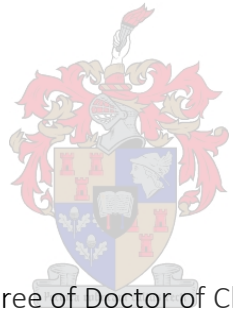
# Methodological issues around the validation of models for predicting diabetes risk in developing countries

Katya Laura Masconi



Dissertation presented for the degree of Doctor of Chemical Pathology in the Faculty of Health Sciences at Stellenbosch University

Supervisors:

Professor Andre Pascal Kengne

Professor Rajiv T Erasmus

Professor Tandi E Matsha

December 2016

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

This dissertation includes 5 original papers published in peer reviewed journals and 1 unpublished publication. The development and writing of the papers (published and unpublished) were the principal responsibility of myself and, for each of the cases where this is not the case, a declaration is included in the dissertation indicating the nature and extent of the contributions of co-authors.

Date: December 2016

ABSTRACT

Background: Multivariable diabetes risk prediction models have the potential to contribute to screening strategies, combining several risk factors to predict undiagnosed diabetes or future risk of developing diabetes. The focus of this study is the prediction of undiagnosed diabetes and diabetes risk prediction in a developing country where no population-specific diabetes risk prediction model currently exists. Existent models have been developed in unrelated populations with different disease prevalence, predictor weightings and methods used for risk factor determination and diabetes diagnosis. For accurate diabetes risk prediction in the mixed ancestry population of Bellville South, Cape Town, methodological issues regarding the validation and performance of these models needs to be addressed.

Methodology: Cross-sectional data from the Cape Town Bellville South cohort was used for this study. Missing data in risk prediction research was investigated through a systematic review and a number of imputation methods were explored to deal with missing data in this dataset. Models were identified via recent systematic reviews and validated in the mixed-ancestry population. Discrimination was assessed and compared using the C-statistic and calibration was assessed via calibration plots. Model recalibration in diabetes risk prediction was investigated through a systematic review. In an effort to improve model performance in the new setting, model recalibration and updating strategies were used and performance was compared before and after implementation.

Results: The study sample consisted of 1256 individuals, of whom 173 were excluded due to previously diagnosed diabetes. Of the final 1083 individuals, 329 (30.4%) had missing data. Deletion resulted in the lowest model performance and simple imputation, the simplest method, resulted in the highest model performance and was employed for further analysis. A systematic review highlighted the gross underreporting and mishandling of missing data in diabetes risk prediction research. Original model performance during validation was poor-to-average, with both over- and underestimation present: Cambridge [C-statistic: 0.67 (0.62-0.72); E/O: 1.81 (1.09-2.52)], Kuwaiti [C-statistic: 0.68 (0.63-0.73); E/O: 0.72 (0.43-1.12)], Omani [C-statistic: 0.66 (0.61-0.70); E/O: 1.28 (0.63-1.93)], Rotterdam [C-statistic: 0.64 (0.59-0.69); E/O: 0.54 (0.50-1.04)] and Simplified Finnish [C-statistic: 0.67 (0.62-0.71); E/O: 0.26 (0.13-0.39)] diabetes risk prediction models. Recalibration, as shown through a systematic

review, was undertaken only in models predicting incident diabetes, and was reported in 22.9% of validation studies, with 77.8% achieving an increase in model performance. Updating results applied to this validation dataset showed an increase in both discrimination and calibration in varying levels across all five models. Overall, the re-estimation of the Cambridge diabetes risk model yielded the best model performance [C-statistic: 0.71 (0.67 – 0.75); E/O: 1.00 (0.86 – 1.17)].

Discussion and conclusion: The frequency of missing data, underreporting and mishandling of missing data, complexity of updating methods and overall model performance of validated models in new settings highlight the challenges in diabetes risk prediction research. This is the first validation study of prevalent diabetes risk prediction models in Sub-Saharan Africa and highlighted important methodological issues. While both simpler imputation and updating methods resulted in similar predictive utility when compared to more complex techniques, model performance was not increased sufficiently to suggest recommendation.

Word count: 500

OPSOMMING

Agtergrond: Diabetesrisikobeoordeling het na vore gekom as ’n eenvoudige manier om intervensiestrategieë af te stem op diegene wat asimptomaties is dog aan ongediagnoseerde diabetes ly, of groot gevaar loop om diabetes te ontwikkel. Meerveranderlike-risikovoorspellingsmodelle kan tot hierdie siftingsproses bydra deur verskeie risikovoorspellers wat in die uitkomsvoorspelling gebruik word te kombineer. Hierdie studie konsentreer op die voorspelling van ongediagnoseerde diabetes en diabetesrisiko in ’n ontwikkelende land waar daar tans geen populasiespesifieke model vir die voorspelling van diabetesrisiko bestaan nie. Bestaande modelle is ontwikkel in nieverwante populasies met verskillende metodes vir risikofaktorbepaling en diabetesdiagnose, die bepaling van siektevoorkoms en die beswaring van voorspellers. Vir akkurate diabetesrisikovoorspelling onder die veelrassige inwoners van Bellville-Suid, Kaapstad, moet die geldigheid en prestasie van hierdie modelle aandag ontvang.

Metodologie: Deursneedata uit die kohort Bellville-Suid, Kaapstad, is vir hierdie studie gebruik. Ontbrekende data in risikovoorspellingsnavorsing is deur middel van ’n stelselmatige oorsig ondersoek, en ’n aantal toerekeningsmetodes is verken om ontbrekende data in hierdie datastel te hanteer. Modelle is deur middel van onlangse stelselmatige ondersoeke geïdentifiseer, en die geldigheid daarvan is onder die veelrassige bevolking bepaal. Diskriminasie is met behulp van C-statistiese en nieparametriese metodes beoordeel en vergelyk, en kalibrering is met kalibreringsgrafieke beoordeel. Om modelprestasie in die nuwe studieomgewing te verbeter, is modelherkalibrering en bywerkingstrategieë gebruik. Modelherkalibrering in diabetesrisikovoorspelling is eers deur ’n stelselmatige oorsig van gepubliseerde geldigheidstudies ondersoek. Daarna is bywerkingstrategieë in hierdie studiepopulasie in werking gestel en is prestasie voor en na inwerkingstelling vergelyk.

Resultate: Die steekproef van die studie het uit 1 256 individue bestaan, van wie 173 weens voorheen gediagnoseerde diabetes uitgesluit is. Van die uiteindelike 1 083 individue, het 329 (30,4%) ontbrekende data gehad. Weglating het tot die laagste modelprestasie gelei, en die eenvoudigste toerekeningsmetode wat die hoogste modelprestasie tot gevolg gehad het, is vir verdere ontleding gebruik. ’n Stelselmatige oorsig het erge onderrapportering en verkeerde hantering van ontbrekende data in navorsing oor diabetesrisikovoorspelling aan

die lig gebring. Oorspronklike modelprestasie gedurende geldigheidsbepaling was gemiddeld, en sowel oor- as onderraming het voorgekom in die diabetesrisikovoorspellingsmodelle van Cambridge [C-statistiek: 0.67 (0.62-0.72); E/O: 1.81 (1.09-2.52)], Koeweit [C-statistiek: 0.68 (0.63-0.73); E/O: 0.72 (0.43-1.12)], Oman [C-statistiek: 0.66 (0.61-0.70); E/O: 1.28 (0.63-1.93)], Rotterdam [C-statistiek: 0.64 (0.59-0.69); E/O: 0.54 (0.50-1.04)] en Finland (vereenvoudig) [C-statistiek: 0.67 (0.62-0.71); E/O: 0.26 (0.13-0.39)]. Herkalibrering, wat slegs onderneem is in modelle wat nuwe diabetesgevalle (insidensie) voorspel, is in 22,9% van geldigheidstudies gerapporteer, en 77,8% het ’n toename in modelprestasie getoon. Bywerkingsresultate wat op hierdie geldigheidsdatastel toegepas is, toon ’n toename in diskriminasie sowel as kalibrering op wisselende vlakke oor ál vyf modelle. Oor die algemeen het die herraming van die Cambridge-diabetesrisikomodel die beste modelprestasie opgelewer [C-statistiek: 0.71 (0.67 – 0.75); E/O: 1.00 (0.86 – 1.17)].

Bespreking en gevolgtrekking: Die frekwensie van ontbrekende data, die onderrapportering en verkeerde hantering van ontbrekende data, die kompleksiteit van bywerkingsmetodes sowel as die algehele modelprestasie van geldige modelle in nuwe studieomgewings beklemtoon die uitdagings van navorsing oor diabetesrisikovoorspelling. Hierdie studie is die eerste geldigheidstudie van bestaande modelle vir diabetesrisikovoorspelling in Afrika suid van die Sahara. Hoewel eenvoudiger toerekening- en bywerkingsmetodes soortgelyke voorspellingsnut as meer komplekse tegnieke tot gevolg gehad het, het modelprestasie nie soveel verbeter dat dit aanbeveling regverdig nie.

Woorde: 519

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisors Professor Kengne, Professor Erasmus and Professor Matsha for their guidance throughout PhD.

Andre, thank you for your patience, motivation, and immense knowledge. Your guidance has encouraged me and allowed me to grow as a researcher.

Tandi, thank you for your continuous support, both professionally and personally. You have been a mentor that every student would be lucky to have. Your encouragement and belief in me has allowed me to get to this point of final hand in. I am forever grateful to you for taking me in as a very naïve Masters student in 2011.

Dinie, thank you for the many professional chats (and gossip sessions), I am always made to feel like I have never left when I come into the CPUT offices. Your no-nonsense attitude has made it easy for me offload all my student problems, the fellow PhD friend I never had.

Many thanks also goes to entire Bellville South team, without who, this thesis would not be possible. Additionally, thank you to the National Research Foundation and The Carl & Emily Fuchs Foundation for funding me during my PhD.

A special thanks to my support system, Emma, Kelly and Amy. Thank you for always being there, for standing by me through the storm. I am forever grateful.

Jason, thank you for absolutely everything. You are my rock.

Mommy, no words can describe how much I owe to you for where I am today. I love you

TABLE OF CONTENTS

# 01

# Introduction

## General introduction

Diabetes mellitus, type 2 diabetes in particular, is a growing epidemic worldwide with developing countries currently paying the highest toll. In addition to improving the detection of those who already suffer from the disease, strategies are needed to improve the detection of those at risk of diabetes such that treatment and prevention measures can be implemented to prevent or delay the onset of the disease and related complications. The use of multivariable risk prediction models has been advocated as a practical and potentially affordable approach for improving the detection of undiagnosed diabetes or screening for future risk of diabetes. Accordingly, guidelines increasingly promote the use of prediction models and derivatives for diabetes risk screening around the world. Despite the availability of many prediction models, a 'one size fits all' approach is unrealistic, and no single model developed from one population can perform well in all other settings. Consequently, many parts of the world, particularly developing countries, have to rely on prediction models developed in other populations. Issues relating to differences in case-mix across populations, inherent to the development of models, can severely affect the applicability of a model in different settings. It is therefore important to identify these issues and attempt to address them with efficient methods, to ensure that diabetes prediction models developed in other populations are used with increased accuracy to improve diabetes prevention and detection in developing countries. Those issues so far have not been investigated in the context of South Africa and Sub-Saharan Africa in general. Thus, this study aims to investigate and address methodological issues which may affect the applicability of existent diabetes risk prediction models in a mixed ancestry population in South Africa. The focus of this study is diabetes risk prediction, in particular the prediction of undiagnosed diabetes. We will use the case of mixed-ancestry population of South Africa as the application population. But issues discussed here, mutatis mutandis, apply to developing countries in general. The literature review will comprise of current diabetes screening information and an expert review on diabetes risk prediction.

## Research questions

1. How do existing undiagnosed diabetes risk prediction models perform in mixed-ancestry South Africans?
2. What are the methodological issues surrounding prediction model validation of diabetes mellitus in developing countries?
3. Can these issues be resolved without the need for deriving a context specific model?

## Objectives

1. Review information on risk predictive modelling in the context of incident and prevalent diabetes
2. Validate existent prevalent diabetes risk prediction models in the Bellville South study participants
3. Review the reporting and handling of missing data in predictive research for prevalent undiagnosed type 2 diabetes mellitus
4. Apply multiple missing data imputation techniques to the Bellville South dataset and investigate the effect these had on the performance of undiagnosed diabetes risk prediction models
5. Review the extent to which recalibration of risk prediction models is undertaken in validation studies of diabetes risk prediction models
6. Apply multiple model updating strategies to the Bellville South dataset and investigate the effect these had on the performance of undiagnosed diabetes risk prediction models

## Hypothesis

Prediction models developed elsewhere generate inaccurate estimates of undiagnosed diabetes risk among mixed-ancestry adult South Africans, which can be substantially improved by efficient application of simple improvement procedures.

## Significance of the study

The burden of diabetes in South Africa is a reality and the issues surrounding diabetes detection and prevention need to be addressed. With mass screening impractical due to cost and resources, the use of multivariable risk models in opportunistic clinical risk assessments is recommended by various guidelines. Diabetes mellitus is a multifaceted disease, with a wide range of attributes that affect the risk of onset. Risk prediction models have the capacity to handle this complexity and numerous models do already exist that predict undiagnosed and future diabetes using clinical information and / or laboratory measurements. These do unfortunately have their limitations, including low utilization rates due to the inclusion of OGTT and complex physician-calculations, and thus an improved method for diabetes risk and diagnosis assessment based on a variety of factors in a format that facilitates use in routine clinical practice is essential.

Despite these numerous prediction models, very few are practically implemented and many have not been externally validated in alternative population groups. Additionally, there is a growing concern that many of the current risk prediction models are poorly developed as they are based on a small and inappropriate selection of the cohort, questionable handling of the continuous risk predictors, inappropriate treatment of missing data, use of flawed or unsuitable statistical methods and ultimately, lack of transparent reporting in the steps taken to derive the model. Methodological issues relating to model development and generalizability need to be better addressed by prediction researchers to help strengthen models and aid in implementation, the true aim of risk prediction models. The prediction of undiagnosed diabetes in the mixed ancestry population of South Africa aims to identify these methodological issues via validation and assessment of common models, which has not been addressed in the Sub-Saharan region of Africa. The possible model enhancement by efficient application of improvement procedures will allow for a model that performs well in the prediction of undiagnosed diabetes risk prediction specific to this community. A suitable final model may be achieved which would improve risk assessment and enable more people to be evaluated and at-risk individuals to be more efficiently identified when implemented. Alternatively, this study will identify the methodological issues involved in diabetes risk

prediction studies, and encourage and aid future research into models specific for South African populations.

It is clear there is a great clinical need for a robust and convenient tool for identifying undiagnosed diabetes and predicating future diabetes easier, quicker and more economically. In the long term, these models can be useful in the implementation of healthcare interventions, lifestyle changes or diabetes preventative programs, specific to this community, which can be targeted towards those at an increased risk of future disease development. While this may seem a far reach for developing countries, screening in these settings will never be considered without the availability and recommendation of affordable methods, something risk prediction models offer. Yet no diabetes prediction research in this region has been undertaken. This study wishes to initiate this, with a methodological approach deemed the best place to start. This study aims to validate existing undiagnosed diabetes risk prediction models in the mixed ancestry population of South Africa while identifying and addressing the methodological issues encountered during the validation process.

# PART I

# BACKGROUND

02

# Screening for type 2 Diabetes

## Epidemiology of diabetes

Type 2 diabetes mellitus is a global health problem with prevalence rates increasing rapidly. In 2015, approximately 400 million individuals suffered from type 2 diabetes, and future predictions indicate that this number will surge to approximately 640 million by 2040 [1]. This rapid rise of diabetes will result in an even greater and more profound burden that developing countries are not equipped to handle. Diabetes is often associated with increased cardiovascular risk and renal complications, which leads to increased morbidity and increased mortality [2]. Diabetes is no longer a condition of the developed, 'industrialised' or 'Western' countries. With a 10 - 16% prevalence of diabetes in South Africa, a 10.4% prevalence for self-reported hypertension, a prevalence as high as 78% in adult South Africans of unknown hypertensive status; and 29.2% and 56.6% overweight or obese men and women, respectively, chronic diseases and subsequently cardiovascular diseases are a startling reality in this country [3-6]. Type 2 diabetes in developing countries is further characterized by the high proportion of people who are undiagnosed, as well as a large number of those without the disease, but who are at risk of developing the condition in the future. The most recent population-based studies have reported age-standardized prevalence rate of diabetes as high as 26% and 13% respectively among mixed-ancestry [7] and Black [8] adult South Africans. The former study indicated that two thirds of the individuals with the disease were undiagnosed at the time of the study [9]. Furthermore, estimates at a national level suggest that the population of adults at risk of developing diabetes in South Africa matches the number of those already living with the disease [10]. This situation is likely to be similar in many other developing countries in Africa and beyond.

The global burden of diabetes can be reduced with prevention, diagnosis and treatment strategies. Diabetes is a chronic disease, largely the result of excess body weight and physical inactivity. Type 2 diabetes symptoms are often asymptomatic, with the International Diabetes Federation (IDF) reporting that up to 62% of people living in Africa are unaware of their condition [11]. The diagnosis may only manifest several years after onset, once complications have already arisen, and life expectancy has been significantly reduced. On average, diabetes reduces life expectancy by 7.5 years in men and 8.2 years in women [12]. Over time, diabetes

damages the heart, blood vessels, eyes, kidneys and nerves. Ensuing cardiovascular diseases, retinopathy, chronic kidney failure and diabetic neuropathy, increase the risk of death by up to 50% in an individual with diabetes when compared to a non-diabetic [13]. Detecting individuals at future risk of diabetes allows for the implementation of education on lifestyle measures that are effective in preventing or delaying the onset of diabetes. Detecting individuals with undiagnosed diabetes allows for effective treatment to control blood glucose levels, preventing or postponing the onset of complications. Interventions in developing countries are both cost saving and feasible, and all efforts should be made to aid this process.

## Screening of undiagnosed diabetes

The purpose of screening is to differentiate an asymptomatic individual at high risk from an individual at low risk for diabetes. This is largely different to diagnostic testing where an individual with visible signs and symptoms is diagnosed. There are various methods for screening, including risk assessment questionnaires, portable capillary blood assessments and laboratory based tests. The ideal screening test should be rapid, simple, safe, reliable, reproducible and affordable [14-17]. Additionally, a screening test should be both highly sensitive, the high probability of detecting or predicting a positive outcome when the subject truly has the disease, and highly specific, a high probability of being negative when the subject does not have the disease. Screening is deemed appropriate when the following seven conditions are met [18-22]:

1. The disease represents an important health problem that imposes a significant burden on the population

Diabetes is currently listed as the seventh highest cause of death worldwide, with gross under-reporting being acknowledged [23]. This is even more profound in minority populations. The nations with the highest prevalence of diabetes are mainly low and middle income countries. By 2030, there is a projected increase of adults diagnosed with diabetes of at least 57% in these low and middle income countries, compared to an only 25% increase in high income countries [24]. Most apt for this study, the greatest proportional increase (90.5%) will occur in sub-Saharan Africa. Diabetic individuals consume health care resources at two to

three times the rate of a non-diabetic patient [25]. This is, in large, due to the major complications of diabetes, namely visual impairment and blindness, end-stage renal disease and contributes to cardiovascular disease, stroke, disability, peripheral vascular disease and premature mortality [26-29]. Additionally, diabetes is a component cause of several non-communicable diseases such as pneumonia [30], bacteraemia [31, 32] and tuberculosis [33], considerably impacting morbidity and mortality in Sub-Saharan African [34-39]. This double burden of disease, along with limited resources, is an important health problem in this region.

**2.** The natural history of the disease is understood

The natural history of type 2 diabetes is well defined, where biological onset due to impaired insulin action or relative insulin deficiency is followed by a phase in which the disease remains undiagnosed and largely asymptomatic [40]. The underlying loss of β-cell function can, in part, be due to factors such as sustained elevated glucose and lipid levels, inflammation, amyloid deposition, and oxidative and endoplasmic reticulum stress [41, 42]. There is an initial postprandial hyperglycemia and subsequent fasting hyperglycemia. As the hyperglycemia increases, the natural course of diabetes involves the development of complications, dependent on the duration and degree of the elevated glucose levels, and ultimately, death [43]. Prior to the onset of diabetes mellitus, hyperglycemia presents in the form of impaired glucose tolerance (IGT) or impaired fasting glucose (IFG), controversially referred to as 'pre-diabetes', which may or may not develop into diabetes. Individuals with IFG have a 20-30% chance of developing diabetes over a 5 – 10 year period. This risk increases if IFG is combined with IGT [44].

**3.** There is a recognizable preclinical (asymptomatic) stage during which the disease can be diagnosed

The asymptomatic pre-diabetes phase has an estimated duration of 8.5-10.3 years and the clear preclinical asymptomatic diabetic stage may be present for up to seven years before diagnosis [45, 46]. This long, latent period can be detected using routine diagnostic tests used in the diagnosis of symptomatic and suspected patients [47]. Importantly, using the same diagnostic criteria as for symptomatic cases, diabetes can be detected in the preclinical stage. The diagnosis of diabetes during this phase should be not be made on a single abnormal

glucose result, but rather confirmed with a random blood sample or oral glucose tolerance test (OGTT) [11]. Both IFG and IGT are conditions characterized by glucose levels higher than normal but lower than the diagnostic cut-point for diabetes, and which can be detected and diagnosed. In addition to the increased risk of diabetes and its complications during this stage, complications are certainly not uncommon even in these early stages of diabetes [47-51].

4. Treatment after early detection yields benefits superior to those obtained when treatment is delayed

The benefits and risks of screening are assessed by comparing short- and long term health outcomes. Many screening analysis studies have been performed however; in-depth studies comparing treatments in a screen and control group have not been undertaken and are unlikely to happen due to feasibility, ethical concerns, and costs [52, 53]. Due to the recommendation of screening by several health organizations, no diagnosis and treatment in the control group once symptoms have already been identified, is seen as unethical. However, although individuals diagnosed with diabetes via screening typically have glucose levels that warrant treatment, evidence supports the benefits of improved glycemic control in type 2 diabetes [54-58]. Additionally, the United Kingdom Prospective Diabetes Study (UKPDS) provided evidence that up to 50% of the newly diagnosed individuals, whether detected conventionally or through screening, had chronic diabetes complications at the time of diagnosis [59-61]. Two large trials have been instrumental in highlighting the benefits of early treatment of diabetes and the related complications [62, 63]. The STENO-2 study showed that long-term and intensified intervention for modifiable risk factors, with conventional diabetes treatment, reduced the risk of cardiovascular and microvascular events in roughly 50% of cases [63]. The ADDITION study relied on screen-detected diabetes only, and although non-significant, results showed a reduction in the incidence of cardiovascular events and death [62].

Data from intervention studies have also been useful in highlighting the effectiveness of diabetes and risk factor treatment. Studies conducted in conventionally diagnosed diabetic populations have compared the effects of individual treatment to lower blood glucose [58], blood pressure [64], and serum cholesterol [65, 66], as well as lifestyle modification [67].

These results can inform early treatment strategies. Additionally, evidence concerning the effect of interventions in IGT has also been established, in particular, interventions aimed at weight reduction, increased physical activity and the use of some pharmacological agents have been shown to be effective in reducing or delaying the transition to diabetes in those with IGT [68]. As most studies based on diabetes treatment have been conducted in Caucasians, and given the increased risk that certain populations in developing countries have to diabetes-related complications and increased progression rate from pre-diabetes to diabetes, this prevalence is likely higher in the mixed-ancestry population in South Africa [69]. With an early detection, patients and medical practitioners can take preventative actions or begin treatment to prevent or delay complications and mortality.

5. Tests are available that can detect the preclinical stage of disease, and the tests are acceptable and reliable

There are two major methods used to screen for preclinical asymptomatic type 2 diabetes: risk prediction through the use of scores or questionnaires, and biochemical tests. Risk prediction is used to assign a person to a higher or lower risk group for diabetes by obtaining self-reported demographic, behavioral, and medical information. Questionnaires or risk scores are popular, better suited for prevalent diabetes screening and less expensive than biochemical tests and will be discussed in detail in Chapter 3. Biochemical measurements of glucose and highly correlated metabolites (e.g., HbA1c and fructosamine levels) have been used extensively for diabetes screening. They may be based on metabolic states including fasting, random, postprandial and glucose load via urine, venous or capillary glucose measurements.

*Urine glucose:* Urine glucose has a low sensitivity, ranging between 21% and 64%, and a low PPV, so despite a specificity of > 98% it is deemed an inappropriate screening test [40, 68]. However, it may have a place in low resource areas where no other test is possible. This may be useful when the prevalence of undiagnosed diabetes is high.

*Random blood glucose*: RBG testing is easy to obtain but also has a lower specificity, limiting it as a screening tool. The cut-off point for RBG is dependent on population characteristics, with a lower value resulting in a significantly lower sensitivity [70].

*Fasting plasma glucose*: FPG is highly correlated with the risk of diabetes complications. The sensitivity ranges between 40% and 65% with a specificity > 90%, this specificity drops to 85.2% when the cut-off is dropped to the recommended optimal cut-off of 7.0mmol/L [68, 70]. However, there is no globally agreed upon cut-off point, and it should be determined via further testing in each population. This test is therefore not as sensitive as the OGTT in IGT identification.

*Glycated haemoglobin*: $HbA_{1c}$ was evaluated as a hyperglycemic test with the desire of replacing the OGTT with a simpler test [71-73]. $HbA_{1c}$ is a stable marker of long-term glycemic level, making it an appealing screening tool. It does not require a fasting sample, may be done on a capillary sample from point-of-care testing and the intra-individual variability is lower than fasting plasma glucose [74-76]. The use of $HbA_{1c}$ for the diagnosis of diabetes has been adopted by the ADA and the WHO, however the WHO recommends $HbA_{1c}$ for diagnosing diabetes as an alternative to plasma glucose measurements only if stringent quality assurance tests are in place and there are no conditions present which prelude the accurate measurement. Both organizations define the cut-off as 6.5%, however a value of less than 6.5% does not exclude diabetes that has been diagnosed using glucose tests [70, 77]. This value was based on the level of $HbA_{1c}$ after which the incidence of retinopathy, a common complication of diabetes that often is present before the actual diagnosis is made, is increased [78]. Unfortunately grey areas exist around this cut-off, and has been shown to be ethnicity dependent [77, 79]. A study conducted in the mixed-ancestry population used in this study found that a level of 6.1% was optimal in diagnostic testing for all age groups [80]. Additionally, $HbA_{1c}$ can be affected by a number of conditions that may be more prevalent in developing countries, including hemoglobinopathies, iron deficiency and chronic kidney disease [81, 82]. The need for global standardization, unavailability of the test and the cost of $HbA_{1c}$ testing still needs to be addressed [83].

*75g oral glucose tolerance test*: The 2-h glucose concentration from the OGTT has served as the reference standard for diabetes screening and diagnosis. It is also the only method that is used to formally detect or diagnose IGT. Due to the practical downsides of the required 8-hour fast before the test, the length of the test and the commitment of the staff administering the test, this test is not the favored screening test [70].

Combinations of biochemical tests in series (with second and subsequent screening tests performed only when the preceding test is positive) can enhance the positive predictive value (PPV) by increasing the prevalence of disease in the population receiving the second screening test. Screening programs can initially use a less expensive and more sensitive test and then use the more complicated, more specific, and more expensive test (e.g. a questionnaire followed by capillary glucose measurement). Strategies that use multiple screening tests will not however, detect more undiagnosed cases (i.e. will not improve sensitivity) [40]. Multiple screening tests, specifically screening tests that involve invasive testing, are not recommended for undiagnosed diabetes screening programs, as this defeats the purpose of the screening methodology.

6. The costs of case finding and treatment are reasonable and are balanced in relation to health expenditures as a whole, and facilities and resources are available to treat newly detected cases

There is limited information concerning the cost of screening, and the cost is dependent on the setting and the screening approach. Three approaches to diabetes screening that have been used are population-based, selective and opportunistic screening. Population based screening is done at a community level and aims to screen each and every person. This is unrealistic, costly and is only potentially efficient in a high diabetes prevalence population. Selective screening targets subgroups of a population where diabetes and its associated risk factors prevalence is high [40, 84, 85]. Finally, opportunistic screening involves screening individuals who come into contact with the health care system or at selected mass gatherings [86]. Selective and opportunistic screening require fewer resources to reach those who are considered high risk but both have poor coverage and have too little control of the number of tests each individual receives – some people are tested too often with too many tests while others miss testing all together [86].

The demand on resources exists in the screening and diagnosis process as well as the additional years of care due to early diagnosis. On the surface, these costs may outweigh the lifetime costs of an individual detected through the current practice, however with the likelihood of complications higher in the latter; these additional costs may rapidly overtake

those calculated in early detection and long-term treatment and management. According to the WHO, diabetes screening programmes, started between the ages of 30 and 45, with screening repeated every 3-5 years (every year for high risk groups), have been recognized to be cost-effective measures [87]. They provide the opportunity to target risk reduction interventions to high risk populations and to reduce the burden of diabetes and its complications through early identification and treatment. These measures should be supported by a healthy diet, regular physical activity, maintenance of a normal body weight and avoidance of tobacco. On the contrary, a number of studies calculating the lifetime costs of diabetes treatment in screening versus none found that the cost was higher with screening [88, 89]. However, the cost per life-year gained and the cost per quality-adjusted life-year (QALY) were significant. Greater benefits and more favorable cost-effectiveness ratios were found if screening was conducted for younger compared with older people (because younger people lived longer with diabetes and had great reductions in lifetime complications) and for African-Americans compared with the general population (primarily because of the higher complication rates among African-Americans). How diabetes screening complements efforts to control other diseases should also be considered. Screening for diabetes can be combined with efforts to detect other conditions, such as hypertension and dyslipidemia [52, 90-92].

**7.** Screening will be a systematic ongoing process and not merely an isolated one-time effort

The high rates of undiagnosed diabetes highlights the nonexistence or ineffectiveness of screening practices in the Sub-Saharan Africa region. To fully address the problem of undiagnosed diabetes, screening programs should be ongoing. For this to occur there needs to be an accurate and simple process of screening. Ideally, opportunistic screening should be made part of routine care, conducted in clinical settings at designated, regular intervals.

Despite population-based and selective screening programs in a community setting demonstrating low yield and poor follow-up with high costs, the periodic screening of high-risk individuals is warranted. Additionally, opportunistic screening can be considered within the healthcare setting. Questions remain about the optimal screening methods and how often screening should be carried out; and the best method for outcome diagnosis with accurate cut-off points for a positive test in each population. However despite the lack of firm

evidence, several health organizations have recommended it for several reasons [68, 93-96]. The already large and continuously growing burden of diabetes is condition enough that screening of diabetes should be supported. The IDF screening and diagnosis recommendations suggest that each health service should decide on programmes to detect undiagnosed diabetes based on the prevalence and the resources available in that region [97]. In areas with limited care, such as developing countries, the detection programmes are suggested to be opportunistic and limited to high-risk individuals. The detection method advocated is a risk assessment questionnaire to identify those individuals at high risk and then use glycemic tests to detect diabetes in these selected individuals. Taking into account the lack of resources or in the least, the misuse of resources, in developing countries, the WHO propose the reorganization of resources to allow for greater availability of supplies for screening and community follow-up [68].

Countries may also implement their own screening guidelines as a strategy to reduce household, public and economic costs. As a developing country, South Africa ranks as a Newly Industrialized Country (NIC) along with India, China, Turkey and Malaysia. Diabetes is a major health problem in India and this country has perhaps implemented and researched undiagnosed diabetes screening the most earnestly. The introduction of the Prevention Awareness Counseling and Evaluation (PACE) Diabetes Project is to date the largest diabetes and non-communicable disease awareness screening and prevention project in India [98]. There were three areas that were focused on, namely the increase of knowledge and awareness of diabetes, the large scale opportunistic blood glucose screening to identify undiagnosed diabetes and finally a community based prevention program in selected communities. This project found screening cost effective and several organizations within India have adopted the PACE diabetes model. Although not without its challenges, this project could serve as a model for similar programs in comparable developing countries. The government of India's national program for Prevention and Control of Diabetes / Cardiovascular Disease and stroke has subsequently implemented the opportunistic screening of all persons over the age of 30 including pregnant women of all ages.

The only South Africa-specific diabetes screening guideline was released by the Society for Endocrinology, Metabolism and Diabetes of South Africa (SEMDSA) in 2012 and follows the American Diabetes Association (ADA) diabetes screening guidelines [96, 99]. The recommendation is for screening to take place in the healthcare setting only, where diagnostic testing is performed in high risk individuals to detect undiagnosed diabetes. The risk of an individual is assessed on the presence of a body mass index (BMI) greater than 25kg/$m^2$ (considered overweight) and the addition of one of the following risk factors: physical inactivity, hypertension, family history of diabetes, dyslipidemia, polycystic ovarian syndrome, high-risk ethnic group, cardiovascular disease history, gestational diabetes or a baby born weighing more than 4 kilograms, previous IFG or IGT or conditions related to insulin resistance. Should none of these risk factors be present, the age of greater than 44 years is considered the additional factor. There is no recommendation on how these risk factors should be determined so it is assumed that it is via a questionnaire or the health professional's inquiry when alerted to risk factors. This leaves much room for error. The interval in which to screen is suggested as every 3 years if the initial diagnostic test is normal, and annually in those with multiple risk factors, IFG or IGT.

The reality is that the expected process of screening for diabetes in developing countries such as South Africa is not undertaken. Even though mass screening is not recommended, patients with risk factors should constantly be encouraged to be screened via non-invasive methods. Unfortunately, the average number of visits for patient care in the diabetic population in Sub-Saharan Africa is low and usually only occurs once complications have arisen [100, 101]. Additionally, information on independent screening is not known nor are the guidelines they choose to follow. New trials to determine the benefits of screening are not necessarily needed in developing countries, but rather the testing of the adaptability of proven diabetes prevention strategies to local settings. Translational studies can be challenging in general and this may be compounded in developing countries where the lack of qualified investigators and infrastructure all contribute to the potential challenges [70, 102]. Cost of financing screening programs along with the required education, health interventions and personal training are not a program most developing countries are able to provide. Additionally, healthcare systems need to be able to handle the possible surge in patients, regardless of the

long term cost and infrastructure benefits. However, a community based approach is feasible in developing countries. With the use of population and cultural specific risk assessment questionnaires, lay persons can be trained to deliver both the risk assessment and the lifestyle intervention and education [102]. This method reduces the overall cost of the screening study and subsequent implementation screening programs. This saving is further enhanced by the use of risk assessment tools that only require laboratory testing for individuals classified as high risk. Risk assessment scores are feasible, cost-effective and can be considered but applicability must be certain, with the required tests available in the area and evidence that the risk score has been validated in the population to be screened.

## Conclusion

Prevention, diagnosis and treatment strategies for diabetes are important in the reduction of the global burden of diabetes. Screening is imperative in targeting diabetes prevention and diagnosis and ultimately treatment of both the disease and the associated complications. Various guidelines have recommended screening for diabetes mellitus as a cost-effective measure. Not without its challenges, there are several compelling reasons to support diabetes screening. Screening provides the opportunity to target risk reduction interventions to high risk populations and to reduce the burden of diabetes and its complications through early identification and treatment. Several reliable biochemical tests are used to detect diabetes however implementation in developing countries is considered unrealistic, when considering the already strained healthcare systems. Risk scores, particularly non-invasive models, are an important tool in diabetes screening in developing countries such as South Africa. These will be explored in the following chapter.

## References


1. IDF. *IDF diabetes atlas, seventh edition*. 2015, International Diabetes Federation: Brussels, Belgium.
2. Wild, S., Roglic, G., Green, A., Sicree, R., and King, H. *Global prevalence of diabetes estimates for the year 2000 and projections for 2030.* Diabetes Care. 2004; **27**(5): p. 1047-1053.
3. Katz, I. *Kidney and kidney related chronic diseases in South Africa and chronic disease intervention program experiences.* Advances in chronic kidney disease. 2005; **12**(1): p. 14-21.
4. Puoane, T., Steyn, K., Bradshaw, D., Laubscher, R., Fourie, J., Lambert, V.*, et al*. *Obesity in South Africa: the South African demographic and health survey.* Obesity research. 2002; **10**(10): p. 1038-1048.
5. Hasumi, T. and Jacobsen, K.H. *Hypertension in South African adults: results of a nationwide survey.* Journal of hypertension. 2012; **30**(11): p. 2098-2104.
6. Lloyd-Sherlock, P., Beard, J., Minicuci, N., Ebrahim, S., and Chatterji, S. *Hypertension among older adults in low-and middle-income countries: prevalence, awareness and control.* International journal of epidemiology. 2014; **43**(1): p. 116-128.
7. Erasmus, R.T., Soita, D.J., Hassan, M.S., Blanco-Blanco, E., Vergotine, Z., Kengne, A.P.*, et al*. *High prevalence of diabetes mellitus and metabolic syndrome in a South African coloured population: Baseline data of a study in Bellville, Cape Town.* SAMJ: South African Medical Journal. 2012; **102**(11): p. 841-844.
8. Peer, N., Steyn, K., Lombard, C., Lambert, E.V., Vythilingum, B., and Levitt, N.S. *Rising diabetes prevalence among urban-dwelling black South Africans.* PloS one. 2012; **7**(9).
9. Soita, D.J. *The prevalence of impaired glucose tolerance and diabetes amongst the middle aged population of Bellville South Community, Cape Town, South Africa*. 2009.
10. IDF. *Global Diabetes Plan 2011-2021*. 2011, Internatinal Diabetes Federation: Brussels, Belgium.
11. IDF. *IDF Diabetes Atlas, 6th edition*. 2013, International Diabetes Federation: Brussels, Belgium.
12. Franco, O.H., Steyerberg, E.W., Hu, F.B., Mackenbach, J., and Nusselder, W. *Associations of diabetes mellitus with total life expectancy and life expectancy with and without cardiovascular disease.* Archives of internal medicine. 2007; **167**(11): p. 1145-1151.
13. Morrish, N., Wang, S.-L., Stevens, L., Fuller, J., Keen, H., and Group, W.M.S. *Mortality and causes of death in the WHO Multinational Study of Vascular Disease in Diabetes.* Diabetologia. 2001; **44**(2): p. S14-S21.
14. Engelgau, M.M., Aubert, R.E., Thompson, T.J., and Herman, W.H. *Screening for NIDDM in nonpregnant adults. A review of principles, screening tests, and recommendations.* Diabetes Care. 1995; **18**(12): p. 1606-1618.
15. Calman, K. *Developing screening in the NHS.* Journal of medical screening. 1994; **1**(2): p. 101-105.
16. Morrison, A.S. *Screening in chronic disease*. 1992: Oxford University Press, USA.
17. Zavon, M.R. *Principles and Practice of Screening for Disease.* Archives of internal medicine. 1969; **123**(3): p. 349-349.
18. Marshall, K.G. *Prevention. How much harm? How much benefit? 1. Influence of reporting methods on perception of benefits.* CMAJ: Canadian Medical Association Journal. 1996; **154**(10): p. 1493.
19. Marshall, K.G. *Prevention. How much harm? How much benefit? 4. The ethics of informed consent for preventive screening programs.* CMAJ: Canadian Medical Association Journal. 1996; **155**(4): p. 377.
20. Trilling, J. *Screening for non-insulin-dependent diabetes mellitus in the elderly.* Clinics in geriatric medicine. 1990; **6**(4): p. 839-848.

21. Cadman, D., Chambers, L., Feldman, W., and Sackett, D. *Assessing the effectiveness of community screening programs.* Jama. 1984; **251**(12): p. 1580-1585.
22. Browder, A.A. *Screening for diabetes.* Preventive medicine. 1974; **3**(2): p. 220-224.
23. Mathers, C.D. and Loncar, D. *Projections of global mortality and burden of disease from 2002 to 2030.* Plos med. 2006; **3**(11): p. e442.
24. Whiting, D.R., Guariguata, L., Weil, C., and Shaw, J. *IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030.* Diabetes research and clinical practice. 2011; **94**(3): p. 311-321.
25. ADA. *Economic consequences of diabetes mellitus in the US in 1997.* Diabetes Care. 1998; **21**(2): p. 296-309.
26. Rettig, B. and Teutsch, S. *The incidence of end-stage renal disease in type I and type II diabetes mellitus.* Diabetic Nephropathy. 1984; **3**(1): p. 26-27.
27. Wetterhall, S.F., Olson, D.R., DeStefano, F., Stevenson, J.M., Ford, E.S., German, R.R.*, et al*. *Trends in diabetes and diabetic complications, 1980–1987.* Diabetes Care. 1992; **15**(8): p. 960-967.
28. Kleinman, J.C., DONAHUE, R.P., HARRIS, M.I., FINUCANE, F.F., MADANS, J.H., and BROCK, D.B. *Mortality among diabetics in a national sample.* American journal of epidemiology. 1988; **128**(2): p. 389-401.
29. Engelgau, M.M., Geiss, L.S., Saaddine, J.B., Boyle, J.P., Benjamin, S.M., Gregg, E.W.*, et al*. *The evolving diabetes burden in the United States.* Annals of Internal Medicine. 2004; **140**(11): p. 945-950.
30. Kornum, J.B., Thomsen, R.W., Riis, A., Lervang, H.-H., Schønheyder, H.C., and Sørensen, H.T. *Diabetes, Glycemic Control, and Risk of Hospitalization With Pneumonia A population-based case-control study.* Diabetes Care. 2008; **31**(8): p. 1541-1545.
31. Thomsen, R.W., Hundborg, H.H., Lervang, H.-H., Johnsen, S.P., Schønheyder, H.C., and Sørensen, H.T. *Risk of community-acquired pneumococcal Bacteremia in patients with Diabetes a population-based case-control study.* Diabetes Care. 2004; **27**(5): p. 1143-1147.
32. Thomsen, R.W., Hundborg, H.H., Lervang, H.-H., Johnsen, S.P., Schønheyder, H.C., and Sørensen, H.T. *Diabetes mellitus as a risk and prognostic factor for community-acquired bacteremia due to enterobacteria: a 10-year, population-based study among adults.* Clinical infectious diseases. 2005; **40**(4): p. 628-631.
33. Jeon, C.Y. and Murray, M.B. *Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies.* Plos med. 2008; **5**(7): p. e152.
34. Rudan, I., Boschi-Pinto, C., Biloglav, Z., Mulholland, K., and Campbell, H. *Epidemiology and etiology of childhood pneumonia.* Bulletin of the world health organization. 2008; **86**(5): p. 408-416B.
35. Seale, A.C., Mwaniki, M., Newton, C.R., and Berkley, J.A. *Maternal and early onset neonatal bacterial sepsis: burden and strategies for prevention in sub-Saharan Africa.* The Lancet infectious diseases. 2009; **9**(7): p. 428-438.
36. Mayanja, B., Todd, J., Hughes, P., Van der Paal, L., Mugisha, J., Atuhumuza, E.*, et al*. *Septicaemia in a population-based HIV clinical cohort in rural Uganda, 1996–2007: incidence, aetiology, antimicrobial drug resistance and impact of antiretroviral therapy.* Tropical Medicine & International Health. 2010; **15**(6): p. 697-705.
37. Reddy, E.A., Shaw, A.V., and Crump, J.A. *Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis.* The Lancet infectious diseases. 2010; **10**(6): p. 417-432.
38. Steen, T., Aruwa, J., and Hone, N. *The epidemiology of adult lung disease in Botswana.* The International Journal of Tuberculosis and Lung Disease. 2001; **5**(8): p. 775-782.
39. Lönnroth, K., Castro, K.G., Chakaya, J.M., Chauhan, L.S., Floyd, K., Glaziou, P.*, et al*. *Tuberculosis control and elimination 2010–50: cure, care, and social development.* The Lancet. 2010; **375**(9728): p. 1814-1829.

40. Engelgau, M.M., Narayan, K., and Herman, W.H. *Screening for type 2 diabetes.* Diabetes Care. 2000; **23**(10): p. 1563-1580.
41. Weyer, C., Bogardus, C., Mott, D.M., and Pratley, R.E. *The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus.* Journal of Clinical Investigation. 1999; **104**(6): p. 787.
42. Butler, A.E., Janson, J., Bonner-Weir, S., Ritzel, R., Rizza, R.A., and Butler, P.C. *β-cell deficit and increased β-cell apoptosis in humans with type 2 diabetes.* Diabetes. 2003; **52**(1): p. 102-110.
43. Ramlo-Halsted, B.A. and Edelman, S.V. *The natural history of type 2 diabetes: practical points to consider in developing prevention and treatment strategies.* Clinical Diabetes. 2000; **18**(2): p. 80-84.
44. Meigs, J.B., Muller, D.C., Nathan, D.M., Blake, D.R., and Andres, R. *The natural history of progression from normal glucose tolerance to type 2 diabetes in the Baltimore Longitudinal Study of Aging.* Diabetes. 2003; **52**(6): p. 1475-1484.
45. Bertram, M.Y. and Vos, T. *Quantifying the duration of pre-diabetes.* Australian and New Zealand journal of public health. 2010; **34**(3): p. 311-314.
46. Patel, P. and Macerollo, A. *Diabetes mellitus: diagnosis and screening.* Diabetes. 2010; **100**: p. 13.
47. Harris, M.I., Klein, R., Welborn, T.A., and Knuiman, M.W. *Onset of NIDDM occurs at least 4–7 yr before clinical diagnosis.* Diabetes Care. 1992; **15**(7): p. 815-819.
48. Harris, M.I. *Impaired glucose tolerance in the US population.* Diabetes Care. 1989; **12**(7): p. 464-474.
49. Eastman, R.C. *Neuropathy in diabetes.* Diabetes in America. 1995; **1**: p. 339-348.
50. Eastman, R.C., Cowie, C.C., and Harris, M.I. *Undiagnosed diabetes or impaired glucose tolerance and cardiovascular risk.* Diabetes Care. 1997; **20**(2): p. 127.
51. Rajala, U., Laakso, M., Qiao, Q., and Keinänen-Kiukaanniemi, S. *Prevalence of retinopathy in people with diabetes, impaired glucose tolerance, and normal glucose tolerance.* Diabetes Care. 1998; **21**(10): p. 1664-1669.
52. Davies, M. and Day, J. *The Cochrane Collaborative Diabetes Review Group.* Diabetic medicine. 1996; **13**(4): p. 391-391.
53. Airey, C. and Williams, D. *Cochrane Collaborative Review Group: Diabetes.* Diabetic medicine. 1995; **12**(5): p. 375-376.
54. Stratton, I.M., Adler, A.I., Neil, H.A.W., Matthews, D.R., Manley, S.E., Cull, C.A.*, et al*. *Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study.* Bmj. 2000; **321**(7258): p. 405-412.
55. Khaw, K.-T. and Wareham, N. *Glycated hemoglobin as a marker of cardiovascular risk.* Current opinion in lipidology. 2006; **17**(6): p. 637-643.
56. Selvin, E., Ning, Y., Steffes, M.W., Bash, L.D., Klein, R., Wong, T.Y.*, et al*. *Glycated hemoglobin and the risk of kidney disease and retinopathy in adults with and without diabetes.* Diabetes. 2011; **60**(1): p. 298-305.
57. Testa, M.A. and Simonson, D.C. *Health economic benefits and quality of life during improved glycemic control in patients with type 2 diabetes mellitus: a randomized, controlled, double-blind trial.* Jama. 1998; **280**(17): p. 1490-1496.
58. Holman, R.R., Paul, S.K., Bethel, M.A., Matthews, D.R., and Neil, H.A.W. *10-year follow-up of intensive glucose control in type 2 diabetes.* New England Journal of Medicine. 2008; **359**(15): p. 1577-1589.
59. UKPDS-Group. *UK Prospective Diabetes Study 6. Complications in newly diagnosed type 2 diabetic patients and their association with different clinical and biochemical risk factors.* Diabetes Res. 1990; **13**(1): p. 1-11.
60. Turner, R., Holman, R., Matthews, D., Bassett, P., Coster, R., Stratton, I.*, et al*. *Hypertension in diabetes study (Hds). 1. Prevalence of hypertension in newly presenting Type-2 diabetic-*

*patients and the association with risk-factors for cardiovascular and diabetic complications.* J HYPERTENS. 1993; **11**(3): p. 309-317.

61. UKPDS-Group. *UK Prospective Diabetes Study VIII: study design, progress and performance.* Diabetologia. 1991; **34**: p. 877-890.
62. Griffin, S.J., Borch-Johnsen, K., Davies, M.J., Khunti, K., Rutten, G.E., Sandbæk, A.*, et al*. *Effect of early intensive multifactorial therapy on 5-year cardiovascular outcomes in individuals with type 2 diabetes detected by screening (ADDITION-Europe): a cluster-randomised trial.* The Lancet. 2011; **378**(9786): p. 156-167.
63. Gæde, P., Vedel, P., Larsen, N., Jensen, G.V., Parving, H.-H., and Pedersen, O. *Multifactorial intervention and cardiovascular disease in patients with type 2 diabetes.* New England Journal of Medicine. 2003; **348**(5): p. 383-393.
64. UKPDS-Group. *Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes.* Bmj. 1998; **317**(7160): p. 703.
65. Costa, J., Borges, M., David, C., and Carneiro, A.V. *Efficacy of lipid lowering drug treatment for diabetic and non-diabetic patients: meta-analysis of randomised controlled trials.* Bmj. 2006.
66. Turnbull, F., Neal, B., Algert, C., Chalmers, J., Chapman, N., Cutler, J.*, et al*. *Effects of different blood pressure-lowering regimens on major cardiovascular events in individuals with and without diabetes mellitus: results of prospectively designed overviews of randomized trials.* Arch Intern Med. 2005; **165**(12): p. 1410-1419.
67. Wing, R.R., Lang, W., Wadden, T.A., Safford, M., Knowler, W.C., Bertoni, A.G.*, et al*. *Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes.* Diabetes Care. 2011; **34**(7): p. 1481-1486.
68. WHO. *Screening for type 2 diabetes: report of a World Health Organization and International Diabetes Federation meeting*. 2003, World Health Organization: Geneva, Switzerland.
69. Hu, F.B. *Globalization of Diabetes The role of diet, lifestyle, and genes.* Diabetes Care. 2011; **34**(6): p. 1249-1257.
70. Echouffo-Tcheugui, J.B., Mayige, M., Ogbera, A.O., Sobngwi, E., and Kengne, A.P. *Screening for hyperglycemia in the developing world: rationale, challenges and opportunities.* Diabetes research and clinical practice. 2012; **98**(2): p. 199-208.
71. Carson, A.P., Reynolds, K., Fonseca, V.A., and Muntner, P. *Comparison of A1C and fasting glucose criteria to diagnose diabetes among US adults.* Diabetes Care. 2010; **33**(1): p. 95-97.
72. Selvin, E., Steffes, M.W., Gregg, E., Brancati, F.L., and Coresh, J. *Performance of A1C for the classification and prediction of diabetes.* Diabetes Care. 2011; **34**(1): p. 84-89.
73. Lu, Z.X., Walker, K.Z., O'Dea, K., Sikaris, K.A., and Shaw, J.E. *A1C for screening and diagnosis of type 2 diabetes in routine clinical practice.* Diabetes Care. 2010; **33**(4): p. 817-819.
74. Rohlfing, C.L., Wiedmeyer, H.-M., Little, R.R., England, J.D., Tennill, A., and Goldstein, D.E. *Defining the relationship between plasma glucose and HbA1c analysis of glucose profiles and HbA1c in the Diabetes Control and Complications Trial.* Diabetes Care. 2002; **25**(2): p. 275-278.
75. Al-Ansary, L., Farmer, A., Hirst, J., Roberts, N., Glasziou, P., Perera, R.*, et al*. *Point-of-care testing for Hb A1c in the management of diabetes: a systematic review and metaanalysis.* Clinical chemistry. 2011; **57**(4): p. 568-576.
76. Jorde, R. and Sundsfjord, J. *Intra-individual variability and longitudinal changes in glycaemic control in patients with Type 1 diabetes mellitus.* Diabetic medicine. 2000; **17**(6): p. 451-456.
77. Bennett, C., Guo, M., and Dharmage, S. *HbA1c as a screening tool for detection of type 2 diabetes: a systematic review.* Diabetic medicine. 2007; **24**(4): p. 333-343.
78. IEC. *International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes.* Diabetes Care. 2009; **32**(7): p. 1327-1334.

79. Christensen, D.L., Witte, D.R., Kaduka, L., Jørgensen, M.E., Borch-Johnsen, K., Mohan, V.*, et al*. *Moving to an A1C-based diagnosis of diabetes has a different impact on prevalence in different ethnic groups.* Diabetes Care. 2010; **33**(3): p. 580-582.
80. Zemlin, A.E., Matsha, T.E., Hassan, M.S., and Erasmus, R.T. *HbA1c of 6.5% to diagnose diabetes mellitus–does it work for us?–the Bellville South Africa study.* PloS one. 2011; **6**(8): p. e22558.
81. Gomez-Perez, F.J., Aguilar-Salinas, C.A., Almeda-Valdes, P., Cuevas-Ramos, D., Garber, I.L., and Rull, J.A. *HbA1c for the diagnosis of diabetes mellitus in a developing country. A position article.* Archives of medical research. 2010; **41**(4): p. 302-308.
82. Herman, W.H. and Cohen, R.M. *Racial and ethnic differences in the relationship between HbA1c and blood glucose: implications for the diagnosis of diabetes.* The Journal of Clinical Endocrinology & Metabolism. 2012; **97**(4): p. 1067-1072.
83. Saudek, C.D., Herman, W.H., Sacks, D.B., Bergenstal, R.M., Edelman, D., and Davidson, M.B. *A new look at screening and diagnosing diabetes mellitus.* The Journal of Clinical Endocrinology & Metabolism. 2013; **93**(7): p. 2447-2453.
84. Davies, M. and Day, J. *Screening for non-insulin-dependent diabetes mellitus (NIDDM): how often should it be performed?* Journal of medical screening. 1994; **1**(2): p. 78-81.
85. Cowie, C.C., Harris, M.I., and Eberhardt, M.S. *Frequency and determinants of screening for diabetes in the US.* Diabetes Care. 1994; **17**(10): p. 1158-1163.
86. Law, M. *"Opportunistic" Screening.* Journal of medical screening. 1994; **1**(4): p. 208-208.
87. Kahn, R., Alperin, P., Eddy, D., Borch-Johnsen, K., Buse, J., Feigelman, J.*, et al*. *Age at initiation and frequency of screening to detect type 2 diabetes: a cost-effectiveness analysis.* The Lancet. 2010; **375**(9723): p. 1365-1374.
88. Herman, W.H. *Economic Analyses of Diabetes Interventions: Rationale, Principles, Findings, and Interpretation.* The Endocrinologist. 1999; **9**(2): p. 113-118.
89. CDC. *The cost-effectiveness of screening for type 2 diabetes.* Jama. 1998; **280**(20): p. 1757-1763.
90. Knowler, W.C. *Screening for NIDDM: opportunities for detection, treatment, and prevention.* Diabetes Care. 1994; **17**(5): p. 445-450.
91. De Courten, M. and Zimmet, P. *Screening for Non-insulin-dependent Diabetes Mellitus: Where to Draw the Line?* Diabetic medicine. 1997; **14**(2): p. 95-98.
92. Donaldson, C. *Using economics to assess the place of screening.* Journal of medical screening. 1994; **1**(2): p. 124-129.
93. Alberti, K.G.M.M. and Zimmet, P.f. *Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation.* Diabetic medicine. 1998; **15**(7): p. 539-553.
94. Paterson, K. *Population screening for diabetes mellitus.* Diabetic medicine. 1993; **10**(8): p. 777-781.
95. Gavin III, J.R., Alberti, K., Davidson, M.B., and DeFronzo, R.A. *Report of the expert committee on the diagnosis and classification of diabetes mellitus.* Diabetes Care. 1997; **20**(7): p. 1183.
96. ADA. *Diagnosis and classification of diabetes mellitus.* Diabetes Care. 2010; **33**(Supplement 1): p. S62-S69.
97. Cho, N., Colagiuri, S., Distiller, L., Dong, B., Dunning, T., Gadsby, R.*, et al*. *International Diabetes Federation: Global guideline for managing older people with type 2 diabetes.* Brussels: International Diabetes Federation. 2013.
98. Somannavar, S., Lanthorn, H., Pradeepa, R., Narayanan, V., Rema, M., and Mohan, V. *Prevention awareness counselling and evaluation (PACE) diabetes project: a mega multi-pronged program for diabetes awareness and prevention in South India (PACE-5).* Journal of the Association of Physicians of India. 2008; **56**: p. 429-435.
99. Amod, A. *The 2012 SEMDSA guideline for the management of type 2 diabetes.* Journal of Endocrinology, Metabolism and Diabetes of South Africa. 2012; **17**(1): p. 61-62.

100. Gning, S., Thiam, M., Fall, F., Ba-Fall, K., Mbaye, P., and Fourcade, L. *Diabetes mellitus in sub-Saharan Africa: epidemiological aspects and management issues.* Medecine tropicale: revue du Corps de sante colonial. 2007; **67**(6): p. 607-611.
101. Otieno, C. and Kariuki, M. *Quality of glycaemic control in ambulatory diabetics at the out-patient clinic of Kenyatta National Hospital, Nairobi.* East African medical journal. 2004; **80**(8): p. 406-410.
102. Ali, M.K., Echouffo-Tcheugui, J.B., and Williamson, D.F. *How effective were lifestyle interventions in real-world settings that were modeled on the Diabetes Prevention Program?* Health Affairs. 2012; **31**(1): p. 67-75.

03

# Predictive modeling for incident and prevalent diabetes risk evaluation: an expert review

Katya Masconi, Justin Basile Echouffo-Tcheugui,
Tandi E. Matsha, Rajiv T. Erasmus and Andre P. Kengne
*Expert Review of Endocrinology & Metabolism* 2015, **10**:3

## SUMMARY

With half of individuals with diabetes undiagnosed worldwide and a projected 55% increase of the population with diabetes by 2035, the identification of undiagnosed and high-risk individuals is imperative. Multivariable diabetes risk prediction models have gained popularity during the past two decades. These have been shown to predict incident or prevalent diabetes through a simple and affordable risk scoring system accurately. Their development requires cohort or cross-sectional type studies with a variable combination, number and definition of included risk factors, and their performance chiefly measured by discrimination and calibration. Models can be used in clinical and public health settings. However, the impact of their use on outcomes in real-world settings needs to be evaluated before widespread implementation.

KEYWORDS: diabetes, incident, prevalent, risk prediction, screening

## Expert commentary

### Rationale for developing diabetes risk models

Worldwide, 382 million individuals are living with diabetes, where 46% are undiagnosed [1]. The prevalence is expected to increase by 55% by 2035, with the highest relative increase occurring in developing countries. Type 2 diabetes mellitus (T2DM), which accounts for 9 in 10 diabetes cases, imposes a substantial burden of morbidity, mortality, suffering and economic cost [2]. The natural history of T2DM is characterized by a silent phase that can last for many years. At diagnosis, micro and macrovascular complications are already present in approximately half of patients [3], with diabetes associated macrovascular complications contributing to most of the morbidity and mortality [4].

All the aforementioned reasons dictate the need for early identification and prevention of diabetes. Indeed, early detection of diabetes can potentially improve morbidity, mortality and quality of life outcomes [5]. Furthermore, the progression to the full stage of diabetes

among individuals with impaired fasting glycaemia (IFG) or impaired glucose tolerance (IGT) has been shown to be reduced by lifestyle changes and/or pharmacotherapy [6-9]. Early identification of individuals with undiagnosed diabetes or future risk of diabetes may be cost-effective; hence, the need for developing strategies to identify individuals at high risk for diabetes and/or with undiagnosed diabetes, which include risk models for diabetes. Indeed, prospective epidemiological studies have highlighted the limitation of the use of IGT as the sole mean for identifying individuals at high risk for T2DM, as only 30 – 40% of individuals with IGT ultimately develop diabetes [10-12]. Additionally, 40% of subjects who develop T2DM have a normal glucose tolerance at baseline [10]. This variability of prognosis among patients prompted the search for alternative methods to accurately predict diabetes risk. Risk prediction models including multiple risk factors have emerged as practical tools to classify and predict diseases. Diabetes results from a complex gene-environment interaction for which several risk factors are well documented [8]. The precise interaction of these risk factors is a complex process that varies both within and across populations [13-15]. Risk models that can handle this complexity but at the same time are sufficiently simple, affordable and implementable, have been developed. These can be roughly divided based on the prediction of prevalent or incident diabetes risk. These are developed in different settings but will be discussed overall, unless specified.

**Historical perspective on risk prediction in diabetes**

Until the late 1970's and early 1980's, there was a lot of controversy surrounding early identification of people with diabetes, treatment of those with a less severe form of the disease, and little or no difference between interventions for diabetes prevention and those for diabetes control. Furthermore, the definition, classification and diagnostic criteria for diabetes varied substantially. In this context, recommendations for diabetes detection, prevention and treatment were very loose, and accordingly little effort was invested in improving diabetes risk screening and stratification [16]. In 1979, the National Diabetes Data Group (NDDG) of the US National Institute of Health proposed the first uniform classification and diagnostic criteria for diabetes mellitus [17], largely adopted by the World Health Organization (WHO) a year later, and subsequently revised a few times by both the American Diabetes Association (ADA) and WHO. One defining event appears to be the change of stance

on community screening for diabetes by the ADA in 1989, in the wake of the accumulating evidence, by issuing a position statement recommending that "*all people with one or more diabetes risk factors or having any diabetes symptoms should be identified and referred for medical evaluation*" [18]. This was followed by a shift of focus towards developing appropriate strategies for diabetes risk screening. In 1993, the ADA issued the risk factor questionnaires for diabetes risk screening [19], thereby embracing the concept of multivariable approaches to diabetes risk screening. The first published multivariable diabetes risk models aimed to prove that conventional diabetes risk factors can predict future diabetes as well as, if not better than, IGT [20]. These models were developed in randomly selected Mexican Americans and non-Hispanic whites within the San Antonio Heart Study cohort; for the overall population and separate for each sex and ethnic group, all containing fasting plasma glucose (FPG) as a variable. Following this, and based largely on the ADA questionnaire, Herman and co-workers used the classification regression tree to develop a model for predicting undiagnosed diabetes based on data from the second Nutrition and Health Examination Survey (NHANES) in 1994 [21]. This marked the beginning of the model development explosion. There has been significantly more focus on incident diabetes risk models, and as expected, risk prediction research has been concentrated in Western countries, with a small spike in developing countries like India, China and Taiwan [22, 23]. These models have been examined in four comprehensive reviews on diabetes risk prediction modelling [22-26].

## Principles of risk prediction applied to diabetes

### *Study design to develop risk models*

Risk scores are ideally developed in large, age-defined populations, measuring baseline risk factors [27]. Hitherto, studies have been mostly undertaken in the USA and European countries on Caucasian individuals [22-26]; however multi-ethnic studies [21, 28, 29] and studies focusing on minorities in a country are available [30-32]. The age range of included participants has mainly encompassed middle aged individuals (40 – 65 years), however some studies did include a younger population (adults over 20 years of age) [22, 23, 26].

Longitudinal study designs, such as the Atherosclerosis Risk in Communities Study (ARIC), are required for the development and validation of incident diabetes risk prediction models [33], where individuals at baseline are without previously diagnosed or undiagnosed diabetes. A cross-sectional study can only be used for the development and validation of prevalent diabetes risk prediction models, for example the Isle of Ely Diabetes Project in Cambridgeshire, where a cross-sectional study was used for the development of the Cambridge Diabetes Risk Score [34]. For models of prevalent diabetes, only individuals with previously diagnosed diabetes are excluded from the study. A single time point collection from a longitudinal collection may be used for prevalent models, illustrated by the FINRISC cohorts, which are comprised of population representative cross-sectional surveys allowing for incident and prevalent diabetes risk prediction [35].

Data may be collected *de novo*, alternatively existent data may be used. Primary data may be collected via surveys, questionnaires or interviews. Readily available data may be in the form of existent databases from longitudinal or cross-sectional studies or unpublished data such as hospital patient records can be used. Lin *et al* recruited individuals for the purpose of risk model validation using interview methods [36]. Primary data has the benefit of control when collecting variables, while it is hindered by high cost and time intensity. Secondary data use is inexpensive and immediately available but missing information may be an issue. The DETECT-2 study is a good example of an existent database, where 34 countries have contributed data of population-based surveys or large cohorts of at least 500 individuals with set required information. This data has been routinely used for model validation [28, 37]. Alternatively, The Health Improvement Network (THIN) cohort is a database comprising of primary care medical records entered by general practitioners in the UK [38].

*Variables included in diabetes risk models*

The final model can vary in the combination and number of risk factors included. Variables included in diabetes risk models have ranged from simple demographic information such as age or family history, to more complex and invasive markers such as triglyceride levels and genetic polymorphisms. The most commonly included risk factors have been age, family

history of diabetes, BMI, hypertension, waist circumference (WC) and gender [22-26]. The selection of the final model variables from preselected diabetes risk factors can follow the stepwise approaches including the backward significance level to stay or the forward significance level to enter [39]; using the full model approach and the all-possible model approach. Bang *et al* used backward elimination, deleting the covariate with a p-value greater than the selected significance level to stay, one at a time, until a final model was reached with statistical significant covariates [40]. Alternatively, forward selection describes the selection of a significance level to enter and variables with a p-value less than this are included in the model one by one, ensuring the model p-value stays below the significance level to enter. In general, stepwise approaches are the most common choice of variable selection [29, 41-44]. Finally, a relative weighting can be assigned to each predictor in the final model to produce a final risk score, making it easier to use.

The definition of a variable included in a model is not uniform across studies. Family history may be limited to parents (e.g. The Framingham Offspring Diabetes Risk Score [45]) or extended to sibling diabetes history as in Cambridge Risk Score [46]. The BMI, if not classified as continuous, can have different cut-offs, depending on the population (e.g. BMI ≥ 28 kg/m$^2$ in the Chinese Diabetes Risk Score was classified as obese [47], while the Cambridge Risk Score gave the highest scoring to a BMI ≥ 30 kg/m$^2$). The same can be said for WC, as in the Atherosclerosis Risk in Communities Diabetes Prediction Model (Men WC cut-off: > 105 cm; Women WC cut-off: > 104cm) [48] compared to the Chinese Risk Score (Men WC cut-off: >100 cm; Women WC cut-off: > 90cm) [49]. Variables such as hypertension may also be based on reported information (e.g. hypertension medication use in Diabetes Risk Score [50]) or based on tested values (e.g. systolic blood pressure in the San Antonio Diabetes Risk Score [51]).

Models that require variables obtained through an interview or questionnaire are generally classified as a non-invasive model; while those containing variable(s) requiring biochemical results are considered invasive models. The use of invasive variables in prevalent diabetes risk prediction modelling is considered impractical, defeating the object of simple and cost-effective screening for undiagnosed diabetes risk. For incident diabetes risk modelling, the

inclusion of diabetes biomarkers in these models is possible, but practically challenging. The added benefits of invasive biomarkers is under question, as it has been shown that the inclusion of conventional circulating biomarkers such as glucose, $HbA_{1C}$, lipids, uric acid or γ-glutamyltransferase improves model performance [52], while a review by Echouffo-Tcheugui *et al* concluded that circulating but more specifically genetic biomarkers did not substantially improve model performance beyond the ability of non-invasive risk models [53]. Variables may have a continuous or categorical form. The conversion of continuous to categorical variables, commonly age, BMI and WC, is not recommended, as it can result in a loss of information. Nevertheless, linearity of the continuous variable cannot be assumed and simple predictor transformations should be tested [54].

*Definition of diabetes in models*

Previously diagnosed diabetes is identified through medical records or current treatment, allowing for immediate exclusion from the study. The identification of diabetes status can vary across studies. Studies may choose to use fasting blood glucose, oral glucose tolerance test, or a combination of both. Longitudinal studies for incident diabetes risk prediction can also rely on self-reporting of a subsequent diabetes diagnosis. The defining cut-offs used in risk model studies have been that of the World Health Organization [5] or the American Diabetes Association [55] As yet, only Simmons *et al* has used $HbA_{1C}$ as a diagnosis for incident diabetes [56].

*Performance of models*

The performance of a newly developed model is measured by the discrimination and calibration. The discrimination describes the reliability of distinguishing between high and low risk individuals, assessed and compared by the concordance (c) statistic [54]. The calibration describes the agreement between the probability of the outcome of interest as estimated by the model, and the observed outcome frequencies [57]. It is assessed graphically by plotting the predicted risk against the observed outcome rate, supplemented with formal statistical tests. However, few studies report the complete measures of predictive performance

including discrimination, calibration with sensitivity and specificity and positive or predicted value for potential cut-offs [22, 24, 25]. The majority of studies report discrimination and less often calibration. The c-statistic in model development samples have range from as low as 0.64 in the Chinese Diabetes Risk Score [49] to as high as 0.92 for the women's Clinical and Biologic Risk Score in the Data from an Epidemiological Study on the Insulin Resistance Syndrome (DESIR) cohort [58]. The performance of non-invasive models ranged from 0.70 to 0.80, while models containing biochemical measures ranged from 0.68 – 0.85, not displaying substantial gain from the inclusion of more invasive variables [25].

*External validation of diabetes risk models*

The good performance of a model during development does not necessarily demonstrate an acceptable model. The performance of developed risk models should be validated both internally (same population as development) and externally (similar but not identical population) to allow for accurate implementation. The San Antonio Diabetes Risk Score had a discrimination performance in the development population in the United States of 0.84 in the clinical model and 0.85 in the full model [59]. However, when applied to different settings (either change in patient ethnicity, from primary to secondary care populations or from adults to children; altering the case mix), this affected the generalizability of the model, reducing the its predictive performance [22, 60]. The external validation of this model in independent populations highlights this reduction in performance, seen in a Mexican population (0.77) [61], an Australian population (0.78) [62], a Finnish population (0.74) [63] and more drastically in a Chinese population (0.68) [64]. The loss of performance during external validation can also be due to the modification of models in the validation population. The Finnish Diabetes Risk Score had a c-statistic of 0.86 in the development population; validation in another Finnish study with a similar methodology reproduced this discrimination (0.86) [50], but was lower in other populations (0.65 – 0.81) [58, 62, 63, 65]. The latter studies altered the models, by adding family history and removing diet and physical activity as a variable. Understandably, the publication of model development has slowed as the validation of the models increases. Well recognized and externally validated incident diabetes models are the ARIC Study Diabetes Risk Score [48, 51, 66, 67], Cambridge Diabetes Risk Score [46, 56, 64, 68, 69] and the Framingham Offspring Diabetes Risk Score [45, 64, 70-73]; and prevalent diabetes risk

models Cambridge Diabetes Risk Score [32, 36, 46, 49, 74-77] and Finnish Risk Score (FINDRISC) [36, 49, 50, 77-81]. The acceleration of validation across diverse settings and specifically in developing countries is necessary. This can be initiated by large research groups who develop or have access to multinational databases, allowing the validation of models across multiple countries and population groups. Although validation of existent risk models can be seen in developing countries, it is significantly less than in developed countries. To encourage risk model validation in developing countries, comprehensive collection of data is required, followed up for risk factors and disease events and used for the validation of risk models. It is the lack of these databases that slow predictive research in developing countries.

**Uses of diabetes risk models**

*Clinical and public health uses*

Diabetes risk prediction scores should be developed with the eventual aim of clinical implementation or public health use. The clinical use of risk prediction models to inform individuals of their current undiagnosed or future diabetes risk, are intended for self-administration via post, email or physically distributed; while others require the assistance of medical practitioners and invasive testing. The usability of a prediction model depends on circumstances, and the variables included in the model need to be relevant or realistic to the clinical practice it is aimed to be implemented [82]. Application of risk models at a population level is potentially beneficial for subsequent public health use, providing insight into the future burden of a disease in an area / country, providing information that can be used for current and future health resource planning and opening the doors for research in the potential health impact of various population-based interventions.

*Impact of risk models on outcomes*

A diabetes risk prediction model has the potential to be a cost and lifesaving tool; however consequences can be adverse if, for example, beneficial treatment is not recommended to an individual due to the low risk classification of the score. Hence, impact studies are highly recommended to determine the extent to which the scores are actually used and have led to

improved outcomes. Unfortunately impact studies are scarce, indeed neither qualitative nor quantitative studies into the impact of a model are common in risk prediction assessment [60]. In addition, no study of prevention of incident diabetes among high risk people identified using a diabetes risk score has been published. However, there are completed intermediate outcome studies and ongoing translational studies [83-85]. Recently, there have been studies assessing the success of implementation of diabetes risk models in primary healthcare. A comprehensive review by Dhippayom *et al* explored the extent of use of diabetes risk assessment tools, showing that implementation mostly took place in Europe in a general practice or healthcare setting, with the American Diabetes risk score being the most frequently used tool [86]. This review illustrates the fact that relatively few prognostic models are currently used in clinical practice, as highlighted by the Prognosis Research Strategy group [87].

The published implementation studies were focused on either the medical practitioner and patient perception of the used diabetes risk score or on the use of the score as a more proactive way of introducing lifestyle modification programs to reduce the set outcome, specifically diabetes risk factors such as weight, blood pressure and glucose levels, dependent on the predicted risk [22, 86]. Surprisingly, none of the published implementation studies recalculated participants' risk scores following the intervention to determine if there was a change in predicted risk [22]. There is a clear lack of systematic approaches post-identification of high risk individuals [86]. The successful use of the models for lifestyle intervention programs has been highlighted by the GOAL study in Finland using the FINDRISC score [83], the Greater Green Triangle Diabetes Prevention Project in Australia using the Diabetes Risk Score [84] and the Active Prevention in High Risk Individuals of Diabetes Type 2 in Eindhoven (APHRODITE) study in the Netherlands, using the adapted FINDRISC score [85].

In most cases, however, the methodological limitations raise doubts as to whether this evidence can be translated into cost-effective large scale community-wide programs [88]. Consequently, the transfer to 'real-world' settings prior to widespread implementation

should be evaluated with a scientifically valid design. It is this transfer to routine clinical practice that remains the greatest challenge in diabetes risk prediction and prevention [89].

*Barriers to implementation of diabetes risk models*

Despite the rise and endorsement of diabetes risk prediction models, barriers to their implementation need to be addressed. The recommendation of a model for implementation can only be given if validation in the new population has shown accurate predictive performance, and with no model for either incident or prevalent diabetes having been validated across all populations, mass-recommendation is not possible. The need for external model validation for generalizability and extrapolation has been highlighted, but what must be stressed is the applicability of these models in clinical practice and community settings, especially in developing countries. Invasive models are limited in their availability of the tests worldwide.

Implementation may be affected due to changes over time, where improvements or changes in biomarker measurements or diagnostics tests will change the prognosis of the patient [82]. Other reasons for the poor implementation of models include the possible complexity for laymen or basic medical professions or the image thereof, limiting their application in routine care. This may also be the case for community screening. Finally, once classified as high-risk, individuals need to be directed to appropriate healthcare and educational centres. Depending on the country of implementation, specific optimal diabetes risk identification would have to be developed based on the unique socio-demographic and risk factor composition of the population.

## Expert commentary

Risk modelling for predicting prevalent undiagnosed or incident diabetes mellitus has mostly developed over the last two decades, and has paralleled the accumulated knowledge on the determinants, natural course and pathophysiology of diabetes mellitus as well as strategies for modifying them through prevention and ongoing management. Many models have essentially been developed over the last two decades to predict the presence of undiagnosed

diabetes, or the future occurrence of the diseases, using varying combinations of history, clinical, biological and at times genetic characteristics related to diabetes risk. In proportion, only a few of these models have been tested in other populations, while an even smaller number have been incorporated in implementation studies, particularly of diabetes prevention. It remains, however, that more external validation studies of existing diabetes prediction models are needed, just like the impact analysis of their introduction in routine care on healthcare providers' behaviour and the outcomes of people at risk of diabetes, or with prevalent undiagnosed diabetes.

## Five year view

We predict that the imbalance between the number of diabetes prediction models developed and those tested in diverse populations will continue to prevail in the coming years. Investigators are often tempted to develop new models with no consideration of what is already available. For those who elect to test existing models, in the presence of poor performance, developing new models is often the preferred alternative to model improvement strategies. We further predict that more results will become available from ongoing diabetes prevention implementation studies that have used prediction models to select participants. Positive or promising results from these studies could accelerate the uptake of prediction models in routine settings worldwide. Finally, we speculate that prediction models, at least for prevalent undiagnosed diabetes and validation studies of existing models, will increasingly emerge from developing countries as the result of the growing worldwide interest in prediction research and personalised medicine.

**Key issues**

- A fast increasing number of multivariable models to predict prevalent undiagnosed or incident diabetes have been developed, but only a few have been tested in diverse settings.
- It is not always obvious from published studies to accurately ascertain how the development of existing models has addressed the critical methodological challenges which may affect the performance of the models both on the derivation sample, and in subsequent external validation studies.
- The complexity of existing models varies substantially and in the absence of head-to-head comparison studies, it will be very difficult in most settings to choose the most appropriate model.
- Existing models are mostly based on glycaemia-defined diabetes, and may not be valid in the context of the recommendations for also using HbA1c for diabetes diagnosis.
- Existing models overwhelmingly originate from developed countries, and have seldom been tested in developing countries that may derive the most benefits from the introduction of those models in routine care.
- Studies to assess the impact of adopting diabetes prediction models in routine care are a very recent development, and little is known on the effect of introducing diabetes prediction models in routine care on the behaviour of healthcare providers and the outcomes of care.

## References


1. Beagley, J., Guariguata, L., Weil, C., and Motala, A.A. *Global estimates of undiagnosed diabetes in adults.* Diabetes research and clinical practice. 2014; **103**(2): p. 150-160.
2. Clarke, P.M., Glasziou, P., Patel, A., Chalmers, J., Woodward, M., Harrap, S.B.*, et al*. *Event rates, hospital utilization, and costs associated with major complications of diabetes: A multicountry comparative analysis.* PLoS medicine. 2010; **7**(2): p. e1000236.
3. Harris, M.I. and Eastman, R.C. *Early detection of undiagnosed diabetes mellitus: A us perspective.* Diabetes/metabolism research and reviews. 2000; **16**(4): p. 230-236.
4. Campbell, I.W. *Type 2 diabetes mellitus:'The silent killer'.* Practical Diabetes International. 2001; **18**(6): p. 187-191.
5. WHO. *Screening for type 2 diabetes: Report of a world health organization and international diabetes federation meeting*. 2003.
6. Pan, X.-R., Li, G.-w., Hu, Y.-H., Wang, J.-X., Yang, W.-Y., An, Z.-X.*, 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**(4): p. 537-544.
7. DPPR-Group. *Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin.* The New England journal of medicine. 2002; **346**(6): p. 393.
8. Tuomilehto, J., Lindström, J., Eriksson, J.G., Valle, T.T., Hämäläinen, H., Ilanne-Parikka, P.*, et al*. *Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance.* New England Journal of Medicine. 2001; **344**(18): p. 1343-1350.
9. DPPR-Group. *Prevention of type 2 diabetes with troglitazone in the diabetes prevention program.* Diabetes. 2005; **54**(4): p. 1150.
10. Unwin, N., Shaw, J., Zimmet, P., and Alberti, K. *Impaired glucose tolerance and impaired fasting glycaemia: The current status on definition and intervention.* Diabetic medicine. 2002; **19**(9): p. 708-723.
11. Edelstein, S.L., Knowler, W.C., Bain, R.P., Andres, R., Barrett-Connor, E.L., Dowse, G.K.*, et al*. *Predictors of progression from impaired glucose tolerance to niddm: An analysis of six prospective studies.* Diabetes. 1997; **46**(4): p. 701-710.
12. Ramachandran, A., Snehalatha, C., Naik, R.A.S., Mohan, V., Shobana, R., and Viswanathan, M. *Significance of impaired glucose tolerance in an asian indian population: A follow-up study.* Diabetes research and clinical practice. 1986; **2**(3): p. 173-178.
13. Meigs, J.B., Shrader, P., Sullivan, L.M., McAteer, J.B., Fox, C.S., Dupuis, J.*, et al*. *Genotype score in addition to common risk factors for prediction of type 2 diabetes.* New England Journal of Medicine. 2008; **359**(21): p. 2208-2219.
14. Talmud, P.J., Hingorani, A.D., Cooper, J.A., Marmot, M.G., Brunner, E.J., 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**: p. b4838.
15. Park, K.S. *The search for genetic risk factors of type 2 diabetes mellitus.* Diabetes & metabolism journal. 2011; **35**(1): p. 12-22.
16. Herron, C.A. *Screening in diabetes mellitus: Report of the atlanta workshop.* Diabetes care. 1979; **2**(4): p. 357-362.
17. National Diabetes Data Group. *Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. National diabetes data group.* Diabetes. 1979; **28**(12): p. 1039-1057.
18. American Diabetes Association. *Screening for diabetes.* Diabetes care. 1989; **12**(8): p. 588-590.
19. American Diabetes Association. *American diabetes alert.* Diabetes forecast. 1993; **46**(3): p. 54-55.
20. Stern, M.P., Morales, P.A., Valdez, R.A., Monterrosa, A., Haffner, S.M., Mitchell, B.D.*, et al*. *Predicting diabetes: Moving beyond impaired glucose tolerance.* Diabetes. 1993; **42**(5): p. 706-714.

21. Herman, W.H., Smith, P.J., Thompson, T.J., Engelgau, M.M., and Aubert, R.E. *A new and simple questionnaire to identify people at increased risk for undiagnosed diabetes.* Diabetes Care. 1995; **18**(3): p. 382-387.
22. Noble, D., Mathur, R., Dent, T., Meads, C., and Greenhalgh, T. *Risk models and scores for type 2 diabetes: Systematic review.* Bmj. 2011; **343**.
23. Brown, N., Critchley, J., Bogowicz, P., Mayige, M., and Unwin, N. *Risk scores based on self-reported or available clinical data to detect undiagnosed type 2 diabetes: A systematic review.* Diabetes research and clinical practice. 2012; **98**(3): p. 369-385.
24. Collins, G.S., Mallett, S., Omar, O., and Yu, L.-M. *Developing risk prediction models for type 2 diabetes: A systematic review of methodology and reporting.* BMC medicine. 2011; **9**(1): p. 103.
25. Buijsse, B., Simmons, R.K., Griffin, S.J., and Schulze, M.B. *Risk assessment tools for identifying individuals at risk of developing type 2 diabetes.* Epidemiologic reviews. 2011: p. mxq019.
26. Thoopputra, T., Newby, D., Schneider, J., and Li, S.C. *Survey of diabetes risk assessment tools: Concepts, structure and performance.* Diabetes/metabolism research and reviews. 2012; **28**(6): p. 485-498.
27. Altman, D.G., Vergouwe, Y., Royston, P., and Moons, K.G. *Prognosis and prognostic research: Validating a prognostic model.* Bmj. 2009; **338**: p. b605.
28. Glümer, C., Vistisen, D., Borch-Johnsen, K., and Colagiuri, S. *Risk scores for type 2 diabetes can be applied in some populations but not all.* Diabetes Care. 2006; **29**(2): p. 410-414.
29. de Sousa, A.G.P., Pereira, A.C., Marquezine, G.F., do Nascimento-Neto, R.M., Freitas, S.N., Nicolato, R.L.d.C.*, et al*. *Derivation and external validation of a simple prediction model for the diagnosis of type 2 diabetes mellitus in the brazilian urban population.* European journal of epidemiology. 2009; **24**(2): p. 101-109.
30. Hanif, M., Valsamakis, G., Dixon, A., Boutsiadis, A., Jones, A., Barnett, A.*, et al*. *Detection of impaired glucose tolerance and undiagnosed type 2 diabetes in uk south asians: An effective screening strategy.* Diabetes, Obesity and Metabolism. 2008; **10**(9): p. 755-762.
31. Bindraban, N.R., van Valkengoed, I.G., Mairuhu, G., Holleman, F., Hoekstra, J.B., Michels, B.P.*, et al*. *Prevalence of diabetes mellitus and the performance of a risk score among hindustani surinamese, african surinamese and ethnic dutch: A cross-sectional population-based study.* BMC Public Health. 2008; **8**(1): p. 271.
32. Spijkerman, A.M., Yuyun, M.F., Griffin, S.J., Dekker, J.M., Nijpels, G., and Wareham, N.J. *The performance of a risk score as a screening test for undiagnosed hyperglycemia in ethnic minority groups data from the 1999 health survey for england.* Diabetes Care. 2004; **27**(1): p. 116-122.
33. Aric-Investigators. *The atherosclerosis risk in community (aric) study: Design and objectives.* American journal of epidemiology. 1989; **129**(4): p. 687-702.
34. Williams, D., Wareham, N., Brown, D., Byrne, C., Clark, P., Cox, B.*, et al*. *Undiagnosed glucose intolerance in the community: The isle of ely diabetes project.* Diabetic medicine. 1995; **12**(1): p. 30-35.
35. Vartiainen, E., Laatikainen, T., Peltonen, M., Juolevi, A., Männistö, S., Sundvall, J.*, et al*. *Thirty-five-year trends in cardiovascular risk factors in finland.* International journal of epidemiology. 2010; **39**(2): p. 504-518.
36. Lin, J.-W., Chang, Y.-C., Li, H.-Y., Chien, Y.-F., Wu, M.-Y., Tsai, R.-Y.*, et al*. *Cross-sectional validation of diabetes risk scores for predicting diabetes, metabolic syndrome, and chronic kidney disease in taiwanese.* Diabetes Care. 2009; **32**(12): p. 2294-2296.
37. Alssema, M., Vistisen, D., Heymans, M., Nijpels, G., Glümer, C., Zimmet, P.*, 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**(5): p. 1004-1012.

38. Lewis, J.D., Schinnar, R., Bilker, W.B., Wang, X., and Strom, B.L. *Validation studies of the health improvement network (thin) database for pharmacoepidemiology research.* Pharmacoepidemiology and drug safety. 2007; **16**(4): p. 393-401.
39. Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman, W. *Applied linear statistical models*. Vol. 4. 1996: Irwin Chicago.
40. Bang, H., Edwards, A.M., Bomback, A.S., Ballantyne, C.M., Brillon, D., Callahan, M.A.*, et al*. *Development and validation of a patient self-assessment score for diabetes risk.* Annals of Internal Medicine. 2009; **151**(11): p. 775-783.
41. Al Khalaf, M., Eid, M., Najjar, H., Alhajry, K., Doi, S., and Thalib, L. *Screening for diabetes in kuwait and evaluation of risk scores.* EMHJ. 2010; **16**(7): p. 725-731.
42. Al-Lawati, J. and Tuomilehto, J. *Diabetes risk score in oman: A tool to identify prevalent type 2 diabetes among arabs of the middle east.* Diabetes research and clinical practice. 2007; **77**(3): p. 438-444.
43. Baan, C.A., Ruige, J.B., Stolk, R.P., Witteman, J., Dekker, J.M., Heine, R.J.*, et al*. *Performance of a predictive model to identify undiagnosed diabetes in a health care setting.* Diabetes Care. 1999; **22**(2): p. 213-219.
44. Glümer, C., Carstensen, B., Sandbæk, A., Lauritzen, T., Jørgensen, T., and Borch-Johnsen, K. *A danish diabetes risk score for targeted screening the inter99 study.* Diabetes Care. 2004; **27**(3): p. 727-733.
45. Wilson, P.W., Meigs, J.B., Sullivan, L., Fox, C.S., Nathan, D.M., and D'Agostino, R.B. *Prediction of incident diabetes mellitus in middle-aged adults: The framingham offspring study.* Archives of internal medicine. 2007; **167**(10): p. 1068-1074.
46. Griffin, S., Little, P., Hales, C., Kinmonth, A., and Wareham, N. *Diabetes risk score: Towards earlier detection of type 2 diabetes in general practice.* Diabetes/metabolism research and reviews. 2000; **16**(3): p. 164-171.
47. Liu, M., Pan, C., and Jin, M. *A chinese diabetes risk score for screening of undiagnosed diabetes and abnormal glucose tolerance.* Diabetes technology & therapeutics. 2011; **13**(5): p. 501-507.
48. Schmidt, M.I., Duncan, B.B., Bang, H., Pankow, J.S., Ballantyne, C.M., Golden, S.H.*, et al*. *Identifying individuals at high risk for diabetes the atherosclerosis risk in communities study.* Diabetes Care. 2005; **28**(8): p. 2013-2018.
49. Gao, W., Dong, Y., Pang, Z., Nan, H., Wang, S., Ren, J.*, et al*. *A simple chinese risk score for undiagnosed diabetes.* Diabetic medicine. 2010; **27**(3): p. 274-281.
50. Lindström, J. and Tuomilehto, J. *The diabetes risk score a practical tool to predict type 2 diabetes risk.* Diabetes Care. 2003; **26**(3): p. 725-731.
51. Stern, M., Williams, K., Eddy, D., and Kahn, R. *Validation of prediction of diabetes by the archimedes model and comparison with other predicting models.* Diabetes Care. 2008; **31**(8): p. 1670-1671.
52. Abbasi, A., Peelen, L.M., Corpeleijn, E., van der Schouw, Y.T., Stolk, R.P., Spijkerman, A.M.*, et al*. *Prediction models for risk of developing type 2 diabetes: Systematic literature search and independent external validation study.* Bmj. 2012; **345**.
53. Echouffo-Tcheugui, J.B., Dieffenbach, S.D., and Kengne, A.P. *Added value of novel circulating and genetic biomarkers in type 2 diabetes prediction: A systematic review.* Diabetes research and clinical practice. 2013; **101**(3): p. 255-269.
54. Moons, K.G., Kengne, A.P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D.G.*, et al*. *Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio) marker.* Heart. 2012; **98**(9): p. 683-690.
55. ADA. *Diagnosis and classification of diabetes mellitus.* Diabetes Care. 2008; **31**(1): p. S55-S60.
56. Simmons, R., Harding, A.H., Wareham, N., and Griffin, S. *Do simple questions about diet and physical activity help to identify those at risk of type 2 diabetes?* Diabetic medicine. 2007; **24**(8): p. 830-835.
57. Steyerberg, E.W. *Clinical prediction models*. 2009, New York: Springer.

58. 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**(10): p. 2056-2061.
59. Stern, M.P., Williams, K., and Haffner, S.M. *Identification of persons at high risk for type 2 diabetes mellitus: Do we need the oral glucose tolerance test?* Annals of Internal Medicine. 2002; **136**(8): p. 575-581.
60. Moons, K.G., Altman, D.G., Vergouwe, Y., and Royston, P. *Prognosis and prognostic research: Application and impact of prognostic models in clinical practice.* Bmj. 2009; **338**: p. b606.
61. Stern, M.P., Williams, K., González-Villalpando, C., Hunt, K.J., and Haffner, S.M. *Does the metabolic syndrome improve identification of individuals at risk of type 2 diabetes and/or cardiovascular disease?* Diabetes Care. 2004; **27**(11): p. 2676-2681.
62. Cameron, A., Magliano, D., Zimmet, P., Welborn, T., Colagiuri, S., Tonkin, A*., et al*. *The metabolic syndrome as a tool for predicting future diabetes: The ausdiab study.* Journal of internal medicine. 2008; **264**(2): p. 177-186.
63. Abdul-Ghani, M.A., Lyssenko, V., Tuomi, T., DeFronzo, R.A., and Groop, L. *Fasting versus postload plasma glucose concentration and the risk for future type 2 diabetes results from the botnia study.* Diabetes Care. 2009; **32**(2): p. 281-286.
64. 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**(3): p. 443-450.
65. Alssema, M., Feskens, E., Bakker, S., Gansevoort, R., Boer, J., Heine, R*., et al*. *Finse vragenlijst redelijk goede voorspeller van het optreden van diabetes in nederland.* Nederlands tijdschrift voor geneeskunde. 2008; **152**(44).
66. Mainous, A.G., Diaz, V.A., and Everett, C.J. *Assessing risk for development of diabetes in young adults.* The Annals of Family Medicine. 2007; **5**(5): p. 425-429.
67. Sun, F., Tao, Q., and Zhan, S. *An accurate risk score for estimation 5-year risk of type 2 diabetes based on a health screening population in taiwan.* Diabetes research and clinical practice. 2009; **85**(2): p. 228-234.
68. Rahman, M., Simmons, R.K., Harding, A.-H., Wareham, N.J., and Griffin, S.J. *A simple risk score identifies individuals at high risk of developing type 2 diabetes: A prospective cohort study.* Family practice. 2008; **25**(3): p. 191-196.
69. Hippisley-Cox, J., Coupland, C., Robson, J., Sheikh, A., and Brindle, P. *Predicting risk of type 2 diabetes in england and wales: Prospective derivation and validation of qdscore.* Bmj. 2009; **338**: p. 812-816.
70. Li, J., Bornstein, S.R., Landgraf, R., and Schwarz, P.E. *Validation of a simple clinical diabetes prediction model in a middle-aged, white, german population.* Archives of internal medicine. 2007; **167**(22): p. 2528-2529.
71. Lyssenko, V., Jonsson, A., Almgren, P., Pulizzi, N., Isomaa, B., Tuomi, T*., et al*. *Clinical risk factors, DNA variants, and the development of type 2 diabetes.* New England Journal of Medicine. 2008; **359**(21): p. 2220-2232.
72. Nichols, G.A. and Brown, J.B. *Validating the framingham offspring study equations for predicting incident diabetes mellitus.* Am J Manag Care. 2008; **14**(9): p. 574-580.
73. Kahn, H.S., Cheng, Y.J., Thompson, T.J., Imperatore, G., and Gregg, E.W. *Two risk-scoring systems for predicting incident diabetes mellitus in us adults age 45 to 64 years.* Annals of Internal Medicine. 2009; **150**(11): p. 741-751.
74. Rathmann, W., Martin, S., Haastert, B., Icks, A., Holle, R., Löwel, H*., et al*. *Performance of screening questionnaires and risk scores for undiagnosed diabetes: The kora survey 2000.* Archives of internal medicine. 2005; **165**(4): p. 436-441.
75. Heldgaard, P.E. and Griffin, S. *Routinely collected general practice data aids identification of people with hyperglycaemia and metabolic syndrome.* Diabetic medicine. 2006; **23**(9): p. 996-1002.

76. Park, P., Griffin, S., Sargeant, L., and Wareham, N. *The performance of a risk score in predicting undiagnosed hyperglycemia.* Diabetes Care. 2002; **25**(6): p. 984-988.
77. Witte, D., Shipley, M., Marmot, M., and Brunner, E. *Performance of existing risk scores in screening for undiagnosed diabetes: An external validation study.* Diabetic medicine. 2010; **27**(1): p. 46-53.
78. Bergmann, A., Li, J., Wang, L., Schulze, J., Bornstein, S., and Schwarz, P. *A simplified finnish diabetes risk score to predict type 2 diabetes risk and disease evolution in a german population.* Hormone and metabolic research. 2007; **39**(09): p. 677-682.
79. Saaristo, T., Peltonen, M., Lindström, J., Saarikoski, L., Sundvall, J., Eriksson, J.G*., et al*. *Cross-sectional evaluation of the finnish diabetes risk score: A tool to identify undetected type 2 diabetes, abnormal glucose tolerance and metabolic syndrome.* Diabetes and Vascular Disease Research. 2005; **2**(2): p. 67-72.
80. Franciosi, M., De Berardis, G., Rossi, M.C., Sacco, M., Belfiglio, M., Pellegrini, F*., et al*. *Use of the diabetes risk score for opportunistic screening of undiagnosed diabetes and impaired glucose tolerance the igloo (impaired glucose tolerance and long-term outcomes observational) study.* Diabetes Care. 2005; **28**(5): p. 1187-1194.
81. Korhonen, P.E., Jaatinen, P.T., Aarnio, P.T., Kantola, I.M., and Saaresranta, T. *Waist circumference home measurement—a device to find out patients in cardiovascular risk.* The European Journal of Public Health. 2009; **19**(1): p. 95-99.
82. Moons, K.G., Kengne, A.P., Grobbee, D.E., Royston, P., Vergouwe, Y., Altman, D.G*., et al*. *Risk prediction models: Ii. External validation, model updating, and impact assessment.* Heart. 2012: p. heartjnl-2011-301247.
83. 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**(8): p. 1418-1420.
84. Laatikainen, T., Dunbar, J.A., 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**(1): p. 249.
85. Vermunt, P., Milder, I., Wielaard, F., van Oers, J., and Westert, G. *An active strategy to identify individuals eligible for type 2 diabetes prevention by lifestyle intervention in dutch primary care: The aphrodite study.* Family practice. 2010; **27**(3): p. 312-319.
86. Dhippayom, T., Chaiyakunapruk, N., and Krass, I. *How diabetes risk assessment tools are implemented in practice: A systematic review.* Diabetes research and clinical practice. 2014.
87. Steyerberg, E.W., Moons, K.G., van der Windt, D.A., Hayden, J.A., Perel, P., Schroter, S*., et al*. *Prognosis research strategy (progress) 3: Prognostic model research.* PLoS medicine. 2013; **10**(2): p. e1001381.
88. Cardona-Morrell, M., Rychetnik, L., Morrell, S.L., Espinel, P.T., and Bauman, A. *Reduction of diabetes risk in routine clinical practice: Are physical activity and nutrition interventions feasible and are the outcomes from reference trials replicable? A systematic review and meta-analysis.* BMC Public Health. 2010; **10**(1): p. 653.
89. Simmons, R., Unwin, N., and Griffin, S. *International diabetes federation: An update of the evidence concerning the prevention of type 2 diabetes.* Diabetes research and clinical practice. 2010; **87**(2): p. 143-149.

# 04

# Bellville South Study

## Ethical considerations

This study was approved by the Faculty of Health and Wellness Sciences Ethics Committee of the Cape Peninsula University of Technology (Project number: CPUT/HW-REC 2008/002 and CPUT/HW-REC 2010/H017) and the University of Stellenbosch Ethics Committee (N10/05/142), and was conducted according to the code of ethics of the World Medical Association (Declaration of Helsinki). All participants signed written informed consent after all the procedures had been fully explained in the language of their choice. In addition, permission was also sought from other relevant authorities such as city and community authorities. These authorities granted permission to operate in the community and also to make use of designated places such as community halls or nearby schools for data and sample collection.

## Research setting and population

The data from the Cape Town Bellville-South cohort will serve as the basis for model validation in this thesis. The Bellville South community in Cape Town, South Africa, has been the center of the Chronic Diseases of Lifestyle study for many years and the data suggested for this study was collected from mid-January 2008 to March 2009 (cohort 1), and from January to November 2011 (cohort 2). Bellville South is located within the northern suburbs of Cape Town, South Africa. It is traditionally a community of a mixed ancestry ethnic population group formed in the late 1950s. The area, often referred to as a township, is an underdeveloped urban residential area that was previously reserved for non-Caucasians, including the mixed ancestry group. The population has ancestry from Khoisan, African, European Caucasian and a small amount of Asian populations [1]. According to the 2001 population census (the latest census done in Bellville), its population stands at approximately 26 758 with the mixed ancestry group accounting for 80.48% (21 536) [2]. The target population for this study were subjects between the ages of 35 – 65 years and their number was estimated to be 6 500, however other age groups were also included and the final range was 16 – 95 years of age [3].

## Research design

The cohort was initially established for follow up for insulin resistance and its sequel in randomly selected mixed ancestry individuals. The random sampling of the population was conducted using a map of Bellville South, where a list of streets from each stratum was prepared. The streets were then classified as short, medium and long, based on the number of houses within each street. Streets with houses ≤ 22 were classified as short, houses 23–40 as medium and long streets were those with > 40 houses. A total of 16 short streets representing approximately 190 houses, 15 medium streets representing approximately 410 houses and 12 long streets representing approximately 400 houses were randomly selected across the different strata. From the selected streets, all household members meeting the selection criteria were eligible to participate in the study. Community authorities requested that participants outside the random selection area also benefit from the study; these were included, but given a different code.

## Recruitment strategy

To publicize the study, information regarding the project was circulated to the community through the use of a local newspaper (the Tygerberger) and radio station (Radio Tygerberg). Brochures and fliers were spread among the residents via school children and taxis. The team responsible for recruitment, unemployed matriculants, was managed by a qualified, retired nurse who lived within the community. Additionally, a local celebrity suffering from diabetes was involved in a roadshow strategy to reach out to potential participants, particularly in the targeted streets. Recruited subjects were visited by the recruitment team the evening before participation and reminded of all the study instructions. The instructions included overnight fasting, abstinence from drinking alcohol or consumption of any fluids in the morning of participation. Since the participants were required to bring in an early morning mid-stream urine sample, they were provided with a sterile container as well as instructions on how to collect the sample Furthermore, participants were encouraged to bring along their medical/clinic cards and/or drugs they were currently using.

## Data collection

To obtain the medical information and samples of the participants, a detailed protocol describing data collection procedures (questionnaires and physical examination) was developed. The team members, consisting of professional nurses and field workers, were trained, and a pilot study in a neighboring community with similar demographics was performed, to validate the questionnaire and to synergize the workflow. A supervisor was allocated for each team to monitor the performance of the personnel and calibrate equipment according to a standard protocol. In addition, weekly meetings were held to assess progress, solve problems and re-train the research team (if necessary). A questionnaire designed to retrospectively obtain information on lifestyle factors such as smoking and alcohol consumption, physical activity, diet, family history of CVD and diabetes mellitus (DM), demographics and medication use was administered by trained personnel. The questionnaire was adapted from several existing standards and recognized sources [4, 5] and was also pre-tested in a neighboring community with similar demographics. Information about medication taken by participants was also obtained through clinic cards and record of drugs that participants brought to the study site. The more detailed the information retrieved, the more accurate and complete the database.

Clinical measurements obtained included: height, weight, hip and waist circumferences and blood pressure. Measurements were carried out by qualified healthcare professionals who underwent training to standardize all measurements prior to the commencement of the study. Blood pressure measurements were performed according to World Health Organization (WHO) guidelines [6]. Measurements were performed using a semi-automatic digital blood pressure monitor (Rossmax MJ90, USA) on the right arm, in sitting and ambulatory position. After a 10 minute rest period, three readings were taken at 5 minute intervals and the lowest of the three readings was taken as the blood pressure. Weight was determined on a Sunbeam EB710 digital bathroom scale, which was calibrated and standardized using a weight of known mass. Weight measurements were recorded to the nearest 0.1 kilograms and taken with each subject in light clothing, without shoes and socks. Height was recorded in centimeters, to one decimal place, using a stadiometer, with subjects standing on a flat surface at a right angle to the vertical board of the stadiometer. Body Mass Index (BMI) was calculated as weight per square meter ($kg/m^2$). Waist circumference was

measured using a non-elastic tape at the level of the narrowest part of the torso as seen from the anterior view. When difficult to observe the waist narrowing, especially in obese subjects, the waist circumference was measured between the ribs and the iliac crest. Hip circumference was measured around the widest segment of the buttocks. All anthropometric measurements were performed three times and the average measurement used for analysis.

All participants, except the self-reported type 2 diabetic subjects (confirmed by either medical card record or drugs in use), underwent a 75g OGTT as prescribed by the WHO, with fasting blood glucose in all participants. Categories of glucose tolerance were defined using the 1998 WHO criteria [6]. Blood samples were transported daily in an ice-pack box for processing at the Metropolis Private Pathology Laboratory (Century City, Cape Town). Serum creatinine was determined using the kinetic-Jaffe reaction (Cobas 6000, Roche Diagnostics, USA) and the result used to determine the glomerular filtration rate using the MDRD and CKD-EPI formulae. Plasma glucose was measured by enzymatic hexokinase method (Cobas 6000, Roche Diagnostics, USA). Glycosylated hemoglobin ($HbA_{1c}$) was assessed by turbidimetric inhibition immunoassay (Cobas 6000, Roche Diagnostics, USA). This method is National Glycohaemoglobin Standardization Programme (NGSP) certified according to Roche Diagnostics. High density lipoprotein cholesterol and triglycerides were estimated by enzymatic colorimetric methods (Cobas 6000, Roche Diagnostics, USA). Low density lipoprotein cholesterol was calculated using Friedwald's formula. Serum cotinine was measured by chemiluminescent assay (Immulite 1000, Siemens, Germany). This laboratory was accredited and performed all the necessary and required daily, weekly and monthly internal and external quality control.

To maintain patient confidentiality, all data captured sheets containing clinical and demographic info of each patient were coded by a study number and the specimen labeled accordingly, thus any info leading to the identity of the subjects was kept separately. All consent forms and questionnaires were stored in confidential files and securely locked away.

## Sample size

This was a pre-collected dataset so sample size calculations for the number of samples to be collected are not relevant in this study. Sample size for model development and validation studies is not well defined, however studies have suggested that a minimum of 100 events and 100 non-events [7], and more recently, a minimum of 100 event and 200 non-events [8], are required for external validation studies. The dataset used in this study met this criteria.

## References


1. Quintana-Murci, L., Harmant, C., Quach, H., Balanovsky, O., Zaporozhchenko, V., Bormans, C., *et al*. *Strong maternal khoisan contribution to the south african coloured population: A case of gender-biased admixture.* The American Journal of Human Genetics. 2010; **86**(4): p. 611-620.
2. CoCT. *2011 census – ward 009*. 2013; Available from: https://www.capetown.gov.za/en/stats/Documents/2011%20Census/Wards/2011_Census_CT_Ward_009_Profile.pdf.
3. CoCT. *Census 2001 – ward 9*. 2006, City of Cape Town: Cape Town.
4. Bradshaw, D., Bourne, D., Schneider, M., and Sayed, R. *Mortality patterns of chronic diseases of lifestyle in south africa.* Chronic diseases of lifestyle in South Africa. 1995; **2005**: p. 9-22.
5. Ewing, J.A. *Detecting alcoholism: The cage questionnaire.* Jama. 1984; **252**(14): p. 1905-1907.
6. Alberti, K.G.M.M. and Zimmet, P.f. *Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus. Provisional report of a who consultation.* Diabetic medicine. 1998; **15**(7): p. 539-553.
7. Vergouwe, Y., et al. *Substantial effective sample sizes were required for external validation studies of predictive logistic regression models.* J Clin Epidemiol. 2005; **58**(5): p. 475-83.
8. Collins, Gary S., Emmanuel O. Ogundimu, and Douglas G. Altman. *Sample size considerations for the external validation of a multivariable prognostic model: a resampling study*. Statistics in medicine. 2016; **35**(2): p. 214-226.

# PART II

# MODEL VALIDATION

05

# Independent external validation and comparison of prevalent diabetes risk prediction models in a mixed-ancestry population of South Africa

Katya Masconi, Tandi E. Matsha, Rajiv T. Erasmus and Andre P. Kengne

*Diabetology & Metabolic Syndrome* 2015, **7**:42

## ABSTRACT

**Background**: Guidelines increasingly encourage the use of multivariable risk models to predict the presence of prevalent undiagnosed type 2 diabetes mellitus worldwide. However, no single model can perform well in all settings and available models must be tested before implementation in new populations. We assessed and compared the performance of five prevalent diabetes risk models in mixed-ancestry South Africans.

**Methods**: Data from the Cape Town Bellville-South cohort were used for this study. Models were identified via recent systematic reviews. Discrimination was assessed and compared using the C-statistic and non-parametric methods, and calibration was assessed via calibration plots, before and after recalibration through intercept adjustment.

**Results:** Seven hundred thirty-seven participants (27% male), mean age, 52.2 years, were included, among whom 130 (17.6%) had prevalent undiagnosed diabetes. The highest C-statistic for the five prediction models was recorded with the Kuwaiti model [C-statistic 0.68: 95% confidence: 0.63–0.73] and the lowest with the Rotterdam model [0. 64 (0.59 – 0.69)]; with no significant statistical differences when the models were compared with each other (Cambridge, Omani and the simplified Finnish models). Calibration ranged from acceptable to good, however over- and underestimation was prevalent. The Rotterdam and the Finnish models showed significant improvement following intercept adjustment.

**Conclusions**: The wide range of performances of different models in our sample highlight the challenges of selecting an appropriate model for prevalent diabetes risk prediction in different settings.

## BACKGROUND

Diabetes mellitus, type 2 diabetes in particular, is a growing epidemic worldwide with developing countries currently paying the highest toll [1]. In 2013 there were approximately 382 million individuals with type 2 diabetes, and this number will surge to approximately 592 million by 2035 [1]. This rapid rise of diabetes will result in an even greater and more profound burden which developing countries are not equipped to handle. Type 2 diabetes in developing countries is further characterized by a low detection rate with a high proportion of people being undiagnosed. Strategies are therefore needed for early detection and risk stratification such that treatment measures can be implemented to prevent the onset or delay the progression of related complications.

The use of multivariable risk prediction models have been advocated as practical and potentially affordable approaches for improving the detection of undiagnosed diabetes. Accordingly, guidelines, including the International Diabetes Federation, increasingly promote the use of reliable, simple and practical risk scoring systems or questionnaires and derivatives for diabetes risk screening around the world [2, 3]. During the last two decades, numerous diabetes prediction models have been developed. However, only a few models have been externally validated, and generally not in developing countries [4, 5]. Consequently, many developing countries have to rely on prediction models developed in other populations and not necessarily validated in their context. However, issues relating to differences in case-mix across populations, inherent to the development of models, can severely affect the applicability of a model in different settings [6, 7].

This study aimed to validate and compare the performance of selected common models for predicting prevalent undiagnosed diabetes based upon non-invasively measured predictors, in mixed ancestry South Africans.

# METHODS

## Study population and design of study

The Cape Town Bellville-South study data served as the basis for model validations [8]. Bellville-South is located within the Northern suburbs of Cape Town, South Africa, and is traditionally a mixed-ancestry township formed in the late 1950s. According to the 2001 population census, its population stands at approximately 26 758 with 80.48% (21 536) consisting of the mixed ancestry individuals [22]. The study was approved by the Ethics Committee of the Cape Peninsula University of Technology (CPUT/HW-REC 2008/002 and CPUT/HW-REC 2010) and Stellenbosch University (N09/05/146).

The Bellville South Study was a cross-sectional study conducted from mid-January 2008 to March 2009 (cohort 1), and from January 2011 to November 2011 (cohort 2). The target population for this study were subject's ≥ 35 years. Using a map of Bellville South obtained from the Bellville municipality, random sampling was approached as follows: first, the area was divided into six strata; second, within each strata the streets were classified as short (≤22 houses), medium (23 to 40 houses) and long (≥40 houses) streets based on the number of houses. Two of each respective streets were randomly selected from each strata. In those instances where the numbers of houses were too few, a short or a medium street were randomly selected and added to such a stratum. The result was a total of 16 short streets representing approximately 190 houses, 15 medium (approximately 410 houses) and 12 long streets (approximately 400 houses). From the selected streets, all household members meeting the selection criteria were invited to participate in the study. One thousand subjects who met the criteria were approached and 642 participated in the study. In addition, community authorities requested that willing participants outside the random selection area should benefit from the study. Therefore volunteers (304 in 2008-2009 [cohort 1]), and 308 in 2011 [cohort2]) from the same community but who were not part of the randomly selected streets or did not meet the age criteria, were also included.

**Recruitment strategy**

Information regarding the project was disseminated to residents through the local radio station, community newspaper, brochures and fliers; the latter bearing information about the project and distributed through school children and taxis by the recruitment team. Additionally, a 'road show' strategy that involved a celebrity suffering from diabetes from the same community was also used, especially in the targeted streets. Recruited subjects were visited by the recruitment team the evening before participation and reminded of all the survey instructions. These included overnight fasting, abstinence from drinking alcohol or consumption of any fluids in the morning of participation. Since the participants were required to bring in an early morning mid-stream urine sample, they were provided with a sterile container as well as instructions on how to collect the sample. Furthermore, participants were encouraged to bring along their medical/clinic cards and/or medication they were currently using.

**Identification of prediction models**

Existing prediction models were obtained from a systematic review by Brown *et al* [9]. The search strategy from Brown's paper was re-ran in PubMed for the time-period up to April 2014, to identify possible new models. The following string search was used, as per Brown *et al*: ((“type 2 diabetes” OR “hyperglycaemia” OR “hyperglycemia”) AND (“risk scores”)). Selected models were only those developed to predict the presence of undiagnosed diabetes. We focused on models developed using non-invasively measured predictors which were available in the Bellville-South cohort database. Models were excluded if they were developed for male and female individuals separately.

**Outcome and predictors' definition and measurements**

The main outcome was newly diagnosed type 2 diabetes from the standard oral glucose tolerance test (OGTT), applying the World Health Organisation (WHO) criteria (i.e. fasting plasma glucose ≥ 7.0 mmol/L and/or 2-hour plasma glucose ≥ 11.1 mmol/L) [10]. At the

baseline evaluation, conducted between 2008 and 2011, participants received a face-to-face interview administered by trained personnel to collect data on personal and family history of diabetes mellitus, cardiovascular disease (CVD) and treatments; habits including smoking, alcohol consumption, physical activity and diet; demographics and education.

Clinical measurements included: height, weight, hip and waist circumferences and blood pressure (BP). BP measurements used a semi-automatic digital blood pressure monitor (Rossmax MJ90, USA) on the right arm, in sitting position, after a 10-minute rest. The lowest value from three consecutive measurements 5 minutes apart was used in the current analysis. Weight to the nearest 0.1 kg was determined on a Sunbeam EB710 digital bathroom scale, with each subject in light clothing, without socks and shoes. Height to the nearest centimetre was measured with a stadiometer, with subjects standing on a flat surface. Body mass index (BMI) was calculated as weight per square meter ($kg/m^2$).

Blood samples were collected and processed for a wide range of biochemical markers. Plasma glucose was measured by enzymatic hexokinase method (Cobas 6000, Roche Diagnostics, USA). High density lipoprotein cholesterol (HDL-c) and triglycerides (TG) were estimated by enzymatic colorimetric methods (Cobas 6000, Roche Diagnostics, USA).

**Assessment of model performance**

The original selected models were validated for the overall data and subsets using the formulas, without any recalibration. The predicted probability of undiagnosed diabetes for each participant was computed using the baseline measured predictors. The performance was expressed in terms of discrimination and calibration. Discrimination describes the ability of the model's performance in distinguishing those at a high risk of developing diabetes from those at low risk [11]. The discrimination was assessed and compared using concordance (C) statistic and non-parametric methods [12].

Calibration describes the agreement between the probability of the outcome of interest as estimated by the model, and the observed outcome frequencies [13]. It was assessed graphically by plotting the predicted risk against the observed outcome rate. The agreement between the expected (E) and observed (O) rates (E/O) was assessed overall and within pre-specified groups of participants. The 95% confidence intervals for the expected/observed

probabilities (E/O) ratio were calculated assuming a Poisson distribution [14]. We also calculated 1) the Yates slope, which is the difference between mean predicted probability of type 2 diabetes for participants with and without prevalent undiagnosed diabetes, with higher values indicating better performance; and 2) the Brier score, which is the squared difference between predicted probability and actual outcome for each participant with values ranging between 0 for a perfect prediction model and 1 for no match in prediction and outcome [11, 13]. To determine optimal cut-off for maximising the potential effectiveness of a model, the Youden's J statistic (Youden's index) was used to determine the best threshold [15], with sensitivity, specificity and percentage of correctly classified individuals determined for each threshold. The main analysis was done for the overall cohort and for subgroups defined by sex, age (<60 vs ≥60 years) and BMI (<25 kg/$m^2$ vs ≥25 g/$m^2$).

**Sensitivity analysis**

To improve performance and eliminate differences in diabetes prevalence between the development population and the test population, models were recalibrated to the test-population-specific prevalence using intercept adjustment [16]. The correction factor calculated is based on the mean predicted risk and the prevalence in the validation set, and is the natural logarithm of the odds ratio of the mean observed prevalence and the mean predicted risk [16]. To assess the potential effect on model performance of validation studies from complete case analysis, we also assess the discrimination of model across five datasets after the application of multiple data imputation procedures to fill missing data.

## Results

### Identification of prediction models

Five non-invasive prevalent diabetes prediction models were selected for validation following the sifting process; the Cambridge Risk Score [17], Kuwaiti Risk Score [18], Omani Diabetes Risk Score [19], Rotterdam Predictive Model 1 [20] and the simplified Finnish Diabetes Risk Score [21] (Fig. 1). Table 1 summarizes the models' characteristics. All models included age as a predictor, while a range of other predictors were variably combined in models. These included: sex, BMI, use of antihypertensive medication, family history of diabetes, waist circumference, past or current smoking and the use of corticosteroids. Additional 1: Table S1 comprises of the full equations for each of the models.

**Fig. 1** – Flow diagram of selected studies

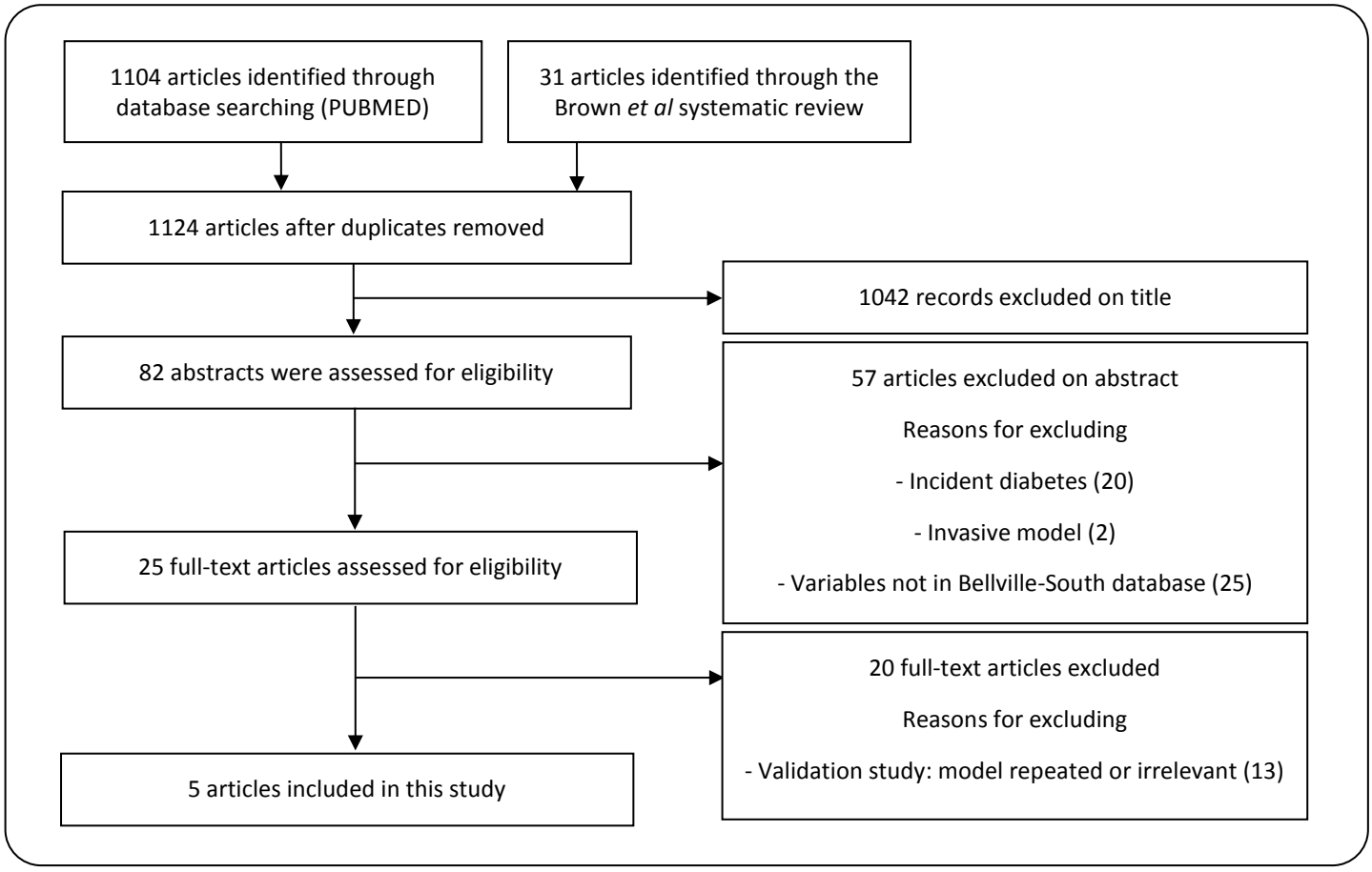


**Table 1** - Overview of the included prevalent diabetes risk prediction models and their performance for the original model and the intercept adjusted model

| | Incident diabetes risk models | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Description** | **Cambridge risk score** | | **Kuwaiti risk score** | | **Omani risk score** | | **Rotterdam Predictive model 1** | | **Simplified Finnish risk score** | | **Bellville South** |
| Authors | Griffin *et al[17]* | | Al Khalaf *et al* [18] | | Al-Lawati & Tuomilehto [19] | | Baan *et al* [20] | | Bergmann *et al* [21] | | - |
| Year published | 2000 | | 2008 | | 2007 | | 1999 | | 2007 | | - |
| Country | UK | | Kuwaiti | | Oman | | Netherlands | | Germany | | South Africa |
| Validation | External [22-28] | | None | | External [26] | | External [27-29] | | External [26-28] | | - |
| Sample size | 1077 | | 460 | | 4881 | | 1016 | | 526 | | 737 |
| Type of study | Cross-sectional | | Cross-sectional | | Cross-sectional | | Cohort | | Cohort | | Cohort |
| Age range | 40 – 79 | | 20 – 40 | | 20 – 80 | | 55 – 75 | | 41 – 79 | | 15-95 |
| Population | Caucasian | | Arab | | Arab | | Caucasian | | Caucasian | | Mixed ancestry |
| Diagnosis of diabetes | FBG ≥ 7.0 mmol/l; 2h glucose ≥ 11.1 mmol/l | | FBG ≥ 7.0 mmol/l; Random glucose ≥ 11.1 mmol/l | | FBG ≥ 7.0 mmol/l; 2h glucose ≥ 11.1 mmol/l | | FBG ≥ 7.0 mmol/l; 2h glucose ≥ 11.1 mmol/l | | FBG ≥ 7.0 mmol/l; 2h glucose ≥ 11.1 mmol/l | | FBG ≥ 7.0 mmol/l; 2h glucose ≥ 11.1 mmol/l |
| Development C-statistic | 0.80 (0.68 – 0.91) | | 0.82 (NS) | | 0.83 (0.82 – 0.84) | | 0.68 (0.64 – 0.72) | | 0.75 (0.68 – 0.81) | | - |
| **Predictors** | | | | | | | | | | | |
| *Age* | Yes | | Yes | | Yes | | Yes | | Yes | | Yes |
| *Sex* | Yes | | No | | No | | Yes | | Yes | | Yes |
| *BMI* | Yes | | No | | Yes | | Yes | | Yes | | Yes |
| *Use of HTN drugs* | Yes | | Yes | | No | | Yes | | Yes | | Yes |
| *Family history* | Yes | | Yes | | Yes | | No | | Yes | | Yes |
| *WC* | No | | Yes | | Yes | | No | | No | | Yes |
| *Smoking* | Yes | | No | | No | | No | | Yes | | Yes |
| *Corticosteroids* | Yes | | No | | No | | No | | Yes | | Yes |
| *Systolic/diastolic* | No | | No | | Yes | | No | | No | | Yes |
| **Performance** | Original | Adjusted | Original | Adjusted | Original | Adjusted | Original | Adjusted | Original | Adjusted | |
| E/O (95% CI) | 1.81 (1.09-2.52) | 1.22 (0.61-1.83) | 0.72 (0.40-1.12) | 0.94 (0.47-1.41) | 1.28 (0.63-1.93) | 1.06 (0.47-1.66) | 0.54 (0.50-1.04) | 0.98 (0.91-1.05) | 0.26 (0.13-0.39) | 0.89 (0.51-1.26) | - |
| Brier score | 0.193 | 0.160 | 0.141 | 0.143 | 0.164 | 0.157 | 0.147 | 0.140 | 0.157 | 0.143 | - |
| Yates slope | 0.379 | 0.379 | 0.496 | 0.496 | 0.392 | 0.392 | 0.971 | 0.971 | 0.491 | 0.491 | - |
| C-statistic (95% CI) | 0.67 (0.62-0.72) | - | 0.68 (0.63- 0.73) | - | 0.66 (0.61- 0.70) | - | 0.64 (0.59- 0.69) | - | 0.67 (0.62-0.71) | - | - |
| Optimal threshold | 0.29 | 0.16 | 0.13 | 0.18 | 0.12 | 0.09 | 0.20 | 0.18 | 0.02 | 0.08 | - |
| Sensitivity | 65 | 65 | 61 | 61 | 85 | 85 | 57 | 57 | 75 | 75 | - |
| Specificity | 61 | 61 | 63 | 63 | 42 | 42 | 65 | 65 | 48 | 48 | - |
| Correctly classified | 62 | 62 | 63 | 63 | 50 | 50 | 64 | 64 | 53 | 53 | - |

*UK, United Kingdom; FBG , fasting blood glucose; OGTT, 2 hour post load oral glucose tolerance test; C-statistic, concordance statistic; NS, not stated; BMI, body mass index; HTN, hypertension; WC, waist circumference; E/O, ratio expected/observed event rate; 95% CI, 95% confidence interval.

**Participants' characteristics**

A total of 1256 participants were examined in the Bellville South studies, including 173 with a history of diagnosed diabetes who were excluded. A further 346 participants were excluded for missing data on predictors or the outcome variable. Therefore the final dataset comprised of 737 participants, of whom 580 (78.70%) were female. In the Additional file 2: Table S2, we compare the profile of participants in the final sample vs. that of participants excluded for missing data. Excluded participants comprised of more men (27.2 vs. 21.3%, p=0.012), were more likely to display a better lifestyle profile for alcohol intake (18.8% vs. 28.1%, p <0.001), smoking (31.8% vs. 43.8%, p<0.001), lower family history of diabetes (all p $\leq$0.001), higher systolic blood pressure (126 vs. 123 mmHg, p=0.009) and lower triglycerides (1.4 vs. 1.5 mmol/l, p=0.043); although absolute differences were mostly clinically trivial.

The baseline profile for men and women included in the study is described in Table 2. The mean baseline age was 51.2 years overall, and 53.5 and 52.1 years, respectively in men and women (p=0.311). The BMI (p<0.001) waist circumference (p=0.024) and fasting blood glucose (p=0.036) were significantly higher in women, while smoking (p <0.001) and alcohol consumption (p <0.001) were more frequent among men.

**Table 2** - Characteristics comparison of participants with valid data

| | Male (157) | Female (580) | p-value | Overall (737) |
|---|---|---|---|---|
| Prevalent undiagnosed DM (%) | 22 (14.0) | 108 (18.6) | 0.220 | 130 (17.3) |
| Age (years) | 53.5 (15.0) | 52.1 (14.3) | 0.311 | 52.2 (14.5) |
| Body mass index (kg/m2) | 25.5 (5.8) | 29.6 (7.0) | <0.001 | 29.4 (7.1) |
| Waist circumference (cm) | 92.5 (15.2) | 95.6 (14.7) | 0.024 | 95.9 (14.9) |
| Hypertensive medication (%) | 43 (27.4) | 208 (35.9) | 0.059 | 251 (34.1) |
| Smoking status (% smoking) | 88 (56.1) | 235 (40.5) | <0.001 | 323 (43.8) |
| Systolic blood pressure (mmHg) | 124.3 (16.6) | 121.6 (19.2) | 0.077 | 122.0 (18.7) |
| Diastolic blood pressure (mmHg) | 75.6 (11.1) | 74.7 (12.1) | 0.365 | 74.7 (11.9) |
| Height (m) | 1.7 (0.1) | 1.6 (0.1) | <0.001 | 1.6 (0.1) |
| Mother having diabetes (%) | 17 (10.8) | 92 (15.9) | 0.147 | 109 (14.8) |
| Father having diabetes (%) | 14 (8.9) | 44 (7.6) | 0.702 | 58 (7.9) |
| Sister having diabetes (%) | 12 (7.6) | 80 (13.8) | 0.053 | 92 (12.5) |
| Brother having diabetes (%) | 9 (5.7) | 49 (8.5) | 0.340 | 58 (7.9) |
| Fasting blood glucose (mmol/L) | 5.4 (1.4) | 5.7 (2.0) | 0.036 | 5.8 (1.9) |
| HDL (mmol/L) | 1.2 (0.4) | 1.3 (0.3) | 0.136 | 1.3 (0.3) |
| Weight (kg) | 72.3 (16.4) | 73.9 (17.7) | 0.290 | 74.1 (17.5) |
| Ever consumed alcohol (%) | 116 (73.9) | 240 (41.4) | <0.001 | 356 (48.3) |
| Current drinking (%) | 80 (51.0) | 127 (21.9) | <0.001 | 207 (28.1) |
| Using Corticosteroid use (%) | 1 (0.6) | 4 (0.7) | >0.99 | 5 (0.7) |
| Triglyceride (mmol/L) | 1.4 (0.9) | 1.4 (0.9) | 0.836 | 1.4 (0.9) |

**Prediction of prevalent undiagnosed diabetes in the overall sample**

A total of 130 participants (17.6%) had prevalent undiagnosed diabetes. This prevalence was similar in men vs. women (14% vs. 18.6%, p=0.220) (Table 2). Table 1 shows the discrimination for the selected prediction models in their original form in the overall sample. Discrimination was modest-to-acceptable and similar between models, with C-statistics (95% CI) ranging from 0.64 (0.59 – 0.69) for the Rotterdam model to 0.68 (0.63 – 0.73) for the Kuwaiti model

(all p>0.05 for C-statistics comparison; Additional file 3: Table S3). At the total population level, the absolute risk of prevalent diabetes was acceptably estimated by the Omani model, overestimated by 81% (9 to 152%) by the Cambridge model, underestimated by 74% (61 to 87%) by the Finnish model and marginally underestimated by the Kuwaiti and Rotterdam models (Table 1). The calibration curves are shown in Fig. 2. There was a systematic risk underestimation across the continuum of predicted probability by the Finnish and Rotterdam models, a selective upper strata risk overestimation by the Cambridge and Omani models, and a combination of both lower strata risk underestimation and upper strata risk overestimation by the Kuwaiti model. Comparison of the C-statistics from the development study and the models' performance in this population showed a drop in performance of all the models. Other performance measures are shown in Table 1.

**Prediction of prevalent undiagnosed diabetes in subgroups**

The performance of the original models across subgroups was parallel to that in the overall dataset (Table 3). When comparing patterns of predictions across complementary subgroups, only stand-alone differences were seen in performance for a subgroup, which was not carried through all performance measures. Estimates of C-statistics were broadly similar across complementary subgroups, except for the Omani and Finnish models across BMI subgroups, whereby lower estimates were always found in the overweight/obese subgroup. The pattern of the overall calibration (E/O) across complementary subgroups varied substantially across models. For instance, across gender subgroups, the overall diabetes risk was acceptably and equally predicted by the Omani model, equally underestimated by the Kuwaiti and Finnish models, equally overestimated by the Cambridge model, but acceptably estimated in men and underestimated in women by the Rotterdam model (Table 3). Other performance measures across subgroups are shown in Table 3.

**Fig. 2** - Calibration curves in the overall cohort for the models before (upper panel) and after the intercept adjustment (lower panel):

A: Cambridge Risk Score, B: Kuwaiti Risk Score, C: Omani Diabetes Risk Score, and D: Simplified Finnish Diabetes Risk Score. Calibration describes the agreement between the probability of undiagnosed diabetes as estimated by the model and the recorded frequencies of the outcome. The ideal calibration is graphically represented by the dotted diagonal line at 45°. Participants are grouped into percentiles across increasing predicted risk. The vertical lines at the bottom of the graph depict the frequency distribution of the calibrated probabilities of diabetes. *E/O, expected/observed ratio.

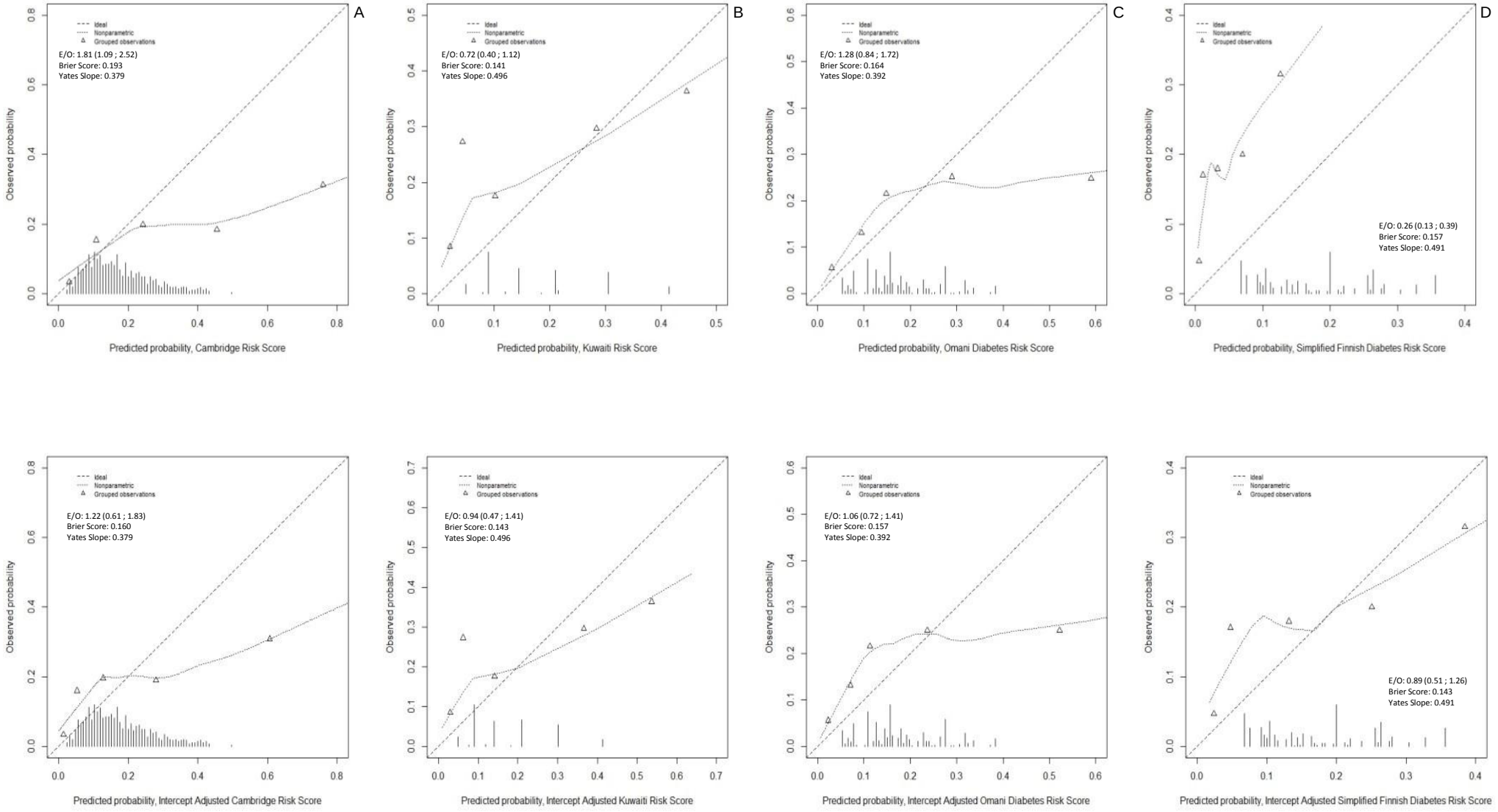


**Table 3:** Discrimination and calibration statistics for diabetes risk model performance in subgroups of participants by gender, age and body mass index (BMI)

| Models | | Male | Female | Age < 60 years | Age ≥ 60 years | BMI < 25 kg/m$^2$ | BMI ≥ 25 kg/m$^2$ |
|---|---|---|---|---|---|---|---|
| Cambridge Diabetes Risk Score [17] | E/O (95% CI) | 2.30 (1.21-3.37) | 1.71 (1.00-1.41) | 1.57 (0.71-2.44) | 2.10 (1.51-2.69) | 1.08 (0.55-1.61) | 1.96 (1.30-2.63) |
| | Brier score | 0.195 | 0.192 | 0.151 | 0.282 | 0.102 | 0.230 |
| | Yates slope | 0.373 | 0.384 | 0.368 | 0.384 | 0.450 | 0.368 |
| | C-statistic (95% CI) | 0.67 (0.56-0.78) | 0.67 (0.62-0.73) | 0.66 (0.60-0.72) | 0.65 (0.56-0.73) | 0.69 (0.58-0.79) | 0.64 (0.59-0.70) |
| Kuwaiti Risk Score [18] | E/O (95% CI) | 0.73 (0.40-1.06) | 0.72 (0.34-1.10) | 0.73 (0.37-1.10) | 0.71 (0.32-1.11) | 0.33 (0.20-0.46) | 0.81 (0.43-1.19) |
| | Brier score | 0.112 | 0.149 | 0.121 | 0.186 | 0.097 | 0.159 |
| | Yates slope | 0.588 | 0.468 | 0.476 | 0.449 | 0.890 | 0.468 |
| | C-statistic (95% CI) | 0.70 (0.58-0.82) | 0.67 (0.61-0.72) | 0.67 (0.61-0.74) | 0.65 (0.57-0.73) | 0.61 (0.51-0.72) | 0.66 (0.60-0.71) |
| Omani Diabetes Risk Score [19] | E/O (95% CI) | 1.33 (0.45-2.20) | 1.32 (0.65-2.00) | 1.26 (0.53-1.99) | 1.40 (0.60-2.20) | 1.16 (0.41-1.92) | 1.36 (0.71-2.01) |
| | Brier score | 0.137 | 0.173 | 0.140 | 0.221 | 0.096 | 0.194 |
| | Yates slope | 0.347 | 0.399 | 0.393 | 0.296 | 0.620 | 0.304 |
| | C-statistic (95% CI) | 0.62 (0.49-0.74) | 0.66 (0.61-0.71) | 0.66 (0.60-0.71) | 0.60 (0.52-0.68) | 0.71 (0.61-0.82) | 0.61 (0.56-0.67) |
| Rotterdam Predictive Model 1 [20] | E/O (95% CI) | 0.84 (-0.38-2.06) | 0.48 (0.45-0.93) | 0.52 (0.44-0.96) | 0.49 (0.39-0.88) | 0.72 (0.34-1.06) | 0.51 (0.45-0.96) |
| | Brier score | 0.117 | 0.155 | 0.125 | 0.199 | 0.096 | 0.168 |
| | Yates slope | 0.913 | 1.154 | 1.135 | 0.838 | 0.791 | 0.886 |
| | C-statistic (95% CI) | 0.62 (0.49-0.75) | 0.66 (0.60-0.72) | 0.62 (0.55-0.69) | 0.61 (0.52-0.69) | 0.61 (0.50-0.72) | 0.63 (0.57-0.69) |
| Simplified Finnish Diabetes Risk score [21] | E/O (95% CI) | 0.22 (0.09-0.35) | 0.32 (0.18-0.45) | 0.34 (0.18-0.50) | 0.26 (0.14-0.37) | 0.11 (0.06-0.15) | 0.34 (0.21-0.48) |
| | Brier score | 0.128 | 0.162 | 0.128 | 0.213 | 0.103 | 0.176 |
| | Yates slope | 0.538 | 0.591 | 0.487 | 0.608 | 1.345 | 0.562 |
| | C-statistic (95% CI) | 0.70 (0.59-0.81) | 0.66 (0.60-0.71) | 0.64 (0.58-0.71) | 0.67 (0.60-0.75) | 0.77 (0.69-0.86) | 0.62 (0.57-0.68) |

*E/O, expected/observed ratio

### Performance of the intercept adjusted models

As expected, intercept adjustment yielded acceptable agreement between predicted and observed prevalent diabetes rates at the total population level. A perfect agreement was also observed across the continuum of the predicted probability by the updated Rotterdam model. However, despite some attenuation, selective upper strata risk overestimations were apparent for other models.

### Model performance at the optimal threshold

The performances of models at the optimal thresholds are shown in Table 1. As anticipated, the optimal threshold probability for our sample varied across models and for the same model between the original and intercept adjusted versions. The sensitivity at the optimal threshold ranged from 61% for the Kuwaiti model to 85% with the Omani model, the specificity from 42% (Omani model) to 65% (Rotterdam model), and the proportion of participants correctly classified from 50% (Omani model) to 64% (Rotterdam model).

### Model performance after multiple imputation of missing data

The discrimination (C-statistic) of models across five datasets obtained after multiple imputation of missing data was very similar: 0.69 (0.64-0.73) for the Cambridge model, 0.69 (0.65-0.74) for the Kuwaiti model, 0.65 (0.61-0.69) for the Omani model, 0.65 (0.60-0.69) for the Rotterdam model and 0.66 (0.62-0.70) for the Finnish model (results not shown). The values were also very similar to those from the validation of models in the dataset comprising only of participants with complete data (Table 1).

## Discussion

To our knowledge, this is the largest and most comprehensive validation study of prevalent diabetes prediction models in a sub-Saharan African population. In the Bellville South cohort, the selected existing prediction models based upon non-invasive measured predictors had modest-to-acceptable discriminatory ability to predict prevalent undiagnosed diabetes, both

overall and within subgroups. Simple intercept adjustment had a mixed effect on the calibration performance of the models, while none of the models was significantly better than other models to be uniquely recommended for use in this setting. At the optimal probability thresholds, the best performing model would correctly classify only about 2/3rds of the population, indicating the existing scope for further improving the models' performance in this setting.

The need for diabetes screening programs is imperative in the reduction of the worldwide burden of complications from diabetes in undiagnosed individuals. In view of the large and continuously growing burden of diabetes, the Centre for Disease Control strongly advocates for diabetes screening programs. In its most recent guidelines for type 2 diabetes screening and diagnosis, the International Diabetes Federation has recommended that each health service should decide on programs to detect undiagnosed diabetes based on the prevalence and the resources available in that region [3]. In areas with limited care, such as developing countries, the detection programs are suggested to be opportunistic and should be limited to high-risk individuals. The World Health Organization African region promotes the screening of at-risk individuals in Africa in healthcare settings and social gatherings [30]. Risk assessment scores are feasible and cost-effective and can be considered, but applicability must be certain, with the required tests available in the area and the validation of that risk score in the population.

With the exception of the Kuwaiti model [18], all other models assessed in our study have been validated externally. The most validated appeared to be the Cambridge model [17], with C-statistics ranging from 0.67 to 0.83 across validation studies [23-25, 27, 28]. With a C-statistic of 0.67 in the Bellville South data set, the Cambridge model performance in this population fell to the bottom end of other validation study results. Similarly, the Finnish model's discrimination performance (C-statistic: 0.67) also compared with lower c-statistic's from validation studies [26-28]. The Rotterdam model mirrored the validation study results (0.64 vs. 0.63 – 0.65) [27-29], while the Omani model underperformed (C-statistic: 0.66) when compared to the only validation study the authors are aware of (C-statistic: 0.72) [26].

Through an attempt to improve calibration with simple intercept adjustment, the E/O ratios for all models were improved. Despite the expected decision that no model was ready for immediate implementation, the Rotterdam Predictive Model 1 showed the best improvement in calibration following this adjustment. A review by Brown et al in 2012 [9] of 17 undiagnosed Type 2 diabetes risk scores, which included all five models discussed here, determined that performance was not associated to the number of predictors in the model. Overall, validation studies showed a drop in model performance when tested in a new population, with the Rotterdam model having the lowest validation performance range, when compared to the other models. This was echoed in our results for the original Rotterdam model validation. The possible reasons to explain the drop in the performance of diabetes prediction models in a new population, some of which apply to our study, have been extensively discussed elsewhere [31].

At the optimal probability threshold, the models tested in our study would at best correctly detect 2/3rds of participants, with diagnostic performance mostly similar to those from published studies [26, 28]. This indicates the existing scope for improving the performance of diabetes prediction models in our setting. This could be done by adopting or developing models enriched with predictors to improve the predictive accuracy. Such an approach however, has to be balanced against the fact that the number of predictors and the complexity and cost of their measurements are severe limitations for their uptake in routine practice [31]. What is probably needed the most in resources limited settings like Africa, is evidence to confirm that the introduction of diabetes prediction models in routine practice will improve early detection of diabetes by healthcare practitioners, and the outcome of those diagnosed with diabetes in the long run.

The results of this study were strengthened by the diagnosis of diabetes based on OGTT, thus limiting the risk of misclassification. The age distribution was wide, including a vast majority of the high-risk population. A potential limitation of the study was the exclusion of some risk scores due to the necessary information being unavailable. The fewer number of males in the final dataset could have played a role in the performance of the models, owing to the

significant difference between the genders in BMI, a predictor in four out of the five models. No power estimation was done, in the absence of consensus methods for sample size estimation in model validation studies. However, studies have suggested that at least 100 events and 100 non-events were the minimum required samples for external validation studies [32]. These requirements were largely met in our main analysis. Our study participants comprised a subset of randomly selected individuals and subset of self-selected participants from the same community. In the absence of any influence on participants' selection of a prior knowledge of the association between relevant study outcomes and predictors included in tested model, any differential effect of the sample selection strategy on the discriminatory performance of tested models, is very unlikely. The prevalence of screen-detected diabetes in our randomly selected participant's alones has been estimated to be 18.1% [33], which is very close to the 17.6% found in the combined sample, suggested the absence of a differential effect on the calibration performance of models. The total number of participants with screen-detected diabetes in the combined sample precluded reliable stratified analyses to investigate and confirm the assumptions above. Finally, a substantial number of participants were excluded from the main analyses due to missing data on predictors included in models or on the status for prevalent undiagnosed diabetes. However, participants with complete data were mostly similar to those with missing data, particularly regarding the distribution of key predictors included in models such as age, gender and measures of adiposity. Therefore, differential effect on the model performance of validation based on complete case analyses, is very unlikely. This was confirmed with multiple imputation of the missing data yielding no difference in model performances. Indeed, in sensitivity analysis, the discriminatory performance of models was very similar across multiple imputed datasets, and not appreciable different from the performance based on complete case analysis. Furthermore, variables with high frequency of missingness were likely to be those that are very difficult to accurately measure in routine setting like family history of diabetes, and therefore, less indicated for uncritical inclusion in models for predicting diabetes across settings [34, 35].

## Conclusions

Our findings highlight how the performance of models differs across different populations, particularly calibration. This low performance can be explained by the obvious lack of transportability due to the differences in development and validation population characteristics and the effect case-mix difference has on model performance. With no model development in the mixed ancestry population of South Africa, selection of generalizable models for validation was limited. There is a great clinical need for a unique, robust and convenient tool for identifying undiagnosed diabetes and predicating future diabetes quicker and more economically in this South African population. Through efficient application of prediction models' improvement procedures, the final model would improve risk assessment specific to this community. With no acceptable validated model, unique model development is possibly the best way forward.

## Competing interests

Nothing to declare

## Acknowledgements

Katya Masconi was supported by a scholarship from the South African National Research Foundation and the Carl & Emily Fuchs Foundation.

## References

1. Guariguata, L., Whiting, D.R., Hambleton, I., Beagley, J., Linnenkamp, U., and Shaw, J.E. *Global estimates of diabetes prevalence for 2013 and projections for 2035.* Diabetes research and clinical practice. 2014; **103**(2): p. 137-149.
2. Alberti, K.G., Zimmet, P., and Shaw, J. *International diabetes federation: A consensus on type 2 diabetes prevention.* Diabetic medicine : a journal of the British Diabetic Association. 2007; **24**(5): p. 451-463.
3. Clinical Guidelines Task Force. *Global guidelines for type 2 diabetes*, ed. International Diabetes Federation. 2012, Geneva. 117.
4. Collins, G.S., Mallett, S., Omar, O., and Yu, L.M. *Developing risk prediction models for type 2 diabetes: A systematic review of methodology and reporting.* BMC medicine. 2011; **9**: p. 103.

5. Buijsse, B., Simmons, R.K., Griffin, S.J., and Schulze, M.B. *Risk assessment tools for identifying individuals at risk of developing type 2 diabetes.* Epidemiologic reviews. 2011; **33**(1): p. 46-62.
6. Moons, K.G., Kengne, A.P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D.G., *et al*. *Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio) marker.* Heart. 2012; **98**(9): p. 683-690.
7. Moons, K.G., Kengne, A.P., Grobbee, D.E., Royston, P., Vergouwe, Y., Altman, D.G., *et al*. *Risk prediction models: Ii. External validation, model updating, and impact assessment.* Heart. 2012; **98**(9): p. 691-698.
8. Matsha, T.E., Hassan, M.S., Kidd, M., and Erasmus, R.T. *The 30-year cardiovascular risk profile of south africans with diagnosed diabetes, undiagnosed diabetes, pre-diabetes or normoglycaemia: The bellville, south africa pilot study.* Cardiovascular journal of Africa. 2012; **23**(1): p. 5-11.
9. Brown, N., Critchley, J., Bogowicz, P., Mayige, M., and Unwin, N. *Risk scores based on self-reported or available clinical data to detect undiagnosed type 2 diabetes: A systematic review.* Diabetes research and clinical practice. 2012; **98**(3): p. 369-385.
10. Alberti, K.G. and Zimmet, P. *Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus provisional report of a who consultation.* Diabetic medicine : a journal of the British Diabetic Association. 1998; **15**: p. 539-553.
11. Kengne, A.P., Masconi, K., Mbanya, V.N., Lekoubou, A., Echouffo-Tcheugui, J.B., and Matsha, T.E. *Risk predictive modelling for diabetes and cardiovascular disease.* Critical reviews in clinical laboratory sciences. 2014; **51**(1): p. 1-12.
12. Moons, K.G., Kengne, A.P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D.G., *et al*. *Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker.* Heart. 2012; **98**(9): p. 683-690.
13. Steyerberg, E.W., Vickers, A.J., Cook, N.R., Gerds, T., Gonen, M., Obuchowski, N., *et al*. *Assessing the performance of prediction models: A framework for traditional and novel measures.* Epidemiology. 2010; **21**(1): p. 128-138.
14. Dobson, A.J., Kuulasmaa, K., Eberle, E., and Scherer, J. *Confidence intervals for weighted sums of poisson parameters.* Stat Med. 1991; **10**(3): p. 457-462.
15. Youden, W.J. *Index for rating diagnostic tests.* Cancer. 1950; **3**(1): p. 32-35.
16. Janssen, K.J., Moons, K.G., Kalkman, C.J., Grobbee, D.E., and Vergouwe, Y. *Updating methods improved the performance of a clinical prediction model in new patients.* Journal of clinical epidemiology. 2008; **61**(1): p. 76-86.
17. Griffin, S.J., Little, P.S., Hales, C.N., Kinmonth, A.L., and Wareham, N.J. *Diabetes risk score: Towards earlier detection of type 2 diabetes in general practice.* Diabetes/metabolism research and reviews. 2000; **16**(3): p. 164-171.
18. Al Khalaf, M.M., Eid, M.M., Najjar, H.A., Alhajry, K.M., Doi, S.A., and Thalib, L. *Screening for diabetes in kuwait and evaluation of risk scores.* East Mediterr Health J. 2010; **16**(7): p. 725-731.
19. Al-Lawati, J.A. and Tuomilehto, J. *Diabetes risk score in oman: A tool to identify prevalent type 2 diabetes among arabs of the middle east.* Diabetes research and clinical practice. 2007; **77**(3): p. 438-444.
20. Baan, C.A., Ruige, J.B., Stolk, R.P., Witteman, J.C., Dekker, J.M., Heine, R.J., *et al*. *Performance of a predictive model to identify undiagnosed diabetes in a health care setting.* Diabetes care. 1999; **22**(2): p. 213-219.
21. Bergmann, A., Li, J., Wang, L., Schulze, J., Bornstein, S.R., and Schwarz, P.E. *A simplified finnish diabetes risk score to predict type 2 diabetes risk and disease evolution in a german population.* Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme. 2007; **39**(9): p. 677-682.

22. Rahman, M., Simmons, R.K., Harding, A.H., Wareham, N.J., and Griffin, S.J. *A simple risk score identifies individuals at high risk of developing type 2 diabetes: A prospective cohort study.* Family practice. 2008; **25**(3): p. 191-196.
23. Spijkerman, A.M., Yuyun, M.F., Griffin, S.J., Dekker, J.M., Nijpels, G., and Wareham, N.J. *The performance of a risk score as a screening test for undiagnosed hyperglycemia in ethnic minority groups: Data from the 1999 health survey for england.* Diabetes care. 2004; **27**(1): p. 116-122.
24. Park, P.J., Griffin, S.J., Sargeant, L., and Wareham, N.J. *The performance of a risk score in predicting undiagnosed hyperglycemia.* Diabetes care. 2002; **25**(6): p. 984-988.
25. Heldgaard, P.E. and Griffin, S.J. *Routinely collected general practice data aids identification of people with hyperglycaemia and metabolic syndrome.* Diabetic medicine : a journal of the British Diabetic Association. 2006; **23**(9): p. 996-1002.
26. Lin, J.W., Chang, Y.C., Li, H.Y., Chien, Y.F., Wu, M.Y., Tsai, R.Y.*, et al*. *Cross-sectional validation of diabetes risk scores for predicting diabetes, metabolic syndrome, and chronic kidney disease in taiwanese.* Diabetes care. 2009; **32**(12): p. 2294-2296.
27. Witte, D.R., Shipley, M.J., Marmot, M.G., and Brunner, E.J. *Performance of existing risk scores in screening for undiagnosed diabetes: An external validation study.* Diabetic medicine : a journal of the British Diabetic Association. 2010; **27**(1): p. 46-53.
28. Gao, W.G., Dong, Y.H., Pang, Z.C., Nan, H.R., Wang, S.J., Ren, J.*, et al*. *A simple chinese risk score for undiagnosed diabetes.* Diabetic medicine : a journal of the British Diabetic Association. 2010; **27**(3): p. 274-281.
29. Lee, Y.H., Bang, H., Kim, H.C., Kim, H.M., Park, S.W., and Kim, D.J. *A simple screening score for diabetes for the korean population: Development, validation, and comparison with other scores.* Diabetes care. 2012; **35**(8): p. 1723-1730.
30. Sambo, B.H. *The diabetes strategy for the who african region: A call to action.* Diabetes Voice. 2007; **52**(4): p. 335-337.
31. Kengne, A.P., Beulens, J.W., Peelen, L.M., Moons, K.G., van der Schouw, Y.T., Schulze, M.B.*, et al*. *Non-invasive risk scores for prediction of type 2 diabetes (epic-interact): A validation of existing models.* Lancet Diabetes Endocrinol. 2014; **2**(1): p. 19-29.
32. Vergouwe, Y., Steyerberg, E.W., Eijkemans, M.J., and Habbema, J.D. *Substantial effective sample sizes were required for external validation studies of predictive logistic regression models.* Journal of clinical epidemiology. 2005; **58**(5): p. 475-483.
33. Erasmus, R.T., Soita, D.J., Hassan, M.S., Blanco-Blanco, E., Vergotine, Z., Kengne, A.P.*, et al*. *High prevalence of diabetes mellitus and metabolic syndrome in a south african coloured population: Baseline data of a study in bellville, cape town.* South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde. 2012; **102**(11 Pt 1): p. 841-844.
34. Wilson, B.J., Qureshi, N., Santaguida, P., Little, J., Carroll, J.C., Allanson, J.*, et al*. *Systematic review: Family history in risk assessment for common diseases.* Annals of Internal Medicine. 2009; **151**(12): p. 878-885.
35. Qureshi, N., Wilson, B., Santaguida, P., Little, J., Carroll, J., Allanson, J.*, et al*. *Family history and improving health.* 2009.

## Additional files

**Supplementary Table 1:** Full equation for risk models to predict prevalent undiagnosed diabetes as applied to the Bellville South cohort

| |
|---|
| **Cambridge logistic regression diabetes risk model** [17] |
| The probability of developing diabetes was calculated as exp(X)/(1 + exp(X)); where<br>X= -6.322 - 0.879 (if female, else 0) + 1.222 (if prescribed antihypertensive medication) + 2.191 (if prescribed steroids) + 0.063 × age in years + 0.699 (if 25 kg/m$^2$ ≤ BMI ≤ 27.49 kg/m$^2$) + 1.970 (if 27.5 kg/m$^2$ ≤ BMI ≤ 29.99 kg/m$^2$) + 2.518 (if BMI ≥ 30 kg/m$^2$) + 0.728 (if parent or sibling has diabetes) + 0.753 (if parent and sibling has diabetes) - 0.218 (if an ex-smoker) + 0.855 (if a current smoker). |
| **Kuwaiti logistic regression diabetes risk model** [18] |
| The probability of developing diabetes was calculated as exp(X)/(1 + exp(X)); where<br>X= -5.018 + 0.979 (if a sibling had a history of diabetes, else 0) + 0.978 (if prescribed antihypertensive medication) + 1.315 (if age ≥ 35 years, else 0) + 1.930 (if the waist circumference ≥ 100cm, else 0). |
| **Omani logistic regression diabetes risk model** [19] |
| The probability of developing diabetes was calculated as exp(X)/(1 + exp(X)); where<br>X= -4.7 + 1.8 (if 40 years ≤ age ≤ 59 years) + 2.3 (if age ≥ 60 years) + 0.38 (if waist circumference ≥ 94cm in men and waist circumference ≥ 80cm in women) + 0.54 (if 25 kg/m$^2$ ≤ BMI < 30 kg/m$^2$) + 0.69 (if BMI ≥ 30 kg/m$^2$) + 1.9 (if parental or sibling history of diabetes) + 0.73 (if if SBP≥140 and/or DBP≥90). |
| **Rotterdam logistic regression diabetes risk model 1** [20] |
| The probability of developing diabetes was calculated as exp(X)/(1 + exp(X)); where<br>X= -3.02 + 0.19 (per 5 year increment from 55 years to >75) + 0.46 (if male, else 0) + 0.42 (if prescribed antihypertensive medication) + 0.51 (if BMI ≥ 30 kg/m$^2$). |
| **Simplified Finnish logistic regression diabetes risk model** [21] |
| The probability of developing diabetes was calculated as exp(X)/(1 + exp(X)); where<br>X= -5.514 + 0.628 (if 45 years ≤ age ≤ 54 years) + 0.892 (if 55 years ≤ age ≤ 64 years) + 0.165 (if 25 kg/m$^2$ ≤ BMI < 30 kg/m$^2$) + 1.096 (if BMI > 30 kg/m$^2$) + 0.857 (if 94cm ≤ waist circumference < 102cm in men and 80cm ≤ waist circumference < 88cm in women) + 1.350 (if waist circumference ≥ 102cm in men and waist circumference ≥ 88cm in women) + 0.711 (if prescribed antihypertensive medication) + 2.139 (if a history of high blood glucose, assumed to be 0 for all participants due to the nature of this study). |

*BMI, body mass index; SBP, systolic blood pressure; DBP, dystolic blood pressure.

**Supplementary Table 2**: Characteristics comparison of participants with valid and missing data

| Characteristics | Valid (737) | Missing (346) | p-value |
|---|---|---|---|
| Male (%) | 157 (21.3) | 94 (27.2) | 0.012 |
| Age (years) | 51.2 (11.9) | 52.8 (18.3) | 0.104 |
| Body mass index (kg/m$^2$) | 29.9 (7.3) | 29.6 (7.0) | 0.515 |
| Waist circumference (cm) | 96.4 (14.9) | 95.0 (16.2) | 0.147 |
| Hypertensive medication (%) | 251 (34.1) | 123 (35.6) | 0.182 |
| Smoking status (% smoking) | 323 (43.8) | 110 (31.8) | <0.001 |
| Systolic blood pressure (mmHg) | 122.8 (17.6) | 126.2 (23.1) | 0.009 |
| Diastolic blood pressure (mmHg) | 75.8 (11.5) | 76.2 (14.5) | 0.669 |
| Height (m) | 1.6 (0.1) | 1.6 (0.1) | 0.522 |
| Mother having diabetes (%) | 109 (14.8) | 15 (4.3) | <0.001 |
| Father having diabetes (%) | 58 (7.9) | 3 (0.9) | <0.001 |
| Sister having diabetes (%) | 92 (12.5) | 11 (3.2) | <0.001 |
| Brother having diabetes (%) | 58 (7.9) | 9 (2.6) | 0.001 |
| Fasting blood glucose (mmol/L) | 5.7 (1.9) | 5.5 (1.5) | 0.048 |
| HDL cholesterol (mmol/L) | 1.3 (0.4) | 1.3 (0.4) | 0.739 |
| Weight (kg) | 75.1 (17.9) | 74.7 (17.6) | 0.727 |
| Ever consumed alcohol (%) | 356 (48.3) | 88 (25.4) | <0.001 |
| Currently drinking (%) | 207 (28.1) | 65 (18.8) | 0.001 |
| Corticosteroid use (%) | 5 (0.7) | 7 (2.0) | 0.097 |
| Triglyceride (mmol/L) | 1.5 (1.0) | 1.4 (0.8) | 0.043 |

*HDL, high-density lipoprotein.

**Supplementary Table 3:** Discrimination values and 95% confidence intervals for selected models and the comparison of the discrimination between each model, expressed using p-value (<0.05 significant).

| | C-statistic | Comparison discrimination | | | | |
|---|---|---|---|---|---|---|
| | | Cambridge [17] | Kuwaiti [18] | Omani [19] | Rotterdam [20] | Finnish [21] |
| Cambridge [17] | 0.67 (0.62-0.72) | - | 0.689 | 0.458 | 0.066 | 0.734 |
| Kuwaiti [18] | 0.68 (0.63-0.73) | - | - | 0.292 | 0.109 | 0.397 |
| Omani [19] | 0.66 (0.61-0.70) | - | - | - | 0.605 | 0.735 |
| Rotterdam [20] | 0.64 (0.59-0.69) | - | - | - | - | 0.320 |
| Finnish [21] | 0.67 (0.62-0.71) | - | - | - | - | - |

# PART III

# MISSING DATA

06

# Reporting and handling of missing data in predictive research for prevalent undiagnosed type 2 diabetes mellitus: a systematic review

Katya Masconi, Tandi E. Matsha, Justin Basile Echouffo-Tcheugui
Rajiv T. Erasmus and Andre P. Kengne

*EPMA Journal* 2015, **6**:7

## Abstract

**Background:** Missing values are common in health research and omitting participants with missing data often leads to a loss of statistical power, biased estimates, and consequently inaccurate inferences.

**Methods:** We critically reviewed the challenges posed by missing data in medical research and approaches to address these. To achieve this more efficiently, these issues were analysed and illustrated through a systematic review on the reporting of missing data and imputation methods (prediction of missing values through relationships within and between variables) undertaken in risk prediction studies of undiagnosed diabetes. Prevalent diabetes risk models were selected based on a recent comprehensive systematic review, supplemented by an updated search of English-language studies published between 1997 and 2014.

**Results:** Reporting of missing data has been limited in studies of prevalent diabetes prediction. Of the 48 articles identified, 62.5% (n=30) did not report any information on missing data or handling techniques. In 21 (43.8%) studies, researchers opted out of imputation, completing case-wise deletion of participants missing any predictor values. Although imputation methods are encouraged to handle missing data and ensure the accuracy of inferences, this has seldom been the case in studies of diabetes risk prediction. Hence, we elaborated on the various types and patterns of missing data, the limitations of case-wise deletion, state-of the-art methods of imputations and their challenges.

**Conclusions:** This review highlights the inexperience or disregard of investigators in the effect of missing data in risk prediction research. Formal guidelines may enhance the reporting and appropriate handing of missing data in scientific journals.

Keywords: risk, prediction, diabetes, missing, imputation

## Background

Missing values on participants' characteristics are common in healthcare research and are often non-optimally handled and/or reported in prediction research. Inappropriate handling of missing data can lead to a poor model performance at the model development stage, and mislabelling of the model at the external validation stage. It is therefore recommended that researchers in predictive research strive to examine the patterns of missing values in their database to aid in missingness classification, use a valid approach to dealing with the missing data and include the description in their final report [1]. Predictive research is an area in which the handling of missing data is of utmost importance. Indeed, simple risk prediction models based upon non-invasively measured predictors are increasingly advocated in population-based strategies for screening prevalent undiagnosed diabetes, particularly in low and middle income countries where undiagnosed diabetes is very common [2]. Accordingly, many prevalent diabetes risk prediction models have been developed over the last decade to convey this new thinking. Available models however, remain specific to the population from which they were developed, until evidence of their good performance during external validations studies in different settings become available [3].

In this paper, we critically review the different patterns of missing data and approaches to dealing with them, with a focus on predictive modelling. For illustrative purpose, we investigated how missing data have been reported and handled in predictive modelling, through a systematic review of studies on the development and/or validation prevalent diabetes risk model. We hypothesized that the level of reporting and extent of imputation in studies of undiagnosed diabetes model development and validation would be poor.

## Methods

Building on a recent comprehensive review article on diabetes risk prediction models by Brown *et al* 2012 [4], additional relevant articles were identified through a search of electronic database PubMed using the key terms 'undiagnosed', 'diabetes', 'risk' and 'score'; and a manual search through reference lists of eligible studies. We selected studies aimed at the development or validation of a risk prediction model. The outcome had to be prevalent

undiagnosed diabetes in adults (aged >18 years). Models excluded were those of incident risk prediction or requiring blood tests (on the grounds that prevalent diabetes risk prediction aims at simple screening). The data extracted included country/setting (including its income classification), population/ethnicity, source of data and if from a questionnaire whether self-administered or not, sample size, age range of participants and the presence of a discussion and action (or lack thereof) on missing data.

We aimed at providing the reader with instances of missing data, the reporting and attempts to handle these, as well as the challenges posed by each method. In some instances, because of the paucity of reports on handling missing data in studies of diabetes risk prediction, we used examples from other fields for greater understanding and clarity on a subject that has not received much attention.

## Results

### Overview of included studies

A total of 48 articles (26 were model development studies and 22 were external validations) were included (Figure 1). These are summarized in Table 1; published between 1997 and 2014 (most appeared in 2005-2010). The number and combination of predictors was variable, with age, sex, body mass index and waist circumference being the most commonly used variables. Models were developed and validated in 24 countries across 5 continents (none from Africa). Participants' ethnicity was not always clearly stated, but a number of studies included minority populations specific to their location (e.g; Asian and Black participants in a study conducted in the Netherlands) [5-10]. Administrative data was the most common source of data (30, 62.5%), from existent healthcare [11, 12], governmental organization [9, 13-15] or research settings [5, 10, 16-37]. The study sample sizes varied from 429 [28] to 68 476 [38]. Finally, the age of participants ranged from 18 to 94 years of age.

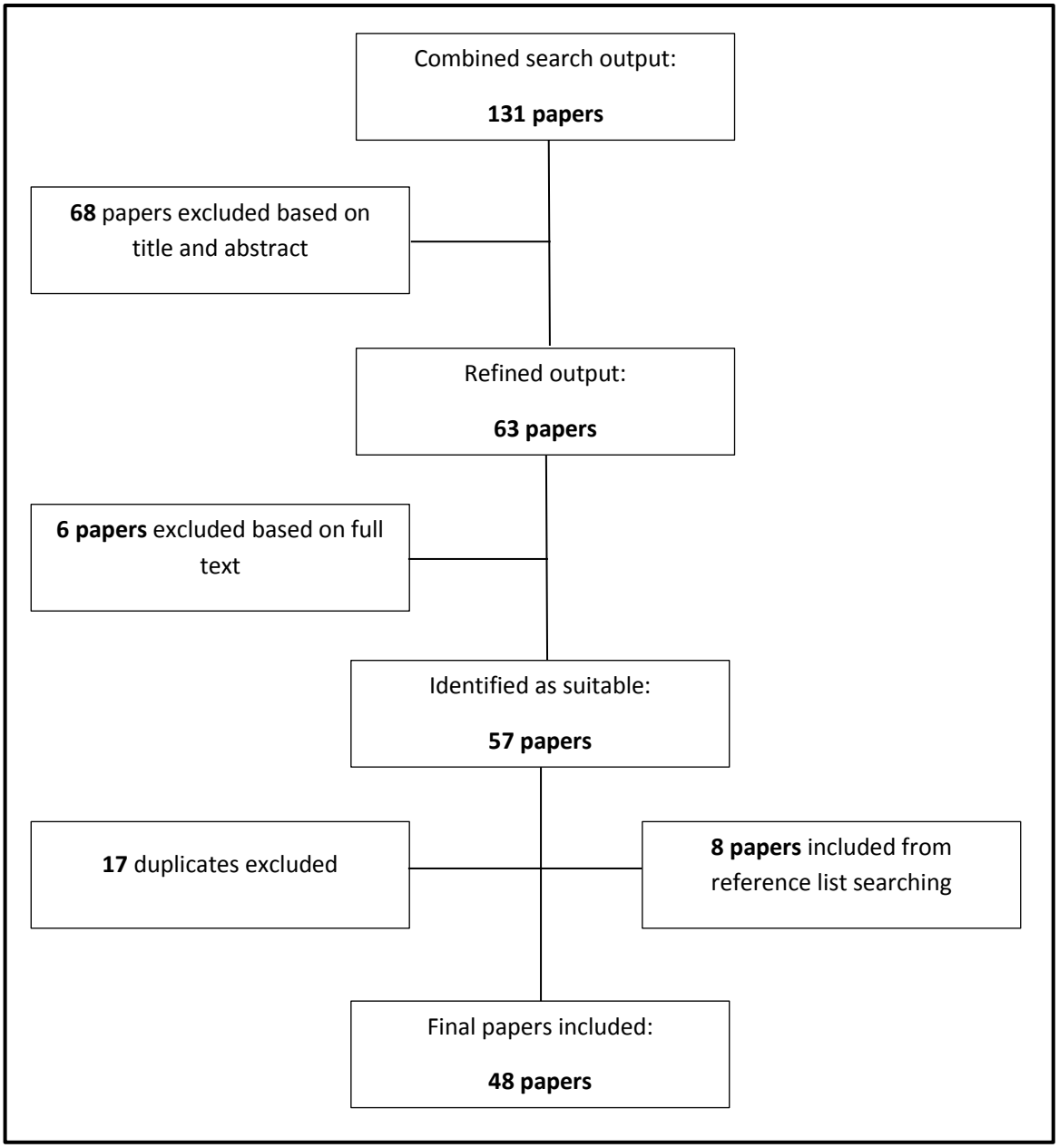



**Figure 1:** Workflow summarising the selection of papers

**Table 1:** Characteristics of 48 included studies of undiagnosed diabetes risk prediction models

| Author | Year | Validation or development | Location of study (Income) | Ethnicity | Time of data collection | Type of Data / Self-administered | Size of study population | Age range | Missing data status | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Reporting | | Handling | | |
| | | | | | | | | | None | % | None | Deletion | Imputation |
| Adhikari et al [39] | 2010 | Validate | India (L/M) | / | Current | | 551 | >20 | X | | X | | |
| Akyil et al [40] | 2014 | Validate | Turkey (L/M) | / | Current | | 702 | / | X | | X | | |
| Al Khalaf et al [41] | 2010 | Develop | Kuwaiti (L/M) | Caucasian | Current | X | 562 | >20 | X | | X | | |
| Al-Lawati et al [16] | 2007 | Develop | Oman (H) | Caucasian | Existing | | 4881 | >20 | X | | | X | |
| Baan et al [17, 42] | 1999 | Develop | Netherlands (H) | / | Existing | X | 1016 | 55-75 | X | | X | | |
| Bang et al [18] | 2009 | Develop | USA (H) | / | Existing | | 5258 | >20 | | X | | X | X |
| Bergmann et al [43] | 2007 | Validate | Germany (H) | / | Current | | 526 | 41-79 | X | | X | | |
| Bindraban et al [5] | 2008 | Develop | Netherlands (H) | Asian, Black, Caucasian | Existing | | 1434 | 35-60 | | X | | X | |
| Chaturvedi et al [19, 44] | 2008 | Develop | India (L/M) | / | Existing | | 4044 | 35-64 | X | | X | | |
| de Leon et al [45] | 2008 | Develop | Canary Islands (H) | Caucasian | Current | | 6237 | 18-75 | X | | X | | |
| de Sousa et al [13] | 2009 | Develop | Brazil (L/M) | Multi-ethnic | Existing | X | 1224 | >35 | X | | X | | |
| Franciosi et al [20] | 2005 | Validate | Italy (H) | / | Existing | X | 1377 | 55-75 | | X | | X | |
| Gao et al [46] | 2010 | Validate | China (L/M) | Asian | Current | | 1986 | 20-74 | | X | | X | |
| Ginde et al [6] | 2007 | Validate | USA (H) | Caucasian, African-American, Hispanic | Current | | 604 | / | X | | | X | |
| Glumer et al [21] | 2004 | Develop | Denmark (H) | / | Existing | | 6784 | 30-60 | | X | | X | |
| Glümer et al [22] | 2005 | Validate | Australia / Denmark (H) | / | Existing | | 7079 / 6270 | 30-60 | | X | | X | |
| Glumer et al [23] | 2006 | Validate | Global | Multi-ethnic | Existing | | 29 758 | / | X | | | X | |
| Gray et al [24] | 2010 | Develop | UK (H) | Caucasian, Asian | Existing | | 6186 | 40-75 | | X | | X | |
| Gray et al [25] | 2013 | Develop | Portugal (H) | / | Existing | | 3435 (18-94) | 18-94 | | X | | | X |
| Griffin et al [11] | 2000 | Develop | UK (H) | Caucasian | Existing | | 1077 | 40-64 | X | | X | | |
| Hanif et al [47] | 2008 | Develop | UK (H) | Asian | Current | | 435 | 20-75 | X | | X | | |
| Heianza et al [26] | 2013 | Develop | Japan (H) | Asian | Existing | | 7477 | 18-88 | | X | | X | |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Heikes et al [27] | 2008 | Develop | USA (H) | Representative of USA population | Existing | | 7029 | >20 | | X | | X | |
| Heldgaard & Griffin [48] | 2006 | Develop | Denmark (H) | / | Current | X | 1355 | 20-69 | X | | X | | |
| Keesukphan et al [28] | 2007 | Develop | Thailand (L/M) | / | Existing | | 429 | 18-81 | X | | X | | |
| Ko et al [12] | 2010 | Develop | China (L/M) | Asian | Existing | | 7695 | | X | | X | | |
| Ku & Kegels [49] | 2013 | Validate | Philippines (L/M) | / | Current | | 1789 | | X | | X | | |
| Lee et al [29] | 2012 | Develop | Korea (L/M) | / | Existing | | 9602 | >20 | | X | | X | |
| Li et al [50] | 2009 | Develop | Germany (H) | / | Current | | 921 | 14-93 | X | | X | | |
| Lin et al [51] | 2009 | Validate | Taiwan (H) | Asian | Current | | 2759 | >18 | X | | X | | |
| Lindstrom et al [14] | 2003 | Develop | Finland (H) | / | Existing | X | 4435 | 35-64 | X | | | X | |
| Liu et al [15] | 2011 | Develop | China (L/M) | / | Existing | | 1851 | 40-90 | X | | X | | |
| Mohan et al [30] | 2005 | Validate | India (L/M) | Asian | Existing | | 2350 | >35 | X | | X | | |
| Park et al [31] | 2002 | Validate | UK (H) | Caucasian | Existing | X | 6567 | 39-78 | X | | | X | |
| Rahman et al [32] | 2008 | Validate | UK (H) | / | Existing | | 25 639 | 40-79 | | X | | X | |
| Ramachandran et al [33] | 2005 | Develop | India (L/M) | Asian | Existing | | 10 003 | >20 | X | | X | | |
| Rathmann et al [34] | 2005 | Validate | Germany (H) | Caucasian | Existing | | 1353 | 55-74 | X | | X | | |
| Robinson et al [7] | 2011 | Develop | Canada (H) | Caucasian, Aboriginal, Asian, Black, Hispanic | Current | | 6475 | 40-74 | | X | | X | X |
| Rolka et al [8] | 2001 | Validate | USA (H) | Hispanics, Caucasian, Black, Native American | Current | | 1471 | >20 | | X | | | X |
| Ruige et al [35] | 1997 | Develop | Netherlands (H) | Caucasian | Existing | X | 2364 | 50-74 | | X | X | | |
| Saaristo et al [52] | 2005 | Validate | Finland (H) | / | Current supplemented with existing | X | 2966 | 45-74 | | X | | | X |
| Spijkerman et al [9] | 2004 | Validate | UK (H) | Black, Asian | Existing | | 803 | 40-75 | X | | | X | |
| Ta et al [53] | 2010 | Validiate | Vietnam (L/M) | / | Current | | 721 | 30-70 | X | | X | | |
| Tankove et al [54] | 2011 | Validate | Bulgaria (L/M) | / | Current | | 2169 | | X | | X | | |
| Winkler et al [38] | 2012 | Validate | Hungary (L/M) | / | Current | | 68 476 | >18 | X | | | X | |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Witte et al [36] | 2010 | Validate | UK (H) | Caucasian | Existing | | 6990 | 35-55 | | X | | X | |
| Zhang et al [10] | 2014 | Validate | USA (H) | Caucasian, Black | Existing | X | 20 633 | >20 | X | | | X | |
| Zhou et al [37] | 2013 | Develop | China (L/M) | / | Existing | | 41 809 | 20-74 | | X | | X | |

**Table 2:** Details of imputation options

| | Theory | Package in R |
|---|---|---|
| **Single imputation methods** | | |
| Simple imputation | In a predictor (X) which is unrelated to all other X's, substitution replaces all missing continuous values with the mean (or median) of all participants who have a valid value or the mode for categorical predictors [55].<br>Simple imputation reduces variability and correlation estimates by ignoring relationships between variables but assumes MCAR. Regression coefficients are biased towards 0 (zero) since the outcome (Y) is not considered [1]. | Mean substitution is easily implemented with the package '*Hmisc*' of R statistical software through the function 'impute (x, fun=mean)' where x is the predictor of interest [56]. |
| Conditional mean imputation | Regression imputation assumes strong relationships between the X to be imputed and the independent X's used in the univariable or multivariable regression formula [1, 57, 58]. An imputation model is made to predict the missing values when X is related to the other X's, this method is far more efficient [59-61]. Conditional mean imputation leads to a weakening of the variance and overestimation of the model fit and correlation estimates. The outcome (Y) should not be included in the imputation model to prevent over exaggeration of the strength of relationship between X and Y [1]. | Conditional mean imputation can be implemented in R through the creation of a regression model and the subsequent inbuilt '*predict*' function. |
| Stochastic regression imputation | An alternative to conditional mean imputation, stochastic regression imputation includes a random element to the prediction of values, highlighting the uncertainty of imputed values [57]. A random draw is taken from the distribution of predicted values, which allows for the inclusion of the outcome in the prediction model. | This can be implemented with the '*mice*' package for R via the command '*mice.impute.norm.nob*' [62]. |
| Hotdecking | Hotdecking replaces the missing value of an individual with a random value from a pool of individuals who are matched to the missing individual by predictors, the 'deck'[63, 64]. These deck predictors may be researcher-determined or a correlation matrix may be used to determine which the most highly correlated predictors are. The standard error is better approximated through the hotdeck procedure than simple imputation. | The command '*hotdeck*' of the R package '*VIM*' can implement the hotdecking [65]. |
| **Multiple imputation methods** | | |
| Markov chain Monte Carlo (MCMC) | Multivariate normal imputation assumes a multivariate distribution and the MCMC algorithm is used to obtain imputed values and allow for uncertainty in the estimated model predictors [66]. MCMC describes a group of methods that use Markov chains to generate pseudorandom draws from probability distributions. | The command '*imp.norm* of the R package '*norm'* can implement MCMC approach to multiple imputation [67]. |

| | | |
|---|---|---|
| Maximum likelihood | The expectation-maximization (EM) algorithm, also called joint modelling, assumes a multivariate distribution. First a set of parameter values that produces the maximum likelihood are identified from the conditional distribution; values that would most likely have resulted in the observed data [62, 68]. New parameter estimates are randomly drawn from a Bayesian posterior distribution, the distribution of unobserved values conditional on observed data [69]. Bootstrap procedures are employed to obtain standard error estimates, correcting for bias associated with non-normality. | The package '*Amelia'* in R implements bootstrapping algorithms to give EM results [70]. |

**Source of missing data in predictive research**

Figure 2 summarizes the reporting and handling of missing data. The chief reasons for missing data were study design, participant characteristics, measurements characteristics, data collection and management, and chance. These may occur alone or simultaneously within a study, with data missing for several different reasons acting additively.

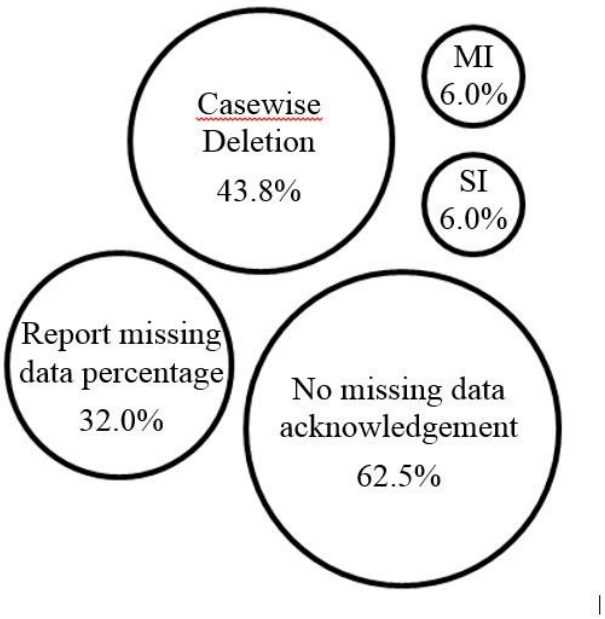



**Figure 2:** Graphical representation of handling of missing data from the 48 selected studies.
*MI, Multiple Imputation; SI, Single Imputation

*Study Design*

The reviewed studies were cross-sectional. No study design can eliminate missing data; however the probability of missing data varies across designs, with longitudinal studies carrying a higher likelihood of missing data than cross-sectional studies. In longitudinal studies, a greater burden on the participants increases the likelihood of missing data, through the duration of the study, multiple repeated measures, long questionnaires and painful procedures. With lengthy and cumbersome procedures, participants are more prone to respond poorly or drop out altogether. Indeed, Rolka *et al* had high missing percentages for the invasive collection of a finger prick, fasting and 2 hour post-load blood collection (0.2%,

26.0% and 27.0% of missing data respectively), as the study design required three invasive and burdensome diagnosis tests.

*Participant characteristics*

Non-response to questions may be associated with personal characteristics of the participants, where the reason may be an inaccuracy of information processing or refusal to provide information. Information processing may be related to the language and comprehension levels of the participant. Beliefs, and the attitude toward the research topic or particular item collected, are important in non-response due to refusal. All studies that reported some form of missing data values were conducted in high income countries; except three studies undertaken in China [37, 46] and Korea [29], all published after 2010.

*Measurement characteristics*

The collection of quantifiable predictors can lead to missing data in a variety of ways. Observations may be lost due to malfunctioning equipment. The complexity, length and invasiveness of the measures may also lead to participants opting out of particular tests (e.g. oral glucose tolerance test). Finally, for predictors that are measured in a laboratory, errors in the pre-analytical sample collection and analytical testing, can result in random missing data (e.g. incorrect blood collection tube selection or extended waiting time before analysing blood glucose sample, where glucose is lost through glycolysis). Demographic or behavioural information may be collected via questionnaires through an interview of or self-administration by the participant. Self-administration is limited by the lack of supervision thus increasing the likelihood of respondent error, ultimately increasing missing data. Only three articles that included self-administrated questionnaires, also reported missing data [20, 35, 52]. Missing data was as high as 9% for body mass index and waist circumference in the study by Saaristo *et al* [52]; and 15.3% for the oral glucose tolerance test and 15.7% for questionnaire data in that of Franciosi *et al* [20].

*Data management*

Poor management of data can result in the loss of data obtained from all participants. This may be due to the data transfer process from one format to another, such as the exclusion of individual values due to unclear writing, unconventional answers or inadvertently missing

questionnaire items. Disorganized or poor data storage can also result in lost data through unsystematic filing and communication, or faulty or non-existent back up files. Of the 17 articles reporting missing data, 13 of these were studies using existing databases, all developed for research [5, 18, 20-22, 24, 26, 27, 29, 32, 35-37]. Although administrative data has its own issues; the reduced response burden, the possibility of a large sample size, and comparatively low costs, make this an increasing popular choice of data collection. *De novo* data collection requires the correct preparation, validation and processing of the survey to limit missing data. The two articles that reported missing data above 20% were based on new data collection [7, 8].

*Chance*

Despite investigators' best efforts to prevent missing data through study design, data collection and measurements and subsequent management of the data; missing data can still occur by chance. This does not produce a bias, however large amounts of data may be missing if multiple chance events occur which produces its own sets of problems, such as a reduction in statistical power [71].

**Reporting of missing data**

Missing data was frequently poorly handled, with 62.5% of the articles not mentioning whether missing data was encountered and, if there was, how it was treated. Sixteen articles (33.3%) stated the missing data percentage, with two testing the effect on the final dataset but not reporting missing data details [24, 27]. However, from the reporting, it is difficult to determine the type of missing data, as this was not investigated.

**Types of missing data**

Missing data can be classified as 'missing completely at random (MCAR)', 'missing at random (MAR)', and 'missing not at random (MNAR)', where the reason for missingness differs [72-76]. Identifying the nature and pattern of missing data allows the researcher to correctly choose a data imputation method, which is based on the assumptions about the patterns of missing data.

*Missing completely at random*

Data is MCAR where the random subset of observations missing will have similar distributions to observed values [72]. The reasons for missing are unrelated to characteristics or responses of the subjects. Missing completely at random is a strict assumption and can be tested for. Little [77] provided a statistical test of the MCAR assumption, where a significant chi-square test indicates that the data are not MCAR. Examples of MCAR include administrative errors or laboratory accidents that occur at random.

*Missing at random*

Missing data is described as MAR when the missing data is conditional. The missing observations commonly depend on observed characteristics not missing, with systematic differences between the missing and observed data [1, 78]. The assumption is fulfilled if the missing values are related only to measured, not unmeasured, values. MAR examples include increased missing data in elderly individuals, subjects from a certain region; or from a different calendar time. This is illustrated by Robinson *et al* [7], where smoking status was only available for selected collection sites, as this question was added to the questionnaire during the last phase of data collection, resulting in a large percentage of item-missing data.

*Missing not at random*

Missing data that are not at random are related to unobserved participant's characteristics [72]. This type of missing data is problematic and imputation is not sufficient. An example of MNAR is the selective non-response by a subject, e.g. sexual orientation or weight where the association with social image may cause people to avoid or underestimate the answer.

**Patterns of missing data**

None of the selected articles on the prediction of prevalent diabetes risk discussed, nor graphically presented, patterns of missing data, nor offered reasons for the missing data. In general, there are three patterns of missing data, namely univariate, monotone, and arbitrary [79].

**Handling of missing data**

In existing studies of diabetes risk prediction, 21 (43.8%) stated all individuals missing data were excluded from the study analysis, conducting complete case analysis. Two articles used simple imputation to overcome missing data [7, 52] and two made use of multiple imputation [8, 25], while a single article undertook both imputation methods [18]. Saaristo *et al*, stated the missing data percentage for the variables which most commonly had missing data (9% for both BMI and waist circumference), both of which were simply imputed with mean substitution [52]. Robinson *et al* used a number of deletion and imputation methods [7]. Waist circumference (6% missing) was imputed with mean substitution, while family history (13%) was dealt with by the substitution of 'no' for unanswered questions. Case-wise deletion was undertaken for all other predictors of missing data, 3.9% of participants were excluded. Finally, smoking was excluded as a predictor all together due to the large percentage of missing data (35.0%).

Bang *et al* used complete case analysis for predictors with missing values as the missing data proportion was considered 'small', although no further details were stated [18]. Multiple imputation was done for family history of diabetes. Perhaps significantly, the studies with low missing data rates or few variables with missing data undertook multiple imputation as a solution. Rolka *et al* reported a full dataset apart from only three predictors with missing data, namely postprandial time (3.0%), fasting blood glucose (26.0%) and oral glucose tolerance test (27.0%) [8]. Finally, Gray *et al* described minimal missing data for the majority of predictors ranging from 0.1% for current hypertension to 1.7% for smoking status, apart from statin use (36%) [25]. The effect of missing data on both the modelling process and the final model chosen, was assessed. Another article did not state the missing data proportion, but rather the overall effect of the missing data, which was to underestimate the prevalence of prediabetes and undiagnosed diabetes by approximately 2% and 1.5% respectively [27]. None of the three studies using multiple imputation stated the details of the method [8, 18, 25], such as the number of imputations or the variables included in the imputation model. We herein discuss the key fundamentals aspects of the various methods to dealing with missing data, which were seldom or inappropriately undertaken as mentioned above.

**Proportion of missing data and impact on the method for handling missing data**

A proportion (considered here as the proportion of subjects having *any* predictors missing) of ≤0.05 is considered the cut-off when deciding if no or simple imputation, without sacrificing results, would be sufficient [80]. A missing data proportion between 0.05 and 0.15 requires some research into predictor relationships. If the predictor with missing values is unrelated to all of the other predictors, simple imputation is considered reasonable; else, conditional mean or stochastic regression is the minimum. Once missingness proportion is ≥0.15, multiple imputation becomes imperative.

**Methods for dealing with missing data**

*Problems with simple alternatives to data imputation*

Common in predictive modelling is the case-wise deletion of individuals with data missing for the required model predictors. Complete case analysis or list-wise deletion, removes all subjects with missing values for any possible predictors to be used in risk models [58, 81]. Alternatively, available case analysis, or pairwise deletion, includes subjects with complete data for the predictors to be included in the final model, but who have missing data for other predictors not considered in the model [1]. List-wise or case-wise deletions lead to reductions in sample size, and as a consequence, a reduction in statistical power, increase in standard error, and bias and imprecision in the regression coefficient estimates is introduced if the data is not MCAR [82-84]. Furthermore, when more than one prevalent diabetes risk prediction models are to be validated in a new population, it is difficult to interpret the results when the number of subjects may vary across the analyses [1].

*Imputation*

Imputation of missing values is the process of replacing these values with accurate parameter estimates [85]. Imputation aims at predicting missing values by obtaining values through relationships within and between variables. In general, individuals should only be discarded if there is a missing predictor of overriding importance that cannot be reliability imputed from other information [1]. Table 2 details the imputation methods available, namely single and multiple imputation, and the implementation in R statistical software. Single imputation (SI)

includes simple imputation, conditional mean imputation, stochastic regression imputation and hotdecking, with each of these methods having its own advantages and drawbacks.

Multiple imputation (MI) describes the production of multiple complete datasets derived from the initial dataset with missing values [86]. Statistical models are used to fill the missing data a number (*m*) of times to generate *m* complete data sets. The multiple datasets add variability, increasing accuracy for both sampling and imputation, and the number of imputed datasets is usually set to 5 or 10 [87]. The datasets are analysed separately using standard procedures, yielding multiple estimates which are then combined using a appropriate statistical method [88]. The first stage requires an imputation algorithm, while the combination of the analysis results of the multiple datasets requires an alternate pooling algorithm. Imputation algorithms may be univariate methods for monotone missing data such as predictive mean matching [89], propensity methods [90] or logistic regression; or for more complicated missing data, the Multiple Imputation by Chained Equations (MICE) or expectation-maximization (EM) algorithm have been proposed. Multiple imputation methods for non-monotone missing patterns using chained equations requires the decision of whether to use Markov chain Monte Carlo (MCMC) or fully conditional specification (FCS) methods. Expectation-maximization has yet to become that popular in medical applications but merits discussion and use.

Multiple imputation is time, labour and computationally intensive, and in case of small amounts of missing data, researchers must decide whether the use of this method or alternative methods [78, 83]. The combination of lack of guidelines, imputer burden, and perhaps lack of knowledge, makes researchers hesitant to undertake MI. This hesitation is encouraged if MI is not going to be carried out successfully, with the failure to combine the final *m* datasets or leaving out of important predictors in the MI model.

*Dealing with missing data in validation studies*

The implementation of a model in an alternative population to that in which the model was developed requires prior validation. Differences between the development and validation datasets can be expected, with predictors possibly missing altogether; hindering validation of the model. This can be handled in a variety of ways, all which will have an effect on model performance or final model selection. Missing predictors can be dealt with by excluding models which contain any predictors not collected in the study. This limits the possibility of finding an existing model that may have suitable performance in the new population. Alternatively, the model may be selected for validation but predictors in the model will be excluded from the model formula. This method could be improved by the substitution of a missing predictor with a reliable proxy variable, preventing model and predictor exclusion. Of the 22 validation studies, 11 (50%) used case-wise deletion of individuals or predictors in dealing with the missing data; with only a single article using mean imputation [52] and another multiple imputation [8].

## Discussion

Dealing with missing data is a complex undertaking, which is not yet common place in medical research. Indeed, for studies on the development and validation of undiagnosed diabetes risk models, we found inconsistent reporting of missing data, with investigators frequently ignoring or failing to handle missing data appropriately. Despite the availability of a wide range of methods for handling missing data, only a handful of studies used the statistical modelling procedures. When imputation was undertaken, the reporting of the imputation procedures was often incomplete. Although multiple imputation is becoming more accessible in research, only three studies used this method, with no details of the method being provided. Despite an increased interest in recent years in the need for understanding and appropriately handling missing data, the scarcity of information on these issues points to the widespread failure to understand the significance of the problem among medical researchers; hence the need to more formally address this issue.

In an effort to understand the lack of reporting and correct data handling in these studies, it must be noted that many imputation methods have mainly been developed theoretically and tested by statisticians. Medical professionals without any experience in statistics may struggle or chose not to undertake imputation procedures for missing data. Suggested reporting guidelines state the inclusion of the number of missing values, along with the reasons for the missing data, and the important differences between individuals with complete and incomplete data [91]. These guidelines can be useful for journal editors and authors alike, as hitherto the full impact of missing data on the research results is not usually considered.

Our review has limitations that are worth considering. Though we aimed to comprehensively review all papers on development and validation of undiagnosed diabetes risk prediction models, given that we relied on a single review article with a simple supplemental search, we may have missed some studies. Furthermore, MI was not widely accessible prior to 1997 (the earliest date of publication of the included article) so papers published immediately after this are more likely to have used complete case analysis or single imputation [92].

## Conclusion

This review highlights the inadequate reporting and handling of missing data in prevalent diabetes prediction research. Appropriate understanding, interpretation and efficient handling of missing data in medical research are essential, as incomplete data and the less than ideal methods in dealing with this can severely affect study estimates and other inferences in general. Publication of formal guidelines on the uniform reporting of missing data and methods for handling them at the analysis stage is warranted. These guidelines should be accessible to all levels of practitioners and researchers to allow for easy implementation, ultimately enhancing the validity of reported results in all spheres of prediction research.

## Competing interests

Nothing to declare

## Acknowledgements

Katya Masconi was supported by a scholarship from the South African National Research Foundation and the Carl & Emily Fuchs Foundation.

## References

1. Steyerberg, E.W. *Clinical prediction models*. 2009, New York: Springer.
2. IDF. *Idf diabetes atlas, 6th edition*. 2013, International Diabetes Federation: Brussels, Belgium.
3. Kengne, A.P., Masconi, K., Mbanya, V.N., Lekoubou, A., Echouffo-Tcheugui, J.B., and Matsha, T.E. *Risk predictive modelling for diabetes and cardiovascular disease.* Critical reviews in clinical laboratory sciences. 2013; **51**(1): p. 1-12.
4. Brown, N., Critchley, J., Bogowicz, P., Mayige, M., and Unwin, N. *Risk scores based on self-reported or available clinical data to detect undiagnosed type 2 diabetes: A systematic review.* Diabetes research and clinical practice. 2012; **98**(3): p. 369-385.
5. Bindraban, N.R., van Valkengoed, I.G., Mairuhu, G., Holleman, F., Hoekstra, J.B., Michels, B.P.*, et al*. *Prevalence of diabetes mellitus and the performance of a risk score among hindustani surinamese, african surinamese and ethnic dutch: A cross-sectional population-based study.* BMC Public Health. 2008; **8**(1): p. 271.
6. Ginde, A.A., Delaney, K.E., Lieberman, R.M., Vanderweil, S.G., and Camargo, C.A. *Estimated risk for undiagnosed diabetes in the emergency department: A multicenter survey.* Academic Emergency Medicine. 2007; **14**(5): p. 492-495.
7. Robinson, C., Agarwal, G., and Nerenberg, K. *Validating the canrisk prognostic model for assessing diabetes risk in canada's multi-ethnic population.* Chronic Dis Inj Can. 2011; **32**(1): p. 19-31.
8. Rolka, D.B., Narayan, K.V., Thompson, T.J., Goldman, D., Lindenmayer, J., Alich, K.*, et al*. *Performance of recommended screening tests for undiagnosed diabetes and dysglycemia.* Diabetes Care. 2001; **24**(11): p. 1899-1903.
9. Spijkerman, A.M., Yuyun, M.F., Griffin, S.J., Dekker, J.M., Nijpels, G., and Wareham, N.J. *The performance of a risk score as a screening test for undiagnosed hyperglycemia in ethnic minority groups data from the 1999 health survey for england.* Diabetes Care. 2004; **27**(1): p. 116-122.
10. Zhang, L., Zhang, Z., Zhang, Y., Hu, G., and Chen, L. *Evaluation of finnish diabetes risk score in screening undiagnosed diabetes and prediabetes among us adults by gender and race: Nhanes 1999-2010.* PloS one. 2014; **9**(5): p. e97865.
11. Griffin, S., Little, P., Hales, C., Kinmonth, A., and Wareham, N. *Diabetes risk score: Towards earlier detection of type 2 diabetes in general practice.* Diabetes/metabolism research and reviews. 2000; **16**(3): p. 164-171.
12. Ko, G., So, W., Tong, P., Ma, R., Kong, A., Ozaki, R.*, et al*. *A simple risk score to identify southern chinese at high risk for diabetes.* Diabetic medicine. 2010; **27**(6): p. 644-649.
13. de Sousa, A.G.P., Pereira, A.C., Marquezine, G.F., do Nascimento-Neto, R.M., Freitas, S.N., Nicolato, R.L.d.C.*, et al*. *Derivation and external validation of a simple prediction model for the diagnosis of type 2 diabetes mellitus in the brazilian urban population.* European journal of epidemiology. 2009; **24**(2): p. 101-109.

14. Lindström, J. and Tuomilehto, J. *The diabetes risk score a practical tool to predict type 2 diabetes risk.* Diabetes Care. 2003; **26**(3): p. 725-731.
15. Liu, M., Pan, C., and Jin, M. *A chinese diabetes risk score for screening of undiagnosed diabetes and abnormal glucose tolerance.* Diabetes technology & therapeutics. 2011; **13**(5): p. 501-507.
16. Al-Lawati, J. and Tuomilehto, J. *Diabetes risk score in oman: A tool to identify prevalent type 2 diabetes among arabs of the middle east.* Diabetes research and clinical practice. 2007; **77**(3): p. 438-444.
17. Baan, C.A., Ruige, J.B., Stolk, R.P., Witteman, J., Dekker, J.M., Heine, R.J.*, et al*. *Performance of a predictive model to identify undiagnosed diabetes in a health care setting.* Diabetes Care. 1999; **22**(2): p. 213-219.
18. Bang, H., Edwards, A.M., Bomback, A.S., Ballantyne, C.M., Brillon, D., Callahan, M.A.*, et al*. *Development and validation of a patient self-assessment score for diabetes risk.* Annals of Internal Medicine. 2009; **151**(11): p. 775-783.
19. Chaturvedi, V., Reddy, K., Prabhakaran, D., Jeemon, P., Ramakrishnan, L., Shah, P.*, et al*. *Development of a clinical risk score in predicting undiagnosed diabetes in urban asian indian adults: A population-based study.* CVD prevention and control. 2008; **3**(3): p. 141-151.
20. Franciosi, M., De Berardis, G., Rossi, M.C., Sacco, M., Belfiglio, M., Pellegrini, F.*, et al*. *Use of the diabetes risk score for opportunistic screening of undiagnosed diabetes and impaired glucose tolerance the igloo (impaired glucose tolerance and long-term outcomes observational) study.* Diabetes Care. 2005; **28**(5): p. 1187-1194.
21. Glümer, C., Carstensen, B., Sandbæk, A., Lauritzen, T., Jørgensen, T., and Borch-Johnsen, K. *A danish diabetes risk score for targeted screening the inter99 study.* Diabetes Care. 2004; **27**(3): p. 727-733.
22. Glümer, C., Borch-Johnsen, K., and Colagiuri, S. *Can a screening programme for diabetes be applied to another population?* Diabetic medicine. 2005; **22**(9): p. 1234-1238.
23. Glümer, C., Vistisen, D., Borch-Johnsen, K., and Colagiuri, S. *Risk scores for type 2 diabetes can be applied in some populations but not all.* Diabetes Care. 2006; **29**(2): p. 410-414.
24. Gray, L., Taub, N., Khunti, K., Gardiner, E., Hiles, S., Webb, D.*, et al*. *The leicester risk assessment score for detecting undiagnosed type 2 diabetes and impaired glucose regulation for use in a multiethnic uk setting.* Diabetic medicine. 2010; **27**(8): p. 887-895.
25. Gray, L.J., Barros, H., Raposo, L., Khunti, K., Davies, M.J., and Santos, A.C. *The development and validation of the portuguese risk score for detecting type 2 diabetes and impaired fasting glucose.* Primary care diabetes. 2013; **7**(1): p. 11-18.
26. Heianza, Y., Arase, Y., Saito, K., Hsieh, S.D., Tsuji, H., Kodama, S.*, et al*. *Development of a screening score for undiagnosed diabetes and its application in estimating absolute risk of future type 2 diabetes in japan: Toranomon hospital health management center study 10 (topics 10).* The Journal of Clinical Endocrinology & Metabolism. 2013; **98**(3): p. 1051-1060.
27. Heikes, K.E., Eddy, D.M., Arondekar, B., and Schlessinger, L. *Diabetes risk calculator a simple tool for detecting undiagnosed diabetes and pre-diabetes.* Diabetes Care. 2008; **31**(5): p. 1040-1045.
28. Keesukphan, P., Chanprasertyothin, S., Ongphiphadhanakul, B., and Puavilai, G. *The development and validation of a diabetes risk score for high-risk thai adults.* Journal of the Medical Association of Thailand. 2007; **90**(1): p. 149.
29. Lee, Y.-h., Bang, H., Kim, H.C., Kim, H.M., Park, S.W., and Kim, D.J. *A simple screening score for diabetes for the korean population development, validation, and comparison with other scores.* Diabetes Care. 2012; **35**(8): p. 1723-1730.
30. Mohan, V., Deepa, R., Deepa, M., Somannavar, S., and Datta, M. *A simplified indian diabetes risk score for screening for undiagnosed diabetic subjects.* Journal of the Association of Physicians of India. 2005; **53**: p. 759-763.

31. Park, P., Griffin, S., Sargeant, L., and Wareham, N. *The performance of a risk score in predicting undiagnosed hyperglycemia.* Diabetes Care. 2002; **25**(6): p. 984-988.
32. Rahman, M., Simmons, R.K., Harding, A.-H., Wareham, N.J., and Griffin, S.J. *A simple risk score identifies individuals at high risk of developing type 2 diabetes: A prospective cohort study.* Family practice. 2008; **25**(3): p. 191-196.
33. Ramachandran, A., Snehalatha, C., Vijay, V., Wareham, N., and Colagiuri, S. *Derivation and validation of diabetes risk score for urban asian indians.* Diabetes research and clinical practice. 2005; **70**(1): p. 63-70.
34. Rathmann, W., Martin, S., Haastert, B., Icks, A., Holle, R., Löwel, H.*, et al*. *Performance of screening questionnaires and risk scores for undiagnosed diabetes: The kora survey 2000.* Archives of internal medicine. 2005; **165**(4): p. 436-441.
35. Ruige, J.B., de Neeling, J.N.D., Kostense, P.J., Bouter, L.M., and Heine, R.J. *Performance of an niddm screening questionnaire based on symptoms and risk factors.* Diabetes Care. 1997; **20**(4): p. 491-496.
36. Witte, D., Shipley, M., Marmot, M., and Brunner, E. *Performance of existing risk scores in screening for undiagnosed diabetes: An external validation study.* Diabetic medicine. 2010; **27**(1): p. 46-53.
37. Zhou, X., Qiao, Q., Ji, L., Ning, F., Yang, W., Weng, J.*, et al*. *Nonlaboratory-based risk assessment algorithm for undiagnosed type 2 diabetes developed on a nation-wide diabetes survey.* Diabetes Care. 2013; **36**(12): p. 3944-3952.
38. Winkler, G., Hídvégi, T., Vándorfi, G., Balogh, S., and Jermendy, G. *Prevalence of undiagnosed abnormal glucose tolerance in adult patients cared for by general practitioners in hungary. Results of a risk-stratified screening based on findrisc questionnaire.* Medical science monitor: international medical journal of experimental and clinical research. 2012; **19**: p. 67-72.
39. Adhikari, P., Pathak, R., and Kotian, S. *Validation of the mdrf-indian diabetes risk score (idrs) in another south indian population through the boloor diabetes study (bds).* JAPI. 2010; **50**: p. 434-436.
40. Akyil, R.C., Miloglu, O., Olgun, N., and Bayrakdar, I.S. *A comparison of three different diabetes screening methods among dental patients in turkey.* Pakistan journal of medical sciences. 2014; **30**(1): p. 65.
41. Al Khalaf, M., Eid, M., Najjar, H., Alhajry, K., Doi, S., and Thalib, L. *Screening for diabetes in kuwait and evaluation of risk scores.* EMHJ. 2010; **16**(7): p. 725-731.
42. Baan, C., Stolk, R., Grobbee, D., Witteman, J., and Feskens, E. *Physical activity in elderly subjects with impaired glucose tolerance and newly diagnosed diabetes mellitus.* American journal of epidemiology. 1999; **149**(3): p. 219-227.
43. Bergmann, A., Li, J., Wang, L., Schulze, J., Bornstein, S., and Schwarz, P. *A simplified finnish diabetes risk score to predict type 2 diabetes risk and disease evolution in a german population.* Hormone and metabolic research. 2007; **39**(09): p. 677-682.
44. Reddy, K.S., Prabhakaran, D., Chaturvedi, V., Jeemon, P., Thankappan, K., Ramakrishnan, L.*, et al*. *Methods for establishing a surveillance system for cardiovascular diseases in indian industrial populations.* Bulletin of the world health organization. 2006; **84**(6): p. 461-469.
45. Cabrera de León, A., Coello, S.D., Rodríguez Pérez, M.d.C., Medina, M.B., Almeida González, D., Díaz, B.B.*, et al*. *A simple clinical score for type 2 diabetes mellitus screening in the canary islands.* Diabetes research and clinical practice. 2008; **80**(1): p. 128-133.
46. Gao, W., Dong, Y., Pang, Z., Nan, H., Wang, S., Ren, J.*, et al*. *A simple chinese risk score for undiagnosed diabetes.* Diabetic medicine. 2010; **27**(3): p. 274-281.
47. Hanif, M., Valsamakis, G., Dixon, A., Boutsiadis, A., Jones, A., Barnett, A.*, et al*. *Detection of impaired glucose tolerance and undiagnosed type 2 diabetes in uk south asians: An effective screening strategy.* Diabetes, Obesity and Metabolism. 2008; **10**(9): p. 755-762.

48. Heldgaard, P.E. and Griffin, S. *Routinely collected general practice data aids identification of people with hyperglycaemia and metabolic syndrome.* Diabetic medicine. 2006; **23**(9): p. 996-1002.
49. Ku, G. and Kegels, G. *The performance of the finnish diabetes risk score, a modified finnish diabetes risk score and a simplified finnish diabetes risk score in community-based cross-sectional screening of undiagnosed type 2 diabetes in the philippines.* Primary care diabetes. 2013; **7**(4): p. 249-259.
50. Li, J., Bergmann, A., Reimann, M., Bornstein, S., and Schwarz, P. *A more simplified finnish diabetes risk score for opportunistic screening of undiagnosed type 2 diabetes in a german population with a family history of the metabolic syndrome.* Hormone and metabolic research. 2009; **41**(02): p. 98-103.
51. Lin, J.-W., Chang, Y.-C., Li, H.-Y., Chien, Y.-F., Wu, M.-Y., Tsai, R.-Y*., et al*. *Cross-sectional validation of diabetes risk scores for predicting diabetes, metabolic syndrome, and chronic kidney disease in taiwanese.* Diabetes Care. 2009; **32**(12): p. 2294-2296.
52. Saaristo, T., Peltonen, M., Lindström, J., Saarikoski, L., Sundvall, J., Eriksson, J.G*., et al*. *Cross-sectional evaluation of the finnish diabetes risk score: A tool to identify undetected type 2 diabetes, abnormal glucose tolerance and metabolic syndrome.* Diabetes and Vascular Disease Research. 2005; **2**(2): p. 67-72.
53. Ta, M., Nguyen, K., Nguyen, N., Campbell, L., and Nguyen, T. *Identification of undiagnosed type 2 diabetes by systolic blood pressure and waist-to-hip ratio.* Diabetologia. 2010; **53**(10): p. 2139-2146.
54. Tankova, T., Chakarova, N., Atanassova, I., and Dakovska, L. *Evaluation of the finnish diabetes risk score as a screening tool for impaired fasting glucose, impaired glucose tolerance and undetected diabetes.* Diabetes research and clinical practice. 2011; **92**(1): p. 46-52.
55. Graham, J.W., Hofer, S.M., and MacKinnon, D.P. *Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures.* Multivariate Behavioral Research. 1996; **31**(2): p. 197-218.
56. Alzola, C.F. and Harrell, F.E. *An introduction to s and the hmisc and design libraries.* URL http://biostat. mc. vanderbilt. edu/twiki/pub/Main/RS/sintro. pdf. 2002; **30**.
57. Campbell, M.J. *Statistics at square two : Understanding modern statistical applications in medicine*. 2nd ed. 2006, Malden, Mass.: Blackwell.
58. Krumholz, H.M., Wang, Y., Mattera, J.A., Han, L.F., Ingber, M.J., Roman, S*., et al*. *An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction.* Circulation. 2006; **113**(13): p. 1683.
59. van der Meulen, J.H., Steyerberg, E.W., van der Graaf, Y., van Herwerden, L.A., Verbaan, C.J., Defauw, J.J*., et al*. *Age thresholds for prophylactic replacement of bjork-shiley convexo-concave heart valves. A clinical and economic evaluation.* Circulation. 1993; **88**(1): p. 156.
60. Austin, P.C. and Tu, J.V. *Bootstrap methods for developing predictive models in cardiovascular research.* Am Stat. 2004; **58**: p. 131/137.
61. Murray, G.D., Butcher, I., McHugh, G.S., Lu, J., Mushkudiani, N.A., Maas, A.I*., et al*. *Multivariable prognostic analysis in traumatic brain injury: Results from the impact study.* J Neurotrauma. 2007; **24**(2): p. 329.
62. Buuren, S. and Groothuis-Oudshoorn, K. *Mice: Multivariate imputation by chained equations in r.* Journal of statistical software. 2011; **45**(3).
63. Sande, I.G. *Hot-deck imputation procedures.* Incomplete data in sample surveys. 1983; **3**: p. 334-350.
64. Little, R. and Rubin, D. *The analysis of social science data with missing values.* Sociological Methods & Research. 1989; **18**(2-3): p. 292-326.

65. Templ, M., Alfons, A., Kowarik, A., and Prantner, B. *Vim: Visualization and imputation of missing values.* R package version. 2011; **2**(3).
66. Schafer, J.L. *Analysis of incomplete multivariate data*. 2010: CRC press.
67. Hadfield, J.D. *Mcmc methods for multi-response generalized linear mixed models: The mcmcglmm r package.* Journal of Statistical Software. 2010; **33**(2): p. 1-22.
68. Dempster, A.P., Laird, N.M., and Rubin, D.B. *Maximum likelihood from incomplete data via the em algorithm.* Journal of the Royal Statistical Society. Series B (Methodological). 1977: p. 1-38.
69. Enders, C.K. *A primer on maximum likelihood algorithms available for use with missing data.* Structural Equation Modeling. 2001; **8**(1): p. 128-141.
70. Honaker, J., King, G., and Blackwell, M. *Amelia ii: A program for missing data.* Journal of Statistical Software. 2011; **45**(7): p. 1-47.
71. Simpson, E.H. *The interpretation of interaction in contingency tables.* Journal of the Royal Statistical Society. Series B (Methodological). 1951: p. 238-241.
72. Rubin, D.B. *Inference and missing data.* Biometrika. 1976; **63**(3): p. 581-592.
73. Fairclough, D. *Methods of analysis for longitudinal studies of health-related quality of life.* Quality of life assessment in clinical trials: methods and practice. 1998: p. 227-247.
74. King, D., King, L., Bachrach, P., and McArdle, J. *Contemporary approaches to missing data: The glass is really half full.* PTSD Res Q. 2001; **12**: p. 1-6.
75. Hair, J.F., Anderson, R.E., Tatham, R.L., and William, C. *Black (1998), multivariate data analysis*. 1998, Upper Saddle River, NJ: Prentice Hall.
76. Curran, D., Fayers, P., Molenberghs, G., and Machin, D. *Incomplete quality of life data in randomized trials: Missing forms.* Statistics in medicine. 1998; **17**(5-7): p. 697-709.
77. Little, R.J. *A test of missing completely at random for multivariate data with missing values.* Journal of the American Statistical Association. 1988; **83**(404): p. 1198-1202.
78. Little, R.J. and Rubin, D.B. *Statistical analysis with missing data.* 2002.
79. Dong, Y. and Peng, C.-Y.J. *Principled missing data methods for researchers.* SpringerPlus. 2013; **2**(1): p. 1-17.
80. Schafer, J.L. *Multiple imputation: A primer.* Statistical methods in medical research. 1999; **8**(1): p. 3-15.
81. Begg, C.B. and Gray, R. *Calculation of polychotomous logistic regression parameters using individualized regressions.* Biometrika. 1984; **71**(1): p. 11.
82. Hawthorne, G. and Elliott, P. *Imputing cross-sectional missing data: Comparison of common techniques.* Aust N Z J Psychiatry. 2005; **39**(7): p. 583-590.
83. Saunders, J.A., Morrow-Howell, N., Spitznagel, E., Doré, P., Proctor, E.K., and Pescarino, R. *Imputing missing data: A comparison of methods for social work researchers.* Social work research. 2006; **30**(1): p. 19-31.
84. Vittinghoff, E., Shiboski, S., and McCulloch, C.E. *Regression methods in biostatistics*. 2005: Springer.
85. Little, R. and Rubin, D.B. *Statistical analysis with missing data.* John A. Wiley & Sons, Inc., New York. 1987.
86. Rubin, D.B. *Multiple imputation for nonresponse in surveys*. 1987, New York: J. Wiley & Sons.
87. Schapire, R.E. *The strength of weak learnability.* Machine Learning. 1990; **5**(2): p. 197–227.
88. Rubin, D.B. *Multiple imputation for nonresponse in surveys. New york: J*. 1987, Wiley & Sons.
89. Rosenbaum, P.R. and Rubin, D.B. *The central role of the propensity score in observational studies for causal effects.* Biometrika. 1983; **70**(1): p. 41-55.
90. Little, R.J. *Missing-data adjustments in large surveys.* Journal of Business & Economic Statistics. 1988; **6**(3): p. 287-296.
91. Sterne, J.A., White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M.G*., et al*. *Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls.* Bmj. 2009; **338**: p. b2393.

92. Karahalios, A., Baglietto, L., Carlin, J.B., English, D.R., and Simpson, J.A. *A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures.* BMC medical research methodology. 2012; **12**(1): p. 96.

07

# Effects of different missing data imputation techniques on the performance of undiagnosed diabetes risk prediction models in a mixed-ancestry population of South Africa

Katya Masconi, Tandi E. Matsha, Rajiv T. Erasmus and Andre P. Kengne

*PLoS One* 2015; **10**:9

## Abstract

**Background:** Imputation techniques used to handle missing data are based on the principle of replacement. It is widely advocated that multiple imputation is superior to other imputation methods, however studies have suggested that simple methods for filling missing data can be just as accurate as complex methods. The objective of this study was to implement a number of simple and more complex imputation methods, and assess the effects on these techniques on the performance of undiagnosed diabetes risk prediction models during external validation.

**Methods:** Data from the Cape Town Bellville-South cohort served as the basis for this study. Imputation methods and models were identified via recent systematic reviews. Models' discrimination was assessed and compared using C-statistic and non-parametric methods, before and after recalibration through simple intercept adjustment.

**Results:** The study sample consisted of 1256 individuals, of whom 173 were excluded due to previously diagnosed diabetes. Of the final 1083 individuals, 329 (30.4%) had missing data. Family history had the highest proportion of missing data (25%). Imputation of the outcome, undiagnosed diabetes, was highest in stochastic regression imputation (163 individuals). Overall, deletion resulted in the lowest model performances while simple imputation yielded the highest C-statistic for the Cambridge Diabetes Risk model, Kuwaiti Risk model, Omani Diabetes Risk model and Rotterdam Predictive model. Multiple imputation only yielded the highest C-statistic for the Rotterdam Predictive model, which was matched by simpler imputation methods.

**Conclusions:** Deletion was confirmed as a poor technique for handling missing data. However, despite the emphasized disadvantages of simpler imputation methods, this study showed that implementing these methods results in similar predictive utility for undiagnosed diabetes when compared to multiple imputation.

Keywords: Imputation, effect, performance, prediction, undiagnosed, diabetes

## Background

Missing data is common in predictive research, and can negatively affect the performance of risk prediction models. In an ideal setting, a subject with missing data on a predictor or outcome variable should be replaced with a randomly selected subject from the source population. However, replacement is burdensome and most often impossible. Instead, researchers can use observed data to make an estimation of the status of the participants for the characteristic with missing value. Imputation techniques are based on the basic principle of replacement, indicating that any conclusion drawn from the study should not depend on the sample that is involved in the study. Should each subject in the chosen sample be replaced by a new subject from the same source population as the original subject, the conclusions should not be compromised [1].

It is widely advocated that imputation of missing data is superior to the overlooking of the missing data, that the indicator method often provides biased results, that conditional mean imputation is better than unconditional implementation, and that multiple imputation method is better than single imputation [1-13]. However, studies have suggested that simple methods for filling missing data can be just as accurate as complex methods, allowing for easier implementation in prediction studies [14, 15]. The type and percentage of missing data are important determining factors for the accuracy of the different imputation methods. Data missing completely at random (MCAR) has a low probability that the observation missing is related to any other patient characteristics and most simple techniques for handling missing data give unbiased results [4]. When the missing data depends on information that is not observed, the missing data is considered missing not at random (MNAR) [3]. Although there is no advocated method available to handle the valuable information that has been lost through MNAR data, multiple imputation can be unbiased for MNAR data [2]. Most often, missing data are neither MCAR nor MNAR [12], but rather missing at random (MAR). This type of missing data is missing at random conditional on the individuals other characteristics that are available at the time of analysis [3]. When missing data are MAR, common and simple techniques used to handle missing data such as complete case and available case analysis,

indicator method and overall mean imputation are likely to introduce selection bias as the database is no longer a random sample of the source population [5, 7, 12, 16].

This study aims to implement a number of simple and more complex imputation methods for filling missing data, and assess the comparative effects on the performance of undiagnosed diabetes risk prediction models during external validation. For this purpose, we use data for mixed-ancestry South Africans who took part in the Bellville-South study in Cape Town.

## Methodology

### Database

Details of the study design and recruitment of the database that served as the basis for all imputation methods implementation have been described below. The Bellville South Study was a cross-sectional study conducted from mid-January 2008 to March 2009 (cohort 1), and from January 2011 to November 2011 (cohort 2). The study was approved by the Ethics Committee of the Cape Peninsula University of Technology and Stellenbosch University. All participants signed written informed consent after all the procedures had been fully explained in the language of their choice.

### Research setting

Bellville-South is located within the Northern suburbs of Cape Town, South Africa and is a traditionally a Coloured township formed in the late 1950s. According to the 2011 population census, its population stands at approximately 29 301 with 76.0% (22 270) consisting of the mixed ancestry individuals [17, 18]. The target population for this study were subjects between the ages of 35 and 65 years and their number was estimated to be 6 500 in the 2001 population census [19].

**Research Design and Study Population**

Using a map of Bellville South, multistage stratified random sampling was approached as follows: From a list of streets of each stratum, the streets were then classified as short, medium and long streets based on the number of houses. Streets with houses ≤ 22 were classified as short, medium; houses 23–40 and long streets were > 40 houses. A total of 16 short streets representing approximately 190 houses, 15 medium streets representing approximately 410 houses and 12 long streets representing approximately 400 houses were randomly selected across the different strata. From the selected streets, all household members meeting the selection criteria were invited to participate in the study. Community authorities requested that participants outside the random selection area also benefit from the study.

**Recruitment Strategy**

Information regarding the project was disseminated to the local residents through the local radio station, community newspaper, brochures and fliers; the latter bearing information about the project and distributed through school children and taxis to the local residents by the recruitment team. Recruited subjects were visited by the recruitment team the evening before participation and reminded of all the survey instructions.

**Data collection**

A detailed protocol describing data-collection procedures (questionnaires and physical examination) was developed. The questionnaire designed to retrospectively obtain information on lifestyle factors such as smoking and alcohol consumption, physical activity, diet, family history of CVD and DM, and demographics was administered by trained personnel. A detailed drug history was obtained by interrogation and by examining the clinic cards as well as the record of drugs that participants brought to the study site. Clinical measurements included height, weight, hip and waist circumferences, body fat measurements and blood pressure.

**Diabetes diagnosis**

All participants, except the self-reported diabetic subjects, confirmed by either medical card record or drugs in use, had blood taken for fasting blood glucose and underwent a 75 g oral glucose tolerance test (OGTT) as prescribed by the WHO. Diabetes was diagnosed according to the WHO 2006 criteria [20].

**Identification of undiagnosed diabetes prediction models**

Existing prediction models were obtained from a systematic review by Brown *et al*, 2012 [21]. Models met the criteria for model selection for this paper if they were developed to predict the presence of undiagnosed diabetes based on predictors measured in the Bellville South study. We focused on models developed from non-invasively measure predictors. Therefore the models retained were: Cambridge Risk model [22], Kuwaiti Risk model [23], Omani Diabetes Risk model [24], Rotterdam Predictive model 1 [25] and the simplified Finnish Diabetes Risk model [26]. Model characteristics and formulas have been published by Masconi *et al* [27]. All models included age as a predictor, while a range of other predictors were variably combined in models. These included: sex, BMI, use of antihypertensive medication, family history of diabetes, waist circumference, past or current smoking and the use of corticosteroids. Table 1 shows the overview of the performance of the prevalent diabetes risk prediction models across the five imputation methods.

**Statistical methods**

*Analysis of missing data*

Data analysis used the R statistical software, version 3.1.2 [28]. Aggregation plots were created using the 'VIM' package to identify of the pattern of missing data for each variable. The corresponding frequencies were tabulated.

*Identification of imputation methods*

A comprehensive search was previously carried out on the imputation methods available [29]. The aim was to compare deletion, single and multiple imputation techniques. To allow for a broad spectrum of techniques, it was decided to compare pair-wise deletion [30], simple imputation [31], conditional mean imputation [8, 30], stochastic regression [8, 32] and multiple imputation for non-monotone missing patterns [16]. Imputation was completed on the outcome and all variables. Where applicable, the outcome and all variables were used as a predictor for the variable being imputed.

*Imputation*

Simple imputation via mean substitution was implemented with the package '*Hmisc*' through the function '*impute (x, fun=mean)*' where x is the predictor of interest [33]. Conditional mean imputation was implemented through the creation of a regression model and the subsequent inbuilt '*predict*' function. Imputation via Stochastic regression used the method '*norm.nob*' of the R package '*mice*' was used [34]. Multiple imputation for non-monotone missing patterns via the Multiple Imputation by Chained Equations (MICE) method, using fully conditional specification (FCS) was implemented using the '*mice*' package [35]. The m imputed datasets were analysed separately, then the estimates and the associated variance from the imputed data sets combined using rules established by Rubin that incorporates the within and between imputation variability [7].

*Model performance*

The original selected models were validated for the overall data and subsets using the formulas, both prior to recalibration and following intercept adjustment to eliminate differences in diabetes prevalence between the development population of the model and this test population. The predicted probability of undiagnosed diabetes for each participant was computed using the baseline measured predictors. The performance was expressed in terms of discrimination and calibration. Discrimination describes the ability of the model's performance in distinguishing those at a high risk of developing diabetes from those at low risk [36]. The discrimination was assessed and compared using concordance (C) statistic and non-parametric methods [37].

Calibration describes the agreement between the probability of the outcome of interest as estimated by the model, and the observed outcome frequencies [30]. It was assessed with formal statistical tests, determining the agreement between the expected (E) and observed (O) rates (E/O). The 95% confidence intervals for the expected/observed probabilities (E/O) ratio were calculated assuming a Poisson distribution [38]. We also calculated 1) the Yates slope, which is the difference between mean predicted probability of type 2 diabetes for participants with and without prevalent undiagnosed diabetes, with higher values indicate better performance; and 2) the Brier score, which is the squared difference between predicted probability and actual outcome for each participant with values ranging between 0 for a perfect prediction model and 1 for no match in prediction and outcome [30, 36].

## Results

### Data available

The study sample consisted of 1256 individuals, of whom 173 were excluded due to previously diagnosed diabetes. Of the final 1083 individuals, 329 (30.4%) had missing data. Table 2 summarises the number of missing values for each variable included in the 5 selected risk prediction models. Additionally, Figures 1 and 2 show the proportion and combinations of missing data respectively. Family history was the variable with the most missing data [mother (25.1%, father (24.9%), sister (25.0%), and brother (25.1%)]. The rest of the variables had a missing proportion of less than 5%, except smoking status (6.1%).

**Table 1:** Overview of the performance of the undiagnosed diabetes risk prediction models across the five imputation methods

| Models | | Deletion | | Simple | | Conditional | | Stochastic | | Multiple | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Original | Adjusted | Original | Adjusted | Original | Adjusted | Original | Adjusted | Original | Adjusted |
| Cambridge Diabetes Risk model | E/O (95% CI) | 1.81 (1.09 ; 2.52) | 1.22 (0.61-1.83) | 2.07 (1.40 ; 2.75) | 1.28 (0.69 – 1.87) | 2.01 (1.28 ; 2.75) | 1.27 (0.64 – 1.90) | 2.17 (1.41 ; 2.93) | 1.27 (0.64 – 1.90) | 2.16 (1.40 ; 2.92) | 1.30 (0.66 – 1.94) |
| | Brier score | 0.193 | | 0.181 | | 0.185 | | 0.186 | | 0.189 | |
| | Yates slope | 0.379 | | -1.401 | | -1.374 | | -1.399 | | -1.441 | |
| | C-statistic (95% CI) | 0.67 (0.62 – 0.72) | | 0.69 (0.65 – 0.73) | | 0.68 (0.63 – 0.72) | | 0.68 (0.64 – 0..73) | | 0.68 (0.64 – 0.72) | |
| Kuwaiti Risk model | E/O (95% CI) | 0.72 (0.40 ; 1.12 ) | 0.94 (0.47-1.41) | 0.79 (0.39 ; 1.18) | 0.96 (0.51 – 1.41) | 0.79 (0.34 ; 1.25) | 0.96 (0.45 – 1.47) | 0.82 (0.44 ; 1.20) | 0.96 (0.45 – 1.47) | 0.82 (0.42 ; 1.22) | 0.96 (0.55 – 1.37) |
| | Brier score | 0.141 | | 0.122 | | 0.126 | | 0.125 | | 0.123 | |
| | Yates slope | 0.496 | | -0.459 | | -0.514 | | -0.473 | | -0.534 | |
| | C-statistic (95% CI) | 0.68 (0.63 – 0.73) | | 0.70 (0.66 – 0.74) | | 0.69 (0.65 – 0.73) | | 0.69 (0.65 – 0.74) | | 0.69 (0.65 – 0.73) | |
| Omani Diabetes Risk model | E/O (95% CI) | 1.28 (0.63 ; 1.93) | 1.06 (0.47-1.66) | 1.40 (0.82 ; 1.98) | 1.08 (0.56 – 1.60) | 1.40 (0.75 ; 2.05) | 1.08 (0.50 – 1.66) | 1.56 (0.81 ; 2.30) | 1.08 (0.50 – 1.66) | 1.54 (0.77 ; 2.31) | 1.11 (0.51 – 1.71) |
| | Brier score | 0.164 | | 0.141 | | 0.149 | | 0.142 | | 0.153 | |
| | Yates slope | 0.392 | | -1.065 | | -1.104 | | -1.049 | | -1.196 | |
| | C-statistic (95% CI) | 0.66 (0.61 – 0.70) | | 0.67 (0.63 – 0.71) | | 0.65 (0.61 – 0.70) | | 0.67 (0.63 – 0.72) | | 0.65 (0.61 – 0.69) | |
| Rotterdam Predictive model | E/O (95% CI) | 0.54 (0.50 ; 1.04) | 0.98 (0.91-1.05) | 0.65 (0.56 ; 0.74) | 0.99 (0.83 – 1.14) | 0.59 (0.48 ; 0.71) | 0.99 (0.93 – 1.04) | 0.65 (0.57 ; 0.74) | 0.99 (0.93 – 1.04) | 0.65 (0.57 ; 0.73) | 0.99 (0.87 – 1.11) |
| | Brier score | 0.147 | | 0.126 | | 0.130 | | 0.129 | | 0.127 | |
| | Yates slope | 0.971 | | 0.558 | | 0.539 | | 0.535 | | 0.498 | |
| | C-statistic (95% CI) | 0.64 (0.59 – 0.69) | | 0.65 (0.61 – 0.70) | | 0.65 (0.60 – 0.69) | | 0.65 (0.60 – 0.70) | | 0.65 (0.61 – 0.70) | |
| Simplified Finnish Diabetes Risk model | E/O (95% CI) | 0.26 (0.13 ; 0.39) | 0.89 (0.51-1.26) | 0.34 (0.17 ; 0.52) | 0.92 (0.53 – 1.31) | 0.34 (0.18 ; 0.50) | 0.92 (0.56 – 1.28) | 0.35 (0.17 ; 0.52) | 0.92 (0.56 – 1.28) | 0.35 (0.17 ; 0.53) | 0.92 (0.53 – 1.32) |
| | Brier score | 0.157 | | 0.133 | | 0.136 | | 0.136 | | 0.133 | |
| | Yates slope | 0.491 | | -0.021 | | 0.080 | | -0.053 | | -0.045 | |
| | C-statistic (95% CI) | 0.67 (0.62 – 0.71) | | 0.66 (0.62 – 0.70) | | 0.67 (0.63 – 0.72) | | 0.66 (0.62 – 0.70) | | 0.66 (0.62 – 0.70) | |

**Table 2:** Missing data analysis

| Variable | % |
|---|---|
| Outcome (prevalent diabetes) | 0.7 |
| Age | 1.4 |
| Gender | 1.6 |
| Body mass index | 3.4 |
| Waist circumference | 2.1 |
| Systolic blood pressure | 1.9 |
| Diastolic blood pressure | 1.9 |
| Mother family history | 25.1 |
| Father family history | 24.9 |
| Sister family history | 25.0 |
| Brother family history | 25.1 |
| Corticosteroid use | 4.3 |
| Hypertensive drugs | 2.5 |
| Smoking status | 6.1 |

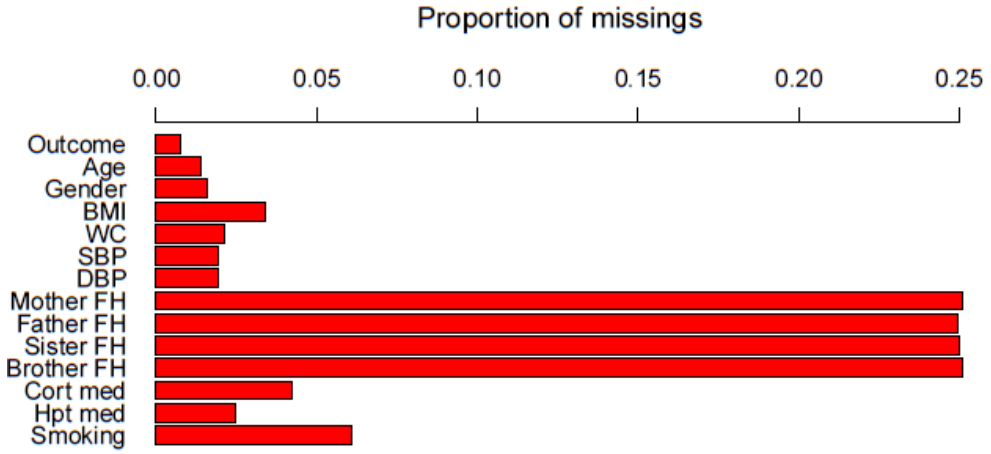



**Fig 1**: Histogram showing the proportion of missing for each variable

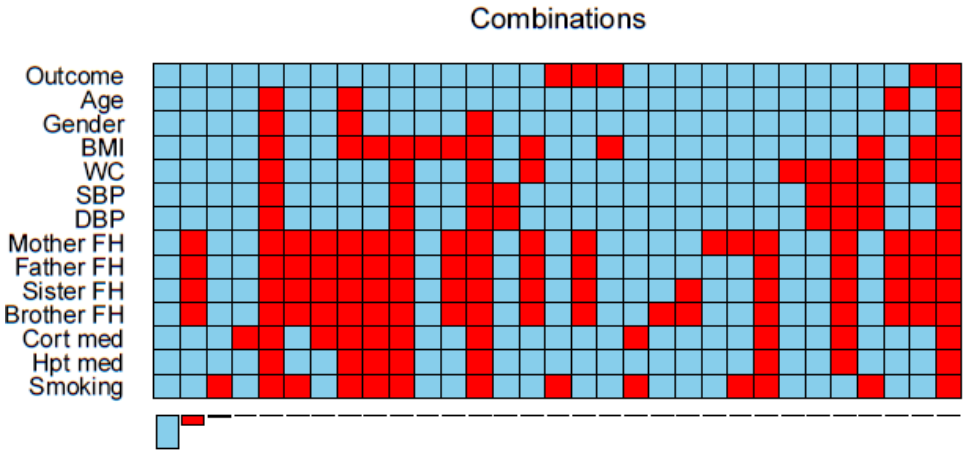



**Fig 2:** Aggregation plot showing all combinations of missing (red) and non-missing (blue) values in the variables, from the highest to lowest frequency.

*BMI, Body Mass Index; WC, Waist Circumference; SBP, Systolic Blood Pressure; DBP, Diastolic Blood Pressure; FH, Family History; Cort, Corticosteroids; med, medication; Hpt, Hypertensive

**Table 3:** Characteristics comparison of participants for the original database and five imputation methods

| | | Imputation methods | | | | |
|---|---|---|---|---|---|---|
| | Original | Pairwise deletion (754) | Simple (1083) | Conditional (Varied) | Stochastic (1083) | Multiple (1083) |
| Prevalent undiagnosed diabetes (Yes/No) | 162/913 | 132/622 | 162/921 | 162/916 | 163/920 | 162/921 |
| Age (years) | 51.9 (15.0) | 52.5 (14.6) | 51.9 (14.9) | 51.9 (15.0) | 51.8 (15.0) | 51.8 (15.1) |
| Body mass index (kg/m2) | 29.7 (7.2) | 29.6 (7.1) | 29.7 (7.0) | 29.7 (7.1) | 29.8 (7.2) | 29.8 (7.2) |
| Gender (Male/Female) | 249/810 | 160/594 | 251/832 | 251/826 | 254/829 | 257/826 |
| Systolic blood pressure (mmHg) | 124.3 (20.2) | 122.0 (18.7) | 124.3 (20.0) | 124.3 (20.2) | 124.3 (20.2) | 124.4 (20.4) |
| Diastolic blood pressure (mmHg) | 76.0 (12.9) | 74.7 (12.0) | 76.0 (12.7) | 76.0 (12.8) | 76.0 (12.9) | 76.1 (14.1) |
| Waist circumference (cm) | 95.8 (15.5) | 95.9 (14.9) | 95.8 (15.3) | 95.8 (15.4) | 95.8 (15.5) | 95.7 (16.9) |
| Hypertensive medication (Yes/No) | 374/682 | 262/492 | 374/709 | 383/688 | 387/696 | 382/701 |
| Using corticosteroids (Yes/No) | 12/1025 | 5/749 | 12/1071 | 12/1050 | 12/1071 | 13/1070 |
| Mother having diabetes (Yes/No) | 124/687 | 114/640 | 124/959 | 124/880 | 182/901 | 165/198 |
| Father having diabetes (Yes/No) | 61/752 | 60/694 | 61/1022 | 61/944 | 73/1010 | 78/1005 |
| Sister having diabetes (Yes/No) | 103/709 | 98/656 | 103/980 | 107/897 | 143/940 | 128/955 |
| Brother having diabetes (Yes/No) | 67/744 | 64/690 | 67/1016 | 67/936 | 79/1004 | 87/996 |
| Smoking status (Current/Past/No) | 433/105/479 | 327/89/338 | 433/105/545 | 437/105/496 | 456/114/513 | 458/113/512 |

**Table 4:** Characteristics comparison of participants for five multiple imputation datasets

| | Multiple imputation datasets | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Prevalent undiagnosed diabetes (Yes/No) | 162/921 | 163/920 | 162/921 | 162/921 | 163/920 |
| Age (years) | 51.9 (15.1) | 51.9 (15.0) | 51.8 (15.0) | 51.9 (15.1) | 51.8 (15.0) |
| Body mass index (kg/m2) | 29.8 (7.2) | 29.8 (7.2) | 29.8 (7.2) | 29.8 (7.2) | 29.7 (7.2) |
| Gender (Male/Female) | 258/825 | 257/826 | 257/826 | 256/827 | 258/825 |
| Systolic blood pressure (mmHg) | 124.5 (20.4) | 124.5 (20.3) | 124.4 (20.3) | 124.5 (20.4) | 124.3 (20.3) |
| Diastolic blood pressure (mmHg) | 76.1 (12.8) | 76.1 (12.8) | 76.2 (13.3) | 76.1 (12.9) | 76.1 (12.9) |
| Waist circumference (cm) | 95.8 (15.9) | 95.7 (15.5) | 95.8 (15.5) | 95.8 (15.4) | 95.7 (15.5) |
| Hypertensive medication (Yes/No) | 383/700 | 378/705 | 381/702 | 382/701 | 384/699 |
| Using corticosteroids (Yes/No) | 13/1070 | 12/1071 | 12/1071 | 13/1070 | 13/1070 |
| Mother having diabetes (Yes/No) | 157/926 | 168/915 | 155/928 | 179/904 | 164/919 |
| Father having diabetes (Yes/No) | 71/1012 | 72/1011 | 83/1000 | 78/1005 | 84/999 |
| Sister having diabetes (Yes/No) | 132/951 | 130/953 | 121/962 | 125/958 | 134/949 |
| Brother having diabetes (Yes/No) | 87/996 | 88/995 | 83/1000 | 88/997 | 88/995 |
| Smoking status (Current/Ex/No) | 464/110/509 | 455/118/510 | 459/115/509 | 452/113/518 | 460/111/512 |

**Table 5:** Overview of the performance of the undiagnosed diabetes risk prediction models across the five multiple imputation datasets

| Multiple imputation datasets | | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Cambridge | E/O (95% CI) | 2.17 (1.35 – 2.99) | 2.13 (1.40 – 2.87) | 2.15 (1.49 – 2.81) | 2.18 (1.34 – 3.01) | 2.16 (1.46 – 3.86) |
| Diabetes Risk model | Brier score | 0.190 | 0.188 | 0.186 | 0.190 | 0.190 |
| | Yates slope | -1.451 | -1.435 | -1.433 | -1.454 | -1.434 |
| | C-statistic (95% CI) | 0.68 (0.64 – 0.72) | 0.68 (0.64 – 0.72) | 0.69 (0.65 – 0.73) | 0.68 (0.64 – 0.73) | 0.69 (0.64 – 0.73) |
| Kuwaiti Risk model | E/O (95% CI) | 0.83 (0.42 – 1.24) | 0.82 (0.40 – 1.23) | 0.82 (0.45 – 1.19) | 0.83 (0.41 – 1.24) | 0.82 (0.44 – 1.19) |
| | Brier score | 0.124 | 0.124 | 0.122 | 0.123 | 0.123 |
| | Yates slope | -0.563 | -0.558 | -0.496 | -0.542 | -0.509 |
| | C-statistic (95% CI) | 0.69 (0.65 – 0.73) | 0.69 (0.64 – 0.73) | 0.70 (0.66 – 0.74) | 0.69 (0.65-0.73) | 0.69 (0.65 – 0.74) |
| Omani Diabetes | E/O (95% CI) | 1.55 (0.76 – 2.33) | 1.54 (0.72 – 2.37) | 1.52 (0.87 – 2.17) | 1.57 (0.78 – 2.37) | 1.54 (0.80 – 2.29) |
| Risk model | Brier score | 0.154 | 0.156 | 0.149 | 0.155 | 0.153 |
| | Yates slope | -1.211 | -1.232 | -1.151 | -1.214 | -1.174 |
| | C-statistic (95% CI) | 0.65 (0.61 – 0.69) | 0.64 (0.60 – 0.68) | 0.66 (0.62 – 0.70) | 0.65 (0.61 – 0.70) | 0.66 (0.61 – 0.70) |
| Rotterdam | E/O (95% CI) | 0.66 (0.57 – 0.75) | 0.65 (0.58 – 0.72) | 0.65 (0.57 – 0.74) | 0.66 (0.57 – 0.75) | 0.65 (0.57 – 0.74) |
| Predictive model | Brier score | 0.126 | 0.127 | 0.126 | 0.127 | 0.127 |
| | Yates slope | 0.486 | 0.539 | 0.526 | 0.479 | 0.461 |
| | C-statistic (95% CI) | 0.65 (0.60 – 0.69) | 0.65 (0.61 – 0.70) | 0.65 (0.60 – 0.70) | 0.65 (0.60 – 0.69) | 0.65 (0.60 – 0.69) |
| Simplified Finnish | E/O (95% CI) | 0.35 (0.17 – 0.52) | 0.34 (0.16 – 0.52) | 0.35 (0.17 – 0.53) | 0.35 (0.17 – 0.52) | 0.34 (0.16 – 0.52) |
| Diabetes Risk model | Brier score | 0.133 | 0.134 | 0.133 | 0.133 | 0.134 |
| | Yates slope | -0.032 | -0.068 | -0.048 | -0.026 | -0.050 |
| | C-statistic (95% CI) | 0.66 (0.62 – 0.71) | 0.66 (0.62 – 0.70) | 0.66 (0.62 – 0.70) | 0.66 (0.62 – 0.71) | 0.66 (0.62 – 0.70) |

**Imputation**

Table 3 shows the variable characteristics of the original database compared to the five methods of dealing with missing data. Pair-wise deletion resulted in a significantly reduced sample size (754) while conditional mean imputation resulted in a varied length of each variables as only missing values with complete cases for every other variables were imputed. Simple imputation, stochastic regression imputation and multiple imputation all imputed all missing data allowing for a full database of 1083 individuals.

Imputation of the outcome, undiagnosed diabetes, was highest in stochastic regression imputation (163 individuals). Pair-wise deletion saw a higher mean age (52.5 years) and lower systolic and diastolic blood pressure (122.0 mmHg and 74.7 mmHg respectively) when compared to the other imputation methods. There was no substantial difference in body mass index and waist circumference between the methods. Stochastic regression imputed a higher prevalence of individuals on hypertensive medication (387 individuals), mother having diabetes (182 individuals), and sister having diabetes (143 individuals). Multiple imputation reported the highest prevalence of father (78 individuals) and brother having diabetes (87 individuals). Variable characteristics across the five imputation datasets is shown in Table 4. Mother, father and sister family history, as well as smoking status had the most variation between the five multiple imputation datasets.

**Model performance**

Most notably, model performance following pair-wise deletion deviated from the model performance from other imputation methods. The discrimination was lower in all five models, however calibration was better in the Cambridge Diabetes Risk model [1.81 (1.09 - 2.52)]. Overall, although not large differences, simple imputation yielded the highest C-statistic for four of the five models; the Cambridge Diabetes Risk model [0.69 (0.65 – 0.73), vs. 0.67 (0.62 – 0.72)], Kuwaiti Risk model [0.70 (0.66 – 0.74), vs. 0.68 (0.63 – 0.73)], Omani Diabetes Risk model [0.67 (0.63 – 0.71) vs. 0.65 (0.61 – 0.69)] and Rotterdam Predictive model [0.65 (0.61 – 0.70) vs. 0.64 (0.59 – 0.69)]. Multiple imputation only yielded the highest C-statistic for the Rotterdam Predictive model [0.65 (0.61 – 0.70), which were matched by simpler imputation

methods. Table 5 details the indifference in model performance across the five datasets produced through multiple imputation.

The pattern of the overall calibration (E/O) did not vary substantially across imputation methods. Uniformly, all imputation methods resulted in the Cambridge and Omani risk models overestimating diabetes risk, while the others showed underestimation. Other performance measures across subgroups, shown in Table 1, did also not show significant differences between imputation methods. When recalibration was performed, all models across all imputation techniques had an improved agreement between predicted and observed rates (Table 1).

## Discussion

The suggested imputation method for the handling of missing data is a hot topic, with strong advocators for multiple imputation and those who propose that simple techniques can be just as effective. Several studies have been done to determine the effect of several imputation methods on the predictive performance of risk models, however these have been largely contradicting. Donders *et al* [1] performed a simulation study in an attempt to illustrate that single imputation yields unbiased estimates with too narrow confidence intervals and multiple imputation indeed yields unbiased estimates with correct standard errors. Both single and multiple imputation produced unbiased estimates of association, and the conclusion was that despite single imputation appearing more precise, multiple imputation produces less bias and more precise results. Alternately, a study by van der Heijden *et al* [15] concluded that the models fitted using the indicator method, a simple method of dealing with missing data, showed higher regression coefficients and predictive accuracy when compared to the models derived from the imputation methods. As confirmed in this study, we did not observe large differences between the models obtained after single unconditional, single conditional and multiple imputation of the missing data. Deletion of individuals with missing data resulting in an expected reduced discriminatory ability of the models. Model calibration was improved across all areas when recalibration was performed. This however, has no influence on the imputation techniques or the discriminatory ability of the models.

What should be noted is that most studies comparing imputation techniques start with a complete data set and introduce missing data to set variables. Our study made use of an existing database which already included missing data on a number of variables. This results in the true underlying value of the missing data being unknown, as well as the true regression coefficients and predictive accuracy of each variable. This lack of reference criterion can be limitation in a study. However, the use of existing databases should be encouraged as this is more translatable to medical research outside of a controlled setting.

Despite recent advances in understanding missing data and imputation methods, most researchers still report deletion, perhaps because of a lack of adequate guidelines for handling missing data. What should be encouraged is the use of more than one method, the results compared and a preferred approach chosen and defended. When data are missing on several variables it is important to use some procedure that imputes them all together, rather than one variable at a time. This ensures that the imputed data are related to each other in the same way as those data that are observed.

## Conclusion

This study aimed to compare the performance results of undiagnosed diabetes risk prediction models across multiple imputation techniques. The results showed a lower model performance when deletion is used to deal with missing data and little difference between simple and more complex methods on the effect of risk prediction model performance. Missing data is an important aspect of predictive research and needs to be handled correctly. Imputation, specifically more complex and time-intensive imputation, can often be avoided by researchers due to preconceived complexity. Simpler imputation methods that allow for similar or better predictive performance are easy to undertake and should encourage researchers of all levels to limit the use of deletion of individuals with missing data. The negligible difference in model performance between simple and multiple imputation allows for the recommendation of single imputation for handling missing data in undiagnosed diabetes predicative research.

## Acknowledgements

KLM was supported by a scholarship from the South African National Research Foundation and the Carl & Emily Fuchs Foundation.

## References

1. Donders, A.R.T., van der Heijden, G.J., Stijnen, T., and Moons, K.G. *Review: A gentle introduction to imputation of missing values.* Journal of clinical epidemiology. 2006; **59**(10): p. 1087-1091.
2. Schafer, J.L. and Graham, J.W. *Missing data: Our view of the state of the art.* Psychological methods. 2002; **7**(2): p. 147.
3. Rubin, D.B. *Inference and missing data.* Biometrika. 1976; **63**(3): p. 581-592.
4. Greenland, S. and Finkle, W.D. *A critical look at methods for handling missing covariates in epidemiologic regression analyses.* Am J Epidemiol. 1995; **142**(12): p. 1255-1264.
5. Little, R.J. *Regression with missing x's: A review.* Journal of the American Statistical Association. 1992; **87**(420): p. 1227-1237.
6. Harrell, F.E. *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. 2001, New York: Springer.
7. Rubin, D.B. *Multiple imputation for nonresponse in surveys*. 1987, New York: J. Wiley & Sons.
8. Little, R. and Rubin, D.B. *Statistical analysis with missing data.* John A. Wiley & Sons, Inc., New York. 1987.
9. Rubin, D.B. and Schenker, N. *Multiple imputation in health-are databases: An overview and some applications.* Statistics in medicine. 1991; **10**(4): p. 585-598.
10. Schafer, J.L. *Multiple imputation: A primer.* Statistical methods in medical research. 1999; **8**(1): p. 3-15.
11. VACH, W. *Some issues in estimating the effect of prognostic factors from incomplete covariate data.* Statistics in medicine. 1997; **16**(1): p. 57-72.
12. Schafer, J.L. *Analysis of incomplete multivariate data*. 2010, Florida: CRC press.
13. Crawford, S.L., Tennstedt, S.L., and McKinlay, J.B. *A comparison of analytic methods for non-random missingness of outcome data.* Journal of clinical epidemiology. 1995; **48**(2): p. 209-219.
14. Baker, J., White, N., and Mengersen, K. *Missing in space: An evaluation of imputation methods for missing data in spatial analysis of risk factors for type ii diabetes.* International journal of health geographics. 2014; **13**(1): p. 47.
15. van der Heijden, G.J., Donders, A.R.T., Stijnen, T., and Moons, K.G. *Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example.* Journal of clinical epidemiology. 2006; **59**(10): p. 1102-1109.
16. van Buuren, S., Boshuizen, H.C., and Knook, D.L. *Multiple imputation of missing blood pressure covariates in survival analysis.* Stat Med. 1999; **18**(6): p. 681.
17. Quintana-Murci, L., Harmant, C., Quach, H., Balanovsky, O., Zaporozhchenko, V., Bormans, C.*, et al*. *Strong maternal khoisan contribution to the south african coloured population: A case of gender-biased admixture.* The American Journal of Human Genetics. 2010; **86**(4): p. 611-620.
18. CoCT. *2011 census – ward 009*. 2013, City of Cape Town.
19. CoCT. *Census 2001 – ward 9*. 2006, City of Cape Town: Cape Town.

20. World Health Organization. *Screening for type 2 diabetes: Report of a world health organization and international diabetes federation meeting.* 2003.
21. Brown, N., Critchley, J., Bogowicz, P., Mayige, M., and Unwin, N. *Risk scores based on self-reported or available clinical data to detect undiagnosed type 2 diabetes: A systematic review.* Diabetes research and clinical practice. 2012; **98**(3): p. 369-385.
22. Griffin, S., Little, P., Hales, C., Kinmonth, A., and Wareham, N. *Diabetes risk score: Towards earlier detection of type 2 diabetes in general practice.* Diabetes/metabolism research and reviews. 2000; **16**(3): p. 164-171.
23. Al Khalaf, M., Eid, M., Najjar, H., Alhajry, K., Doi, S., and Thalib, L. *Screening for diabetes in kuwait and evaluation of risk scores.* EMHJ. 2010; **16**(7): p. 725-731.
24. Al-Lawati, J. and Tuomilehto, J. *Diabetes risk score in oman: A tool to identify prevalent type 2 diabetes among arabs of the middle east.* Diabetes research and clinical practice. 2007; **77**(3): p. 438-444.
25. Baan, C.A., Ruige, J.B., Stolk, R.P., Witteman, J., Dekker, J.M., Heine, R.J.*, et al*. *Performance of a predictive model to identify undiagnosed diabetes in a health care setting.* Diabetes Care. 1999; **22**(2): p. 213-219.
26. Bergmann, A., Li, J., Wang, L., Schulze, J., Bornstein, S., and Schwarz, P. *A simplified finnish diabetes risk score to predict type 2 diabetes risk and disease evolution in a german population.* Hormone and metabolic research. 2007; **39**(09): p. 677-682.
27. Masconi, K., Matsha, T.E., Erasmus, R.T., and Kengne, A.P. *Independent external validation and comparison of prevalent diabetes risk prediction models in a mixed-ancestry population of south africa.* Diabetology & Metabolic Syndrome. 2015.
28. Team, R.C. *A language and environment for statistical computing. Vienna, austria: R foundation for statistical computing.* URL http://www. R-project. org. 2014.
29. Masconi, K.L., Matsha, T.E., Echouffo-Tcheugui, J.B., Erasmus, R.T., and Kengne, A.P. *Reporting and handling of missing data in predictive research for prevalent undiagnosed type 2 diabetes mellitus: A systematic review.* EPMA Journal. 2015; **6**(1): p. 7.
30. Steyerberg, E.W. *Clinical prediction models*. 2009, New York: Springer.
31. Graham, J.W., Hofer, S.M., and MacKinnon, D.P. *Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures.* Multivariate Behavioral Research. 1996; **31**(2): p. 197-218.
32. Brick, J.M. and Kalton, G. *Handling missing data in survey research.* Stat Methods Med Res. 1996; **5**(3): p. 215-238.
33. Alzola, C.F. and Harrell, F.E. *An introduction to s and the hmisc and design libraries.* URL http://biostat. mc. vanderbilt. edu/twiki/pub/Main/RS/sintro. pdf. 2002; **30**.
34. Templ, M., Alfons, A., Kowarik, A., and Prantner, B. *Vim: Visualization and imputation of missing values.* R package version. 2011; **2**(3).
35. Buuren, S. and Groothuis-Oudshoorn, K. *Mice: Multivariate imputation by chained equations in r.* Journal of statistical software. 2011; **45**(3).
36. Kengne, A.P., Masconi, K., Mbanya, V.N., Lekoubou, A., Echouffo-Tcheugui, J.B., and Matsha, T.E. *Risk predictive modelling for diabetes and cardiovascular disease.* Critical reviews in clinical laboratory sciences. 2013; **51**(1): p. 1-12.
37. Moons, K.G., Kengne, A.P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D.G.*, et al*. *Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio) marker.* Heart. 2012; **98**(9): p. 683-690.
38. Dobson, A.J., Kuulasmaa, K., Eberle, E., and Scherer, J. *Confidence intervals for weighted sums of poisson parameters.* Statistics in medicine. 1991; **10**(3): p. 457-462.

# PART IV

# MODEL UPDATING

08

# Recalibration in validation studies of diabetes risk prediction models: A systematic review

Katya Masconi, Tandi E. Matsha, Rajiv T. Erasmus and Andre P. Kengne

*International Journal of Statistics in Medical Research* 2015; **4(4)**:347-369

## Abstract

**Background:** Poor performance of risk prediction models in a new setting is common. Recalibration methods aim to improve the prediction performance of a model in a validation population, however the extent of its application in the validation of diabetes risk prediction models is not yet known.

**Methods:** We critically reviewed published validation studies of diabetes prediction models, selected from five recent comprehensive systematic reviews and database searches. Common recalibration techniques applied were described and the extent to which recalibration and impacts were reported analysed.

**Results:** Of the 236 validations identified, 22.9% (n = 54) undertook recalibration on existent models in the validation population. The publication of these studies was consistent from 2008. Only incident diabetes risk prediction models were validated, and the most commonly validated Framingham offspring simple clinical risk model was the most recalibrated of the models, in 4 studies (7.4%).

**Conclusions:** This review highlights the lack of attempt by validation studies to improve the performance of the existent models in new settings. Model validation is a fruitless exercise if the model is not recalibrated or updated to allow for greater accuracy. This halts the possible implementation of an existent model into routine clinical care. The use of recalibration procedures should be encouraged in all validation studies, to correct for the anticipated drop in model performance.

**Keywords:** risk prediction, diabetes, update, recalibration, validation

## Background

The use of risk prediction models in a validation population is expected to have an effect on the performance of the model (usually a drop in the performance) due to the differences between development and validation populations, particularly the variances in outcome frequency between the populations, case-mix and measurements used for the variables and outcome determination [1]. In an effort to improve the performance of a model in a new setting, updating strategies have been proposed [2, 3]. The updating strategies range from simple adjustment of models' parameters to more complex model alterations. Simple updating methods, termed re-calibration, describes the re-estimation of the model intercept (or baseline risk parameter) with or without re-estimation of the regression coefficients.

The recalibration of risk prediction models is encouraged, where the resulting updated model combines the prediction information that was captured in model development with the information of the new population. This lends to the concept that risk prediction models should be based on as many individuals' data as possible. Too often, existent models are externally validated and when performance is disappointing, a new prediction model is developed. This results in a large number of models available, which are all poorly externally validated [4]. For illustration, a systematic review by Noble and co-workers [5] found that between 1993 and 2011, over 145 models were developed to predict prevalent or incidents diabetes, of which only a few were externally validated. This is of concern, considering the use of accurate and validated risk models is increasingly advocated as a basis for risk screening in strategies to prevent the occurrence of diabetes among those at high risk, to promote early detection among those with prevalent undiagnosed diabetes, and tailoring the complexity and intensity of the management among those with diagnosed diabetes, to the risk of subsequent complications. Indeed, with diabetes mellitus growing to the epidemic proportions around the world, and considering the complexity of the interaction of factors contributing to diabetes occurrence and related complications, the ability of risk prediction models to incorporate a multitude of risk factors, accounting for this complexity, cements their importance in diabetes prevention and control

strategies. Beyond the field of diabetes and non-communicable diseases in general, with the opening era of personalised healthcare, prediction models will be increasingly used to assist clinical decision making. Efforts to limit the number of prediction models through careful updating of existing models to work in various settings, have a potential to improve their uptake in routine practice.

A recent validation study applied simple updating methods to diabetes risk prediction models, and reported some improvement, although non-optimal, of models performance [6]. However, the extent of the application of recalibration strategies in the validation of diabetes risk prediction models is not yet known. In this paper, we critically review the level of reporting, method of choice and extent of use of recalibration methods in validation studies, through a systematic review of studies on the validation of incident and prevalent diabetes risk prediction models, in an attempt to make conclusions on the extent of recalibration in diabetes risk prediction research.

## Methodology

Building on the five most comprehensive review articles on both incident and prevalent diabetes risk prediction models by Buijsse *et al* (2011) [7], Collins *et al* (2011) [8], Noble *et al* (2011) [5], Thoopputra *et al* (2012) [9], and Brown *et al* (2012) [10], additional relevant articles were identified through a systematic literature review according to the PRISMA guidelines, where necessary [11]. We searched PubMed for all published studies aimed at validating diabetes risk prediction models using the following string search: ((“diabetes” OR “diabetes mellitus” OR “type 2 diabetes”) AND (“risk score” OR “prediction model” OR “predictive model” OR “predicting” OR “prediction rule” OR “risk assessment” OR “algorithm”) AND (“validation” OR ‘validate”)).

Studies were included if they validated risk scores, models or questionnaires and the outcome was prevalent undiagnosed or incident diabetes in adults (aged >18 years). Studies undertaking internal validation were excluded as model recalibration should not be required at this early stage. Additionally, studies aimed at validating guidelines in new populations were excluded. Models that were developed outside of the logistic, cox or Weibull development methods were excluded due to the inability to validate these models (e.g. classification tree analysis method). There was no restriction on the variables included in the models, both non-invasive and invasive models were included. Additionally, there was no restriction on sample size or country. The data extracted included country/setting, name of the models validated, whether the study aimed at validation alone or with development of a model and the presence of a discussion and action (or lack thereof) on the recalibration of models. We reviewed the included studies with the aim of providing the reader with a comprehensive list of validated models, instances and prevalence of model recalibration, as well as the possible increase in performance of the updated model.

## Results

### Overview of included studies

Following the sifting process, a total of 94 articles were included (Figure **1**). These articles included 70 models, and 236 validations were conducted. Figure **2** depicts the distributions of risk prediction model validation. Included published studies undertook the validation of existent diabetes risk prediction model/s, where validation refers to the process of evaluating the performance of a model. Studies were focussed on external validation which goes beyond the assessment of model performance in all or a portion of the developmental datasets by assessing the performance in an independent dataset. The validation of a model can be grouped by a hierarchy proposed by Justice *et al* (1999) [12], according to the reproducibility and historic, geographic, methodologic, spectrum and follow-up period transportability (Text Box **1**). Additionally, one paper can report on the validation of more than one model. Many studies undertook the validation of a model(s) as an added section to the development of a model in

their population group (48.8%). Details of the included studies are provided in Table **1**; published between 1997 and 2014, however most appeared in 2005-2011. Articles reporting recalibration of existent models only appeared from 2008 onwards, with the most appearing in 2010. The number and combination of predictors was variable, with age, sex, body mass index and waist circumference being the most commonly used variables. The study setting was highly heterogeneous; models were validated in 31 countries across 5 continents (only 1 in Africa). Models predicting incident diabetes were more commonly validated (62.7%) when compared to prevalent diabetes risk prediction. The development, recalibration and use of incident and prevalent risk prediction models vary and will therefore be discussed separately.

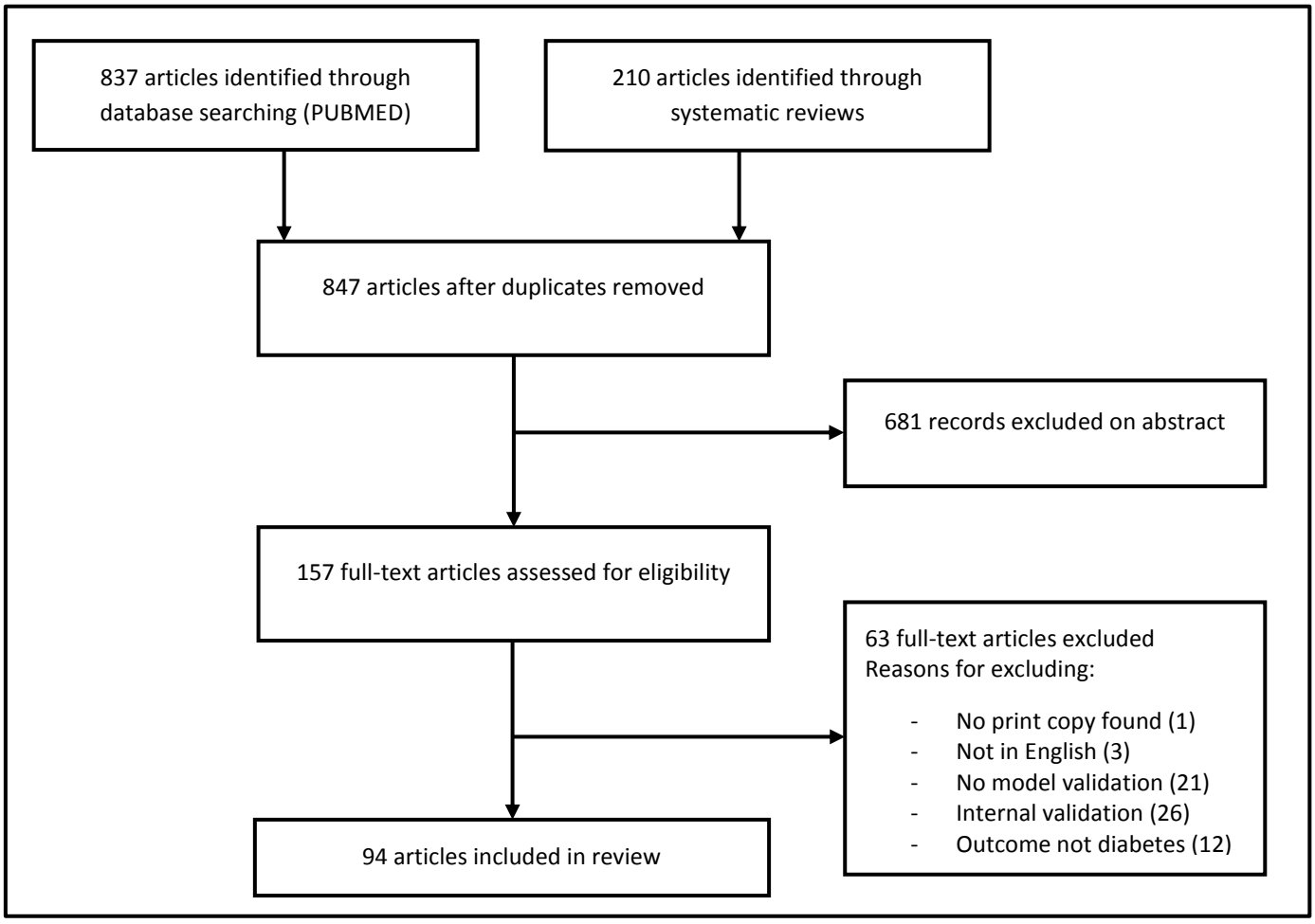



**Figure 1**: Flow diagram of selected studies

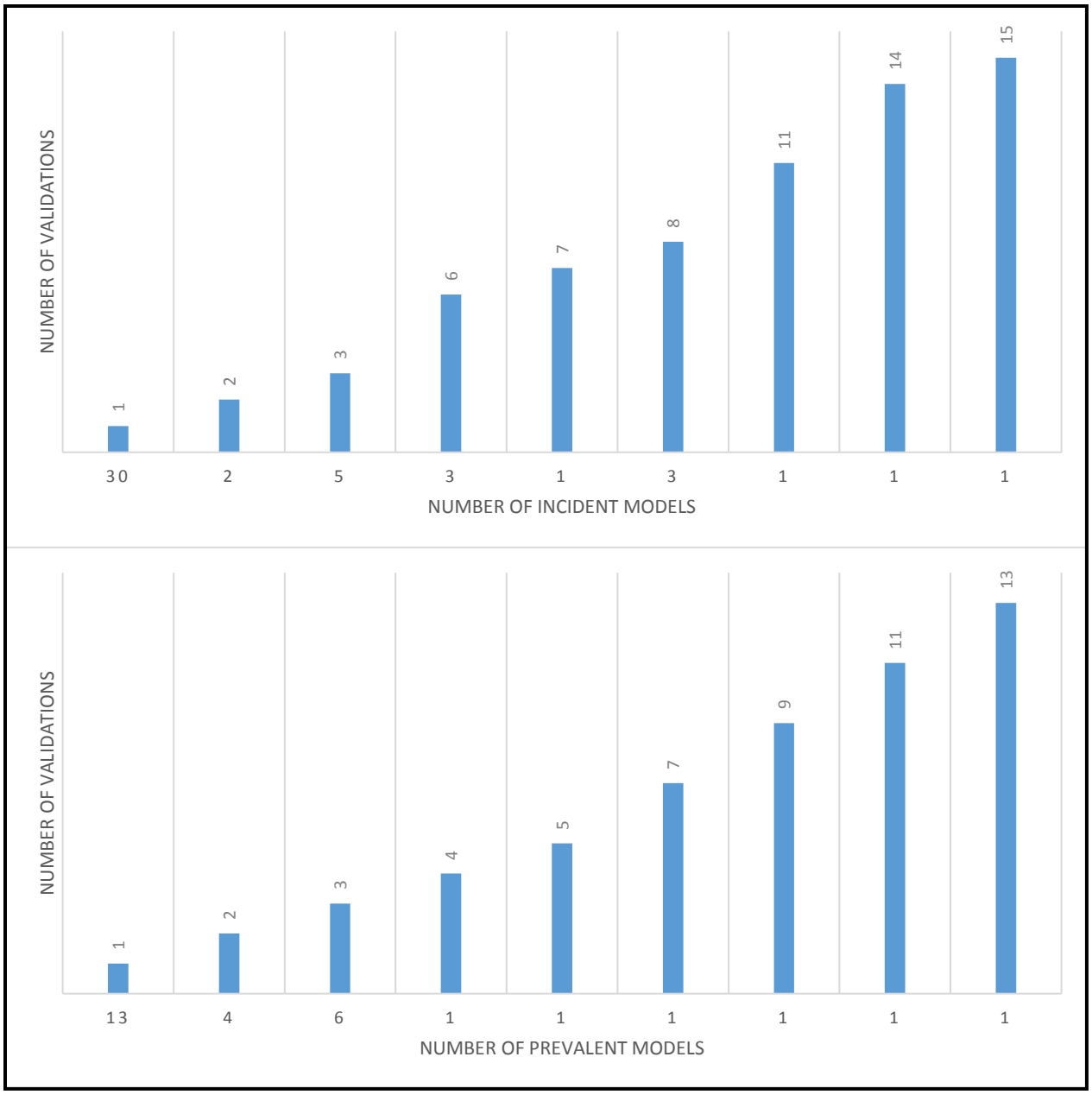



**Figure 2**: Illustration of incident (top) and prevalent (bottom) model validation distribution. Top 3 validated incident diabetes models: 15 times – Framingham offspring simple clinical diabetes model; 14 times – San Antonio clinical risk model; 11 times – Cambridge diabetes risk score. Top 3 validated prevalent diabetes models: 13 times – Full prevalent FINDRISC risk model; 11 times – Rotterdam risk predicative model 1; 9 times – Cambridge diabetes risk model.

Text box 1: A hierarchy of external validation of predictive systems – adapted from Justice *et al* [12]

| Level of validation | Cumulative generalizability evaluated |
|---|---|
| **0: Internal validation** | Reproducibility |
| **1: Prospective validation** | Level 0 + historic transportability |
| **2: Independent validation** | Level 1 + geographic transportability, methodologic transportability, spectrum transportability |
| **3: Multisite validation** | Level 2 at multiple sites |
| **4: Multiple independent validation** | Level 3 by multiple investigators |
| **5: Multiple independent validations with life-table analyses** | Level 4 + follow-up period transportability |

**Incident diabetes risk prediction models**

The most commonly validated model was the Framingham offspring simple clinical risk model (10.1%) [13] followed by the San Antonio clinical risk model (9.5%) [14]. Validations were ranked according to the levels of transportability. There was no evidence of level 4 or 5 diabetes risk prediction validation. The most common form of validation (level 2) tested the models' geographic, methodologic and spectrum transportability in addition to the reproducibility and historic transportability (62.8%). This included models which were validated in the same country as their development but a different city or cohort to development, as well as validation of a model for a different outcome.

**Prevalent diabetes risk prediction models**

The Finnish diabetes full risk model was the most frequently validated prevalent diabetes risk prediction model (14.8%) [15, 16], followed by the Rotterdam predictive model 1 (12.5%). As with incident risk models, hierarchy level 2 was the most common level of validation (81.8%), with no level 4 or 5 validation.

**Recalibration methods**

Multiple updating methods exist [2, 3, 17-19], varying in the complexity and the number of parameters that are adjusted or re-estimated. The term 'recalibration' is used to describe basic techniques to fit a predictive model to a new setting. The development of the model dictates the recalibration methods available. The mathematical model chosen for development may follow logistic regression, cox or Weibull principles. The intercept, or equivalent, of risk models is determined by the prevalence of the outcome in the population in which the model was developed and the updating of this intercept aims to solve the discrepancy between the mean predicted risk and mean observed risk resulting in better calibration. To be noted, recalibration, through either method, does not change the discriminatory ability of the risk prediction model as the relative ranking of the predicted probabilities remain the same [20].

Logistic regressions are the most commonly used for risk prediction research. Recalibration methods, described by Steyerberg [1] and Janssen *et al* [3], aim to update the intercept of logistic models to better account for the prevalence in the validation population. The intercept can be updated by fitting a logistic regression model with a linear predictor as the only covariate in the updating set or by calculating a correction factor that is based on the mean predicted risk and observed outcome frequency in the validation population. When the outcome frequency is not particularly low or high, the correction factor will equal the calibration intercept. The final correction factor is simply added to the intercept of the original model. This is considered the most basic form of logistic model updating. An additional method, termed logistic calibration, fits a logistic regression model with a linear predictor as the only covariate in the updating set [3]. The calibration slope is used to recalibrate (multiply by) the original regression coefficients. The closer the calibration slope is to 1, the less adjustment the original regression coefficients required. The intercept is also updated by adding the calibration intercept to the intercept of the original prediction model.

Survival models available for risk prediction research depend on the distribution assumptions that can be made. Weibull models are generalised exponential models with the inclusion of shape (survival rate), allowing for more flexibility on the types of data that the model can fit. The model has a hazard function which measures how likely the outcome/event will take place as a function of the length of observation [21]. While exponential distribution has a constant hazard function, the Weibull distribution hazard rate can increase or decrease in relation to time. The Weibull model is a popular method for parametric data. When distribution assumptions of the survival time (time until diagnosis) cannot be met, the Cox proportional hazard model can be used. Additionally, cox models are used when the risk factors have a multiplicative effect on the hazard function and can be extended for multiple regression situations [22]. Cox models do not have an intercept but rather an equivalent, the 'baseline survival function' or 'baseline risk'. This baseline information is almost never given by authors of published medical articles that report a cox model, however it can be recalculated [23]. Cox models are often referred to as semi-parametric, as the baseline hazard function is non-parametric, while the linear predictor in the cox model is fully parametric.

The incorporation of diagnosis time in both of these models allows for them only to be used for the development of incident diabetes risk prediction models (as opposed to prevalent diabetes prediction). The choice of model is researcher dependent and each come with their own advantages, parametric models are more precise with smaller standard errors, while it is easier, and can prevent biases, not having to make assumptions of the underlying hazard function nature or shape with semi-parametric models. The recalibration of all survival analysis models uses Kaplan-Meier to determine the average incidence rates and update the model to the validation population incidence rate [17]. Additionally available, the mean values of each variable within the model which were derived from the validation population is replaced by the mean values of the same variables from the validation population. These methods are described in more detail by D' Agostino (2001) [17]. Text box **2** details the components of the various models that are altered during recalibration.

Text box 2: Mathematical formula for key models illustrating change before and after recalibration – adapted from Janssen *et al* [3] and Houwelingen [23]

| Model | Formula | Components | Recalibration change |
|---|---|---|---|
| **Logistic** | $1/\{1 + EXP[-(ß_0 + ß_1 x\ predictor_1 + \ldots + ßn\ x\ predictor_n)]\}$ | Intercept: $ß_0$<br>Variable coefficient: $ß_1$- ßn | Update intercept: $ß_0$ + correction factor |
| | | | Update intercept: $ß_0$ + correction factor<br>Coefficient: linear predictor x $\beta_{calibration}$ |
| **Cox** | $H_o(t)EXP(x\beta)$ where $x\beta = \beta_1(x_1 - M_{1+\ldots+}\beta_n(x_n - M_n)$ | Baseline hazard function: $H_0(t)$<br>Prognostic index: xβ<br>Regression coefficient: β<br>Mean of risk factor: M | Update incidence rate of validation cohort: $H_0(t)$ |
| | | | Update mean value of variable in validation cohort: β |
| **Weibull** | $(\beta_0 + \beta_1 \ln(t))EXP(x\beta)$ where $x\beta = \beta_1(x_1 - M_{1+\ldots+}\beta_n(x_n - M_n)$ | Hazard function: $\beta_0 + \beta_1 \ln(t)$<br>Prognostic index: xβ<br>Regression coefficient: β<br>Mean of risk factor: M | Update incidence rate of validation cohort: $\beta_0$ of $(\beta_0 + \beta_1 \ln(t))$ |
| | | | Update mean value of variable in validation cohort: β |

**Reporting of recalibration**

Of the 236 validations of diabetes risk prediction models in alternate populations, 54 (22.9%) reported the use of recalibration methods in an effort to increase performance of the existent models. The reporting of the recalibration method was clear, the only article to not report the method of recalibration was Bozorgmanesh *et al* (2011) [24]. Forty two of these studies (77.8%) reported an increase in model performance following the recalibration of the original model (seven studies did not report the original or recalibrated model performance [25-28]). Every recalibration was carried out on an incident diabetes risk prediction model, with most of them being logistic regression models (75.9%). Additionally, 68.5% of recalibrations were carried out in level 3 calibrations. There was no one model that was recalibrated significantly more often than others. The Framingham offspring simple clinical model was recalibrated four times (7.4%) [14, 24, 28, 29], while the DPoRT, concise Finnish, German, KORA base, KORA clinical, QDScore and San Antonio clinical diabetes risk models were recalibrated three times (5.6%).

Table 1 – Characteristics of validation/updating studies of diabetes prediction models

| Author | Year | Location of study | Model/s | Incident or prevalent model | Validation with development | Level of validation | Recalibration | Increase in calibration | Alteration to model (if any) |
|---|---|---|---|---|---|---|---|---|---|
| Abbasi *et al* [30] | 2012 | Netherlands | KORA base model (model 1) – logistic [31] | Incident | No | 2 | Yes | No | Addition of WC following recalibration |
| | | | KORA clinical model (model 2) – logistic [31] | | | | | Yes | Exclusion of HbA1c and uric acid; Addition of WC following recalibration |
| | | | KORA clinical model (model 3) – logistic [31] | | | | | Yes | Exclusion of HbA1c and OGTT; Addition of WC following recalibration |
| Abbasi *et al* [32] | 2012 | Netherlands | DETECT-2 model – logistic [33] | Incident | No | 3 | Yes | Yes | / |
| | | | BRHS simple clinical model – logistic [34] | | | | | | |
| | | | BRHS fasting biomarker model – logistic [34] | | | | | | |
| | | | BRHS non-fasting biomarker model – logistic [34] | | | | | | |
| | | | KORA base model (model 1) – logistic [31] | | | | | | |
| | | | KORA clinical model (model 2) – logistic [31] | | | | | | |
| | | | AUSDRISK – logistic [35] | | | | | | Self-reported prevalent cases of diabetes excluded - history of high blood glucose variable unavailable, therefore set to zero |
| | | | DPoRT – weibull [27] | | | | | | / |
| | | | Tromso – cox [36] | | | | | | |
| | | | ARIC basic model – weibull [37] | | | | | | |
| | | | ARIC enhanced model – weibull [37] | | | | | | |
| | | | QDScore – cox [38] | | | | | | Z score combination of education levels and occupation status as proxy for social economic status |
| | | | DESIR clinical risk model – logistic [39] | | | | | | / |
| | | | DESIR clinical and biological risk model – logistic [39] | | | | | | |
| | | | Framingham offspring simple clinical categorical model 1 – logistic [13] | | | | | | |
| | | | Framingham offspring simple clinical categorical model 2 – logistic [13] | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Framingham offspring simple clinical categorical model 3 – logistic [13] | | | | | | |
| | | | EPIC-Norfolk – logistic [40] | | | | | | |
| | | | German diabetes risk score – cox [41] | | | | | | |
| | | | Finnish diabetes risk score concise – logistic [15] | | | | | | Self-reported prevalent cases of diabetes excluded - history of high blood glucose variable unavailable, therefore set to zero |
| | | | Finnish diabetes risk score full – logistic [15] | | | | | | Self-reported prevalent cases of diabetes excluded - history of high blood glucose variable unavailable, therefore set to zero |
| | | | San Antonio risk clinical model – logistic [42] | | | | | | |
| | | | PROCAM risk model – logistic [43] | | | | | | / |
| | | | San Antonio reduced model – logistic [44] | | | | | | |
| Adbul-Ghani *et al* [45] | 2009 | Finland | San Antonio risk clinical model – logistic [42] | Incident | No | 2 | No | / | Addition of 1-hour plasma glucose |
| | | | California scoring model – logistic [46] | | | | | | |
| | | | Finnish diabetes risk score– logistic [15] (concise/full model not stated) | | | | | | |
| Adhikari *et al* [47] | 2010 | India | Indian diabetes risk score – logistic [48] | Prevalent | No | 2 | No | / | Incident to prevalent |
| Akyil *et al* [49] | 2014 | Turkey | Finnish diabetes risk score full – logistic [15, 16] | Prevalent | No | 2 | No | / | / |
| Al Khalaf *et al* [50] | 2010 | Kuwait | American Diabetes Association risk assessment questionnaire [51] | Prevalent | Yes | 2 | No | / | / |
| | | | Rotterdam predictive model 1 – logistic [52] | | | | | | |
| | | | Cambridge diabetes risk score – logistic [53] | | | | | | |
| | | | Finnish diabetes risk score full – logistic [15, 16] | | | | | | |
| | | | Danish risk score – logistic [54] | | | | | | |
| | | | Indian diabetes risk score – logistic [48] | | | | | | |
| | | | Thai simple risk model – logistic [55] | | | | | | |
| | | | Omani risk score – logistic [56] | | | | | | |
| Al-Lawati *et al* [56] | 2007 | Oman | Rotterdam predictive model 1 – logistic [52] | Prevalent | Yes | 2 | No | / | / |
| | | | Thai simple risk model – logistic [55] | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Finnish diabetes risk score full – logistic [15, 16]<br>Danish risk score – logistic [54] | | | | | | |
| Alssema *et al* [33] | 2011 | Netherlands, Denmark, Sweden, UK, Australia, Mauritius | Finnish diabetes risk score concise – logistic [15] | Incident | No | 3 | Yes | Yes | History of high blood glucose swopped for gestational diabetes |
| Alssema *et al* [57] | 2012 | Netherlands | Finnish diabetes risk score concise – logistic [15] | Incident | Yes | 2 | No | / | / |
| Baan *et al* [52] | 1999 | Netherlands | Rotterdam predictive model 1 – logistic [52]<br>Rotterdam predictive model 2 – logistic [52] | Prevalent | Yes | 1 | No | / | / |
| Balkau *et al* [39] | 2008 | France | San Antonio risk clinical model – logistic [42]<br>Finnish diabetes risk score full – logistic [15]<br>DESIR clinical risk model – logistic [39]<br>DESIR clinical and biological risk model – logistic [39]<br>DESIR clinical, biological and genetic risk model – logistic [39] | Incident | Yes | 2<br>1 | No | / | / |
| Bang *et al* [58] | 2009 | USA | Rotterdam predictive model 1 – logistic [52]<br>American Diabetes Association risk assessment questionnaire [51] | Prevalent | Yes | 2 | No | / | / |
| Bergmann *et al* [59] | 2007 | Germany | Finnish diabetes risk score concise – logistic [15]<br>Finnish diabetes risk score concise – logistic [15, 16] | Incident<br>Prevalent | No | 2 | No | / | / |
| Bhadoria *et al* [60] | 2014 | India | Indian diabetes risk score – logistic [48] | Prevalent | No | 2 | No | / | / |
| Bozorgmanesh *et al* [61] | 2010 | Iran | ARIC enhanced model – Weibull [37] | Incident | No | 2 | No | / | / |
| Bozorgmanesh *et al* [62] | 2010 | Iran | San Antonio risk clinical model – logistic [42]<br>San Antonio reduced model – logistic [44] | Incident | No | 2 | Yes<br>No | No<br>/ | Addition of OGTT |
| Bozorgmanesh *et al* [24] | 2011 | Iran | Framingham offspring simple clinical model – logistic [13] | Incident | Yes | 2 | Yes | Yes | / |
| Chaturvedi *et al* [63] | 2008 | India | Urban Asian Indian risk score – logistic [63] | Prevalent | Yes | 1 | No | / | / |
| Cameron *et al* [64] | 2007 | Mauritius | San Antonio risk clinical model – logistic [42] | Incident | No | 2 | No | / | / |
| Cameron *et al* [65] | 2008 | Australia | Finnish diabetes risk score full – logistic [15]<br>San Antonio risk clinical model – logistic [42] | Incident | No | 2 | No | / | History of high blood glucose excluded<br>Family history only included parental history |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Chen *et al* [35] | 2010 | Australia | AUSDRISK – logistic [35] | Incident | Yes | 1 | No | / | / |
| Chien *et al* [66] | 2009 | Taiwan | Framingham offspring simple clinical model – logistic [13] | Incident | Yes | 2 | No | / | / |
| | | | San Antonio risk clinical model – logistic [42] | | | | | | |
| | | | Cambridge diabetes risk score – logistic [53] | | | | | | |
| | | | PROCAM risk model – logistic [43] | | | | | | |
| Collins *et al* [8] | 2011 | United Kingdom | QD Score – cox [38] | Incident | No | 2 | No | / | Continuous Townsend score replaced by categorical proxy |
| Farran *et al* [67] | 2013 | Kuwait | US screening score – logistic [58] | Incident | Yes | 2 | No | / | / |
| Franciosi *et al* [68] | 2005 | Italy | Finnish diabetes risk score full – logistic [15, 16] | Prevalent | No | 2 | No | / | / |
| Gao *et al* [69] | 2010 | China | Qingdao diabetes risk score – logistic [69] | Prevalent | Yes | 1 | No | / | / |
| | | | Rotterdam predictive model 1 – logistic [52] | | | 2 | | | |
| | | | Cambridge diabetes risk score – logistic [53] | | | | | | |
| | | | Finnish diabetes risk score full – logistic [15, 16] | | | | | | |
| | | | Danish risk score – logistic [54] | | | | | | |
| | | | Asian Indian diabetes risk score – logistic [70] | | | | | | |
| | | | Thai simple risk model – logistic [55] | | | | | | |
| | | | DESIR clinical risk model – logistic [39] | | | | | | |
| Ginde *et al* [71] | 2007 | USA | American Diabetes Association risk assessment questionnaire [51] | Prevalent | No | 2 | No | / | / |
| Glümer *et al* [54] | 2004 | Denmark | Danish risk score – logistic [54] | Prevalent | Yes | 1 | No | / | / |
| Glümer *et al* [72] | 2005 | Australia / Denmark | Danish risk score – logistic [54] | Prevalent | No | 2 | No | / | Physical activity excluded |
| Glümer *et al* [73] | 2006 | Global | Rotterdam predictive model 1 – logistic [52] | Prevalent | No | 3 | No | / | / |
| Gray *et al* [74] | 2010 | UK | Leicester Risk assessment score – logistic [74] | Prevalent | Yes | 1 | No | / | / |
| Gray et al [75] | 2012 | UK | Leicester practice risk score – logistic [75] | Prevalent | Yes | 1 | No | / | / |
| Gray *et al* [76] | 2014 | South Asians in UK | Leicester Risk assessment score – logistic [74] | Prevalent | No | 2 | No | / | / |
| | | | Leicester practice risk score – logistic [75] | | | | | | |
| Griffin *et al* [53] | 2000 | UK | Cambridge diabetes risk score – logistic [53] | Incident | Yes | 1 | No | / | / |
| Guasch-Ferré *et al* [77] | 2012 | Spain | PREDIMED personal model – cox [77] | Incident | Yes | 1 | No | / | / |
| | | | Finnish diabetes risk score full – logistic [15] | | | 2 | | | |
| | | | German diabetes risk score – cox [41] | | | | | | |
| Guerrero-Romero *et al* [78] | 2010 | Mexico | Mexican diabetes model – cox [78] | Incident | Yes | 1 | No | / | / |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Hanley *et al* [79] | 2004 | USA | San Antonio risk clinical model – logistic [42] | Incident | Yes | 2 | No | / | / |
| Hartwig *et al* [25] | 2013 | Germany | German diabetes risk score – cox [41] | Incident | No | 2 | Yes | NS | Addition of HbA1c, blood glucose, triglycerides, HDL, alanine aminotransferase and gammaglutamyltransferase |
| He *et al* [80] | 2012 | China | DESIR clinical risk model – logistic [39] | Incident | No | 2 | No | / | / |
| | | | Finnish diabetes risk score concise – logistic [15] | | | | | | |
| | | | Framingham offspring simple clinical model – logistic [13] | | | | | | |
| | | | Thai simple risk model – logistic [55] | | | | | | |
| | | | Taiwan scoring concise model – cox [66] | | | | | | |
| | | | San Antonio risk clinical model – logistic [42] | | | | | | |
| | | | PROCAM risk model – logistic [43] | | | | | | |
| | | | California scoring model – logistic [46] | Prevalent | No | 2 | No | / | / |
| | | | Cambridge diabetes risk score – logistic [53] | | | | | | |
| | | | Omani risk score – logistic [56] | | | | | | |
| | | | Asian Indian diabetes risk score – logistic [70] | | | | | | |
| | | | Urban Asian Indian risk score – logistic [63] | | | | | | |
| Heizana *et al* [81] | 2013 | Japan | TOPICS diabetes categorical screening score –cox [81] | Prevalent | Yes | 1 | No | / | Prevalent to incident |
| | | | TOPICS diabetes continuous screening score –cox [81] | | | | | | |
| | | | Rotterdam predictive model 1 – logistic [52] | | | 2 | | | / |
| | | | Danish risk score – logistic [54] | | | | | | |
| | | | Omani risk score – logistic [56] | | | | | | |
| | | | US screening score – logistic [58] | | | | | | |
| | | | Asian Indian diabetes risk score – logistic [70] | | | | | | |
| | | | Qingdao diabetes risk score – logistic [69] | | | | | | |
| | | | Leicester Risk assessment score – logistic [74] | | | | | | |
| | | | Screening tool in ADDITION-Leicester – logistic [75] | | | | | | |
| Heldgaard & Griffin [82] | 2006 | Denmark | Cambridge diabetes risk score – logistic [53] | Prevalent | No | 2 | No | / | / |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Hippisley-Cox *et al* [38] | 2009 | UK | QDScore – cox [38] | Incident | Yes | 1 | No | / | / |
| | | | Cambridge diabetes risk score – logistic [53] | | | 2 | | | |
| Hippisley-Cox *et al* [26] | 2014 | UK | QDScore – cox [38] | Incident | No | 1 | Yes | NS | / |
| Kahn *et al* [37] | 2009 | USA | Framingham offspring simple clinical model – logistic [13] | Incident | Yes | 2 | No | / | / |
| | | | DESIR clinical risk model – logistic [39] | | | | | | |
| Kanaya *et al* [46] | 2005 | USA | California scoring model – logistic [46] | Incident | Yes | 1 | No | / | / |
| Keesukphan *et al* [83] | 2007 | Thailand | Thailand diabetes risk model – logistic [83] | Incident | Yes | 1 | No | / | / |
| Kengne *et al* [6] | 2014 | Europe | ARIC clinical only model – logistic [84] | Incident | No | 3 | Yes | Yes | |
| | | | ARIC enhanced model – weibull [37] | | | | | No | |
| | | | AUSDRISK – logistic [35] | | | | | Yes | Blood pressure medication replaced by proxy 'any hypertension' |
| | | | Cambridge diabetes risk score – logistic [53] | | | | | No | Family history of diabetes excluded<br>Blood pressure medication replaced by proxy 'any hypertension' |
| | | | DESIR clinical risk model– logistic [39] | | | | | Yes | |
| | | | DPoRT – weibull [27] | | | | | No | |
| | | | Finnish diabetes risk score concise – logistic [15] | | | | | Yes | Blood pressure medication replaced by proxy 'any hypertension' |
| | | | Finnish diabetes risk score full – logistic [15] | | | | | Yes | Categorical daily intake of fruit, vegetables or berries replaced by continuous proxy<br>Blood pressure medication replaced by proxy 'any hypertension' |
| | | | Framingham offspring complex clinical model 1 – logistic [13] | | | | | Yes | |
| | | | KORA base model (model 1) – logistic [31] | | | | | Yes | |
| | | | German diabetes risk score – cox [41] | | | | | Yes | |
| | | | QDScore – cox [38] | | | | | Yes | Continuous Townsend score replaced by categorical education proxy |
| Ko *et al* [85] | 2010 | China | Southern Chinese diabetes risk model – logistic [85] | Incident | Yes | 1 | No | / | / |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Ku and Kegels [86] | 2013 | Philippines | Finnish diabetes risk score full – logistic [15, 16] | Prevalent | No | 2 | No | / | Removal of variables |
| Lee *et al* [87] | 2012 | Korea | Korean diabetes prediction score – logistic [87] | Prevalent | Yes | 1 | No | / | / |
| | | | Qingdao diabetes risk score – logistic [69] | | | 2 | | | |
| | | | Thai simple risk model – logistic [55] | | | | | | |
| | | | Rotterdam predictive model 1 – logistic [52] | | | | | | |
| | | | US screening score – logistic [58] | | | | | | |
| Li *et al* [88] | 2007 | Germany | Framingham offspring simple clinical model – logistic [13] | Incident | No | 2 | No | / | / |
| Li *et al* [89] | 2009 | Germany | Finnish diabetes risk score full – logistic [15, 16] | Prevalent | No | 2 | No | / | Removal of variables |
| Li *et al* [90] | 2011 | Taiwan | American Diabetes Association risk tool [91] | Incident | No | 2 | No | / | Addition of family history of hyperlipidaemia, education levels, TV hours, history of cardiovascular disease and hypertriglyceride |
| Lin *et al* [92] | 2009 | Taiwan | ARIC clinical only model – logistic [84] | Incident | No | 2 | No | / | / |
| | | | Asian Indian diabetes risk score – logistic [70] | | | | | | |
| | | | Cambridge diabetes risk score – logistic [53] | | | | | | |
| | | | Danish risk score – logistic [54] | | | | | | |
| | | | DESIR clinical risk model – logistic [39] | | | | | | |
| | | | Finnish diabetes risk score concise – logistic [15] | | | | | | |
| | | | Rotterdam predictive model – logistic [52] (model 1/2 not stated) | | | | | | |
| | | | Omani risk score – logistic [56] | | | | | | |
| | | | QDScore – cox [38] | | | | | | |
| | | | Thai simple risk model – logistic [55] | | | | | | |
| Lindstrom *et al* [15] | 2003 | Finland | Finnish diabetes risk score concise – logistic [15] | Incident | Yes | 1 | No | / | / |
| | | | Finnish diabetes risk score full – logistic [15] | | | | | | |
| Luo et al [93] | 2014 | China | Chinese risk assessment model – logistic [93] | Incident | Yes | 1 | No | / | / |
| Lui et al [94] | 2011 | China | Chinese diabetes risk model 1 – logistic [94] | Prevalent | Yes | 1 | No | / | / |
| | | | Chinese diabetes risk model 2 – logistic [94] | | | | | | |
| | | | Chinese diabetes risk model 3 – logistic [94] | | | | | | |
| Lyssenko *et al* [95] | 2008 | Sweden and Finland | Framingham offspring simple clinical model – logistic [13] | Incident | Yes | 2 | No | / | / |
| Lyssenko *et al* [96] | 2012 | Denmark | PreDX diabetes risk score – logistic [97] | Incident | No | 2 | No | / | / |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | San Antonio risk clinical model – logistic [42] | | | | | | |
| | | | Framingham offspring simple clinical model – logistic [13] | | | | | | |
| Mainous *et al* [98] | 2007 | USA | ARIC clinical only model – logistic [84] | Incident | No | 2 | No | / | / |
| Makrilakis *et al* [99] | 2010 | Greece | Finnish diabetes risk score full – logistic [15, 16] | Prevalent | No | 2 | No | / | / |
| Mann *et al* [14] | 2010 | USA | Framingham offspring simple clinical model – logistic [13] | Incident | No | 2 | Yes | Yes | / |
| | | | San Antonio risk clinical model – logistic [42] | | | | | | |
| | | | ARIC clinical only model – logistic [84] | | | | | | |
| McNeely *et al* [100] | 2003 | USA | San Antonio risk clinical model – logistic [42] | Incident | No | 2 | No | / | Addition of multiple variables |
| Mühlenbruch *et al* [101] | 2014 | Germany | German diabetes risk score – cox [41] | Incident | No | 2 | No | / | Sibling history of diabetes not available and all individuals assigned 0.5 Minor variable changes |
| Nicols *et al* [28] | 2008 | USA | Framingham offspring personal model – logistic [13] | Incident | No | 2 | Yes | N/S | / |
| | | | Framingham offspring simple clinical model – logistic [13] | | | | | Yes | / |
| | | | Framingham offspring complex clinical model 1 – logistic [13] | | | | | N/S | / |
| | | | Framingham offspring complex clinical model 2 – logistic [13] | | | | | N/S | |
| | | | Framingham offspring complex clinical model 3 – logistic [13] | | | | | N/S | |
| Park *et al* [102] | 2002 | UK | Cambridge diabetes risk score – logistic [53] | Prevalent | Yes | 1 | No | / | / |
| | | | Rotterdam predictive model 1 – logistic [52] | | | 2 | | | |
| Phillips *et al* [103] | 2013 | Ireland | DESIR clinical risk model – logistic [39] | Incident | No | 2 | No | / | / |
| | | | Cambridge diabetes risk score – logistic [53] | | | | | | |
| | | | ARIC clinical only model – logistic [84] | | | | | | |
| | | | ARIC enhanced model – weibull [37] | | | | | | |
| | | | Finnish diabetes risk score full – logistic [15] | | | | | | |
| | | | German diabetes risk score – cox [41] | | | | | | |
| | | | Framingham offspring simple clinical model – logistic [13] | | | | | | |
| Rahman *et al* [104] | 2008 | UK | Cambridge diabetes risk score – logistic [53] | Incident | Yes | 2 | No | / | Prevalent to incident |
| Ramachandran *et al* [70] | 2005 | India | Cambridge diabetes risk score – logistic [53] | Prevalent | Yes | 2 | No | / | / |
| Rathmann *et al* [105] | 2005 | Germany | Cambridge diabetes risk score – logistic [53] | Prevalent | No | 2 | No | / | / |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Rotterdam predictive model 1 – logistic [52]<br>San Antonio risk clinical model – logistic [42]<br>Finnish diabetes risk score full – logistic [15, 16] | | | | | | |
| Riaz *et al* [106] | 2012 | Pakistan | RAPID model – logistic [106] | Incident | Yes | 1 | No | / | / |
| Rolka *et al* [107] | 2001 | USA | American Diabetes Association risk assessment questionnaire [51] | Prevalent | No | 2 | No | / | / |
| Rosella et al [27] | 2010 | Canada | DPoRT – weibull [27] | Incident | Yes | 1 | Yes | N/S | Recalibrated for ethnicity |
| Rowe *et al* [108] | 2012 | USA | PreDX diabetes risk score – logistic [97] | Incident | No | 2 | No | / | / |
| Ruige *et al* [109] | 1997 | Netherlands | American Diabetes Association risk assessment questionnaire [51]<br>American Diabetes Association risk tool [91] | Prevalent | Yes | 2 | No | / | / |
| Saaristo *et al* [16] | 2005 | Finland | Finnish diabetes risk score full – logistic [15, 16] | Prevalent | No | 2 | No | / | Incident to prevalent |
| Schmidt *et al* [110] | 2012 | Switzerland | ARIC basic model – weibull [37]<br>ARIC enhanced model – weibull [37]<br>DESIR clinical risk model – logistic [39]<br>Cambridge diabetes risk score – logistic [53]<br>Finnish diabetes risk score full – logistic [15]<br>Framingham offspring simple clinical model – logistic [13] | Incident | No | 2 | No | / | / |
| Schmidt *et al* [111] | 2012 | Switzerland | ARIC enhanced model – weibull [37] | Incident | No | 2 | No | / | Addition of genetic variables |
| Schulze *et al* [41] | 2007 | Germany | German diabetes risk score – cox [41] | Incident | Yes | 3 | No | / | / |
| Simmons *et al* [40] | 2007 | UK | Cambridge diabetes risk score – logistic [53] | Incident | Yes | 2 | No | / | Addition of a physical activity and diet |
| Spijkerman *et al* [112] | 2004 | Minorities in UK | Cambridge diabetes risk score – logistic [53] | Prevalent | No | 2 | No | / | / |
| Stern *et al* [113] | 2004 | Mexico | San Antonio risk clinical model – logistic [42]<br>Framingham offspring simple clinical model – logistic [13] | Incident | No | 2 | No | / | Addition of metabolic syndrome |
| Stern *et al* [114] | 2008 | USA | San Antonio risk clinical model – logistic [42]<br>ARIC clinical only model – logistic [84] | Incident | No | 2 | No | / | / |
| Sun *et al* [115] | 2009 | Taiwan | ARIC clinical model plus FBG – logistic [84]<br>ARIC clinical model plus FBG and lipids – logistic [84] | Incident | Yes | 2 | No | / | / |
| Tabaebi *et al* [116] | 2002 | Egypt | Egyptian diabetes risk model – logistic [116] | Prevalent | Yes | 1 | No | / | / |
| Talmud *et al* [117] | 2010 | UK | Cambridge diabetes risk score – logistic [53] | Incident | Yes | 2 | No | / | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Framingham offspring simple clinical model – logistic [13] | | | | | | Addition of genetic variables |
| Tankova *et al* [118] | 2011 | Bulgaria | Finnish diabetes risk score full – logistic [15, 16] | Prevalent | No | 2 | No | / | Incident to prevalent |
| Tuomilehto *et al* [119] | 2010 | Global | STOP-NIDDM risk score – cox [119] | Incident | Yes | 2 | No | / | / |
| Urdea *et al* [120] | 2009 | Denmark | PreDX diabetes risk score – logistic [97] | Incident | No | 2 | No | / | / |
| Wannamethee *et al* [121] | 2005 | UK | Framingham offspring simple clinical model – logistic [13] | Incident | No | 2 | No | / | Waist circumference replaced by BMI proxy |
| Witte *et al* [122] | 2010 | UK | Cambridge diabetes risk score – logistic [53] | Prevalent | No | 2 | No | / | / |
| | | | Rotterdam predictive model 1 – logistic [52] | | | | | | |
| | | | Rotterdam predictive model 2 – logistic [52] | | | | | | |
| | | | Finnish diabetes risk score full – logistic [15, 16] | | | | | | |
| | | | Danish risk score – logistic [54] | | | | | | |
| | | | Hoorn study risk model – logistic [109] | | | | | | |
| Xu *et al* [29] | 2014 | China | Framingham offspring simple clinical model – logistic [13] | Incident | Yes | 2 | Yes | Yes | / |
| Zhang *et al* [123] | 2014 | USA | Finnish diabetes risk score full – logistic [15, 16] | Prevalent | No | 2 | No | / | / |
| | | | US screening score – logistic [58] | | | | | | |
| Zhou *et al* [124] | 2013 | China | New Chinese Diabetes Risk Score - logistic [124] | Prevalent | Yes | 2 | No | / | / |

## Discussion

The validation of existent models in a new population is highly encouraged, preventing the availability of numerous models, where few have been externally validated. The common method of developing and validating models simultaneously in a database in which previous risk prediction research has not been, defeats this purpose. Ideally, should a database suitable for diabetes risk prediction research be available, models should first be validated in an attempt to find an existent model that can perform at an optimum discrimination and calibration. Should a model show systematic overestimation or underestimation of risk, and the performance be too low to allow for accurate prediction and successful implementation, recalibration techniques can be employed in an effort to increase the performance of the model.

The aim of this study was to determine the extent to which model recalibration was undertaken in validation of diabetes risk models. This review of available published literature on the validation of diabetes risk prediction models showed that although validation of existent models is occurring, the attempt to fit these models to the new setting is poor. Additionally, we wished to determine if this recalibration was successful in increasing model performance when incorporating information for the validation population. Most studies that undertook the recalibration of models were able to show that model performance can be increased with basic recalibration techniques. The new models retain their importance in a new setting, taking into account the underlying incidence of the outcome and the variable relative importance of each risk factor from the development to the validation population. This increase in performance through simple recalibration is important in the effort to encourage the updating of models during validation. The statistical effort in recalibrating a model is slight and the final product of a model better fitted to the population in question and increased model performance worth the added step.

Although we aimed to comprehensively review all published papers on development and validation of undiagnosed diabetes risk prediction models, it should be highlighted that we

may have missed some published validation studies. However, the overall result would not be expected to differ significantly with the possible inclusion of more model validation studies.

## Conclusion

Without recalibration in the validation of a diabetes risk prediction model, the ability of these models to generate an accurate point estimate of an individual's diabetes risk may be inadequate. The importance of generalizability and validation of current models is repeatedly emphasized in literature, however this is fruitless if extra efforts are not taken to fit the model as best as possible to the new setting. Unfortunately, only a relatively small number of validation studies have included recalibration in their methodologies. Additionally, no prevalent diabetes risk prediction models used recalibration in an attempt to better fit the model to the validation population. An increased focus on the validation, and particularly recalibration, of existent models will improve the generalizability of the models and likely lead to greater application of diabetes risk prediction models in daily clinical practice. The question that remains is, when is a model ruled sufficiently validated and recalibration / updated? Future research should address this question and allow for the determination of how many validation studies, what type of adjustments need to be made and most importantly, what is optimum performance to justify the implementation of the risk prediction model into clinical practice.

## Acknowledgements

KLM was supported by a scholarship from the South African National Research Foundation and the Carl & Emily Fuchs Foundation.

## References


1. Steyerberg, E.W. *Clinical prediction models*. 2009, New York: Springer.
2. Steyerberg, E.W., Borsboom, G.J., van Houwelingen, H.C., Eijkemans, M.J., and Habbema, J.D.F. *Validation and updating of predictive logistic regression models: A study on sample size and shrinkage.* Statistics in medicine. 2004; **23**(16): p. 2567-2586.
3. Janssen, K., Moons, K., Kalkman, C., Grobbee, D., and Vergouwe, Y. *Updating methods improved the performance of a clinical prediction model in new patients.* Journal of clinical epidemiology. 2008; **61**(1): p. 76-86.
4. Collins, G.S., de Groot, J.A., Dutton, S., Omar, O., Shanyinde, M., Tajar, A.*, et al*. *External validation of multivariable prediction models: A systematic review of methodological conduct and reporting.* BMC medical research methodology. 2014; **14**(1): p. 40.
5. Noble, D., Mathur, R., Dent, T., Meads, C., and Greenhalgh, T. *Risk models and scores for type 2 diabetes: Systematic review.* Bmj. 2011; **343**.
6. Kengne, A.P., Beulens, J.W., Peelen, L.M., Moons, K.G., van der Schouw, Y.T., Schulze, M.B.*, et al*. *Non-invasive risk scores for prediction of type 2 diabetes (epic-interact): A validation of existing models.* The Lancet Diabetes & Endocrinology. 2014; **2**(1): p. 19-29.
7. Buijsse, B., Simmons, R.K., Griffin, S.J., and Schulze, M.B. *Risk assessment tools for identifying individuals at risk of developing type 2 diabetes.* Epidemiologic reviews. 2011; **33**(1): p. 46-62.
8. Collins, G.S., Mallett, S., Omar, O., and Yu, L.-M. *Developing risk prediction models for type 2 diabetes: A systematic review of methodology and reporting.* BMC medicine. 2011; **9**(1): p. 103.
9. Thoopputra, T., Newby, D., Schneider, J., and Li, S.C. *Survey of diabetes risk assessment tools: Concepts, structure and performance.* Diabetes/metabolism research and reviews. 2012; **28**(6): p. 485-498.
10. Brown, N., Critchley, J., Bogowicz, P., Mayige, M., and Unwin, N. *Risk scores based on self-reported or available clinical data to detect undiagnosed type 2 diabetes: A systematic review.* Diabetes research and clinical practice. 2012; **98**(3): p. 369-385.
11. Moher, D., Liberati, A., Tetzlaff, J., and Altman, D.G. *Preferred reporting items for systematic reviews and meta-analyses: The prisma statement.* Annals of Internal Medicine. 2009; **151**(4): p. 264-269.
12. Justice, A.C., Covinsky, K.E., and Berlin, J.A. *Assessing the generalizability of prognostic information.* Annals of Internal Medicine. 1999; **130**(6): p. 515-524.
13. Wilson, P.W., Meigs, J.B., Sullivan, L., Fox, C.S., Nathan, D.M., and D'Agostino, R.B. *Prediction of incident diabetes mellitus in middle-aged adults: The framingham offspring study.* Archives of internal medicine. 2007; **167**(10): p. 1068-1074.
14. Mann, D.M., Bertoni, A.G., Shimbo, D., Carnethon, M.R., Chen, H., Jenny, N.S.*, et al*. *Comparative validity of 3 diabetes mellitus risk prediction scoring models in a multiethnic us cohort the multi-ethnic study of atherosclerosis.* American journal of epidemiology. 2010; **171**(9): p. 980-988.
15. Lindström, J. and Tuomilehto, J. *The diabetes risk score a practical tool to predict type 2 diabetes risk.* Diabetes Care. 2003; **26**(3): p. 725-731.
16. Saaristo, T., Peltonen, M., Lindström, J., Saarikoski, L., Sundvall, J., Eriksson, J.G.*, et al*. *Cross-sectional evaluation of the finnish diabetes risk score: A tool to identify undetected type 2 diabetes, abnormal glucose tolerance and metabolic syndrome.* Diabetes and Vascular Disease Research. 2005; **2**(2): p. 67-72.
17. D'Agostino Sr, R.B., Grundy, S., Sullivan, L.M., and Wilson, P. *Validation of the framingham coronary heart disease prediction scores: Results of a multiple ethnic groups investigation.* Jama. 2001; **286**(2): p. 180-187.

18. Hosmer Jr, D.W. and Lemeshow, S. *Applied logistic regression*. 2004, New Jersey: John Wiley & Sons.
19. Ivanov, J., Tu, J.V., and Naylor, C.D. *Ready-made, recalibrated, or remodeled? Issues in the use of risk indexes for assessing mortality after coronary artery bypass graft surgery.* Circulation. 1999; **99**(16): p. 2098-2104.
20. Toll, D., Janssen, K., Vergouwe, Y., and Moons, K. *Validation, updating and impact of clinical prediction rules: A review.* Journal of clinical epidemiology. 2008; **61**(11): p. 1085-1094.
21. Nelson, W. *Applied life data analysis, 1982.* Statistics, Published by John Wiley & Sons, lnc., New York, NY.
22. Cox, D. *Regression models and life-tables.* Journal of the Royal Statistical Society. 1972; **34**(2): p. 187-220.
23. van Houwelingen, H.C. *Validation, calibration, revision and combination of prognostic survival models.* Statistics in medicine. 2000; **19**(24): p. 3401-3415.
24. Bozorgmanesh, M., Hadaegh, F., Ghaffari, S., Harati, H., and Azizi, F. *A simple risk score effectively predicted type 2 diabetes in iranian adult population: Population-based cohort study.* The European Journal of Public Health. 2011; **21**(5): p. 554-559.
25. Hartwig, S., Kuss, O., Tiller, D., Greiser, K., Schulze, M., Dierkes, J*., et al*. *Validation of the german diabetes risk score within a population-based representative cohort.* Diabetic medicine. 2013; **30**(9): p. 1047-1053.
26. Hippisley-Cox, J., Coupland, C., and Brindle, P. *The performance of seven qprediction risk scores in an independent external sample of patients from general practice: A validation study.* BMJ open. 2014; **4**(8): p. e005809.
27. Rosella, L.C., Manuel, D.G., Burchill, C., and Stukel, T.A. *A population-based risk algorithm for the development of diabetes: Development and validation of the diabetes population risk tool (dport).* Journal of epidemiology and community health. 2011; **65**(7): p. 613-620.
28. Nichols, G.A. and Brown, J.B. *Validating the framingham offspring study equations for predicting incident diabetes mellitus.* Am J Manag Care. 2008; **14**(9): p. 574-580.
29. Xu, L., Jiang, C., Schooling, C., Zhang, W., Cheng, K., and Lam, T. *Prediction of 4-year incident diabetes in older chinese: Recalibration of the framingham diabetes score on guangzhou biobank cohort study.* Preventive medicine. 2014; **69**: p. 63-68.
30. Abbasi, A., Corpeleijn, E., Peelen, L.M., Gansevoort, R.T., de Jong, P.E., Gans, R.O*., et al*. *External validation of the kora s4/f4 prediction models for the risk of developing type 2 diabetes in older adults: The prevend study.* European journal of epidemiology. 2012; **27**(1): p. 47-52.
31. 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.* Diabetic medicine. 2010; **27**(10): p. 1116-1123.
32. Abbasi, A., Peelen, L.M., Corpeleijn, E., van der Schouw, Y.T., Stolk, R.P., Spijkerman, A.M*., et al*. *Prediction models for risk of developing type 2 diabetes: Systematic literature search and independent external validation study.* Bmj. 2012; **345**.
33. Alssema, M., Vistisen, D., Heymans, M., Nijpels, G., Glümer, C., Zimmet, P*., 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**(5): p. 1004-1012.
34. Wannamethee, S., Papacosta, O., Whincup, P., Thomas, M., Carson, C., Lawlor, D*., 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.* Diabetic medicine. 2011; **28**(1): p. 23-30.
35. Chen, L., Magliano, D.J., Balkau, B., Colagiuri, S., Zimmet, P.Z., Tonkin, A.M*., et al*. *Ausdrisk: An australian type 2 diabetes risk assessment tool based on demographic, lifestyle and simple anthropometric measures.* Medical Journal of Australia. 2010; **192**(4): p. 197.

36. Joseph, J., Svartberg, J., Njølstad, I., and Schirmer, H. *Incidence of and risk factors for type-2 diabetes in a general population: The tromsø study.* Scandinavian journal of public health. 2010; **38**(7): p. 768-775.
37. Kahn, H.S., Cheng, Y.J., Thompson, T.J., Imperatore, G., and Gregg, E.W. *Two risk-scoring systems for predicting incident diabetes mellitus in us adults age 45 to 64 years.* Annals of Internal Medicine. 2009; **150**(11): p. 741-751.
38. Hippisley-Cox, J., Coupland, C., Robson, J., Sheikh, A., and Brindle, P. *Predicting risk of type 2 diabetes in england and wales: Prospective derivation and validation of qdscore.* Bmj. 2009; **338**: p. 812-816.
39. 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**(10): p. 2056-2061.
40. Simmons, R., Harding, A.H., Wareham, N., and Griffin, S. *Do simple questions about diet and physical activity help to identify those at risk of type 2 diabetes?* Diabetic medicine. 2007; **24**(8): p. 830-835.
41. Schulze, M.B., Hoffmann, K., Boeing, H., Linseisen, J., Rohrmann, S., Möhlig, 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**(3): p. 510-515.
42. Stern, M.P., Williams, K., and Haffner, S.M. *Identification of persons at high risk for type 2 diabetes mellitus: Do we need the oral glucose tolerance test?* Annals of Internal Medicine. 2002; **136**(8): p. 575-581.
43. Von Eckardstein, A., Schulte, H., and 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.* The Journal of Clinical Endocrinology & Metabolism. 2000; **85**(9): p. 3101-3108.
44. Stern, M.P., Morales, P.A., Valdez, R.A., Monterrosa, A., Haffner, S.M., Mitchell, B.D.*, et al*. *Predicting diabetes: Moving beyond impaired glucose tolerance.* Diabetes. 1993; **42**(5): p. 706-714.
45. Abdul-Ghani, M.A., Lyssenko, V., Tuomi, T., DeFronzo, R.A., and Groop, L. *Fasting versus postload plasma glucose concentration and the risk for future type 2 diabetes results from the botnia study.* Diabetes Care. 2009; **32**(2): p. 281-286.
46. Kanaya, A.M., Fyr, C.L.W., De Rekeneire, N., Shorr, R.I., Schwartz, A.V., Goodpaster, B.H.*, et al*. *Predicting the development of diabetes in older adults the derivation and validation of a prediction rule.* Diabetes Care. 2005; **28**(2): p. 404-408.
47. Adhikari, P., Pathak, R., and Kotian, S. *Validation of the mdrf-indian diabetes risk score (idrs) in another south indian population through the boloor diabetes study (bds).* JAPI. 2010; **50**: p. 434-436.
48. Mohan, V., Deepa, R., Deepa, M., Somannavar, S., and Datta, M. *A simplified indian diabetes risk score for screening for undiagnosed diabetic subjects.* Journal of the Association of Physicians of India. 2005; **53**: p. 759-763.
49. Akyil, R.C., Miloglu, O., Olgun, N., and Bayrakdar, I.S. *A comparison of three different diabetes screening methods among dental patients in turkey.* Pakistan journal of medical sciences. 2014; **30**(1): p. 65.
50. Al Khalaf, M., Eid, M., Najjar, H., Alhajry, K., Doi, S., and Thalib, L. *Screening for diabetes in kuwait and evaluation of risk scores.* EMHJ. 2010; **16**(7): p. 725-731.
51. Herman, W.H., Smith, P.J., Thompson, T.J., Engelgau, M.M., and Aubert, R.E. *A new and simple questionnaire to identify people at increased risk for undiagnosed diabetes.* Diabetes Care. 1995; **18**(3): p. 382-387.

52. Baan, C.A., Ruige, J.B., Stolk, R.P., Witteman, J., Dekker, J.M., Heine, R.J.*, et al*. *Performance of a predictive model to identify undiagnosed diabetes in a health care setting.* Diabetes Care. 1999; **22**(2): p. 213-219.
53. Griffin, S., Little, P., Hales, C., Kinmonth, A., and Wareham, N. *Diabetes risk score: Towards earlier detection of type 2 diabetes in general practice.* Diabetes/metabolism research and reviews. 2000; **16**(3): p. 164-171.
54. Glümer, C., Carstensen, B., Sandbæk, A., Lauritzen, T., Jørgensen, T., and Borch-Johnsen, K. *A danish diabetes risk score for targeted screening the inter99 study.* Diabetes Care. 2004; **27**(3): p. 727-733.
55. 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**(8): p. 1872-1877.
56. Al-Lawati, J. and Tuomilehto, J. *Diabetes risk score in oman: A tool to identify prevalent type 2 diabetes among arabs of the middle east.* Diabetes research and clinical practice. 2007; **77**(3): p. 438-444.
57. Alssema, M., Newson, R.S., Bakker, S.J., Stehouwer, C.D., Heymans, M.W., Nijpels, G.*, et al*. *One risk assessment tool for cardiovascular disease, type 2 diabetes, and chronic kidney disease.* Diabetes Care. 2012; **35**(4): p. 741-748.
58. Bang, H., Edwards, A.M., Bomback, A.S., Ballantyne, C.M., Brillon, D., Callahan, M.A.*, et al*. *Development and validation of a patient self-assessment score for diabetes risk.* Annals of Internal Medicine. 2009; **151**(11): p. 775-783.
59. Bergmann, A., Li, J., Wang, L., Schulze, J., Bornstein, S., and Schwarz, P. *A simplified finnish diabetes risk score to predict type 2 diabetes risk and disease evolution in a german population.* Hormone and metabolic research. 2007; **39**(09): p. 677-682.
60. Bhadoria, A.S., Kasar, P.K., and Toppo, N.A. *Validation of indian diabetic risk score in diagnosing type 2 diabetes mellitus against high fasting blood sugar levels among adult population of central india.* Biomedical journal. 2014; **38**(4): p. 359-360.
61. Bozorgmanesh, M., Hadaegh, F., and 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 diabetologica. 2013; **50**(2): p. 175-181.
62. Bozorgmanesh, M., Hadaegh, F., Zabetian, A., and Azizi, F. *San antonio heart study diabetes prediction model applicable to a middle eastern population? Tehran glucose and lipid study.* International journal of public health. 2010; **55**(4): p. 315-323.
63. Chaturvedi, V., Reddy, K., Prabhakaran, D., Jeemon, P., Ramakrishnan, L., Shah, P.*, et al*. *Development of a clinical risk score in predicting undiagnosed diabetes in urban asian indian adults: A population-based study.* CVD prevention and control. 2008; **3**(3): p. 141-151.
64. Cameron, A., Zimmet, P., Soderberg, S., Alberti, K., Sicree, R., Tuomilehto, J.*, et al*. *The metabolic syndrome as a predictor of incident diabetes mellitus in mauritius.* Diabetic medicine. 2007; **24**(12): p. 1460-1469.
65. Cameron, A., Magliano, D., Zimmet, P., Welborn, T., Colagiuri, S., Tonkin, A.*, et al*. *The metabolic syndrome as a tool for predicting future diabetes: The ausdiab study.* Journal of internal medicine. 2008; **264**(2): p. 177-186.
66. 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**(3): p. 443-450.
67. Farran, B., Channanath, A.M., Behbehani, K., and Thanaraj, T.A. *Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: Machine-learning algorithms and validation using national health data from kuwait—a cohort study.* BMJ open. 2013; **3**(5): p. e002457.
68. Franciosi, M., De Berardis, G., Rossi, M.C., Sacco, M., Belfiglio, M., Pellegrini, F.*, et al*. *Use of the diabetes risk score for opportunistic screening of undiagnosed diabetes and impaired*

*glucose tolerance the igloo (impaired glucose tolerance and long-term outcomes observational) study.* Diabetes Care. 2005; **28**(5): p. 1187-1194.

69. Gao, W., Dong, Y., Pang, Z., Nan, H., Wang, S., Ren, J.*, et al*. *A simple chinese risk score for undiagnosed diabetes.* Diabetic medicine. 2010; **27**(3): p. 274-281.
70. Ramachandran, A., Snehalatha, C., Vijay, V., Wareham, N., and Colagiuri, S. *Derivation and validation of diabetes risk score for urban asian indians.* Diabetes research and clinical practice. 2005; **70**(1): p. 63-70.
71. Ginde, A.A., Delaney, K.E., Lieberman, R.M., Vanderweil, S.G., and Camargo, C.A. *Estimated risk for undiagnosed diabetes in the emergency department: A multicenter survey.* Academic Emergency Medicine. 2007; **14**(5): p. 492-495.
72. Glümer, C., Borch-Johnsen, K., and Colagiuri, S. *Can a screening programme for diabetes be applied to another population?* Diabetic medicine. 2005; **22**(9): p. 1234-1238.
73. Glümer, C., Vistisen, D., Borch-Johnsen, K., and Colagiuri, S. *Risk scores for type 2 diabetes can be applied in some populations but not all.* Diabetes Care. 2006; **29**(2): p. 410-414.
74. Gray, L., Taub, N., Khunti, K., Gardiner, E., Hiles, S., Webb, D.*, et al*. *The leicester risk assessment score for detecting undiagnosed type 2 diabetes and impaired glucose regulation for use in a multiethnic uk setting.* Diabetic medicine. 2010; **27**(8): p. 887-895.
75. Gray, L., Davies, M., Hiles, S., Taub, N., Webb, D., Srinivasan, B.*, et al*. *Detection of impaired glucose regulation and/or type 2 diabetes mellitus, using primary care electronic data, in a multiethnic uk community setting.* Diabetologia. 2012; **55**(4): p. 959-966.
76. Gray, L., Khunti, K., Wilmot, E., Yates, T., and Davies, M. *External validation of two diabetes risk scores in a young uk south asian population.* Diabetes research and clinical practice. 2014; **104**(3): p. 451-458.
77. Guasch-Ferré, M., Bulló, M., Costa, B., Martínez-Gonzalez, M.Á., Ibarrola-Jurado, N., Estruch, R.*, et al*. *A risk score to predict type 2 diabetes mellitus in an elderly spanish mediterranean population at high cardiovascular risk.* PloS one. 2012; **7**(3): p. e33437.
78. Guerrero-Romero, F. and Rodríguez-Morán, M. *Validation of an instrument for screening cases of type 2 diabetes and monitoring at-risk individuals in mexico.* Revista Panamericana de Salud Pública. 2010; **27**(3): p. 181-186.
79. Hanley, A.J., Festa, A., D'Agostino, R.B., Wagenknecht, L.E., Savage, P.J., Tracy, R.P.*, et al*. *Metabolic and inflammation variable clusters and prediction of type 2 diabetes factor analysis using directly measured insulin sensitivity.* Diabetes. 2004; **53**(7): p. 1773-1781.
80. He, S., Chen, X., Cui, K., Peng, Y., Liu, K., Lv, Z.*, et al*. *Validity evaluation of recently published diabetes risk scoring models in a general chinese population.* Diabetes research and clinical practice. 2012; **95**(2): p. 291-298.
81. Heianza, Y., Arase, Y., Saito, K., Hsieh, S.D., Tsuji, H., Kodama, S.*, et al*. *Development of a screening score for undiagnosed diabetes and its application in estimating absolute risk of future type 2 diabetes in japan: Toranomon hospital health management center study 10 (topics 10).* The Journal of Clinical Endocrinology & Metabolism. 2013; **98**(3): p. 1051-1060.
82. Heldgaard, P.E. and Griffin, S. *Routinely collected general practice data aids identification of people with hyperglycaemia and metabolic syndrome.* Diabetic medicine. 2006; **23**(9): p. 996-1002.
83. Keesukphan, P., Chanprasertyothin, S., Ongphiphadhanakul, B., and Puavilai, G. *The development and validation of a diabetes risk score for high-risk thai adults.* Journal of the Medical Association of Thailand. 2007; **90**(1): p. 149.
84. Schmidt, M.I., Duncan, B.B., Bang, H., Pankow, J.S., Ballantyne, C.M., Golden, S.H.*, et al*. *Identifying individuals at high risk for diabetes the atherosclerosis risk in communities study.* Diabetes Care. 2005; **28**(8): p. 2013-2018.
85. Ko, G., So, W., Tong, P., Ma, R., Kong, A., Ozaki, R.*, et al*. *A simple risk score to identify southern chinese at high risk for diabetes.* Diabetic medicine. 2010; **27**(6): p. 644-649.

86. Ku, G. and Kegels, G. *The performance of the finnish diabetes risk score, a modified finnish diabetes risk score and a simplified finnish diabetes risk score in community-based cross-sectional screening of undiagnosed type 2 diabetes in the philippines.* Primary care diabetes. 2013; **7**(4): p. 249-259.
87. Lee, Y.-h., Bang, H., Kim, H.C., Kim, H.M., Park, S.W., and Kim, D.J. *A simple screening score for diabetes for the korean population development, validation, and comparison with other scores.* Diabetes Care. 2012; **35**(8): p. 1723-1730.
88. Li, J., Bornstein, S.R., Landgraf, R., and Schwarz, P.E. *Validation of a simple clinical diabetes prediction model in a middle-aged, white, german population.* Archives of internal medicine. 2007; **167**(22): p. 2528-2529.
89. Li, J., Bergmann, A., Reimann, M., Bornstein, S., and Schwarz, P. *A more simplified finnish diabetes risk score for opportunistic screening of undiagnosed type 2 diabetes in a german population with a family history of the metabolic syndrome.* Hormone and metabolic research. 2009; **41**(02): p. 98-103.
90. Li, C.-I., Chien, L., Liu, C.-S., Lin, W.-Y., Lai, M.-M., Lee, C.-C*., et al*. *Prospective validation of american diabetes association risk tool for predicting pre-diabetes and diabetes in taiwan–taichung community health study.* PloS one. 2011; **6**(10): p. e25906.
91. ADA. *Screening for type 2 diabetes.* Diabetes Care. 2004; **27**: p. S11.
92. Lin, J.-W., Chang, Y.-C., Li, H.-Y., Chien, Y.-F., Wu, M.-Y., Tsai, R.-Y*., et al*. *Cross-sectional validation of diabetes risk scores for predicting diabetes, metabolic syndrome, and chronic kidney disease in taiwanese.* Diabetes Care. 2009; **32**(12): p. 2294-2296.
93. Luo, S., Han, L., Zeng, P., Chen, F., Pan, L., Wang, S*., et al*. *A risk assessment model for type 2 diabetes in chinese.* PloS one. 2014; **9**(8): p. e104046.
94. Liu, M., Pan, C., and Jin, M. *A chinese diabetes risk score for screening of undiagnosed diabetes and abnormal glucose tolerance.* Diabetes technology & therapeutics. 2011; **13**(5): p. 501-507.
95. Lyssenko, V., Jonsson, A., Almgren, P., Pulizzi, N., Isomaa, B., Tuomi, T*., et al*. *Clinical risk factors, DNA variants, and the development of type 2 diabetes.* New England Journal of Medicine. 2008; **359**(21): p. 2220-2232.
96. Lyssenko, V., Jørgensen, T., Gerwien, R.W., Hansen, T., Rowe, M.W., McKenna, M.P*., et al*. *Validation of a multi-marker model for the prediction of incident type 2 diabetes mellitus: Combined results of the inter99 and botnia studies.* Diabetes and Vascular Disease Research. 2012; **9**(1): p. 59-67.
97. Kolberg, J.A., Jørgensen, T., Gerwien, R.W., Hamren, S., McKenna, M.P., 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**(7): p. 1207-1212.
98. Mainous, A.G., Diaz, V.A., and Everett, C.J. *Assessing risk for development of diabetes in young adults.* The Annals of Family Medicine. 2007; **5**(5): p. 425-429.
99. Makrilakis, K., Liatis, S., Grammatikou, S., Perrea, D., Stathi, C., Tsiligros, P*., et al*. *Validation of the finnish diabetes risk score (findrisc) questionnaire for screening for undiagnosed type 2 diabetes, dysglycaemia and the metabolic syndrome in greece.* Diabetes & metabolism. 2011; **37**(2): p. 144-151.
100. McNeely, M.J., Boyko, E.J., Leonetti, D.L., Kahn, S.E., and Fujimoto, W.Y. *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**(3): p. 758-763.
101. Mühlenbruch, K., Ludwig, T., Jeppesen, C., Joost, H.-G., Rathmann, W., Meisinger, C*., et al*. *Update of the german diabetes risk score and external validation in the german monica/kora study.* Diabetes research and clinical practice. 2014; **104**(3): p. 459-466.
102. Park, P., Griffin, S., Sargeant, L., and Wareham, N. *The performance of a risk score in predicting undiagnosed hyperglycemia.* Diabetes Care. 2002; **25**(6): p. 984-988.

103. Phillips, C.M., Kearney, P.M., McCarthy, V.J., Harrington, J.M., Fitzgerald, A.P., and Perry, I.J. *Comparison of diabetes risk score estimates and cardiometabolic risk profiles in a middle-aged irish population.* PloS one. 2013; **8**(11): p. e78950.
104. Rahman, M., Simmons, R.K., Harding, A.-H., Wareham, N.J., and Griffin, S.J. *A simple risk score identifies individuals at high risk of developing type 2 diabetes: A prospective cohort study.* Family practice. 2008; **25**(3): p. 191-196.
105. Rathmann, W., Martin, S., Haastert, B., Icks, A., Holle, R., Löwel, H.*, et al*. *Performance of screening questionnaires and risk scores for undiagnosed diabetes: The kora survey 2000.* Archives of internal medicine. 2005; **165**(4): p. 436-441.
106. Riaz, M., Basit, A., Hydrie, M.Z.I., Shaheen, F., Hussain, A., Hakeem, R.*, et al*. *Risk assessment of pakistani individuals for diabetes (rapid).* Primary care diabetes. 2012; **6**(4): p. 297-302.
107. Rolka, D.B., Narayan, K.V., Thompson, T.J., Goldman, D., Lindenmayer, J., Alich, K.*, et al*. *Performance of recommended screening tests for undiagnosed diabetes and dysglycemia.* Diabetes Care. 2001; **24**(11): p. 1899-1903.
108. Rowe, M.W., Bergman, R.N., Wagenknecht, L.E., and Kolberg, J.A. *Performance of a multi-marker diabetes risk score in the insulin resistance atherosclerosis study (iras), a multi-ethnic us cohort.* Diabetes/metabolism research and reviews. 2012; **28**(6): p. 519-526.
109. Ruige, J.B., de Neeling, J.N.D., Kostense, P.J., Bouter, L.M., and Heine, R.J. *Performance of an niddm screening questionnaire based on symptoms and risk factors.* Diabetes Care. 1997; **20**(4): p. 491-496.
110. Schmid, R., Vollenweider, P., Bastardot, F., Waeber, G., and Marques-Vidal, P. *Validation of 7 type 2 diabetes mellitus risk scores in a population-based cohort: Colaus study.* Archives of internal medicine. 2012; **172**(2): p. 188-189.
111. Schmid, R., Vollenweider, P., Bastardot, F., Vaucher, J., Waeber, G., and Marques-Vidal, P. *Current genetic data do not improve the prediction of type 2 diabetes mellitus: The colaus study.* The Journal of Clinical Endocrinology & Metabolism. 2012; **97**(7): p. E1338-E1341.
112. Spijkerman, A.M., Yuyun, M.F., Griffin, S.J., Dekker, J.M., Nijpels, G., and Wareham, N.J. *The performance of a risk score as a screening test for undiagnosed hyperglycemia in ethnic minority groups data from the 1999 health survey for england.* Diabetes Care. 2004; **27**(1): p. 116-122.
113. Stern, M.P., Williams, K., González-Villalpando, C., Hunt, K.J., and Haffner, S.M. *Does the metabolic syndrome improve identification of individuals at risk of type 2 diabetes and/or cardiovascular disease?* Diabetes Care. 2004; **27**(11): p. 2676-2681.
114. Stern, M., Williams, K., Eddy, D., and Kahn, R. *Validation of prediction of diabetes by the archimedes model and comparison with other predicting models.* Diabetes Care. 2008; **31**(8): p. 1670-1671.
115. Sun, F., Tao, Q., and Zhan, S. *An accurate risk score for estimation 5-year risk of type 2 diabetes based on a health screening population in taiwan.* Diabetes research and clinical practice. 2009; **85**(2): p. 228-234.
116. Tabaei, B.P. and Herman, W.H. *A multivariate logistic regression equation to screen for diabetes development and validation.* Diabetes Care. 2002; **25**(11): p. 1999-2003.
117. Talmud, P.J., Hingorani, A.D., Cooper, J.A., Marmot, M.G., Brunner, E.J., 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**: p. b4838.
118. Tankova, T., Chakarova, N., Atanassova, I., and Dakovska, L. *Evaluation of the finnish diabetes risk score as a screening tool for impaired fasting glucose, impaired glucose tolerance and undetected diabetes.* Diabetes research and clinical practice. 2011; **92**(1): p. 46-52.
119. Tuomilehto, J., Lindström, 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 research and clinical practice. 2010; **87**(2): p. 267-274.

120. 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).* Journal of diabetes science and technology. 2009; **3**(4): p. 748-755.
121. Wannamethee, S.G., Shaper, A.G., Lennon, L., and Morris, R.W. *Metabolic syndrome vs framingham risk score for prediction of coronary heart disease, stroke, and type 2 diabetes mellitus.* Archives of internal medicine. 2005; **165**(22): p. 2644-2650.
122. Witte, D., Shipley, M., Marmot, M., and Brunner, E. *Performance of existing risk scores in screening for undiagnosed diabetes: An external validation study.* Diabetic medicine. 2010; **27**(1): p. 46-53.
123. Zhang, L., Zhang, Z., Zhang, Y., Hu, G., and Chen, L. *Evaluation of finnish diabetes risk score in screening undiagnosed diabetes and prediabetes among us adults by gender and race: Nhanes 1999-2010.* PloS one. 2014; **9**(5): p. e97865.
124. Zhou, X., Qiao, Q., Ji, L., Ning, F., Yang, W., Weng, J.*, et al*. *Nonlaboratory-based risk assessment algorithm for undiagnosed type 2 diabetes developed on a nation-wide diabetes survey.* Diabetes Care. 2013; **36**(12): p. 3944-3952.

09

# Effect of model updating strategies on the performance of prevalent diabetes risk prediction models in mixed-ancestry population of South Africa

Katya Masconi, Tandi E. Matsha, Rajiv T. Erasmus and Andre P. Kengne

*Submitted*

## Abstract

**Background:** Many prediction models perform poorly when applied to populations different to the one in which they were developed. In an attempt to limit endless models development, model validation and updating methods to incorporate characteristics of the validation population into an existent models are encouraged. We assessed the impact of updating techniques on the accuracy of prevalent diabetes prediction models.

**Methods:** Data from the Cape Town Bellville South cohort served as the basis for this study. Model updating techniques and models were identified via recent systematic reviews. Models' discrimination was assessed and compared using C-statistic and non-parametric methods.

**Results:** The study sample consisted of 1256 individuals, of whom 173 were excluded due to previously diagnosed diabetes. Updating methods increased the discrimination of four models (Cambridge, Omani, Rotterdam and Simplified Finnish diabetes risk models), while the Kuwaiti Risk model's performance was maintained [C-statistic: 0.70 (0.66 – 0.74)]. Calibration was improved for all five models with logistic calibration and held throughout the remaining updating methods. Overall, the re-estimation of the Cambridge diabetes risk model yielded the best model performance [C-statistic: 0.71 (0.67 – 0.75); E/O: 1.00 (0.86 – 1.17)].

**Conclusions:** Model updating techniques increased both discrimination and calibration in varying levels across all five models. These methods can potentially be employed when validating an existent diabetes risk model in an effort to better fit the model to the validation population.

***Word count*** – 224

***Keywords*****:** risk, prediction, diabetes, update, performance

## Background

Updating methods aim to improve the prediction performance of a model in a new setting [1]. Prediction performance is often decreased when a model is tested in a population different to that in which the model was developed. To limit the number of models redeveloped in smaller datasets due to poor performance of existent models, the updating of model is encouraged. This allows for the information captured during the development of the model to be incorporated with characteristics of the validation population [1-5].

Several updating methods are available in the statistical literature [1, 4-6]. These methods vary in the extent to which the model is adjusted, and range from simple recalibration in which only the intercept of the model may be update, to more extensive updating where all the model parameters are re-estimated and new variables are considered. There is no advocated method to use, however the most extensively used approach in the updating of an existent risk prediction model, simple intercept correcting, does not account for the difference in strength of the individual variables in the validation population; while the re-estimation of the regression coefficients to replace unbiased estimates and fit the model to the validation outcome prevalence, can be unreliable [6]. In addition, model aggregation allows for the classical paradigm of model updating to be extended to allow for the use of evidence from multiple potentially useful models.

Updating methods are, however, not a remedy against poorly conceived and underpowered prediction research, nor do they guarantee complete bridging of the gaps due to large differences between development and validation datasets. How these methods alter the performance of the existent models during the validation of prevalent diabetes risk prediction in empirical data has not yet been investigated.

In this study, we applied the updating methods presented by Janssen *et al* [6], adapted from Steyerberg *et al* [7], and model averaging presented by Debray *et al* [8], in a validation dataset from South Africa, where population specific diabetes risk prediction models are not available. Existent models developed in vastly different populations were validated and updated with

multiple methods, in an attempt to document if performance can be improved enough to allow recommendation for use.

## Methodology

### Database

Details of the study design and recruitment of the database that served as the basis for all updating methods implementation are described below. The study was approved by the Ethics Committee of the Cape Peninsula University of Technology and Stellenbosch University.

### Research setting

Bellville South is located within the Northern suburbs of Cape Town, South Africa and is a traditionally a Coloured township formed in the late 1950s. According to the 2011 population census, its population stands at approximately 29 301 with 76.0% (22 270) consisting of mixed ancestry individuals [9, 10]. The target population for this study were subjects between the ages of 35 and 65 years and their number was estimated to be 6 500 in the 2001 population census [11].

### Research Design and Study Population

The data was collected during January 2008 to March 2009. Using a map of Bellville South, multistage stratified random sampling was approached as follows: From a list of streets from each stratum, the streets were then classified as short, medium and long streets, based on the number of houses. Streets with houses ≤ 22 were classified as short, medium; houses 23–40 and long streets were > 40 houses. A total of 16 short streets representing approximately 190 houses, 15 medium streets representing approximately 410 houses and 12 long streets representing approximately 400 houses, were randomly selected across the different strata. From the selected streets, all household members meeting the selection criteria were invited to participate in the study. Community authorities requested that participants outside the random selection area should benefit from the study.

**Recruitment Strategy**

Information regarding the project was disseminated to the local residents through the local radio station, community newspaper, brochures and fliers; the latter bearing information about the project and distributed through school children and taxis to the local residents by the recruitment team. Recruited subjects were visited by the recruitment team the evening before participation and reminded of all the survey instructions.

**Data collection**

A detailed protocol describing data-collection procedures (questionnaires and physical examination) was developed. The questionnaire designed to retrospectively obtain information on lifestyle factors such as smoking and alcohol consumption, physical activity, diet, family history of cardiovascular disease (CVD) and diabetes mellitus (DM), and demographics was administered by trained personnel. A detailed drug history was obtained by interrogation and by examining the clinic cards as well as the record of drugs that participants brought to the study site. Clinical measurements included height, weight, hip and waist circumferences, body fat measurements and blood pressure.

**Diabetes diagnosis**

All participants, except self-reported diabetic subjects, confirmed by either medical card record or drugs in use, had blood taken for fasting blood glucose and underwent a 75 g oral glucose tolerance test (OGTT) as prescribed by the World Health Organisation (WHO). Diabetes was diagnosed according to the WHO 2006 criteria [12].

**Identification of prevalent diabetes prediction models**

Existing prediction models were obtained from a systematic review by Brown *et al,* 2012 [13]. Models met the criteria for model selection for this paper if they were developed to predict the presence of undiagnosed diabetes based on predictors measured in the Bellville South study. We focused on models developed from non-invasively measure predictors. Therefore the models retained were: Cambridge Risk model [14], Kuwaiti Risk model [15], Omani Diabetes Risk model

[16], Rotterdam Predictive model 1 [17] and the simplified Finnish Diabetes Risk model [18]. Model characteristics, formulas and base performance in this dataset are available elsewhere [19]. All models included age as a predictor, while a range of other predictors were variably combined in models. These included: sex, body mass index (BMI), use of antihypertensive medication, family history of diabetes, waist circumference, past or current smoking and the use of corticosteroids.

**Statistical methods**

***Analysis of missing data***

The proportion of missing data for each variable was determined. Family history was the variable with the most missing data [mother (25.1%), father (24.9%), sister (25.0%), and brother (25.1%)]. The rest of the variables had a missing proportion of less than 5%, except smoking status (6.1%). During the comparison of several imputation methods on this dataset, simple imputation resulted in higher predictive utility and was hence used to handle missing data in this study, before the implementation of updating methods [20].

***Updating methods***

Updating methods ranged in extent to which the original model is altered and in the dataset requirements [6, 7, 21]. This study naturally did not have access to the development datasets of the validated prevalent diabetes risk prediction models, therefore excluding updating methods requiring the merging of both development and validation datasets. Updating methods which required the addition of variables were also not considered as the aim of this study was to focus on updating methods using the core structure of existing models in a new setting in an attempt to compare the change in model performance. The models were initially run without adjustment, termed Method 0. This is termed the 'reference method' to which all other methods were compared. The updating techniques explained by Janssen *et al* [6] were used to update the prevalent diabetes risk prediction models in this study. Methods 1 and 2 refer to recalibration. Method 1 updated only the intercept using a correction factor to correct for the difference in disease prevalence between the development and validation population. Method 2 updated

both the intercept and the regression coefficients of the variables using the intercept and calibration slope from Method 1 respectively. Method 3 and 4 are more comprehensive revision methods. Method 3 tested weather the effect of each variable is different in the updating dataset, following the calibration of Method 2. Variables were individually added as an offset, calculating a deviation from the recalibrated regression coefficient based on Method 2. Likelihood ratio tests were used to test whether this deviation has added predictive value. The deviation was added to the regression coefficients of variables with statistically significant differences. Finally, Method 4 was the complete re-estimation of the intercept and the regression coefficients, fitting the variables from the original models in the validation dataset. All novel modelling used the *lrm* function of the R package '*rms'*, which fits binary logistic regression models using the maximum likelihood estimation. The '*lmtest'* package was used to perform likelihood ratio tests in Method 3.

***Model aggregation and development***

As a reference for comparison of the model performance of each of the updating techniques, model aggregation and development was done. Model averaging combines the predictions from updated literature models by means of a weighted average. A meta-model that combines the best performing models was developed, using a Bayesian model averaging (BMA) method adapted by Debray *et al* [8]. A logistic regression prevalent diabetes risk prediction model was developed using forward stepwise regression. To reduce model overfitting, lasso regression was used to shrink the regression coefficients with the function *penalized* in the R package '*penalized'*.

***Model performance***

The original selected models were validated for the overall data and subsets using the formulas, without any recalibration. The predicted probability of undiagnosed diabetes for each participant was computed using the baseline measured predictors. The performance was expressed in terms of discrimination and calibration. Discrimination describes the ability of the model's performance in distinguishing those at a high risk of developing diabetes from those at low risk [22]. The

discrimination was assessed and compared using concordance (C) statistic and non-parametric methods [23].

Calibration describes the agreement between the probability of the outcome of interest as estimated by the model, and the observed outcome frequencies [1]. It was assessed by computing the expected (E) over observed (O) ratio (E/O); with the 95% confidence intervals calculated assuming a Poisson distribution [24]. We also calculated 1) the Yates slope, which is the difference between mean predicted probability of type 2 diabetes for participants with and without prevalent undiagnosed diabetes, with higher values indicating better performance; and 2) the Brier score, which is the squared difference between predicted probability and actual outcome for each participant with values ranging between 0 for a perfect prediction model and 1 for no match in prediction and outcome [1, 22]. The R package '*rms*' and the '*pROC*' package were used for model performance measures.

## Results

### Updating dataset

The study sample consisted of 1256 individuals, of whom 173 were excluded due to previously diagnosed diabetes. Of the final 1083 individuals, 329 (30.4%) had missing data, which were imputed using simple imputation. The baseline profile for men and women included in the study is described in Table 1. The mean age was 51.9 (14.9) years and a total of 162 (15%) individuals had undiagnosed diabetes. The database was made up of 832 (76.8%) females. A comparison between the genders showed a significant difference for BMI, systolic and diastolic blood pressure, waist circumference, smoking status and a sibling with a history of diabetes.

### Models parameters

Supplementary Table 1 shows the estimates of the main parameters of the various updating methods. The regression model intercept and regression coefficients of all the variables, per updated model, are presented in supplementary Table 2. In Method 1, too highly predicted risks

by the Cambridge and Omani diabetes risk models in the updating dataset required the intercept to be decreased, while the predicted risk was too low and the intercept increased for the Kuwaiti, Rotterdam and Simplified Finnish diabetes risk models. Method 2 shows additional adjustment to the intercept of all models. The Cambridge and Omani diabetes risk prediction models were still predicting too high a risk and the intercepts were further reduced, both by more than the initial adjustment [Cambridge: -0.415, - 0.986; Omani: -0.308, -0.752]. The Rotterdam diabetes risk prediction model was also further reduced, however only by 0.007. However, on the contrary to Method 1, the remaining models required the predicted risk to be lowered, as opposed to following the calibration to increase the predicted risk in Method 1. The regression coefficients of the original models are multiplied by the calibration slopes from method 0. All models required the weighting of the variables to be decreased, with the exception of the Rotterdam predictive model which had a calibration slope of 1.067.

Table 1: Characteristics of the updating dataset, following simple imputation

| Variables | | Updating dataset | | |
|---|---|---|---|---|
| | Overall (1083) | Male (251) | Female (832) | P value |
| Prevalent undiagnosed diabetes (Yes) | 162 (15.0) | 28 (11.2) | 134 (16.1) | 0.068 |
| Age (years) | 51.9 (14.9) | 53.8 (16.1) | 51.3 (14.5) | 0.031 |
| Body mass index (kg/m2) | 29.7 (7.0) | 26.0 (5.9) | 30.9 (6.9) | <0.001 |
| Systolic blood pressure (mmHg) | 124.3 (20.0) | 127.5 (19.4) | 123.3 (20.1) | 0.003 |
| Diastolic blood pressure (mmHg) | 76.0 (12.7) | 77.6 (13.9) | 75.5 (12.3) | 0.035 |
| Waist circumference (cm) | 95.8 (15.3) | 92.7 (14.5) | 96.7 (15.4) | < 0.001 |
| Hypertensive medication (Yes) | 374 (34.5) | 74 (29.5) | 300 (36.1) | 0.065 |
| Using corticosteroids (Yes) | 12 (1.1) | 2 (0.8) | 10 (1.2) | 0.847 |
| Mother having diabetes (Yes) | 124 (11.5) | 20 (8.0) | 104 (12.5) | 0.062 |
| Father having diabetes (Yes) | 61 (5.6) | 15 (6.0) | 46 (5.5) | 0.910 |
| Sister having diabetes (Yes) | 103 (9.5) | 12 (4.8) | 91 (10.9) | 0.005 |
| Brother having diabetes (Yes) | 67 (6.2) | 9 (3.6) | 58 (7.0) | 0.072 |
| Smoking status (Current) | 433 (40.0) | 123 (49.0) | 310 (37.3) | 0.001 |

The likelihood ratio test results from Method 3 showed significantly different effects in at least one variable across all the models, except the Simplified Finnish diabetes risk model (Supplementary Table 1). Additionally, a calibration slope of 1, from Method 2 for the Simplified Finnish diabetes risk model, resulted in no change in the regression coefficients, and an intercept of 0 meant no change in the intercept of the final model for Method 3. The remaining Method 2 calibration slopes were all above 1, varying in the increasing effect on the linear predictor. The intercept adjustments for Method 3 followed the pattern from Method 1 to 2 for Cambridge, Kuwaiti and Omani diabetes risk prediction models, where the risk was lowered even further. The Rotterdam diabetes risk prediction model intercept was slightly lowered from -2.311 to -2.240. Finally, the re-estimation of the models (Method 4) yielded an intercept closer to 0 (when compared to the original model) for all the models. The coefficients showed no pattern of increase or decrease across the models, however comparison of the final intercepts and regression coefficients of the updated models showed variability, Supplementary Table 2.

**Model performance**

Table 2 shows that the model performance in the updating dataset. As expected recalibration Methods 1 and 2 did not change the discriminative ability of the models. The Kuwaiti diabetes risk model maintained a C-statistic of 0.70 (0.66 – 0.74) throughout all the updating methods. Method 3 increased the C-statistic for the remaining models, excluding the Simplified Finnish diabetes risk model which remained at 0.66 (0.62 – 0.70), while the re-estimation resulted in the highest C-statistic for the Omani [0.70 (0.65 – 0.74)] and Simplified Finnish diabetes risk models [0.68 (0.63 – 0.72)]. Method 4 was able to increase the discrimination for the Omani, Simplified Finnish and Cambridge diabetes risk models, while holding the discrimination for the Rotterdam diabetes risk model [0.67 (0.62 – 0.72)]. The optimal calibration results were reached with Method 2 [E/O: 1.00 (0.86 – 1.17)] for all the models and was maintained through to Method 4. Both the final model developed through model averaging and complete model development matched the performance of the model re-estimation of the Cambridge model [C-statistic: 0.71 (0.67 – 0.75)]. Both models' parameters and performance is shown in Supplementary Table 3.

Table 2: Overview of the performance of the prevalent diabetes risk prediction models across the updating techniques

| Models | | Original (0) | Intercept adjustment (1) | Logistic calibration (2) | Revision (3) | Revision (4) |
|---|---|---|---|---|---|---|
| Cambridge Diabetes Risk Model | E/O (95% CI) | 0.48 (0.41 ; 0.56) | 0.78 (0.67 ; 0.91) | 1.00 (0.86 ; 1.17) | 1.00 (0.86 ; 1.17) | 1.00 (0.86 ; 1.17) |
| | Brier score | 0.181 | 0.140 | 0.120 | 0.119 | 0.117 |
| | Yates slope | 0.174 | 0.135 | 0.060 | 0.066 | 0.081 |
| | C-statistic (95% CI) | 0.69 (0.65 – 0.73) | 0.69 (0.65 – 0.73) | 0.69 (0.65 – 0.73) | 0.70 (0.66 – 0.74) | 0.71 (0.67 – 0.75) |
| Kuwaiti Risk model | E/O (95% CI) | 1.27 (1.09 ; 1.48) | 1.04 (0.90 ; 1.22) | 1.00 (0.86 ; 1.17) | 1.00 (0.86 ; 1.17) | 1.00 (0.86 ; 1.17) |
| | Brier score | 0.122 | 0.123 | 0.118 | 0.118 | 0.118 |
| | Yates slope | 0.097 | 0.111 | 0.070 | 0.075 | 0.074 |
| | C-statistic (95% CI) | 0.70 (0.66 – 0.74) | 0.70 (0.66 – 0.74) | 0.70 (0.66 – 0.74) | 0.70 (0.66 – 0.74) | 0.70 (0.66 – 0.74) |
| Omani Diabetes Risk model | E/O (95% CI) | 0.70 (0.60 ; 0.82) | 0.92 (0.79 ; 1.08) | 1.00 (0.86 ; 1.17) | 1.00 (0.86 ; 1.17) | 1.00 (0.86 ; 1.17) |
| | Brier score | 0.142 | 0.132 | 0.122 | 0.121 | 0.119 |
| | Yates slope | 0.110 | 0.090 | 0.047 | 0.053 | 0.062 |
| | C-statistic (95% CI) | 0.67 (0.63 – 0.71) | 0.67 (0.63 – 0.71) | 0.67 (0.63 – 0.71) | 0.68 (0.64 – 0.72) | 0.70 (0.65 – 0.74) |
| Rotterdam Predictive model | E/O (95% CI) | 1.62 (1.38 ; 1.88) | 1.01 (0.87 ; 1.18) | 1.00 (0.86 ; 1.17) | 1.00 (0.86 ; 1.17) | 1.00 (0.86 ; 1.17) |
| | Brier score | 0.126 | 0.123 | 0.122 | 0.120 | 0.120 |
| | Yates slope | 0.024 | 0.035 | 0.037 | 0.055 | 0.055 |
| | C-statistic (95% CI) | 0.66 (0.61 – 0.70) | 0.65 (0.61 – 0.70) | 0.66 (0.61 – 0.70) | 0.67 (0.63 – 0.72) | 0.67 (0.62 – 0.72) |
| Simplified Finnish Diabetes Risk model | E/O (95% CI) | 2.92 (2.51 ; 3.41) | 1.09 (0.93 ; 1.27) | 1.00 (0.86 ; 1.17) | 1.00 (0.86 ; 1.17) | 1.00 (0.86 ; 1.17) |
| | Brier score | 0.133 | 0.125 | 0.122 | 0.122 | 0.121 |
| | Yates slope | 0.026 | 0.063 | 0.041 | 0.041 | 0.051 |
| | C-statistic (95% CI) | 0.66 (0.62 – 0.70) | 0.66 (0.62 – 0.70) | 0.66 (0.62 – 0.70) | 0.66 (0.62 – 0.70) | 0.68 (0.64 – 0.72) |

## Discussion

The aim of this study was to compare the effects of different updating techniques on the performance of existent diabetes risk prediction models. The performance of the original models was not sufficient to recommend implementation and the updating methods were intended to better fit a model to the new setting, therefore improving performance. Discrimination was increased slightly across four of the five models with full model re-estimation, and calibration was significantly improved. To aid discussion of the model performance across the updating methods, model averaging and development was undertaken. Model re-estimation of the Cambridge diabetes risk model achieved the highest accuracy in predicting undiagnosed diabetes, when compared to other updating methods.

The over or under estimated prediction of risk models in a new settings may often be due to variables of characteristics that are not incorporated into the model, but do have an effect on the final model parameters. With large disparities between the development and updating populations, as in this study, simple recalibration methods (Methods 1 and 2) are not anticipated to be able to fully adjust for the variability. The total re-estimation in the updating dataset (Method 4) is often undertaken in this situation, however revision methods with more simple adjustments (Method 3) can also achieve the incorporation of this new information in the model. Although discrimination was improved across the models, this was slight. Calibration was largely improved, with Method 2, 3 and 4 being able to achieve the best agreement between predicted and observed outcome frequencies. Comparison of the final intercepts and regression coefficients of the updated models showed large variability across the methods, highlighting the different effect each method had on the model parameters and the predictive ability of the variables. In large, Method 4 showed the best adjustment to the disparity between development and validation population. This was only matched by model averaging methods and the model developed through stepwise regression.

The results of this study allow us to expand on the possible limitations. Two updating methods were excluded from the methodology, one was not possible (merging of the development and updating datasets), while the other was the retention of the original model predictors while

assessing the added value of new predictors. This method was omitted as the aim was to keep the core structure of the original models and this is a far more extensive model revision method, however, it is the predictive effect of predictors that were not considered that may have increased model calibration and discrimination. Additionally, a limitation on model averaging must be mentioned. The final meta-model may include multiple classes and categories of the same base variable due to the nature of including all the variables from the highly weighted models. It is of note that model updating/revision results in new models, which must undergo external validation before recommendation for implementation in the new population. This extra step was deemed unnecessary in the current study, considering magnitude of improvement in models' discrimination which was only modest.

In conclusion, extensive updating methods on models validated in empirical data were superior over more simple methods. Total model re-estimation achieved good calibration while also increasing the discrimination of the model. Diabetes risk prediction research in Africa is poorly developed, and the largely diverse population setting makes existent models possibly too different for even the most complex of updating methods. Model validation and updating is advocated to prevent added models to the already saturated literature. The increase in model performance and comparison to model development of this study support model updating strategies in the role of diabetes risk prediction research.

## Acknowledgments

KLM was supported by a scholarship from the South African National Research Foundation and the Carl & Emily Fuchs Foundation.

## References


1. Steyerberg, E.W. *Clinical prediction models*. 2009, New York: Springer.
2. Poses, R.M., Cebul, R.D., Collins, M., and Fager, S.S. *The importance of disease prevalence in transporting clinical prediction rules: The case of streptococcal pharyngitis.* Annals of Internal Medicine. 1986; **105**(4): p. 586-591.
3. Wigton, R.S., Connor, J.L., and Centor, R.M. *Transportability of a decision rule for the diagnosis of streptococcal pharyngitis.* Archives of internal medicine. 1986; **146**(1): p. 81-83.
4. Hosmer Jr, D.W. and Lemeshow, S. *Applied logistic regression*. 2004, New Jersey: John Wiley & Sons.
5. Ivanov, J., Tu, J.V., and Naylor, C.D. *Ready-made, recalibrated, or remodeled? Issues in the use of risk indexes for assessing mortality after coronary artery bypass graft surgery.* Circulation. 1999; **99**(16): p. 2098-2104.
6. Janssen, K., Moons, K., Kalkman, C., Grobbee, D., and Vergouwe, Y. *Updating methods improved the performance of a clinical prediction model in new patients.* Journal of clinical epidemiology. 2008; **61**(1): p. 76-86.
7. Steyerberg, E.W., Borsboom, G.J., van Houwelingen, H.C., Eijkemans, M.J., and Habbema, J.D.F. *Validation and updating of predictive logistic regression models: A study on sample size and shrinkage.* Statistics in medicine. 2004; **23**(16): p. 2567-2586.
8. Debray, T., Koffijberg, H., Nieboer, D., Vergouwe, Y., Steyerberg, E.W., and Moons, K.G. *Meta-analysis and aggregation of multiple published prediction models.* Statistics in medicine. 2014; **33**(14): p. 2341-2362.
9. Quintana-Murci, L., Harmant, C., Quach, H., Balanovsky, O., Zaporozhchenko, V., Bormans, C*., et al*. *Strong maternal khoisan contribution to the south african coloured population: A case of gender-biased admixture.* The American Journal of Human Genetics. 2010; **86**(4): p. 611-620.
10. CoCT. *2011 census – ward 009*. 2013, City of Cape Town: Cape Town.
11. CoCT. *2001 census – ward 009*. 2006, City of Cape Town: Cape Town.
12. World Health Organization. *Screening for type 2 diabetes: Report of a world health organization and international diabetes federation meeting.* 2003.
13. Brown, N., Critchley, J., Bogowicz, P., Mayige, M., and Unwin, N. *Risk scores based on self-reported or available clinical data to detect undiagnosed type 2 diabetes: A systematic review.* Diabetes research and clinical practice. 2012; **98**(3): p. 369-385.
14. Griffin, S., Little, P., Hales, C., Kinmonth, A., and Wareham, N. *Diabetes risk score: Towards earlier detection of type 2 diabetes in general practice.* Diabetes/metabolism research and reviews. 2000; **16**(3): p. 164-171.
15. Al Khalaf, M., Eid, M., Najjar, H., Alhajry, K., Doi, S., and Thalib, L. *Screening for diabetes in kuwait and evaluation of risk scores.* EMHJ. 2010; **16**(7): p. 725-731.
16. Al-Lawati, J. and Tuomilehto, J. *Diabetes risk score in oman: A tool to identify prevalent type 2 diabetes among arabs of the middle east.* Diabetes research and clinical practice. 2007; **77**(3): p. 438-444.
17. Baan, C.A., Ruige, J.B., Stolk, R.P., Witteman, J., Dekker, J.M., Heine, R.J*., et al*. *Performance of a predictive model to identify undiagnosed diabetes in a health care setting.* Diabetes Care. 1999; **22**(2): p. 213-219.
18. Bergmann, A., Li, J., Wang, L., Schulze, J., Bornstein, S., and Schwarz, P. *A simplified finnish diabetes risk score to predict type 2 diabetes risk and disease evolution in a german population.* Hormone and metabolic research. 2007; **39**(09): p. 677-682.

19. Masconi, K., Matsha, T.E., Erasmus, R.T., and Kengne, A.P. *Independent external validation and comparison of prevalent diabetes risk prediction models in a mixed-ancestry population of south africa.* Diabetology & Metabolic Syndrome. 2015; **7**(42).
20. Masconi, K.L., Matsha, T.E., Erasmus, R.T., and Kengne, A.P. *Effects of different missing data imputation techniques on the performance of undiagnosed diabetes risk prediction models in a mixed-ancestry population of south africa.* PloS one. 2015; **10**(9): p. e0139210.
21. Toll, D., Janssen, K., Vergouwe, Y., and Moons, K. *Validation, updating and impact of clinical prediction rules: A review.* Journal of clinical epidemiology. 2008; **61**(11): p. 1085-1094.
22. Kengne, A.P., Masconi, K., Mbanya, V.N., Lekoubou, A., Echouffo-Tcheugui, J.B., and Matsha, T.E. *Risk predictive modelling for diabetes and cardiovascular disease.* Critical reviews in clinical laboratory sciences. 2013; **51**(1): p. 1-12.
23. Moons, K.G., Kengne, A.P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D.G.*, et al*. *Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio) marker.* Heart. 2012; **98**(9): p. 683-690.
24. Dobson, A.J., Kuulasmaa, K., Eberle, E., and Scherer, J. *Confidence intervals for weighted sums of poisson parameters.* Statistics in medicine. 1991; **10**(3): p. 457-462.

Supplementary Table 1: Estimated parameters of the updating methods 1-3 and shrinkage factor for updating methods 3

| | | Method 1 | Method 2 | Method 3 |
|---|---|---|---|---|
| Cambridge Diabetes Risk model | Correction factor (1) / Calibration intercept (2-4) | -0.939 | - 1.401 | -0.387 |
| | Calibration slope | - | 0.415 | 1.015 |
| | Deviation from recalibration regression coefficient | | | |
| | Female gender | - | - | 0.520 |
| Kuwaiti Risk model | Correction factor (1) / Calibration intercept (2-4) | 0.276 | -0.459 | 0.944 |
| | Calibration slope | - | 0.555 | 1.366 |
| | Deviation from recalibration regression coefficient | | | |
| | Waist circumference > 100 cm | - | - | -0.647 |
| Omani Diabetes Risk model | Correction factor (1) / Calibration intercept (2-4) | -0.434 | -1.060 | -0.352 |
| | Calibration slope | - | 0.467 | 1.091 |
| | Deviation from recalibration regression coefficient | | | |
| | WC ≥ 94cm in men and ≥ 80cm in women | - | - | 0.723 |
| | Parent or sibling history of diabetes | - | - | -0.385 |
| Rotterdam Predictive model | Correction factor (1) / Calibration intercept (2-4) | 0.544 | 0.709 | 0.521 |
| | Calibration slope | - | 1.067 | 1.195 |
| | Deviation from recalibration regression coefficient | | | |
| | Male gender | - | - | -0.879 |
| Simplified Finnish Diabetes Risk model | Correction factor (1) / Calibration intercept (2-4) | 1.183 | -0.021 | 0.000 |
| | Calibration slope | - | 0.535 | 1.000 |
| | Deviation from recalibration regression coefficient | | | |
| | - | - | - | - |

* Method 1: correction factor updated intercept; Method 2: both the intercept and the regression coefficients of the variables using the intercept and calibration slope from Method 1; Method 3: Extra adjustment of predictors with a different effect in the updating set compared to the derivation set, after recalibration by Method 2

Supplementary Table 2: Intercept and regression coefficients of the updated models per existing model updated

| | | Method 0 | Method 1 | Method 2 | Method 3 | Method 4 (p-value) |
|---|---|---|---|---|---|---|
| Cambridge Diabetes Risk model | Intercept | -6.322 | -6.737 | -7.723 | -8.224 | -4.966 |
| | Female gender | -0.879 | -0.879 | -0.365 | 0.150 | 0.025 (0.309) |
| | Prescribed antihypertensive medication | 1.222 | 1.222 | 0.507 | 0.515 | 0.412 (0.035) |
| | Prescribed steroids | 2.191 | 2.191 | 0.910 | 0.924 | -0.054 (0.947) |
| | Age | 0.063 | 0.063 | 0.026 | 0.026 | 0.036 (<0.001) |
| | 25 kg/m$^2$ ≤ BMI ≥ 27.49 kg/m$^2$ | 0.699 | 0.699 | 0.290 | 0.294 | 0.471 (0.166) |
| | 27.5 kg/m$^2$ ≤ BMI ≤ 29.99 kg/m$^2$ | 1.970 | 1.970 | 0.818 | 0.830 | 0.369 (0.248) |
| | BMI ≥ 30 kg/m$^2$ | 2.518 | 2.518 | 1.046 | 1.062 | 0.898 (0.001) |
| | Parent or sibling has diabetes | 0.728 | 0.728 | 0.303 | 0.308 | 0.358 (0.099) |
| | Parent and sibling has diabetes | 0.753 | 0.753 | 0.313 | 0.318 | 0.779 (0.027) |
| | Ex-smoker | -0.218 | -0.218 | -0.091 | -0.092 | 0.272 (0.355) |
| | Current smoker | 0.855 | 0.855 | 0.355 | 0.360 | 0.406 (0.052) |
| Kuwaiti Risk model | Intercept | -5.018 | -4.634 | -5.477 | -6.538 | -3.837 |
| | Sibling history of diabetes | 0.979 | 0.979 | 0.544 | 0.743 | 0.791 (<0.001) |
| | Prescribed antihypertensive medication | 0.978 | 0.978 | 0.543 | 0.742 | 0.617 (0.001) |
| | Age ≥ 35 years | 1.315 | 1.315 | 0.730 | 0.997 | 1.406 (0.007) |
| | Waist circumference > 100 cm | 1.930 | 1.930 | 1.071 | 0.816 | 0.824 (<0.001) |
| Omani Diabetes Risk model | Intercept | -4.700 | -5.008 | -5.760 | -6.366 | -3.902 |
| | 40 years ≤ age ≤ 59 years | 1.800 | 1.800 | 0.840 | 0.916 | 0.973 (0.003) |
| | Age ≥ 60 years | 2.300 | 2.300 | 1.073 | 1.170 | 1.476 (<0.001) |
| | WC ≥ 94cm in men and ≥ 80cm in women | 0.380 | 0.380 | 0.177 | 0.916 | 1.034 (0.005) |
| | 25 kg/m$^2$ ≤ BMI < 30 kg/m$^2$ | 0.540 | 0.540 | 0.252 | 0.275 | 0.143 (0.658) |
| | BMI ≥ 30 kg/m$^2$ | 0.690 | 0.690 | 0.322 | 0.351 | 0.294 (0.337) |
| | Parental or sibling history of diabetes | 1.900 | 1.900 | 0.887 | 0.583 | 0.519 (0.006) |
| | SBP≥140 and/or DBP≥90 | 0.730 | 0.730 | 0.341 | 0.372 | 0.064 (0.753) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Rotterdam Predictive model | Intercept | -3.020 | -2.318 | -2.311 | -2.240 | -2.474 |
| | Age per 5 year increment from 55 years to >75 | 0.190 | 0.190 | 0.203 | 0.243 | 0.249 (<0.001) |
| | Male gender | 0.460 | 0.460 | 0.491 | -0.292 | -0.303 (0.196) |
| | Prescribed antihypertensive medication | 0.420 | 0.420 | 0.448 | 0.535 | 0.536 (0.004) |
| | BMI ≥ 30 kg/m$^2$ | 0.510 | 0.510 | 0.544 | 0.650 | 0.623 (0.001) |
| Simplified Finnish Diabetes Risk model | Intercept | -5.514 | -4.015 | -5.535 | -5.535 | -3.010 |
| | 45 years ≤ age ≤ 54 years | 0.628 | 0.628 | 0.336 | 0.336 | -0.056 (0.807) |
| | 55 years ≤ age ≤ 64 years | 0.892 | 0.892 | 0.478 | 0.478 | 0.481 (0.021) |
| | 25 kg/m$^2$ ≤ BMI < 30 kg/m$^2$ | 0.165 | 0.165 | 0.088 | 0.088 | -0.430 (0.200) |
| | BMI > 30 kg/m$^2$ | 1.096 | 1.096 | 0.587 | 0.587 | -0.133 (0.703) |
| | 94cm ≤ WC < 102cm in men<br>80cm ≤ WC < 88cm in women | 0.857 | 0.857 | 0.459 | 0.459 | 0.965 (0.015) |
| | WC ≥ 102cm in men and ≥ 88cm in women | 1.350 | 1.350 | 0.723 | 0.723 | 1.369 (0.001) |
| | Prescribed antihypertensive medication | 0.711 | 0.711 | 0.381 | 0.381 | 0.637 (<0.001) |
| | History of high blood glucose, assumed to be 0 for all participants due to the nature of this study | - | - | - | - | - |

* Method 0: original risk model; Method 1: correction factor updated intercept; Method 2: both the intercept and the regression coefficients of the variables using the intercept and calibration slope from Method 1; Method 3: Extra adjustment of predictors with a different effect in the updating set compared to the derivation set, after recalibration by Method 2; Method 4: complete re-estimation of the intercept and the regression coefficients, fitting the variables from the original models in the validation dataset

Supplementary Table 3: Intercept and regression coefficients of averaged and developed models

| | | Estimates | P value | Model performance | |
|---|---|---|---|---|---|
| Averaged model | Intercept | -4.903 | | E/O (95% CI) | 1.00 (0.86 ; 1.17) |
| | Female gender | 0.233 | < 0.001 | Brier score | 0.117 |
| | Prescribed antihypertensive medication | 0.418 | < 0.001 | Yates slope | 0.079 |
| | Prescribed steroids | -0.052 | < 0.001 | C-statistic (95% CI) | 0.71 (0.67 – 0.75) |
| | Age in years | 0.041 | < 0.001 | | |
| | 25 kg/m$^2$ ≤ BMI ≥ 27.49 kg/m$^2$ | 0.448 | < 0.001 | | |
| | 27.5 kg/m$^2$ ≤ BMI ≤ 29.99 kg/m$^2$ | 0.350 | < 0.001 | | |
| | BMI ≥ 30 kg/m$^2$ | 0.855 | < 0.001 | | |
| | Parent or sibling has diabetes | 0.343 | < 0.001 | | |
| | Parent and sibling has diabetes | 0.740 | < 0.001 | | |
| | Ex-smoker | 0.258 | < 0.001 | | |
| | Current smoker | 0.385 | < 0.001 | | |
| | Sibling history of diabetes | 0.036 | < 0.001 | | |
| | Age ≥ 35 years | 0.063 | < 0.001 | | |
| | Waist circumference > 100 cm | 0.036 | < 0.001 | | |
| | 40 years ≤ age ≤ 59 years | 0.0008 | < 0.001 | | |
| | Age ≥ 60 years | 0.002 | 0.076 | | |
| | WC ≥ 94cm in men and ≥ 80cm in women | 0.006 | < 0.001 | | |
| | SBP≥140 and/or DBP≥90 | 0.001 | 0.028 | | |
| Developed model | Intercept | -3.955 | | E/O (95% CI) | 1.00 (0.86 ; 1.17) |
| | Age in years | 0.034 | < 0.001 | Brier score | 0.118 |
| | Sibling history of diabetes | 0.296 | < 0.001 | Yates slope | 0.062 |
| | Waist circumference > 100 cm | 0.655 | < 0.001 | C-statistic (95% CI) | 0.71 (0.67 – 0.75) |
| | Prescribed antihypertensive medication | 0.096 | < 0.001 | | |

* Averaged model: Bayesian model averaging to combine all five prediction models to form a meta model. BMI categories from Omani diabetes risk model removed from averaged model due to collinearity: '25 kg/m$^2$ ≤ BMI < 30 kg/m$^2$' and 'BMI ≥ 30 kg/; Developed model: stepwise regression with shrinkage to develop to new model

# 10

# Discussion

The worldwide burden of diabetes can be reduced with prevention, diagnosis and treatment strategies. Targeting groups with higher prevalence of diabetes for screening, by utilizing potential risk factors for diabetes that are population specific, will enhance diagnosis, treatment and prevent complications and lower comorbidities. However, before these end benefits of screening can be achieved, the chosen method of screening needs to be effective in the target population. Screening in developing countries will never be considered without appropriate affordable methods, and the identification of possible methodological risks involved in the development and implementation of the advocated method. Clinical prediction models are developed to facilitate prognostic or diagnostic probability estimations in daily medical practice [1-4]. These have been advocated as valuable diabetes screening tools. Diabetes risk prediction models are typically developed by associating multiple risk factors with the outcome in a development dataset [1-3, 5]. Well known examples of diabetes risk prediction models include the Cambridge diabetes risk model [6] and the simplified FINDRISC model [7]. However these models are specific to the populations in which they were developed in and accuracy is not guaranteed in the target population to be screened. Chapters 2 and 3 provide the details of the background information on the outcome, screening and the use of the risk prediction models to achieve diabetes screening.

In acknowledgement of the vast differences in populations around the world, and to reduce the requirement of developing a model for each population and sub-population group, the value of a prediction model depends on its performance outside of the development sample [8-12]. This implies that the predictive performance of the model should remain sufficiently accurate across new samples from the same development population group (model reproducibility) or from different but related target population (model transportability) [10, 11, 13]. This ability to uphold model performance across multiple populations refers to model generalizability and is commonly assessed in external validation studies [1, 3-5, 8-11, 14]. Validation studies aim to quantify the predictive performance of a previously developed model in individuals that were not used to develop the model. The extent to which these studies move away from the development population characteristics, and in the case of good model performance being upheld, the transportability of the model is increased [8, 15]. The

differences in development and validation settings can range from temporal or geographical differences, to validations across different medical settings with increasingly different case-mix and discrepancies in predictor and outcome definitions.

The transportability of a model therefore relates to what extent the differences in the development and the validation population affect the model performance. This is where the research community comes together to find a balance between the differences that are too great to allow for possible accurate outcome prediction and a new model will need to be developed, while still encouraging the validation of existing models. External validation of diabetes prediction models is widely advocated, particularly when there are a large number of existing models available which have been developed, and validated, in many population groups. However before validation can be fully appreciated, importantly, when there is a lack of diabetes risk prediction research in the validation population, the methodological issues around the validation of risk prediction models needs to be addressed.

Unfortunately, there is a lack of clear guidelines for the process of model validation, particularly when reporting or publishing. The transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement has recently been published in an attempt to improve the reporting of prediction modelling studies of all types. However, although comprehensive, and certainly filling the needed for all-inclusive guidelines in this field, this publication is still too new to be both fully circulated or implemented [16]. As a consequence, it is has often been unclear to what extent researchers should do or not do regarding these methodological issues when undertaking the validation of a risk prediction model. This has impeded transparent interpretation of results from external validation studies, which may add to the circle of continuous model development in small datasets. This dissertation therefore aimed to validate existing prevalent diabetes risk prediction models in a mixed-ancestry population in South Africa, which has yet to be done; while investigating and describing methodological issues encountered during the validation process. This is presented through a number of published articles, from the initial validation of existing

models to the final updating techniques employed in an attempt to increase model performance.

The results of this thesis enable the validation of existing diabetes risk prediction models to be introduced in Sub-Saharan Africa, which was previously lacking, and encompasses five main findings, through part II, III and IV. Chapter 5 presents the basic external validation of five selected models, which were not developed in Africa, but rather the UK, Germany, Netherlands, Oman and Kuwaiti. While the original model performance in the development dataset was good for the Cambridge, Omani and Kuwaiti diabetes risk models (C-statistic greater than 0.80), and acceptable for the FINDRSIC model (C-statistic greater than 0.70), it was fairly low for the Simplified Rotterdam risk model. To our knowledge, this is the largest and most comprehensive validation study of prevalent diabetes prediction models in a sub-Saharan African population and with the exception of the Kuwaiti model [17], all other models assessed in this study have been validated externally elsewhere. Our first main result from this study was the presentation of the poor performance of these models in this mixed-ancestry population, explained by the lack of transportability of these models to this new setting. At the optimal probability thresholds, the best performing model would correctly classify only about 2/3rds of the population. While overall, validation studies showed a drop in model performance when tested in a new population, the C-statistic of all the models either mirrored or underperformed in this population when compared to the other external validation results. Unfortunately, with no model development in the mixed ancestry population of South Africa, selection of generalizable models for validation was limited. The poor performance of these models in their original form was therefore in no way surprising and has opened the door for further prediction research.

Dataset quality is an important aspect of prediction research. The clear definitions of the variables and outcomes, the uniform collection of these within the study and the comprehensiveness of the dataset all play an important role in the validity of the study results and ultimately the accuracy of risk prediction models, whether developed or validated. Missing data is generally the first hurdle for prediction research, where researchers are faced

with removal or imputation, covered in Part III. Deletion is incredibly common in this setting (Chapter 6), with some controversy. Deletion as a method of handling missing data has been deemed suitable if no significant difference is found between the dataset created through complete case analysis and that without the exclusion of individuals. However multiple studies show that bias is introduced into model performance results. There are many imputation methods available to researchers, from simple imputation to more complex multiple imputation. We critically reviewed the different patterns of missing data and approaches to dealing with them, and through a systematic review of studies on the development and/or validation of diabetes risk models, we investigated how missing data has been reported and handled to date. Our second main finding was the inconsistent reporting of missing data, with investigators frequently ignoring or failing to handle missing data appropriately. Dealing with missing data can be acknowledged as a complex task, which is not yet commonly undertaken in medical research. However, the availability of a wide range of methods for handling missing data should make even simple methods more routine in prediction research, yet only a handful of studies used the statistical procedures available.

Chapter 7 illustrated some of the imputation methods available to deal with missing data using the empirical Bellville South data. Comparison analysis for missing data between the complete case analysis and full dataset in the initial validation of the existent risk prediction models (Chapter 5) showed a statistically significant difference for a number of variables, however these differences were deemed clinical trivial. This subsequent analysis of model performance across a number of imputation methods available to researchers did however highlight a considerable effect on performance when compared to other imputation methods. Models developed or validated with biased data are likely to yield meaningless models that do not perform well when applied in any local circumstances. The second objective of this chapter was to compare the effect of simple and complex imputation methods on the performance of the same existing diabetes risk models, as validated in Chapter 5. Literature is divided as to whether simpler methods are too broad in their approach, not taking into account other individual characteristics and ignoring patient clustering, or if multiple imputation techniques are unnecessarily complex and too labour

intensive for non-statistical researchers. It is important to investigate under which conditions model performance can be increased. Empirical data research comes with the limitation that the true missing value will never be known, so which imputation technique produces a value closest to the true value cannot be answered. However, the advantage is that missing data is an empirical data research problem and studies as close to 'real world' research as possible are translatable to investigators in prediction research outside of a controlled setting. This can be addressed by comparing the characteristics of the imputed datasets. These were shown to be similar across all methods, resulting in little difference in the models' performances, with the third finding of this study highlighting the similar effect of simple over complex imputation techniques when assessing model performance. This is of great assistance for researchers with little statistical expertise or insufficient time, as the more complex methods are labour and time intensive, with multiple imputation requiring the pooling of the estimated parameters from each imputed dataset to effectively have a single combined dataset for which to run the models.

The updating of models may considerably improve the performance of previously developed models by accounting for differences between development and validation datasets. The methods presented in Chapter 9 illustrate how models can be updated through a series of statistical steps. However, these methods were shown to be poorly undertaken and reported in diabetes risk prediction research (Chapter 8). Our fourth finding was that published validation studies rarely attempted to better fit a model, with only a handful of studies performing the most basic updating techniques, recalibration of the intercept. The importance of generalizability and validation of current models is repeatedly emphasized in literature, however this is fruitless if extra efforts are not taken to fit the model as best as possible to the new setting. Unfortunately, only a relatively small number of validation studies have included recalibration in their methodologies. Additionally, no prevalent diabetes risk prediction models were recalibrated in an attempt to better fit the model to the validation population. An increased focus on the validation, and particularly recalibration, of existing models will improve the generalizability of the models and likely lead to greater application of diabetes risk prediction models in daily clinical practice.

Fortunately, this study gave evidence (Chapter 9) that adjusting for the validation population outcome prevalence may suffice to adjust existent diabetes risk prediction models. The fifth and final conclusion of the study demonstrated that the more complex updating methods were no more successful in increasing performance of the models, when compared to simpler updating methods. Discrimination was increased slightly across four of the five models with full model re-estimation, and calibration was significantly improved. This result was disappointing, however it must be noted that although extensive updating strategies may further improve a model's performance in local circumstances, their implementation requires a substantial amount of statistical knowledge and may not be an option for all risk prediction investigators. And although updating techniques deserve a fundamental role in diabetes risk prediction research, and have been proven to be successful in improving model performance in validation populations, they are not in use extensively and, from this study, are not always successful in achieving an increase in performance great enough to allow for recommendation.

Diabetes risk prediction research in Sub-Saharan Africa has been lacking, and although these results highlight that existing models may not be the right fit for our population group, it may increase the research undertaken in the future and improve the level at which methodological issues are handled. The future prospective is to facilitate greater validation of models, allowing for identification of models that are of limited value and implementation of genuinely useful models, aiding diabetes screening in developing countries. Final model performance increase in either discrimination or calibration may not be attainable when studies are too distantly related. This situation may arise when study populations differ too much and predictor-outcome associations strongly vary across studies. However, this should not discourage model validation and attempts to better fit the model to the validation dataset.

In summary, five existent prevalent diabetes risk prediction models were validated in a mixed-ancestry population in South Africa. This studies research questions asked if a model with sufficiently accurate predictive ability could be identified and recommended for use, while

identifying and discussing the methodological issues in the search for this model. The performance of these models in their original format was poor. Did the deletion of individuals with missing data affect this performance? Yes. Could this poor performance be improved by employing imputation techniques to deal with the missing data? Yes. Was there a difference between the multiple methods? Marginally, which allows for the recommendation that simple imputation with mean and mode, which are easily implemented and interpreted, can be just as effective as more complex, labour and time intensive methods. Was this increase in performance considered great enough for recommendation? No. Could model updating strategies increase the model performance? Yes, however the increase in discrimination in the more complex updating methods was at the expense of the calibration of the models. Can any model be recommended in this population group? No, simple intercept adjustment yielded the best performance across all models, however the models were still not accurate enough for recommendation. Our hypothesis; prediction models developed elsewhere generate inaccurate estimates of diabetes risk among adult South Africans, which can be substantially improved by efficient application of simple prediction models' improvement procedures; is therefore rejected as the performance of these models was not substantially improved and no single model can be recommended for use.

The future recommendations of this study can be separated into the risk prediction research community as a whole, as well as locally in Sub Saharan Africa. On a larger scale, this study highlighted the dire need for guidelines for both the handling and reporting of missing data, subsequently handled with the publication of the TRIPOD study. Despite recent advances in understanding missing data and imputation methods, most researchers still report deletion, perhaps because of the previous lack of adequate guidelines for handling missing data. It is hoped that these guidelines will provide a single, clear statement which will allow for uniform publication and interpretation. These guidelines are still fresh and require time to allow for circulation and recommendation from large health organizations, ultimately being accessible to all levels of practitioners and researchers to allow for easy implementation, enhancing the validity of reported results in all spheres of prediction research. What should be encouraged is the use of more than one method, the results compared and a preferred approach chosen

and defended. When data are missing on several variables it is important to use some procedure that imputes them all together, rather than one variable at a time. This ensures that the imputed data are related to each other in the same way as those data that are observed.

Additionally, we have shown that the model validation results are often taken at face value. Few investigators attempt to better fit a model to the population in which they are investigating. These results show a poor effect on the model calibration with more extensive updating methods, however this is in contrast to most research. Future research should be done on empirical data using these updating methods to determine their use in risk prediction on this type of data. It can be concluded that updating methods should become an integral part of model validation. A framework is therefore needed to identify whether early adaptation of a prediction model is justified, and to decide upon the extensiveness of updating methods. Furthermore, it may be helpful to identify when a prediction model has been sufficiently validated, and subsequent updating is no longer required or possible. Future research should address this question and allow for the determination of how many validation studies, what type of adjustments need to be made and most importantly, what is optimum performance to justify the implementation of the risk prediction model into clinical practice.

Additional to the research required on methodological issues, more external validation studies of existing diabetes prediction models are needed, along with the impact analysis of their introduction in routine care on healthcare providers' behaviour and the outcomes of people at risk of diabetes, or with prevalent undiagnosed diabetes. Positive or promising results from these studies could accelerate the uptake of prediction models in routine settings worldwide. Importantly, this work need to increasingly emerge from developing countries as the result of the growing worldwide interest in prediction research and personalised medicine.

On a local level, a risk prediction model is still required for this population, where the effect of non-invasive screening will have long term benefits on both an individual and community level. This is the only active cohort (cross-sectional data used for this study) in this community, and therefore limits the idea of validating other existent diabetes models which use variables not collected in this study. This could be a long term plan, incorporating these missing variables in future collections in this community. However, this will take time and it must be acknowledged that this database used is extensive in size, and covers a multitude of possible diabetes risk factors relative in this population. Therefore, the development of a new model may be recommended on the basis that existent models were validated and, with the aid of updating techniques, not able to accurately predict undiagnosed diabetes in the mixed-ancestry population of Bellville South, South Africa.

## References


1. Harrell, F.E. *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. 2001, New York: Springer.
2. Moons, K.G., Kengne, A.P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D.G.*, et al*. *Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio) marker.* Heart. 2012; **98**(9): p. 683-690.
3. Steyerberg, E.W. *Clinical prediction models*. 2009, New York: Springer.
4. Steyerberg, E.W., Moons, K.G., van der Windt, D.A., Hayden, J.A., Perel, P., Schroter, S.*, et al*. *Prognosis research strategy (progress) 3: Prognostic model research.* PLoS medicine. 2013; **10**(2): p. e1001381.
5. Moons, K.G., Altman, D.G., Vergouwe, Y., and Royston, P. *Prognosis and prognostic research: Application and impact of prognostic models in clinical practice.* Bmj. 2009; **338**: p. b606.
6. Griffin, S., Little, P., Hales, C., Kinmonth, A., and Wareham, N. *Diabetes risk score: Towards earlier detection of type 2 diabetes in general practice.* Diabetes/metabolism research and reviews. 2000; **16**(3): p. 164-171.
7. Bergmann, A., Li, J., Wang, L., Schulze, J., Bornstein, S., and Schwarz, P. *A simplified finnish diabetes risk score to predict type 2 diabetes risk and disease evolution in a german population.* Hormone and metabolic research. 2007; **39**(09): p. 677-682.
8. Altman, D.G., Vergouwe, Y., Royston, P., and Moons, K.G. *Prognosis and prognostic research: Validating a prognostic model.* Bmj. 2009; **338**: p. b605.
9. Altman, D.G. and Royston, P. *What do we mean by validating a prognostic model?* Statistics in medicine. 2000; **19**(4): p. 453-473.
10. Justice, A.C., Covinsky, K.E., and Berlin, J.A. *Assessing the generalizability of prognostic information.* Annals of Internal Medicine. 1999; **130**(6): p. 515-524.
11. Moons, K.G., Kengne, A.P., Grobbee, D.E., Royston, P., Vergouwe, Y., Altman, D.G.*, et al*. *Risk prediction models: Ii. External validation, model updating, and impact assessment.* Heart. 2012; **98**(9): p. 691-698.
12. Reilly, B.M. and Evans, A.T. *Translating clinical research into clinical practice: Impact of using prediction rules to make decisions.* Annals of Internal Medicine. 2006; **144**(3): p. 201-209.

13. Vergouwe, Y., Moons, K.G., and Steyerberg, E.W. *External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients.* American journal of epidemiology. 2010; **172**(8): p. 971-980.
14. Moons, K.G., Royston, P., Vergouwe, Y., Grobbee, D.E., and Altman, D.G. *Prognosis and prognostic research: What, why, and how?* Bmj. 2009; **338**.
15. Toll, D., Janssen, K., Vergouwe, Y., and Moons, K. *Validation, updating and impact of clinical prediction rules: A review.* Journal of clinical epidemiology. 2008; **61**(11): p. 1085-1094.
16. Collins, G.S., Reitsma, J.B., Altman, D.G., and Moons, K.G. *Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): The tripod statement.* BMC medicine. 2015; **13**(1): p. 1.
17. Al Khalaf, M.M., Eid, M.M., Najjar, H.A., Alhajry, K.M., Doi, S.A., and Thalib, L. *Screening for diabetes in kuwait and evaluation of risk scores.* East Mediterr Health J. 2010; **16**(7): p. 725-731.

# 11

# Appendix

## APPENDIX A – Declaration for Chapter 3 and 6

**Declaration by the candidate:**

With regard to Chapters 3 (page 23) and 6 (page 73) in this dissertation, the nature and scope of my contribution were as follows:

| Nature of contribution | Extent of contribution (%) |
|---|---|
| Data extraction | 100 |
| Compiled results | 80 |
| Preparation of first draft | 100 |
| Manuscript revision | 20 |

The following co-authors have contributed to Chapters 3 (pages 23) and 6 (page 73) in this dissertation:

| Name | e-mail address | Nature of contribution | Extent of contribution (%) |
|---|---|---|---|
| AP Kengne | andre.kengne@mrc.ac.za | Study conception | 80 |
| | | Duplicate data extraction | 40 |
| | | Compiled results | 20 |
| | | Manuscript revision | 30 |
| JB Echouffo-Tcheugui | jechouffotcheugui@partners.org | Study conception | 20 |
| | | Duplicate data extraction | 40 |
| | | Manuscript revision | 30 |
| TE Matsha | matshat@cput.ac.za | Duplicate data extraction | 10 |
| | | Manuscript revision | 10 |
| RT Erasmus | rte@sun.ac.za | Duplicate data extraction | 10 |
| | | Manuscript revision | 10 |

Signature of candidate: 

Date: 8 April 2016

**Declaration by co-authors:**

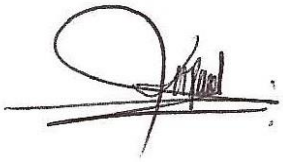
The undersigned hereby confirm that

1. the declaration above accurately reflects the nature and extent of the contributions of the candidate and the co-authors to Chapters 3 (pages 23) and 6 (page 73) in this dissertation,

2. no other authors contributed to Chapters 3 (pages 23) and 6 (page 73) in this dissertation,

besides those specified above, and

3. potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapters 3 (pages 23) and 6 (page 73) of this dissertation.

| **Signature** | **Institutional affiliation** | **Date** |
|---|---|---|
| AP Kengne  | Non-Communicable Diseases Research Unit, South African Medical Research Council, Cape Town, South Africa; Department of Medicine, University of Cape Town, Cape Town, South Africa | 8 April 2016 |
| JB Echouffo-Tcheugui  | Hubert Department of Public Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA; Department of Medicine, MedStar Health System, Baltimore, MD, USA | 8 April 2016 |
| TE Matsha  | Department of Biomedical Technology, Faculty of Health and Wellness Sciences, Cape Peninsula University of Technology, Cape Town, South Africa | 8 April 2016 |
| RT Erasmus  | Division of Chemical Pathology, Faculty of Health Sciences, National Health Laboratory Service (NHLS) and University of Stellenbosch, Cape Town, South Africa | 8 April 2016 |

## APPENDIX B – Declaration for Chapter 5, 7, 8 and 9

**Declaration by the candidate:**

With regard to Chapters 5 (page 48), 7 (page 93), 8 (page 116) and 9 (page 148) in this dissertation, the nature and scope of my contribution were as follows:

| Nature of contribution | Extent of contribution (%) |
|---|---|
| Data analysis | 100 |
| Data interpretation | 70 |
| Preparation of first draft | 100 |
| Manuscript revision | 30 |

The following co-authors have contributed to Chapters 5 (page 48), 7 (page 93), 8 (page 116) and 9 (page 148) in this dissertation:

| Name | e-mail address | Nature of contribution | Extent of contribution (%) |
|---|---|---|---|
| AP Kengne | andre.kengne@mrc.ac.za | Study conception | 100 |
| | | Data acquisition | 30 |
| | | Data interpretation | 30 |
| | | Manuscript revision | 50 |
| TE Matsha | matshat@cput.ac.za | Data acquisition | 35 |
| | | Manuscript revision | 10 |
| RT Erasmus | rte@sun.ac.za | Data acquisition | 35 |
| | | Manuscript revision | 10 |

Signature of candidate: 

Date: 8 April 2016

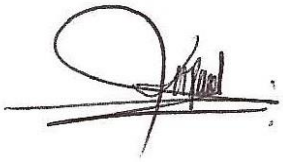
**Declaration by co-authors:**

The undersigned hereby confirm that

1. the declaration above accurately reflects the nature and extent of the contributions of the candidate and the co-authors to Chapters 5 (page 48), 7 (page 93), 8 (page 116) and 9 (page 148) in this dissertation,

2. no other authors contributed to Chapters 5 (page 48), 7 (page 93), 8 (page 116) and 9 (page 148) in this dissertation, besides those specified above, and

3. potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapters 5 (page 48), 7 (page 93), 8 (page 116) and 9 (page 148) of this dissertation.

| **Signature** | **Institutional affiliation** | **Date** |
|---|---|---|
| AP Kengne  | Non-Communicable Diseases Research Unit, South African Medical Research Council, Cape Town, South Africa; Department of Medicine, University of Cape Town, Cape Town, South Africa | 8 April 2016 |
| TE Matsha  | Department of Biomedical Technology, Faculty of Health and Wellness Sciences, Cape Peninsula University of Technology, Cape Town, South Africa | 8 April 2016 |
| RT Erasmus  | Division of Chemical Pathology, Faculty of Health Sciences, National Health Laboratory Service (NHLS) and University of Stellenbosch, Cape Town, South Africa | 8 April 2016 |