 Juarez LD, Gonzalez JS, Agne AA, Kulczycki A, Pavela G, Carson AP *et al.* Diabetes risk scores for Hispanics living in the United States: a systematic review. *Diabetes Res Clin Pract* 2018; **142**: 120–129.
- 14 Alaofe H, Asaolu I, Ehiri J, Moretz H, Asuzu C, Balogun M *et al.* Community health workers in diabetes prevention and management in developing countries. *Ann Global Health* 2017; **83**: 661–675.
- 15 Ballard M, Montgomery P. Systematic review of interventions for improving the performance of community health workers in low-income and middle-income countries. *BMJ Open* 2017; **7**: e014216.
- 16 Perry HB, Zulliger R, Rogers MM. Community health workers in low-, middle-, and high-income countries: an overview of their history, recent evolution, and current effectiveness. *Ann Rev Public Health* 2014; **35**: 399–421.
- 17 Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009; **6**: e1000100.
- 18 Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB *et al.* A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017; **356**: i6460.
- 19 Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG *et al.* Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014; **11**: e1001744.
- 20 Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan – a web and mobile app for systematic reviews. *Syst Rev* 2016; **5**: 210.
- 21 Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS *et al.* PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med* 2019; **170**: W1–W33.
- 22 Wolff RF, Moons KM, Riley RD, Whiting PF, Westwood M, Collins GS *et al.* PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019; **170**: 51–58.
- 23 Bernabe-Ortiz A, Perel P, Miranda JJ, Smeeth L. Diagnostic accuracy of the Finnish Diabetes Risk Score (FINDRISC) for undiagnosed T2DM in Peruvian population. *Prim Care Diabetes* 2018; **12**: 517–525.
- 24 Bernabe-Ortiz A, Smeeth L, Gilman RH, Sanchez-Abanto JR, Checkley W, Miranda JJ; Study Group CC. Development and validation of a simple risk score for undiagnosed type 2 diabetes in a resource-constrained setting. *J Diabetes Res* 2016; **2016**: 8790235.
- 25 Félix-Martínez GJ, Godínez-Fernández JR. Screening models for undiagnosed diabetes in Mexican adults using clinical and self-reported information. *Endocrinol Diabetes Nutr* 2018; **65**: 603–610.
- 26 Guerrero-Romero F, Rodríguez-Morán M. Validación de un instrumento para el tamizaje de casos de diabetes tipo 2 y la vigilancia de personas en riesgo en México. *Rev Panam Salud Publica* 2010; **27**: 181–186.
- 27 Pires de Sousa AG, Pereira A, Costa R, Marquezine GF, Marques do Nascimento-Neto R, Freitas SN *et al.* Derivation and external validation of a simple prediction model for the diagnosis of type 2 diabetes mellitus in the Brazilian urban population. *Eur J Epidemiol* 2009; **24**: 101–109.
- 28 Steyerberg EW, Uno H, Ioannidis JPA, van Calster B. Poor performance of clinical prediction models: the harm of commonly applied methods. *J Clin Epidemiol* 2018; **98**: 133–143.
- 29 Riley RD, van der Windt D, Croft P, Moons KGM. *Prognosis Research in Healthcare Concepts: Methods, and Impact*. Oxford: Oxford University Press, 2019.
- 30 Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016; **35**: 214–226.
- 31 Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ *et al.* Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA* 2017; **318**: 1377–1384.
- 32 Grant SW, Collins GS, Nashef SAM. Statistical Primer: developing and validating a risk prediction model. *Eur J Cardiothorac Surg* 2018; **54**: 203–208.
- 33 Wynants L, Collins GS, Van Calster B. Key steps and common pitfalls in developing and validating risk models. *BJOG* 2017; **124**: 423–432.
- 34 Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG *et al.* Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)-marker. *Heart* 2012; **98**: 683–690.
- 35 Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG *et al.* Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012; **98**: 691–698.
- 36 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014; **35**: 1925–1931.
- 37 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Diabet Med* 2015; **32**: 146–154.
- 38 Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; **162**: W1–W73.
- 39 Rojas-Martínez R, Escamilla-Núñez C, Gomez-Velasco DV, Zarate-Rojas E, Aguilar-Salinas CA. [Development and validation of a screening score for prediabetes and undiagnosed diabetes.]. *Salud Publica Mex* 2018; **60**: 500–509.
- 40 Barengo NC, Tamayo DC, Tono T, Tuomilehto J. A Colombian diabetes risk score for detecting undiagnosed diabetes and impaired glucose regulation. *Prim Care Diabetes* 2017; **11**: 86–93.
- 41 Gomez-Arbelaiz D, Alvarado-Jurado L, Ayala-Castillo M, Forero-Naranjo L, Camacho PA, Lopez-Jaramillo P. Evaluation of the Finnish Diabetes Risk Score to predict type 2 diabetes mellitus in a Colombian population: a longitudinal observational study. *World J Diabetes* 2015; **6**: 1337–1344.
- 42 Muñoz-González M, Lima-Martínez M, Nava A, Trerotola G, Paoli M, Cabrera-Rego JO *et al.* FINDRISC modified for Latin America as a screening tool for persons with impaired glucose metabolism in Ciudad Bolívar, Venezuela. *Med Princ Pract* 2019; **28**: 324–332.
- 43 Milton EC, Herman WH, Aiello AE, Danielson KR, Mendoza-Avelarez MO, Piette JD. Validation of a type 2 diabetes screening tool in rural Honduras. *Diabetes Care* 2010; **33**: 275–277.
- 44 Association Latinoamericana de Diabetes. Guías ALAD sobre el Diagnóstico, Control y Tratamiento de la Diabetes Mellitus Tipo 2 con Medicina Basada en Evidencia, Edición 2013. Available at https://issuu.com/alad-diabetes/docs/guias_alad_2013 Last accessed 26 August 2019.
- 45 American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes – 2019. *Diabetes Care* 2019; **42**: S13–S28.

- 46 Diabetes Canada Clinical Practice Guidelines Expert Committee. Diabetes Canada 2018 clinical practice guidelines for the prevention and management of diabetes in Canada. *Can J Diabetes* 2018; 42(Suppl 1): S1–S325.
- 47 Hajifathalian K, Ueda P, Lu Y, Woodward M, Ahmadvand A, Aguilar-Salinas CA *et al.* A novel risk score to predict cardiovascular disease risk in national populations (GloboRisk): a pooled analysis of prospective cohorts and health examination surveys. *Lancet Diabetes Endocrinol* 2015; 3: 339–355.
- 48 Ueda P, Woodward M, Lu Y, Hajifathalian K, Al-Wotayan R, Aguilar-Salinas CA *et al.* Laboratory-based and office-based risk scores and charts to predict 10-year risk of cardiovascular disease in 182 countries: a pooled analysis of prospective cohorts and health surveys. *Lancet Diabetes Endocrinol* 2017; 5: 196–213.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Search terms used in MEDLINE, Embase and Global Health (through Ovid).

Appendix S2. Search terms used in Scopus.

Appendix S3. Search terms used in LILACS.

Appendix S4. Data extraction by domain.

Appendix S5. Risk of bias assessment (PROBAST).