

## ORIGINAL ARTICLE

# Comparative analysis of screening models for undiagnosed diabetes in Mexico



Gerardo Jorge Félix-Martínez<sup>a,b,\*</sup>, José Rafael Godínez-Fernández<sup>b</sup>

<sup>a</sup> Cátedras CONACYT (Consejo Nacional de Ciencia y Tecnología, México), Mexico

<sup>b</sup> Departamento de Ingeniería Eléctrica, Universidad Autónoma Metropolitana, Unidad Iztapalapa, Mexico

Received 18 April 2019; accepted 30 August 2019

Available online 30 November 2019

### KEYWORDS

Undiagnosed diabetes;  
Mexico;  
Screening models

### Abstract

**Background:** It is estimated that 37% of Mexican adults have undiagnosed diabetes, and are therefore at high risk of developing the severe and devastating complications associated to it. In recent years, a variety of screening tools based on the characteristics of the adult Mexican population have been proposed in order to reduce the negative effects of the disease.

**Objectives:** To assess the performance of screening models to diagnose diabetes in the Mexican adult population and to propose a screening model based on HbA1c measurements.

**Materials and methods:** Data from the 2016 Halfway National Health and Nutrition Survey (NHNS) were used to assess the screening models and to develop and validate the proposed 2016 NHNS model, built using a multivariate logistic regression model. Explanatory variables included in the 2016 NHNS 2016 model were selected through a stepwise backward procedure, using sensitivity and specificity as performance indicators.

**Results:** Of the screening models assessed, only the model based on the 2006 NHNS survey showed a performance consistent with previous reports. The proposed 2016 NHNS model included age, waist circumference, and systolic blood pressure as explanatory variables and showed a sensitivity of 0.72 and a specificity of 0.80 in the validation data set.

**Conclusions:** Age, waist circumference, and systolic blood pressure are variables of special importance for early detection of undiagnosed diabetes in Mexican adults. Based on the consistent performance of the 2006 NHNS model in different data sets, its use as a screening tool for adults with undiagnosed diabetes in Mexico is recommended.

© 2019 SEEN y SED. Published by Elsevier España, S.L.U. All rights reserved.

\* Corresponding author.

E-mail addresses: [gfelix@conacyt.mx](mailto:gfelix@conacyt.mx), [gfelix@xanum.uam.mx](mailto:gfelix@xanum.uam.mx), [gjfelix2005@gmail.com](mailto:gjfelix2005@gmail.com) (G.J. Félix-Martínez).

**PALABRAS CLAVE**

Diabetes no diagnosticada;  
México;  
Modelos de cribado

**Análisis comparativo de modelos de cribado de la diabetes no diagnosticada en México****Resumen**

*Antecedentes:* Se estima que el 37% de adultos mexicanos que padecen diabetes no han sido diagnosticados, teniendo un alto riesgo de desarrollar las graves y devastadoras complicaciones asociadas. En los últimos años, con el objetivo de reducir los efectos negativos de la enfermedad, se han propuesto herramientas de detección basadas en las características de la población adulta mexicana.

*Objetivos:* Evaluar el funcionamiento de los modelos de detección de diabetes no diagnosticada en la población adulta mexicana y proponer un modelo de detección basado en mediciones de HbA1c.

*Material y métodos:* Se utilizaron datos de la Encuesta Nacional de Salud y Nutrición (NHNS) de Medio Camino 2016 para evaluar los modelos de detección y para desarrollar y validar el modelo NHNS 2016, construido usando un modelo de regresión logística multivariada. Las variables incluidas en el modelo NHNS 2016 se seleccionaron mediante un procedimiento de pasos hacia atrás, utilizando la sensibilidad y la especificidad como indicadores de funcionamiento.

*Resultados:* De los modelos de detección evaluados, únicamente el modelo basado en la encuesta NHNS 2006 mostró un funcionamiento consistente con lo reportado anteriormente. Las variables seleccionadas para el modelo propuesto (NHNS 2016) fueron edad, circunferencia de cintura y presión sistólica como variables explicativas, mostrando una sensibilidad de 0,72 y una especificidad de 0,80 en el conjunto de datos de validación.

*Conclusiones:* Edad, circunferencia de cintura y presión sistólica son variables de relevancia especial para la detección temprana de diabetes no diagnosticada en adultos mexicanos. Dada la consistencia en el funcionamiento mostrado por el modelo NHNS 2006 en diferentes conjuntos de datos, se recomienda su uso como herramienta de detección de adultos con diabetes no diagnosticada en México.

© 2019 SEEN y SED. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

**Introduction**

In 2017, the International Diabetes Federation (IDF) estimated that 50% of people with diabetes worldwide (>200 million) remain undiagnosed.<sup>1</sup> Likewise, it is estimated that in Mexico, at least 37% of individuals with diabetes are unaware of their disease.<sup>1</sup> These numbers are alarming given the fact that type 2 diabetes (T2D) can remain asymptomatic for several years before diagnosis,<sup>2</sup> thus increasing the risk of developing the associated complications.<sup>3</sup> For this reason, the American Diabetes Association<sup>4</sup> recommends the use of screening or risk assessment tools in asymptomatic adults for the opportune detection and treatment of those subjects with or at high risk of developing T2D. In this regard, screening models have been developed within specific populations or regions, taking into account the possible effect of ethnic, racial, cultural or/and environmental differences on the risk of T2D (reviewed in<sup>5-7</sup>).

In Mexico, one of the countries with a highest number of people with diabetes (according to the 2017 IDF Diabetes Atlas<sup>1</sup>), a variety of screening models for undiagnosed diabetes have been proposed in the last years, aiming to provide low cost and simple-to-use tools to identify the population with undiagnosed diabetes or at high risk of developing the disease. Rojas-Martínez et al.<sup>8</sup> evaluated waist circumference cutoff values as potential explanatory variable to identify Mexican adults with undiagnosed

diabetes. Although this approach was capable of correctly identifying ~85% of subjects with diabetes when tested in a validation dataset, it also yielded a high percentage of false positives (~80%). In a later study, Rojas-Martínez et al.<sup>9</sup> used data from a cohort of Mexican adults obtained in the *Instituto Nacional de Ciencias Médicas y Nutrición* Salvador Zubirán (INCMNSZ) (National Institute of Medical Sciences and Nutrition), to design a model that identified correctly 53.9 and 41.7% of Mexican adult males and females with undiagnosed diabetes, respectively, in an external validation dataset, with only 32.5 and 15.1% of false positives. On the other hand, in a previous report, we proposed two models based on the NHNS 2006 and 2012 surveys<sup>10</sup> that correctly identified ~74 and 76% of Mexican adults with undiagnosed diabetes in an external validation dataset, with 38 and 45% of false positives, respectively. Finally, there is also a risk score distributed and implemented by the Mexican Secretariat of Health<sup>11</sup> which, to our knowledge, has been neither developed nor validated using data from Mexican adults.

The main objective of this work is to evaluate these models using recent data from the halfway National Health and Nutrition Survey (NHNS) 2016<sup>12</sup> in order to assess how they perform in external validation scenarios. In addition, using the same NHNS 2016 dataset, we propose a new screening model for undiagnosed diabetes in the Mexican adult population based on glycated hemoglobin (HbA1c) measurements.

## Material and methods

### Dataset: National Health and Nutrition Survey 2016 (NHNS 2016)

Data from Mexican adults interviewed in the context of the halfway National Health and Nutrition Survey 2016 (NHNS 2016)<sup>12</sup> was used to evaluate the screening models as well as to develop the NHNS 2016 model. The NHNS 2016 survey was implemented nationwide in Mexico to assess the general state of health and nutrition of the Mexican population with a particular interest in aspects related to the prevention and control of obesity and diabetes.<sup>12</sup> Household interviews were conducted in which individuals were asked about several subjects related to chronic diseases, physical activity, diet diversity, quality of care for patients, among others.<sup>12</sup> In addition, anthropometric measurements and blood samples were taken to determine the evolution of obesity, hypertension and diabetes.<sup>12,13</sup> Methodological details of the halfway NHNS 2016 survey can be found in Ref.<sup>13</sup> From the 9406 Mexican adults interviewed for the NHNS 2016 survey, we selected the subsample from which blood samples were taken and all the variables listed in Table 1 were available. Neither subjects with a previous diagnosis of diabetes nor those with missing information were included in the final dataset used for the evaluation, development, optimization or validation of the screening models. The resulting dataset was therefore composed of 2386 observations (whose characteristics are summarized in Table 1). The fact that the NHNS 2016 dataset includes HbA1c measurements, in addition to all the relevant variables used by all the models evaluated, constitutes an important advantage since HbA1c measurements are a better indicator of glycemic exposure and risk for long-term complications.<sup>14</sup> Moreover, HbA1c assays have important advantages when compared to measurements of the fasting plasma glucose or the oral glucose tolerance test, such as less day-to-day or intraindividual variability, less preanalytic instability and the fact that fasting is not required for HbA1c tests.<sup>1,14</sup> For this reason, undiagnosed diabetes was determined in the NHNS 2016 dataset adopting the criteria established by the American Diabetes Association,<sup>4</sup> according to which a measurement of glycated hemoglobin (HbA1c)  $\geq 6.5\%$  indicate the presence of the disease. As a result, only 82 of the 2386 observations contained in the NHNS 2016 dataset used in this work ( $\sim 3.4\%$ ) were classified as having undiagnosed diabetes. Important differences between the subjects with undiagnosed diabetes and those considered as normal were observed in most of the variables listed in Table 1.

### Screening models for undiagnosed diabetes in the Mexican adult population

Screening models for undiagnosed diabetes developed specifically for the Mexican adult population were evaluated using the NHNS 2016 dataset described in the last section. In the following, the five models evaluated are briefly described (see also Appendix A). Performance of the models was determined by their capability to correctly identify subjects with undiagnosed diabetes while yielding the lowest

number of false positives. Sensitivity (i.e. the proportion or percentage of people with undiagnosed diabetes correctly identified by the models) and specificity (i.e. the proportion or percentage of healthy people correctly identified by the models) were used to compare the models performance.

### Waist circumference model

The first effort to detect adults with undiagnosed diabetes in Mexico using minimal and simple obtainable explanatory variables was made by Rojas et al.,<sup>8</sup> who evaluated the use of cutoff values for abdominal obesity (i.e. waist circumference) as indicators of undiagnosed diabetes using data from the NHNS 2006 survey. The authors evaluated the waist circumference cutoff values proposed, on the one hand, by the American Heart Association and the National Heart, Lung and Blood Institute (AHA/NHLBI,  $\geq 102$  for males and  $\geq 82$  for females),<sup>15</sup> and on the other, by the International Diabetes Federation (IDF,  $\geq 90$  for males and  $\geq 80$  for females).<sup>16</sup> In short, according to this model, a subject was classified as at high risk of having undiagnosed diabetes if her or his waist circumference measurement was above either of the two cutoff values evaluated. This approach was capable of identifying correctly  $\sim 85\%$  of subjects with diabetes, although at the expense of an extremely high percentage ( $\sim 80\%$ ) of false positives.

### Mexican Secretariat of Health (SSA) risk score

A simple risk score distributed by the Mexican Secretariat of Health (SSA)<sup>11</sup> was also evaluated in this work. Surprisingly, there is no information about how it was developed or validated, or how it has performed as a tool to identify Mexican subjects with undiagnosed diabetes. According to Rojas-Martínez et al.,<sup>9</sup> the SSA score is based on a score developed by the American Diabetes Association<sup>17</sup> and has not been validated for the Mexican population. The explanatory variables used in the SSA risk score are body mass index, waist circumference, physical activity, age, family history of diabetes and birth weight as an indicator of possible gestational diabetes in the female population. The SSA risk score is reproduced in Appendix A.1.

### INCMNSZ risk score

Recently, Rojas-Martínez et al.<sup>9</sup> developed a risk score for detecting prediabetes and undiagnosed diabetes in Mexican adults. The risk score is based on data from a cohort of Mexican adults obtained in the *Instituto Nacional de Ciencias Médicas y Nutrición Salvador Suvirán* (INCMNSZ) and was developed using multiple logistic regression. Explanatory variables considered in the INCMNSZ risk score were age, family history of diabetes, waist circumference, body mass index, hypertension status and physical activity. External validation was performed using a dataset from the National Health and Nutrition Survey 2012.<sup>18</sup> As a result, they proposed two different scores for the male and female populations (described in detail in Appendix A.2) and reported a sensitivity and specificity in the external validation dataset

**Table 1** Characteristics of the NHNS 2016 dataset used in this work. Continuous variables are presented as means (SD) and categorical variables are presented as number of observations and percentages.

	Total	Normal	UDM	P-value <sup>a</sup>
<i>n</i>	2386	2304	82	
%	100	96.6	3.4	
Sex (%)	F: 66.0 M: 34.0	F: 66.0 M: 34.0	F: 65.9 M: 34.1	
Age (y)	41.2 (12.8)	40.9 (12.7)	49.8 (11.1)	**
Weight (kg)	71.3 (15.3)	71.0 (15.1)	77.9 (17.1)	**
Height (cm)	157.3 (9.0)	157.4 (9.1)	155.4 (8.0)	*
BMI (kg/m <sup>2</sup> )	28.79 (5.5)	28.67 (5.46)	32.15 (6.0)	**
Waist (cm)	94.8 (12.9)	94.4 (12.7)	105.4 (12.8)	**
Family history of DM (%)	36.0	35.7	37.8	
Hypertension (%)	18.0	17.4	43.9	**
SBP (mmHg)	117.8 (17.1)	117.3 (16.8)	131.1 (20.4)	**
DBP (mmHg)	72.54 (10.6)	72.3 (10.4)	78.8 (13.0)	**
Hb1Ac (%)	5.41 (0.8)	5.3 (0.4)	8.7 (2.0)	**

<sup>a</sup> Statistical significances between normal and UDM groups were assessed using the Student's *t*-test for the interval/ratio variables (Age, Weight, Height, BMI, Waist, SBP, DBP and Hb1Ac) and the  $\chi^2$ -test for the categorical variables (Sex, Family history of DM and Hypertension). A *P*-value < 0.05 was considered to be statistically significant (\**P* < 0.05, \*\**P* < 0.001).

of 53.9 and 67.5% for the female population, and 53.2 and 69.1% for the male population, respectively.

## NHNS 2006 and 2012 models

In a previous study,<sup>10</sup> we developed two screening models for undiagnosed diabetes based on data from two National Health and Nutrition Surveys carried out in Mexico in 2006 and 2012.<sup>18,19</sup> The models were developed using multiple logistic regression with a minimal set of explanatory variables determined by a stepwise backward procedure (see Appendix A.3). The explanatory variables used in the NHNS 2006 model were age, waist circumference and systolic blood pressure, while in the NHNS 2012 model the explanatory variables used were age, waist circumference, height and family history of diabetes. The models proposed achieved a sensitivity and specificity of 74 and 62% for the NHNS 2006 model and 76% and 55% for the NHNS 2012 model in an external validation dataset obtained from the National Health Survey performed in 2000.<sup>20</sup>

## New screening model for undiagnosed diabetes in Mexican adults based on HbA1c measurements

In addition to evaluating the models described above, a new screening model for undiagnosed diabetes in Mexican adults was developed using data from the halfway National Health and Nutrition Survey 2016.<sup>12</sup> The methodology followed to develop the NHNS 2016 model was similar to that used in a previous work for the development of the NHNS 2006 and 2012 models<sup>10</sup> and consisted in deriving a multivariate logistic regression model to calculate the probability *P* of having

undiagnosed diabetes (UDM) based on selected explanatory variables  $X_i$  as:

$$P(\text{UDM} = 1 | X_1, X_2, \dots) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i X_i)}} \quad (1)$$

Once the probability is calculated, each observation from the dataset is classified as healthy or at high risk of having undiagnosed diabetes according to a probability cutoff value determined through an optimization procedure using a separate optimization dataset.

The dataset used to derive the NHNS 2016 model (described in Section 2.1) was divided in development ( $n=1193$ ), optimization ( $n=596$ ) and validation ( $n=597$ ) datasets comprising 50, 25 and 25% of the complete NHNS 2016 dataset ( $n=2386$ ), respectively. The characteristics of the resulting datasets are presented in Table 2, where it can be seen that, in general, the differences between normal observations and those classified as having undiagnosed diabetes are also present in most of the variables listed, as observed in Table 1 for the complete NHNS 2016 dataset.

Explanatory variables used in the final NHNS 2016 model were determined from the development dataset by a stepwise backward procedure based in the Akaike Information Criterion (AIC) using 10-fold crossvalidation repeated five times. Sex, age, weight, height, waist circumference, family history of diabetes, hypertension status and systolic and diastolic blood pressure were initially evaluated as potential explanatory variables. In this procedure, a model containing all these explanatory variables is iteratively fitted, removing variables from the model in order to choose only those variables that contribute significantly to the model according to the AIC, calculated at each iteration. Afterwards, an optimization procedure was performed using the resulting model in order to determine the optimal probability cutoff value that maximize both the sensitivity and specificity in the optimization dataset. Finally, the

**Table 2** Characteristics of the development, optimization and validation datasets (NHNS 2016). Continuous variables are presented as means (SD) and categorical variables are presented as number of observations and percentages.

	Development				Optimization				Validation			
	Total	Normal	UDM	<i>P</i> -value <sup>a</sup>	Total	Normal	UDM	<i>P</i> -value <sup>a</sup>	Total	Normal	UDM	<i>P</i> -value <sup>a</sup>
<i>n</i>	1193	1154	39		596	576	20		597	574	23	
%	100	96.7	3.3		100	96.6		3.4	100	96.1	3.9	
Sex (%)	F: 65.4 M: 34.6	F: 65.3 M: 34.7	F: 69.2 M: 30.8		F: 67.4 M: 32.6	F: 67.5 M: 32.5	F: 65 M: 35		F: 65.6 M: 34.4	F: 65.9 M: 34.1	F: 60.9 M: 39.1	
Age (y)	41.3 (12.7)	41.0 (12.3)	49.3 (12.2)	***	41.6 (12.9)	41.2 (12.9)	51.2 (8.6)	***	40.8 (12.7)	40.4 (12.7)	49.4 (11.5)	**
Weight (kg)	71.9 (15.3)	71.7 (15.1)	77.1 (18.8)		71.1 (15.6)	70.8 (5.5)	80.7 (14.2)		70.3 (14.9)	70.0 (14.7)	76.6 (17.0)	**
Height (cm)	157.5 (9.2)	157.6 (9.1)	155.7 (9.3)		157.4 (9.0)	157.5 (9.0)	155.2 (7.0)		156.8 (8.9)	156.8 (8.9)	154.9 (6.7)	
BMI (kg/m <sup>2</sup> )	28.96 (5.6)	28.87 (5.5)	31.71 (6.5)	**	28.68 (5.6)	28.52 (5.6)	33.44 (5.3)	***	28.56 (5.3)	28.43 (5.3)	31.77 (5.8)	*
Waist (cm)	95.3 (13.0)	95.0 (12.9)	104.4 (12.4)	***	94.5 (13.0)	94.0 (12.8)	107.7 (12.8)	***	94.2 (12.5)	93.7 (12.3)	105.0 (13.6)	***
Family history of DM (%)	37.7	37.6	41.0		34.2	34.0	40.0		33.3	33.4	30.4	
Hypertension (%)	18.7	17.8	46.2	***	15.6	14.9	35.0	*	20.1	19.0	47.8	**
SBP (mmHg)	117.9 (17.2)	117.5 (16.9)	129.7 (21.2)	**	117.1 (16.1)	116.6 (15.9)	131.4 (15.4)	**	118.2 (17.7)	117.6 (17.2)	133.2 (23.4)	**
DBP (mmHg)	72.5 (10.3)	72.4 (10.1)	77.1 (12.9)	*	72.3 (10.6)	71.9 (10.4)	82.5 (12.0)	***	72.9 (11.2)	72.7 (11.1)	78.6 (13.7)	
Hb1ac (%)	5.4 (0.8)	5.3 (0.4)	8.7 (2.0)	**	5.4 (0.7)	5.3 (0.4)	8.1 (1.7)	***	5.4 (0.9)	5.3 (0.4)	9.1 (2.1)	*

<sup>a</sup> Statistical significances between normal and UDM groups were assessed using the Student's *t*-test for the interval/ratio variables (Age, Weight, Height, BMI, Waist, SBP, DBP and Hb1Ac) and the  $\chi^2$ -test for the categorical variables (Sex, Family history of DM and Hypertension). A *P*-value < 0.05 was considered to be statistically significant (\**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001).

performance of the final model, composed of the selected explanatory variables and the optimized cutoff probability value, was evaluated in the validation dataset in terms of the sensitivity, specificity and area under the receiver operation characteristic (ROC) curve (AUC).

### Computational details

R 3.5.1<sup>21</sup> was used for the evaluation, development and validation of the screening models. Fit and validation of the models were performed using the Caret package.<sup>22</sup> The area under the ROC curves and optimal probability cutoff values were calculated using the ROCR package.<sup>23</sup>

## Results and discussion

### Evaluation of screening models for undiagnosed diabetes in Mexican adults

Models described above were evaluated using the NHNS 2016 dataset. The main objective was to determine which model behaves the best in terms of its capability of correctly identifying the 82 subjects with undiagnosed diabetes while yielding the lowest number of false positives (i.e. high sensitivity and specificity). Performance parameters of all the models evaluated are shown in Table 3. Most of the models, including the waist circumference models (IDF and AHA/NHLBI), the NHNS 2012 model and the SSA risk score, showed a high sensitivity (between 0.88 and 0.99), although also showing a very low specificity, in the range of 0.17–0.36. This means that practically all the subjects with undiagnosed diabetes were identified correctly, with the drawback that around 75–80% of normal subjects would be incorrectly classified as at high risk of having the disease. The opposite scenario occurred with the INCMNSZ score risk,<sup>9</sup> which was only capable of identifying 17% of adults with undiagnosed diabetes, while classifying correctly more than 70% of normal or healthy subjects.

Imagining a scenario where these models were used as an initial step of a stepwise program for the early detection of subjects at high risk of having undiagnosed diabetes, in which those subjects detected by the screening model as at high risk of having the disease would be further evaluated by blood test analyses and/or by a specialist physician, a model with a high number of false positives would increase considerably the cost of detection of those who do have undiagnosed diabetes. This is no worse than the opposite scenario in which a screening model yields a high number of false negatives, implying that people with undiagnosed diabetes would remain undetected, which would put them at risk of developing the devastating long-term complications of diabetes, with the economic consequences associated for the patients, their families and the health system as a whole.

The only model that showed consistent balanced results was the NHNS 2006 model.<sup>10</sup> As it can be seen in Table 3, the NHNS 2006 model showed a high sensitivity of 0.82 and a specificity of 0.63. These numbers resemble those obtained by the NHNS 2006 model in the optimization and validation datasets, where the model showed a sensitivity and specificity of 0.74 and 0.62, respectively (see the original work in Ref.<sup>10</sup>). In contrast to the other models evaluated, the NHNS

2006 model was the only model that showed a consistent performance in the development (NHNS 2006), validation (NHNS 2000) and external datasets (NHNS 2016).

### NHNS 2016 model

From the stepwise backward procedure in the development dataset ( $n=1193$ , see Table 2), only age, waist circumference and systolic blood pressure were considered in the final NHNS 2016 model. Model parameters  $\alpha$  and  $\beta_i$ , as well as the corresponding odds ratios are shown in Table 4. According to the  $\beta$  coefficients, an increase of one unit in age (in years), waist circumference (in centimeters) and systolic blood pressure (in mmHg) increases the odds of having undiagnosed diabetes by 3, 4 and 2%, respectively. Interestingly, explanatory variables selected by the stepwise backward procedure in the NHNS 2016 model are exactly the same than those selected through the same procedure for the NHNS 2006 model which, as mentioned in the previous section, was the only model that showed a stable performance when evaluated in different datasets.

The performance of the NHNS 2016 model in the optimization and validation datasets (see Table 2) is summarized in Table 5. The NHNS 2016 model performed similarly in the optimization and validation datasets (using the cutoff probability value of 0.042), achieving a sensitivity of 0.75 and 0.72, respectively, and a specificity of 0.80 in both cases. These numbers translate to an AUC of 0.85 and 0.82 in the optimization and validation datasets, respectively. In practice, these numbers mean that the NHNS 2016 model correctly identified about 80% of Mexican adults with undiagnosed diabetes in the NHNS 2016 dataset, with only between 18 to 15% of false positives. In comparison to the NHNS 2006 model, which achieved specificity values of 61–63% (in the optimization and external validation datasets), the NHNS 2016 model was better at identifying healthy subjects, as indicated by its greater specificity (between 82 and 85%). Regarding the sensitivity, both models showed similar values (0.72–0.75 for the NHNS 2016 model and 0.74–0.75 for the NHNS 2006 model). The main difference between these two models is therefore that the NHNS 2016 showed a higher specificity, which may be reflecting the fact that HbA1c measurements were used in the NHNS 2016 dataset to determine which observations were classified as having undiagnosed diabetes, in contrast to the NHNS 2006 model in which measurements of fasting plasma glucose were used. In spite of these optimistic results, external validation of the NHNS 2016 model must be performed in order to determine if it shows a stable and balanced performance in different datasets.

A comparison between the performance of the NHNS 2016 model and other models developed for other countries and/or populations is shown in Table 6. In terms of the capability of the models to correctly identify subjects with undiagnosed diabetes (i.e. sensitivity), the NHNS 2016 model showed a slightly lower value (0.72) with respect to the mean of the models included in Table 6 (0.76), although also showing a considerably higher specificity (0.80) if compared to the average value of the specificity achieved by the other models (0.61). In general, as can be seen in Table 6, the models developed for the Mexican

**Table 3** Comparison between screening models for undiagnosed diabetes in Mexican adults.

Model	Authors	Ref.	Sens.	Spec
Waist circumference (IDF)	Rojas-Martínez et al.	8	0.98	0.2
Waist circumference (AHA/NHLBI)	Rojas-Martínez et al.	8	0.88	0.36
SSA risk score (2015)	Mexican Secretariat of Health	11	0.98	0.19
INCMNSZ risk score	Rojas-Martínez et al.	9	0.26	0.71
NHNS 2006	Félix-Martínez and Godínez-Fernández	10	0.82	0.63
NHNS 2012	Félix-Martínez and Godínez-Fernández	10	0.99	0.17
NHNS 2016	Félix-Martínez and Godínez-Fernández	*	0.72	0.80

\* This work.

**Table 4** Explanatory variables included in the NHNS 2016 model, odds ratios and model parameters  $\alpha$ ,  $\beta$  used in Eq. (1) to calculate the probability of having undiagnosed diabetes.

Variable	$\beta$ coefficient ( $\pm$ SE)	OR (95% CI)
$\alpha$	-10.89 ( $\pm$ 1.44)	1.86E-5 (1.1E-6-3.1E-4)
Age (y)	0.03 ( $\pm$ 0.01)	1.03 (1.00-1.06)
Waist circumference (cm)	0.04 ( $\pm$ 0.01)	1.04 (1.02-1.06)
SBP (mmHg)	0.02 ( $\pm$ 0.01)	1.02 (1.01-1.03)

**Table 5** Performance of the NHNS 2016 model. Cutoff probability value = 0.042.

Cutoff optimization dataset		
Sens.	Spec.	AUC
0.75	0.80	0.85
Validation dataset		
Sens.	Spec.	AUC
0.72	0.80	0.82

**Table 6** Comparison between the NHNS 2016 model and similar models developed for other countries, regions or populations.

Population	Sensitivity	Specificity	Ref.
Mexico (NHNS 2016)	0.72	0.80	This work
Mexico (NHNS 2012)	0.76	0.55	10
Mexico (NHNS 2006)	0.74	0.62	10
Oman	0.63	0.78	26
India	0.73	0.56	27
China	0.84	0.40	28
Denmark	0.76	0.72	29
Thailand	0.87	0.38	30
Brazil	0.86	0.45	31
Netherlands	0.72	0.56	32
USA	0.79	0.67	33
Egypt	0.62	0.96	34
China	0.73	0.64	35
Peru	0.85	0.47	36
Europe	0.68	0.81	37
USA (Hispanic)	0.80	0.61	38
Colombia	0.73	0.67	39
Slovenia	0.76	0.73	40
United Arab Emirates	0.75	0.70	41
Sri Lanka	0.78	0.66	42
Average	0.76	0.61	

adult population perform within the range of the models developed for other countries or populations. It is worth noting that the NHNS 2006, 2012 and 2016 models were designed as screening tools for undiagnosed diabetes, thus leaving aside the possibility of identifying those subjects with prediabetes. This is an important aspect to consider in future models given the fact that the prevalence of prediabetes in the Mexican adult population is also high, as indicated by a study performed by Kumar et al.,<sup>24</sup> who calculated a prevalence of prediabetes of 44.2% from a subsample of the 2012 Mexican Health and Aging Study, in which Mexican adults older than 50 years of age were surveyed in both rural and urban areas of Mexico to study the aging process and its impact on their health state.<sup>25</sup>

## Conclusions

Screening models for undiagnosed diabetes in Mexican adults were evaluated using data from the Mexican halfway National Health and Nutrition Survey 2016. According to our results, the NHNS 2006 model was the only model that showed similar performance indicators than those reported previously in the optimization and external validation datasets,<sup>10</sup> as indicated by its capability of identifying correctly 82% of subjects with undiagnosed diabetes with a reasonable percentage of false positives of 37%.

A new screening model for undiagnosed diabetes in the Mexican adult population, derived using recent data from the NHNS 2016 survey, was also proposed. As the NHNS 2006 model, the NHNS 2016 model considered age, waist circumference and systolic blood pressure as explanatory variables and correctly identified 72% of subjects with undiagnosed diabetes with a low percentage of false positives of 20%. The fact that the stepwise backward procedure in both the NHNS 2006 and 2016 models yielded the same explanatory variables to determine the probability of having the disease, allow us to conclude that age, waist circumference and systolic blood pressure are of special importance for the early detection of Mexican adults with undiagnosed diabetes. According to our results, the use of HbA1c measurements to classify subjects as normal or as having undiagnosed diabetes for the development of the NHNS 2016 model contributed to achieve a higher specificity (0.8), in comparison to the NHNS 2006 and 2012 models, in which subjects were classified as normal or with undiagnosed diabetes based on measurements of the fasting plasma glucose. In spite of this, classification using HbA1c for the development of the NHNS 2016 model did not lead to an improvement of the sensitivity achieved by our previous models (NHNS 2006 and 2012<sup>10</sup>).

Despite the fact that the NHNS 2016 model showed a better performance than the NHNS 2006 model (when evaluated in the NHNS 2016 dataset), the NHNS 2016 model must be tested in different datasets in order to confirm the results here presented. For this reason, in accordance with the results of the comparative analysis performed in this work, the use of the NHNS 2006 screening model is recommended for the detection of Mexican adults with undiagnosed diabetes.

## Conflict of interests

The authors declare no conflict of interest.

## Acknowledgments

G.J. Félix-Martínez (CVU 302514) acknowledges CONACYT (National Council of Science and Technology, México) for the support given to this project.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.plantsci.2004.08.011](https://doi.org/10.1016/j.plantsci.2004.08.011)

## References

1. International Diabetes Federation, Brussels, Belgium. *Diabetes Atlas* (8th ed.); 2017.
2. Harris MI, Klein R, Welborn TA, Knudman MW. Onset of NIDDM occurs at least 4–7 yr before clinical diagnosis. *Diabetes Care*. 1992;15:815–9.
3. Forbes JM, Cooper ME. Mechanisms of diabetic complications. *Physiol Rev*. 2013;93:137–88.
4. American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2019. *Diabetes Care*. 2019; 42(Suppl. 1):S13.
5. Brown N, Critchley J, Bogowicz P, Mayige M, Unwin N. Risk scores based on self-reported or available clinical data to detect undiagnosed type 2 diabetes: a systematic review. *Diabetes Res Clin Pract*. 2012;98:369–85.
6. Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. *BMJ*. 2011;343:d7163.
7. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med*. 2011;9:103.
8. Rojas-Martínez R, Aguilar-Salinas CA, Jiménez-Corona A. Optimal cutoff points for the detection of undiagnosed type 2 diabetes, hypertension and metabolic syndrome in Mexican adults. *Salud Pública de México*. 2012;54:13–9.
9. Rojas-Martínez R, Escamilla-Núñez C, Gómez-Velasco DV, Zárate-Rojas E, Aguilar-Salinas CA, Grupo Colaborador de la Cohorte para Estimar la Incidencia del Síndrome Metabólico. Diseño y validación de un score para detectar adultos con prediabetes y diabetes no diagnosticada. *Salud Pública de México*. 2018;60:500–9.
10. Félix-Martínez GJ, Godínez-Fernández JR. Screening models for undiagnosed diabetes in Mexican adults using clinical and self-reported information. *Endocrinol Diabetes Nutr (English ed)*. 2018;65:603–10.
11. Cuestionario de Factores de Riesgo. <https://www.gob.mx/salud/documentos/cuestionario-de-factores-de-riesgo.2015> [accessed 23.03.19].
12. Shamah T, Cuevas L, Rivera J, Hernández M. Encuesta Nacional de Salud y Nutrición de Medio Camino 2016: Informe final de resultados. Cuernavaca, México: Instituto Nacional de Salud Pública; 2016. p. 2016.
13. Romero-Martínez M, Shamah-Levy T, Cuevas-Nasu L, Gómez-Humarán IM, Gaona-Pineda EB, Gómez-Acosta LM, et al. Diseño metodológico de la Encuesta Nacional de Salud y Nutrición de Medio Camino 2016. *Salud Pública de México*. 2017;59:299–305.



14. Committee IE, et al. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care*. 2009;32:1327–34.
15. Grundy SM, Cleeman JI, Daniels SR, Donato KA, Eckel RH, Franklin BA, et al. Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart Lung, and Blood Institute scientific statement. *Circulation*. 2005;112:2735–52.
16. Alberti KGM, Zimmet P, Shaw J. The metabolic syndrome—a new worldwide definition. *Lancet*. 2005;366:1059–62.
17. Herman WH, Smith PJ, Thompson TJ, Engelgau MM, Aubert RE. A new and simple questionnaire to identify people at increased risk for undiagnosed diabetes. *Diabetes Care*. 1995;18:382–7.
18. Gutierrez JP, Rivera-Dommarco J, Shamah-Levy T, Villalpando-Hernández S, Franco A, Cuevas-Nasu L, et al. Encuesta Nacional de Salud y Nutrición 2012. Resultados Nacionales Cuernavaca, 1. México: Instituto Nacional de Salud Pública; 2012 (1.48).
19. Abúndez CO, Cázares GN, Cordero JFR, Zetina DAD, Angona SR, de Voghel Gutiérrez S, et al. Encuesta Nacional de Salud y Nutrición [2006] Cuernavaca. México: Instituto Nacional de Salud Pública; 2006.
20. Olaiz G, Barquera S, Shamah-Levy T, Aguilar-Salinas C, Sepúlveda J, Tapia-Conyer R, et al. Encuesta Nacional de Salud 2000. La salud de los adultos Cuernavaca. México: Instituto Nacional de Salud Pública; 2003.
21. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2018. Available from: <https://www.R-project.org/>
22. Kuhn M. Building predictive models in R using the caret package. *J Stat Soft*. 2008;28:1–26. Available from: <https://www.jstatsoft.org/v028/i05>
23. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCr: visualizing classifier performance in R. *Bioinformatics*. 2005;21:3940–1.
24. Kumar A, Wong R, Ottenbacher KJ, Al Snih S. Prediabetes, undiagnosed diabetes, and diabetes among Mexican adults: findings from the Mexican Health and Aging Study. *Ann Epidemiol*. 2016;26:163–70.
25. Wong R, Michaels-Obregon A, Palloni A. Cohort profile: the Mexican health and aging study (MHAS). *Int J Epidemiol*. 2017;46:e2.
26. Al-Lawati J, Tuomilehto J. Diabetes risk score in Oman: a tool to identify prevalent type 2 diabetes among Arabs of the Middle East. *Diabetes Res Clin Pract*. 2007;77:438–44.
27. 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 Prevent Control*. 2008;3:141–51.
28. Gao W, Dong Y, Pang Z, Nan H, Wang S, Ren J, et al. A simple Chinese risk score for undiagnosed diabetes. *Diabetic Med*. 2010;27:274–81.
29. Glümer C, Carstensen B, Sandbæk A, Lauritzen T, Jørgensen T, Borch-Johnsen K. A Danish diabetes risk score for targeted screening: the Inter99 study. *Diabetes Care*. 2004;27:727–33.
30. Keesukphan P, Chanprasertyothin S, Ongphiphadhanakul B, Puavilail G. The development and validation of a diabetes risk score for high-risk Thai adults. *J-Med Assoc Thailand*. 2007;90:149.
31. De Sousa AGP, Pereira AC, Marquezine GF, do Nascimento-Neto RM, Freitas SN, Nicolato RLdC, et al. Derivation and external validation of a simple prediction model for the diagnosis of type 2 diabetes mellitus in the Brazilian urban population. *Eur J Epidemiol*. 2009;24:101.
32. Ruige JB, De Neeling JND, Kostense PJ, Bouter LM, Heine RJ. Performance of an NIDDM screening questionnaire based on symptoms and risk factors. *Diabetes Care*. 1997;20:491–6.
33. Bang H, Edwards AM, Bomback AS, Ballantyne CM, Brillon D, Callahan MA, et al. Development and validation of a patient self-assessment score for diabetes risk. *Ann Internal Med*. 2009;151:775–83.
34. Tabaei BP, Herman WH. A multivariate logistic regression equation to screen for diabetes: development and validation. *Diabetes Care*. 2002;25:1999–2003.
35. Liu M, Pan C, Jin M. A Chinese diabetes risk score for screening of undiagnosed diabetes and abnormal glucose tolerance. *Diabetes Technol Therap*. 2011;13:501–7.
36. Bernabe-Ortiz A, Perel P, Miranda JJ, Smeeth L. Diagnostic accuracy of the Finnish Diabetes Risk Score (FINDRISC) for undiagnosed T2DM in Peruvian population. *Primary Care Diabetes*. 2018;12:517–25.
37. Mavrogianni C, Lambrinou CP, Androutsos O, Lindström J, Kivela J, Cardon G, et al. Evaluation of the Finnish Diabetes Risk Score as a screening tool for undiagnosed type 2 diabetes and dysglycaemia among early middle-aged adults in a large-scale European cohort The Feel4Diabetes-study. *Diabetes Res Clin Pract*. 2019;150:99–110.
38. Juarez LD, Gonzalez JS, Agne AA, Kulczycki A, Pavela G, Carson AP, et al. Diabetes risk scores for Hispanics living in the United States: a systematic review. *Diabetes Res Clin Pract*. 2018;142:120–9.
39. Barengo NC, Tamayo DC, Tono T, Tuomilehto J. A Colombian diabetes risk score for detecting undiagnosed diabetes and impaired glucose regulation. *Primary Care Diabetes*. 2017;11:86–93.
40. Stiglic G, Kocbek P, Cilar L, Fijačko N, Stožer A, Zaletel J, et al. Development of a screening tool using electronic health records for undiagnosed Type 2 diabetes mellitus and impaired fasting glucose detection in the Slovenian population. *Diabetic Med*. 2018;35:640–9.
41. Sulaiman N, Mahmoud I, Hussein A, Elbadawi S, Abusnana S, Zimmet P, et al. Diabetes risk score in the United Arab Emirates: a screening tool for the early detection of type 2 diabetes mellitus. *BMJ Open Diabetes Res Care*. 2018;6:e000489.
42. Katulanda P, Hill N, Stratton I, Sheriff R, De Silva S, Matthews D. Development and validation of a Diabetes Risk Score for screening undiagnosed diabetes in Sri Lanka (SLDRISK). *BMC Endocrine Disord*. 2016;16:42.