

# External Validation of a Periodontal Prediction Model for Identification of Diabetes among Saudi Adults

Arwa A Talakey<sup>1</sup>, Francis Hughes<sup>2</sup>, Eduardo Bernabé<sup>3</sup>

## ABSTRACT

**Aim and objective:** To externally validate the performance of a novel periodontal prediction model (PPM) for identification of diabetes among Saudi adults.

**Materials and methods:** The study was carried out among 150 adults attending primary care clinics in Riyadh (Saudi Arabia). The study adopted a temporal external validation approach, where the performance of the PPM was evaluated in the same location as the development study, but at a later time to allow for some variation between samples. A case-control approach was adopted, where diabetes status was first ascertained, followed by the completion of the Finnish Diabetes Risk Score (FINDRISC), Canadian Diabetes Risk (CANRISK) tools, and periodontal examinations.

**Results:** The area under the curve (AUC) of the PPM (based on the number of missing teeth, the proportion of sites with pocket probing depth  $\geq 6$  mm, and mean pocket probing depth) was 0.514 (95% CI: 0.385, 0.642). The FINDRISC and CANRISK tools had AUC values of 0.871 (95% CI: 0.811–0.931) and 0.927 (95% CI: 0.884–0.971), respectively. The addition of the PPM did not improve the AUC of FINDRISC ( $p = 0.479$ ) or CANRISK ( $p = 0.920$ ). The decision curve analysis showed that there was no clinical benefit in adding the PPM to either tool. The PPM was updated with an overall adjustment factor for all existing predictors and three more periodontal measures.

**Conclusion:** In an external sample, the PPM had poor performance for identification of diabetes and no added value when combined with FINDRISC and CANRISK. The performance of the PPM improved after recalibration and extension.

**Clinical significance:** The results underscore the value of externally validating prediction models before applying them in clinical dental practice.

**Keywords:** Diabetes, Diagnostic study, Periodontal disease, Prediction, Validation.

*The Journal of Contemporary Dental Practice* (2020): 10.5005/jp-journals-10024-2952

## INTRODUCTION

A few promising studies have evaluated the value of using periodontal measurements for identification of individuals with diabetes.<sup>1–4</sup> The first of those studies found that the probability of having undiagnosed diabetes among American adults was greater for those with periodontal disease (defined as having two or more sites with clinical attachment loss [CAL]  $\geq 6$  mm and one or more sites with pocket probing depth [PPD]  $\geq 5$  mm in one of these sites) than those without periodontal disease.<sup>1</sup> The authors also found clearer associations with undiagnosed diabetes when using PPD than when using CAL.<sup>1</sup>

Another American study showed that a prediction model based on the percentage of sites with PPD  $\geq 5$  mm and number of missing teeth had a moderate performance to identify unrecognized or prediabetes [area under the curve (AUC): 0.65; 95% confidence interval (CI): 0.60–0.70] among dental patients with at least one of four diabetes risk factors.<sup>2</sup> The performance of the prediction model dropped to 0.58 (95% CI: 0.53, 0.62) when assessed in a subsequent sample recruited from the same clinic, and to 0.64 (95% CI: 0.58, 0.71) when samples from both studies were combined.<sup>3</sup>

Recently, a periodontal prediction model (PPM) for identification of diabetes was developed among Saudi adults visiting primary care clinics.<sup>4</sup> The PPM, which was based on three periodontal measures (number of missing teeth, proportion of sites with PPD  $\geq 6$  mm, and mean PPD), showed an AUC of 0.69 (95% CI: 0.61–0.78). What is more, the addition of the PPM significantly improved the performance of the Finnish Diabetes Risk Score (FINDRISC) but not of the Canadian Diabetes Risk questionnaire (CANRISK). The authors used decision curve analysis to show that adding the PPM to both tools would result in greater net benefits than using any of the tools alone at probability scores lower than 70%.<sup>4</sup>

<sup>1</sup>Faculty of Dentistry, Oral and Craniofacial Sciences, King's College London, London, UK; Department of Periodontics and Community Dentistry, Faculty of Dentistry, King Saud University, Riyadh, Kingdom of Saudi Arabia

<sup>2,3</sup>Faculty of Dentistry, Oral and Craniofacial Sciences, King's College London, London, UK

**Corresponding Author:** Arwa A Talakey, Faculty of Dentistry, Oral and Craniofacial Sciences, King's College London, London, UK; Department of Periodontics and Community Dentistry, Faculty of Dentistry, King Saud University, Riyadh, Kingdom of Saudi Arabia, Phone: +44 (0) 20 3299 3022, e-mail: arwa.talakey@kcl.ac.uk

**How to cite this article:** Talakey AA, Hughes F, Bernabé E. External Validation of a Periodontal Prediction Model for Identification of Diabetes among Saudi Adults. *J Contemp Dent Pract* 2020;21(10): 1176–1181.

**Source of support:** Nil

**Conflict of interest:** None

The value of prediction models depends on their performance beyond the development sample.<sup>5</sup> Therefore, it is recommended to assess their performance in different samples from the same target populations (external validation).<sup>5–7</sup> Although the importance of external validation of prediction models is widely recognized,<sup>6,8,9</sup> only a few existing prediction models for diabetes identification have been externally validated.<sup>10,11</sup> This study was set out to externally validate the performance of our PPM for identification of individuals with diabetes among Saudi adults.

## MATERIALS AND METHODS

This study followed the standards for reporting of diagnostic accuracy studies (STARD)<sup>12</sup> and the transparent reporting of a

multivariable prediction model for individual prognosis or diagnosis (TRIPOD).<sup>13</sup> It was approved by the Research Ethics Committees of King's College London (HR-17/18-8281) and King Saud University (E-18-3386). All participants signed a written informed consent before participation.

The study adopted a case-control design, where cases were defined as individuals diagnosed with diabetes while controls were defined as nondiabetic individuals, and the PPM results were determined afterward.<sup>14</sup> A total of 1,531 participants who visited the primary care clinics at King Khalid University Hospital, King Saud University (Riyadh, Saudi Arabia) between July and November 2019 were initially approached. Of them, 298 were eligible and 150 (50.3%) (33 diabetic and 117 controls) agreed to participate. The participants' diabetic status was confirmed from medical records and defined as fasting plasma glucose of  $\geq 126$  mg/dL or hemoglobin A1c (HbA1c)  $\geq 6.5\%$  according to the American Diabetics Association guidelines<sup>15</sup> and the hospital protocol. A minimum sample size of 124 participants (31 cases and 93 controls) was required to estimate an AUC for the PPM of 0.65, with a margin of error of 0.10, 95% confidence level, 80% statistical power, and a case/control ratio of 1/3.<sup>16</sup>

Participants were included if they were from Saudi nationality and aged 30 years or older. Cases were diagnosed with type 2 diabetes during the past 12 months (incident cases) while controls were free from the condition. Participants were excluded if they had type 1 or gestational diabetes, were edentulous, wore fixed orthodontic appliances, or had any contraindications to carry out a periodontal examination (congenital heart disease, congenital heart murmurs, bacterial endocarditis, valvular heart disease, pacemaker or prosthetic valve replacement, or going through surgery in the next 6 months for pacemaker implantation or valve replacements).

### Data Collection

Data were collected using questionnaires, body measurements, and periodontal examinations, all undertaken at the dental clinics of the College of Dentistry, King Saud University. First, participants completed a self-administrated questionnaire to provide information on demographic factors (sex and age), socioeconomic position (education and occupation), six perceived periodontal measurements,<sup>17</sup> the 8 items of FINDRISC,<sup>18</sup> and the 12 items of CANRISK.<sup>19</sup> After revising that all questions had been answered, participants' body measurements were taken in duplicate by a dental assistant. Participants' weight (kg) and height (m) were measured using a portable scale and stadiometer, respectively. Waist circumference (cm) was measured using a measuring tape. Body measurements were taken in addition to participants' self-reports. Body measurements were used for analysis as there were some participants with missing values (particularly for waist circumference) from self-reports.

Periodontal examinations were carried out by two trained and calibrated dentists who were blinded to the case/control status of participants and supported by a dental assistant. A full-mouth periodontal assessment was conducted, including PPD, CAL, and bleeding on probing (BoP) at six sites (mesiobuccal, mid-buccal, distobuccal, distolingual, mid-lingual, mesiolingual) per tooth, excluding third molars.<sup>20</sup> The William's probe was used and measurements rounded to the lowest whole millimeter.<sup>21,22</sup> Duplicate examinations in 10% of the participants (one random quadrant) were conducted to assess intra- and interexaminer reliability. The intraexaminer reliability values (intraclass correlation

coefficients) were 0.87 (PPD), 0.91 (CAL), and 0.77 (BoP) for dentist 1 and 0.89 (PPD), 0.94 (CAL), and 0.80 (BoP) for dentist 2. The interexaminer reliability values were 0.89, 0.93, and 0.87 for PPD, CAL, and BoP, respectively. The following clinical periodontal measures were derived: number of missing teeth, proportion of sites with BoP, PPD (cut-offs:  $\geq 4$  and  $\geq 6$  mm), CAL (cut-offs:  $\geq 3$  and  $\geq 5$  mm), mean PPD, and mean CAL, which are the current standards for reporting periodontal disease in epidemiology surveys.<sup>23</sup> Two case definitions of periodontitis were also assessed, namely those for population-based surveillance<sup>24</sup> and clinical monitoring of periodontitis.<sup>25</sup> For each participant, the predicted probability of having diabetes (PPM score) was calculated using the following formula:  $p(\text{diabetes}) = 1/(1 + \exp(-1.060 + \text{number of missing teeth} \times 0.143 + \% \text{ sites with PPD} \geq 6 \text{ mm} \times 0.273 + \text{mean PPD (mm)} \times -1.195))$ .<sup>4</sup>

### Data Analysis

All analyzes were run in Stata 15 (Stata Corp., College Station, TX, USA). Cases and controls were compared in terms of their sociodemographic characteristics (sex, age, education, and employment status), clinical periodontal measurements, as well as PPM, FINDRISC, and CANRISK scores using the chi-squared test for categorical variables and the Student's *t*-test for numerical variables.

The performance of the PPM was assessed in terms of calibration and discrimination.<sup>13</sup> Calibration was assessed via calibration plots and the Hosmer-Lemeshow goodness-of-fit test. Discrimination was assessed by measuring the AUC. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated at the optimal cut-off point of 0.175 in the PPM score as identified in the development sample.<sup>4</sup> Finally, the incremental value of adding the PPM to FINDRISC and CANRISK was assessed using reclassification tables with three categories (preset FINDRISC and CANRISK risk groups) and decision curve analysis. The preset categories were low or slightly elevated ( $<15$ ), moderate (15–20), and high or very high risk of diabetes ( $>20$ ) in FINDRISC<sup>18</sup> and low ( $<21$ ), moderate (21–32), and high risk of diabetes ( $>32$ ) in CANRISK.<sup>19</sup> The three-category and continuous net reclassification improvement (NRI) as well as the integrated discrimination improvement (IDI) were used to quantify the extent of disease reclassification.<sup>13,26,27</sup> The clinical benefit of adding the PPM to FINDRISC and CANRISK was assessed in the decision curve analysis.

Six updating methods were considered to improve the performance of the PPM: method 1 recalibrated the PPM by adjusting the intercept; method 2 recalibrated the PPM by also adjusting the intercept and all predictor regression coefficients by one overall adjustment factor; method 3 revised the PPM by extra adjustment of regression coefficients for predictors with different strength in the validation sample compared with the development sample; method 4 revised the model by reestimating all predictor regression coefficients, using the data of the validation sample only; methods 5 and 6 extended methods 3 and 4, respectively, with the selection of additional predictors.<sup>13,28,29</sup>

### RESULTS

The characteristics of the sample are shown in Table 1. The proportion of male, older, and less-educated participants was significantly higher among cases than controls. Although cases had more missing teeth, greater proportion of sites with PPD  $\geq 6$  mm, and greater mean PPD than controls, these differences were not significant. Similarly, the PPM score was higher (albeit not significantly) in cases than controls. On the other hand, FINDRISC

and CANRISK scores were significantly higher among cases than controls (Table 1).

The PPM showed slightly poor calibration with some predictions (especially those at each end of the distribution) diverging from the 45° line of agreement (Fig. 1). However, differences between observed and expected probabilities were not significant (Hosmer-Lemeshow goodness-of-fit test,  $p = 0.052$ ). As for discrimination, the AUC of the PPM was 0.514 (95% CI: 0.385, 0.642). The optimal cut-off point of 0.175 in the PPM score classified correctly 64.0% of participants, yielding a sensitivity of 42.4% (95% CI: 25.5–60.8) and a specificity of 70.1% (95% CI: 60.9–78.2).

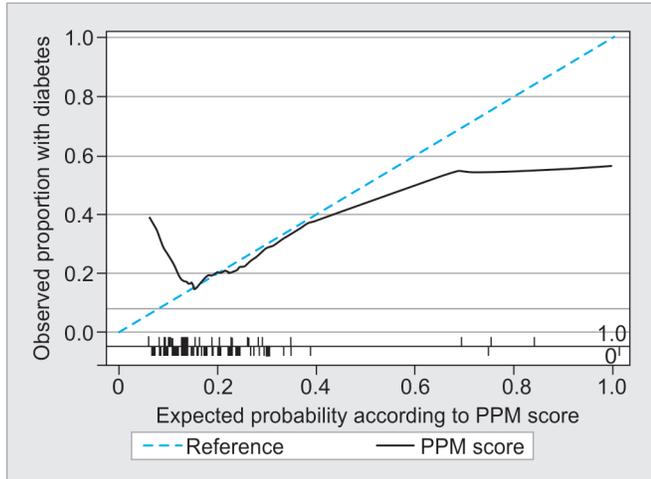


Fig. 1: Calibration plot of periodontal predictive model (PPM) score

The AUC values for the FINDRISC and CANRISK scores were 0.871 (95% CI: 0.811, 0.931) and 0.927 (95% CI: 0.884, 0.971), respectively. The addition of the PPM to each tool slightly improved the AUC for the FINDRISC to 0.877 (95% CI: 0.818, 0.936) but not for the CANRISK AUC 0.927 (95% CI: 0.883, 0.971) (Fig. 2). However, this improvement in performance for FINDRISC was not significant (DeLong test,  $p = 0.479$ ).

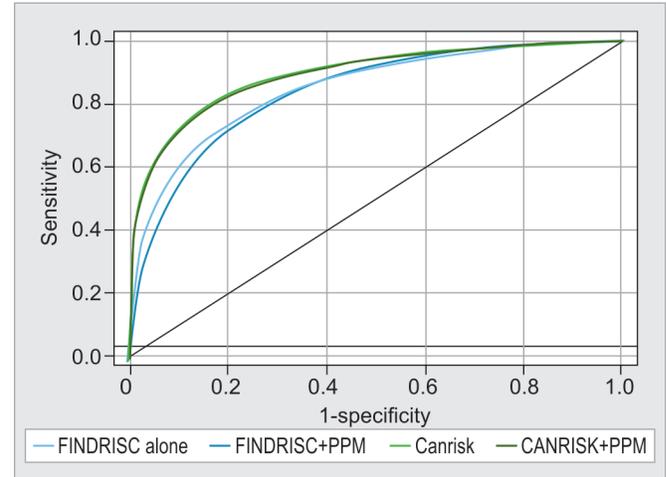


Fig. 2: Receiver operating characteristic (ROC) curves of different prediction models for identification of diabetes. The area under the ROC curve (AUC) was 0.871 (95% CI: 0.811–0.931) for the FINDRISC alone, 0.877 (95% CI: 0.818, 0.936) for FINDRISC + PPM, 0.927 (95% CI: 0.884–0.971) for CANRISK alone, and 0.927 (95% CI: 0.883, 0.971) for CANRISK + PPM. CANRISK, Canadian Diabetes Risk; FINDRISC, Finnish Diabetes Risk Score; PPM, periodontal prediction model

Table 1: Comparison of sociodemographic characteristics, screening tools scores, and periodontal measurements by diabetes status ( $n = 150$ )

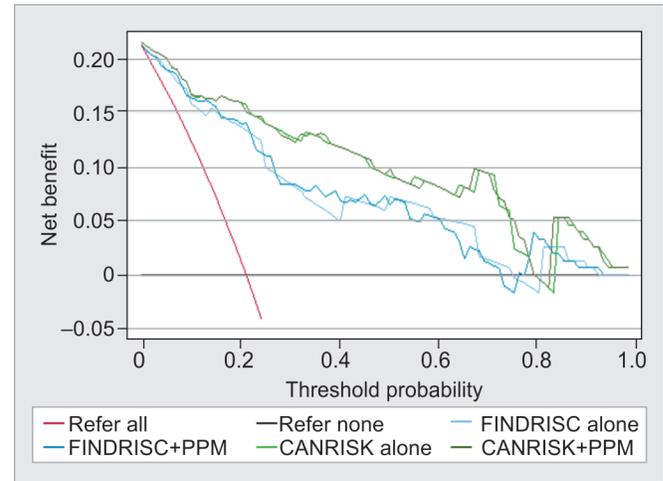
	Control group ( $n = 117$ )		Diabetic group ( $n = 33$ )		$p$ value <sup>a</sup>
Sex, $n$ %					0.005
Male	17	14.5	12	36.4	
Female	100	85.5	21	63.6	
Mean age $\pm$ SD, years	42.9 $\pm$ 9.8		50.8 $\pm$ 8.6		<0.001
Education, $n$ %					0.022
Less than high school	15	12.8	10	30.3	
High school	17	14.5	7	21.2	
Higher education	85	72.6	16	48.5	
Occupation, $n$ %					0.134
Employed	77	65.8	17	51.5	
Unemployed	40	34.2	16	48.5	
Number of missing teeth	2.8 $\pm$ 3.0		3.7 $\pm$ 4.2		0.171
% sites with PPD $\geq$ 6 mm $\pm$ SD	0.7 $\pm$ 3.8		1.1 $\pm$ 3.3		0.524
Mean PPD $\pm$ SD (mm)	2.7 $\pm$ 0.3		2.8 $\pm$ 0.4		0.226
Mean PPM score $\pm$ SD	0.17 $\pm$ 0.11		0.22 $\pm$ 0.19		0.094
Mean FINDRISC score $\pm$ SD	10.6 $\pm$ 4.7		17.7 $\pm$ 3.8		<0.001
Mean CANRISK score $\pm$ SD	23.1 $\pm$ 12.0		46.3 $\pm$ 9.4		<0.001

<sup>a</sup>Chi-squared test used to compare proportions and  $t$ -test to compare means. CANRISK, Canadian Diabetes Risk; FINDRISC, Finnish Diabetes Risk Score; PPD, pocket probing depth; PPM, periodontal prediction model

Adding the PPM to FINDRISC yielded four cases reclassified up and three cases reclassified down (with a net gain in the reclassification proportion of 0.03) as well as two controls reclassified up and zero controls reclassified down (with a net loss in reclassification proportion of 0.02) (Table 2). Therefore, the three-category NRI of the FINDRISC + PPM was 0.013 ( $p = 0.871$ ). The continuous NRI and IDI were 0.267 (95% CI: -0.474, 0.901) and 0.000 (95% CI: -0.020, 0.100), respectively. The addition of the PPM to CANRISK yielded a net loss in the reclassification proportion of 0 among cases and a net gain in the reclassification proportion of -0.02 among controls, with a 3-category NRI of 0.017 ( $p = 0.157$ ) (Table 2). The continuous NRI and IDI for the CANRISK + PPM were -0.598 (95% CI: -0.810, 1.024) and -0.006 (95% CI: -0.019, 0.020), respectively. Finally, the decision curve analysis showed that FINDRISC + PPM had greater net benefits than FINDRISC alone at probability thresholds of 0.20, between 0.30 and 0.40, and at 0.80 whereas no differences in net benefit were found between CANRISK and CANRISK+PPM along the entire range of probability thresholds (Fig. 3).

Given the poor performance of the PPM in the validation sample, six updating methods were used to improve its diagnostic performance. All updating methods improved calibration but the greatest improvements in discrimination were achieved with methods 5 and 6. Method 5 was preferred to update the PPM as it was less complex. Table 3 shows the updated PPM model, as derived from method 5, which had an AUC of 0.740 (95% CI: 0.635, 0.845).

The optimal cut-off point (0.260) in the updated PPM score classified correctly 75.3% of participants, with a sensitivity of 60.6% (95% CI: 42.1–77.1) and specificity of 79.5% (95% CI: 71.0–86.4).



**Fig. 3:** Decision curve analysis showing the net benefit of referring none of the patients for further testing, referring all patients for further testing, FINDRISC alone, FINDRISC + PPM, CANRISK alone, and CANRISK + PPM across all possible probability thresholds. CANRISK, Canadian Diabetes Risk; FINDRISC, Finnish Diabetes Risk Score; PPM, periodontal prediction model

**Table 2:** Reclassification tables for the Finnish Diabetes Risk Score (FINDRISC) and Canadian Diabetes Risk (CANRISK) tools before and after addition of the periodontal prediction model (PPM)

	FINDRISC + PPM						Total	
	<15 <sup>a</sup>		15–20 <sup>a</sup>		>20 <sup>a</sup>		n	(%)
FINDRISC alone	n	(%)	n	(%)	n	(%)	n	(%)
Participants who were diabetic (n = 33)								
<15 <sup>a</sup>	8	100.0	2	13.3	0	0.0	10	30.3
15–20 <sup>a</sup>	0	0.0	10	66.7	2	20.0	12	36.4
>20 <sup>a</sup>	0	0.0	3	20.0	8	80.0	11	33.3
Total	8	100.0	15	100.0	10	100.0	33	100.0
Participants who were nondiabetic (n = 177)								
<15 <sup>a</sup>	93	100.0	1	4.8	0	0.0	94	80.3
15–20 <sup>a</sup>	0	0.0	20	95.2	1	33.3	21	18.0
>20 <sup>a</sup>	0	0.0	0	0.0	2	66.7	2	1.7
Total	93	100.0	21	100.0	3	100.0	117	100.0

	CANRISK + PPM						Total	
	<21 <sup>a</sup>		21–32 <sup>a</sup>		>32 <sup>a</sup>		n	%
CANRISK alone	n	%	n	%	n	%	n	%
Participants who were diabetic (n = 33)								
<21 <sup>a</sup>	1	100.0	0	0.0	0	0.0	1	3.0
21–32 <sup>a</sup>	0	0.0	3	100.0	0	0.0	3	9.1
>32 <sup>a</sup>	0	0.0	0	0.0	29	100.0	29	87.9
Total	1	100.0	3	100.0	29	100.0	33	100.0
Participants who were nondiabetic (n = 177)								
<21 <sup>a</sup>	75	100.0	0	0.0	0	0.0	75	64.1
21–32 <sup>a</sup>	0	0.0	10	83.3	0	0.0	10	8.5
>32 <sup>a</sup>	0	0.0	2	16.7	30	100.0	32	27.4
Total	75	100.0	12	100.0	30	100.0	117	100.0

<sup>a</sup>The original cut-off points of <15, 15–20, and >20 in FINDRISC score<sup>18</sup> and <21, 21–32, and >32 in CANRISK score<sup>19</sup> correspond to probability scores of <0.26, 0.26–0.68, and >0.68, and <0.07, 0.07–0.12, and >0.12, respectively. The net reclassification improvement (NRI) for the addition of PPM to FINDRISC was [(0.12–0.09)–(0.02–0.0)] = 0.01 and to CANRISK was [(0.0–0.0)–(0.0–0.02)] = 0.02

**Table 3:** Predictors included in the updated periodontal prediction model (PPM) (*n* = 150)

Predictors	OR <sup>a</sup>	[95% CI]	<i>p</i> value
PPM score	5.55	[0.22–137.77]	0.295
Mean clinical attachment loss (CAL)	2.48	[1.28–4.81]	0.007
Self-reported gum disease			
No	1.00	[Reference]	
Yes	2.34	[0.97–5.65]	0.059
Self-rated oral health			
Excellent/very good/good	1.00	[Reference]	
Fair/poor	0.23	[0.08–0.68]	0.008
Intercept	0.02	[0.00–0.10]	<0.001

The predicted probability of having diabetes can be calculated using the following formula:  $p(\text{diabetes}) = 1 / (1 + \exp(-(-4.502 + \text{number of missing teeth} \times 0.025 + \% \text{ sites with PPD} \geq 6 \text{ mm} \times 0.468 + \text{mean PPD} \times -2.048 + \text{self-reported gum disease [no} = 0, \text{yes} = 1] \times 0.851 + \text{self-rated oral health [excellent/very good/good} = 0, \text{fair/poor} = 1] \times -1.474 + \text{mean CAL} \times 0.907)))$

<sup>a</sup>Logistic regression was fitted and odd ratios (OR) reported

## DISCUSSION

We found that the performance of our novel PPM dropped after external validation. Although the percentage of participants correctly classified in this external sample (64%) was close to the 62.4% reported in the development sample,<sup>4</sup> the AUC of the PPM dropped from 0.694 to 0.514.<sup>4</sup> The poorer performance of the PPM in an external sample agrees with that of the US study where performance of a prediction model, based on percentage of sites with PPD  $\geq 5$  mm and number of missing teeth, dropped from 0.65 to 0.58 when validated in a subsequent sample.<sup>3</sup> More generally, our findings coincide with a systematic review concluding that risk prediction models often had poor performance after external validation.<sup>10</sup>

The poorer performance of the PPM in the validation sample could be due to underlying differences between the development and validation samples. Our participants had significantly lower PPM score (0.18 vs 0.21) and higher mean PPD (2.8 vs 2.6 mm) than those in the development sample. However, the fact remains that for any prediction model to be useful in clinical practice it must be applicable to other populations or settings. Furthermore, the differences observed between cases and controls in this sample were somewhat expected as diabetes is more common among male, older, and less educated adults.<sup>30</sup> Indeed, these risk factors are included in conventional tools for diabetes screening.<sup>18,19</sup>

No evidence on the value of adding periodontal measurements to conventional diabetes screening tool was found either. The FINDRISC + PPM reclassified correctly 1.3% of participants into the three risk groups of FINDRISC, with an upper bound of 26% (continuous NRI) and no difference in mean predicted probability between controls and cases (continuous IDI). Results were worse for the CANRISK + PPM where only 1.7% of participants were reclassified correctly into the three risk groups of CANRISK, with an upper bound of -60% and a difference in mean predicted probability of -1%. These findings may be attributed to the fact that CANRISK includes four more diabetes risk factors (sex, education, ethnicity, and giving birth to a large baby) than FINDRISC.<sup>19</sup> Furthermore, the decision curve analysis showed

that the PPM adds only minimal benefit to FINDRISC at very specific thresholds. At a probability threshold of 0.35, the net benefit would be 0.08, which is slightly higher than 0.07 for FINDRISC alone, suggesting that FINDRISC + PPM can correctly refer one extra-patient of 100 suspected cases without having an unnecessary referral (false-positive) when compared to no referral and no benefit to CANRISK along the entire probability range.

Updating methods were used to improve the performance of the PPM in the validation sample. Of the six methods recommended for model updating,<sup>26</sup> a combination of model recalibration (adjustment of predictors' weights) and extension (adding new predictors) improved both calibration and discrimination with the fewest number of modifications. This method also had the key advantage of using information from both the development and validation samples; therefore, it built up on information from the development study rather than discarded it. It must be noted though that updating the PPM would adjust the PPM to the circumstances of the validation sample and would therefore require further evaluation in other external samples.

The study findings have some implications for practice and future research. They confirm the general idea that internal validation neither replace validation in an external sample nor guarantee transportability.<sup>13</sup> What is more, they support the value of externally validating prediction models before applying them in clinical practice.<sup>13,31</sup> Further studies with stronger (longitudinal) designs and larger samples in alternative settings would be welcome to demonstrate the added value of periodontal measurements when combined with conventional diabetes screening tools.

The study has some limitations that need to be addressed. The study adopted a case-control design where a prospective cohort design is considered the optimal approach to develop and validate prediction models. However, the cross-sectional design is considered reasonable when building a body of evidence before utilizing more costly designs.<sup>14</sup> Additionally, we recruited participants from a single site in Saudi Arabia, which is not representative of the entire Saudi population. Therefore, the results cannot be generalized beyond the study sample.

## CONCLUSION

When evaluated in a temporal external sample of Saudi adults in primary care, a recently developed PPM based on number of missing teeth, proportion of sites with PPD  $\geq 6$  mm, and mean PPD showed poor performance (both in terms of calibration and discrimination) for identification of diabetes and no added value when combined with conventional diabetes screening tools. The performance of the PPM improved after recalibration and extension.

## CLINICAL SIGNIFICANCE

The results underscore the value of externally validating prediction models before applying them in clinical dental practice.

## ACKNOWLEDGMENTS

We thank Drs Hani Almoharib and Mansour Al-Askar for their substantial contribution during the data collection for this study.

## REFERENCES

- Borrell LN, Kunzel C, Lamster I, et al. Diabetes in the dental office: using NHANES III to estimate the probability of undiagnosed disease. *J Periodontol Res* 2007;42(6):559–565. DOI: 10.1111/j.1600-0765.2007.00983.x.

2. Lalla E, Kunzel C, Burkett S, et al. Identification of unrecognized diabetes and pre-diabetes in a dental setting. *J Dent Res* 2011;90(7):855–860. DOI: 10.1177/0022034511407069.
3. Lalla E, Cheng B, Kunzel C, et al. Dental findings and identification of undiagnosed hyperglycemia. *J Dent Res* 2013;92(10):888–892. DOI: 10.1177/0022034513502791.
4. Talakey A, Hughes F, Almoharib H, et al. The added value of periodontal measurements for identification of diabetes among Saudi adults. *J Periodontol* 2020. DOI: 10.1002/JPER.20-0118.
5. Debray TP, Vergouwe Y, Koffijberg H, et al. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 2015;68(3):279–289. DOI: 10.1016/j.jclinepi.2014.06.018.
6. Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* 2014;14(1):40. DOI: 10.1186/1471-2288-14-40.
7. Wynants L, Collins GS, Van Calster B. Key steps and common pitfalls in developing and validating risk models. *BJOG* 2017;124(3):423–432. DOI: 10.1111/1471-0528.14170.
8. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMC Med* 2015;13(1):1. DOI: 10.1186/s12916-014-0241-z.
9. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;98(9):691–698. DOI: 10.1136/heartjnl-2011-301247.
10. Siontis GC, Tzoulaki I, Castaldi PJ, et al. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol* 2015;68(1):25–34. DOI: 10.1016/j.jclinepi.2014.09.007.
11. Abbasi A, Peelen LM, Corpeleijn E, et al. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. *BMJ* 2012;345(sep18 2):e5900. DOI: 10.1136/bmj.e5900.
12. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6(11):e012799. DOI: 10.1136/bmjopen-2016-012799.
13. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162(1):W1–W73. DOI: 10.7326/M14-0698.
14. Linnet K, Bossuyt PM, Moons KG, et al. Quantifying the accuracy of a diagnostic test or marker. *Clin Chem* 2012;58(9):1292–1301. DOI: 10.1373/clinchem.2012.182543.
15. American Diabetes Association. 2. Classification and diagnosis of diabetes: Standards of medical care in diabetes—2018. *Diabetes Care* 2018;41(Suppl. 1):S13–S27. DOI: 10.2337/dc18-S002.
16. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143(1):29–36. DOI: 10.1148/radiology.143.1.7063747.
17. Eke PI, Dye BA, Wei L, et al. Self-reported measures for surveillance of periodontitis. *J Dent Res* 2013;92(11):1041–1047. DOI: 10.1177/0022034513505621.
18. Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 2003;26(3):725–731. DOI: 10.2337/diacare.26.3.725.
19. Kaczorowski J, Robinson C, Nerenberg K. Development of the CANRISK questionnaire to screen for prediabetes and undiagnosed type 2 diabetes. *Canadian J Diabetes* 2009;33(4):381–385. DOI: 10.1016/S1499-2671(09)34008-3.
20. Newman MG, Takei H, Klokkevold PR, et al. Carranza's clinical periodontology-e-book: Expert consult: Online. Elsevier Health Sciences; 2014.
21. Beltrán-Aguilar ED, Eke PI, Thornton-Evans G, et al. Recording and surveillance systems for periodontal diseases. *Periodontol* 2000 2012;60(1):40–53. DOI: 10.1111/j.1600-0757.2012.00446.x.
22. Control CfD, Prevention. National Health and Nutrition Examination Survey (NHANES). 2013–14. Retrieved August. 2016.
23. Holtfrete B, Albandar JM, Dietrich T, et al. Standards for reporting chronic periodontitis prevalence and severity in epidemiologic studies: Proposed standards from the joint EU/USA periodontal epidemiology working group. *J Clin Periodontol* 2015;42(5):407–412. DOI: 10.1111/jcpe.12392.
24. Page RC, Eke PI. Case definitions for use in population-based surveillance of periodontitis. *J Periodontol* 2007;78(Suppl 75):1387–1399. DOI: 10.1902/jop.2007.060264.
25. Tonetti MS, Greenwell H, Kornman KS. Staging and grading of periodontitis: framework and proposal of a new classification and case definition. *J Clin Periodontol* 2018;45(Suppl 20):S149–S161. DOI: 10.1111/jcpe.12945.
26. Moons KG, de Groot JA, Linnet K, et al. Quantifying the added value of a diagnostic test or marker. *Clin Chem* 2012;58(10):1408–1417. DOI: 10.1373/clinchem.2012.182550.
27. Pencina MJ, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27(2):157–172. DOI: 10.1002/sim.2929.
28. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;35(29):1925–1931. DOI: 10.1093/eurheartj/ehu207.
29. Steyerberg EW. *Clinical prediction models*. Springer International Publishing; 2019.
30. Kyrou I, Tsigos C, Mavrogianni C, et al. Sociodemographic and lifestyle-related risk factors for identifying vulnerable groups for type 2 diabetes: a narrative review with emphasis on data from Europe. *BMC Endocr Disord* 2020;20(1):1–13. DOI: 10.1186/s12902-019-0463-3.
31. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338(may28 1):b605. DOI: 10.1136/bmj.b605.