Can J Diabetes 39 (2015) 152–156

Contents lists available at ScienceDirect

Canadian Journal of Diabetes

journal homepage: www.canadianjournalofdiabetes.com

Original Research

Evaluating the Performance of the Framingham Diabetes Risk Scoring Model in Canadian Electronic Medical Records

Morteza Mashayekhi MBA [a,*], Franklyn Prescod PhD [a], Bharat Shah MBA, P Eng [a], Linying Dong PhD [a], Karim Keshavjee MD, MBA [b,c], Aziz Guergachi PhD [a]

[a] Ted Rogers School of Information Technology Management, Ryerson University, Toronto, Ontario, Canada
[b] InfoClin Inc, Toronto, Ontario, Canada
[c] University of Victoria, School of Health Informatics, Victoria, British Columbia, Canada

ARTICLE INFO

Article history:
Received 5 August 2014
Received in revised form
6 October 2014
Accepted 7 October 2014

Keywords:
Canadian population
diabetes mellitus
electronic medical records
Framingham offspring study
prediabetes

ABSTRACT

Objective: The objective of this study was to evaluate the performance of the Framingham Diabetes Risk Scoring Model (FDRSM) in a Canadian population, using the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) database.
Methods: We analyzed the records of 571 631 patients, between the ages of 45 and 64, between 2002 and 2005, by extracting the most recent laboratory and examination results, including age, sex, body mass index, fasting blood glucose, high-density lipoprotein, triglycerides and blood pressure. We calculated the risk scores of these patients based on the FDRSM. We tracked these patients for 8 years to find out whether or not they were diagnosed with diabetes. We used the area under the receiver operating characteristics curve (AROC) to estimate the discrimination capability of the FDRSM on our study sample and compared it with the AROC reported in the original Framingham diabetes study.
Results: The AROC for our main research sample of 1970 patients for whom all risk factors and follow-up data were available was 78.6% compared to the AROC of 85% reported in the FDRSM. We found that 70.1% of our main sample had risks lower than 3%; 16.3% had risks between 3% and 10%; and 13.6% had risks greater than 10% for diabetes over the following 8-year period.
Conclusions: The discrimination capability of the FDRSM Canadian electronic medical records is fair. However, building a more accurate model for predicting diabetes based on the characteristics of Canadian patients is highly recommended.

Mots clés :
population canadienne
diabète sucré
dossiers médicaux électroniques
étude Framingham offspring study
prédiabète

RÉSUMÉ

Objectif : L'objectif de cette étude était d'évaluer la performance du modèle de scores de risque de diabète de Framingham (MSRDF) auprès d'une population canadienne en utilisant la banque de données du Réseau canadien de surveillance sentinelle en soins primaires (RCSSSP).
Méthodes : Nous avons analysé les dossiers de 571 631 patients de 45 à 64 ans au cours des années 2002 à 2005 en extrayant les plus récents résultats de laboratoire et d'examens, y compris l'âge, le sexe, l'indice de masse corporelle, la glycémie à jeun, les lipoprotéines de haute densité, les triglycérides et la pression artérielle. Nous avons calculé les scores de risque de ces patients selon le MSRDF. Nous avons suivi ces patients durant 8 ans pour savoir s'ils allaient recevoir ou non un diagnostic de diabète. Nous avons utilisé la surface sous la courbe caractéristique d'efficacité du récepteur (ROC) pour estimer la capacité de discrimination du MSRDF sur notre échantillon d'étude et l'avons comparé à la surface sous la courbe ROC rapportée dans l'étude originale de Framingham sur le diabète.
Résultats : La surface sous la courbe ROC de notre principal échantillon de recherche qui regroupait 1970 patients pour lesquels les facteurs de risque et les données de suivi étaient disponibles a été de 78,6 % comparativement à la surface sous la courbe de 85 % rapportée dans le MSRDF. Nous avons observé qu'au cours des 8 années subséquentes 70,1 % de notre échantillon principal avaient des risques de diabète inférieurs à 3 %, 16,3 % avaient des risques entre 3 % et 10 % et 13,6 % avaient des risques supérieurs à 10 %.

* Address for correspondence: Morteza Mashayekhi, MBA, Suite 2802, 7171 Yonge Street, Thornhill, Ontario L3T 0C5, Canada.
E-mail address: Morteza.mashayekhi@ryerson.ca

1499-2671/$ – see front matter © 2015 Canadian Diabetes Association
http://dx.doi.org/10.1016/j.jcjd.2014.10.006

Conclusions : La capacité de discrimination des dossiers médicaux électroniques du Canada selon le MSRDF est honnête. Cependant, l'élaboration d'un modèle plus précis de prédiction du diabète s'appuyant sur les caractéristiques des patients canadiens est fortement recommandée.

Introduction

According to the Canadian Diabetes Association, between 2010 and 2020, the number of people diagnosed with diabetes in Canada is expected to rise from 2.5 million to about 3.7 million, which brings the proportion of the total population with diabetes from 7.3% in 2010 to 9.9% by 2020 (1–3). The economic burden of such an increase in the prevalence of diabetes in Canada is significant. In 2010, the direct economic cost of diabetes in Canada was about $12.2 billion, accounting for about 3.5% of public healthcare spending in Canada (1). It is expected that the cost of diabetes in Canada will rise by another $4.7 billion by 2020 and will reach $16.9 billion, given the expected increase in the number of people with diabetes in Canada (1). Primary prevention interventions to halt or delay the onset of diabetes are the best ways to reduce the costs of diabetes but are dependent on the extent to which physicians are able to identify readily patients at high risk for developing diabetes.

In this respect, electronic medical records have been reported to play an important role in the early diagnosis of various diseases (4–6). A British study using data from electronic medical records (EMRs) showed that women diagnosed with ovarian cancer had complained of symptoms weeks to months before the diagnoses were made (4). A study conducted by Woods et al (5) of EMRs showed that as much as 36% of preventable adverse events were related to diagnostic errors. Undoubtedly, the uptake of electronic medical records in recent years presents a novel opportunity to review and mine previous records easily for clues to future diagnoses. In contrast to administrative databases and survey-based, self-reported data, which usually suffer from low sensitivity, poor accuracy and lack of clinical data, EMR datasets have rich clinical information as well as fair sensitivity and accuracy (6–8). With access to historical EMR data, it may be possible to develop algorithms that can identify diseases earlier, even before physicians might otherwise suspect that patients have disease; this could lead to improved and more cost-effective patient care (9–11). Such disease-detection algorithms might decrease the need for tests, provide a comprehensive knowledge base for physicians to perform diagnoses, and improve service delivery to patients. From patients' perspectives, the development of these algorithms could provide faster diagnosis, allowing for treatment to be administered sooner. This type of research is increasingly done using EMRs (7,12–15).

In this study, we used data extracted from the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) database. The CPCSSN is Canada's first multidisease EMR-based surveillance system. It includes 10 practice-based primary care research networks in 8 provinces across Canada. Data from all participating networks, provided by family physicians and other primary care providers, are aggregated in a single national database (16,17; http://cpcssn.ca/).

The main objective of this study was to evaluate the performance of the Framingham Diabetes Risk Scoring Model (FDRSM) in a Canadian primary care EMR data set. There are various scoring models for diabetes that aim to identify patients at high risk for developing diabetes (18,19). Among them, the FDRSM is well known and widely used in a variety of populations (20–22). The Framingham offspring study was published in 2007 and proposed different models for the prediction of incident diabetes mellitus in middle-aged adults (45 to 64 years of age) (23). The Framingham simple clinical model included 8 factors (age, gender, blood pressure, fasting blood glucose, triglycerides, high-density lipoprotein [HDL], body mass index [BMI], and parental history of diabetes) in order to predict the 8-year risk for developing diabetes (23). The FDRSM excludes age and gender and calculates the overall risk for diabetes by using blood pressure, fasting blood glucose, triglycerides, HDL, BMI and parental history of diabetes. The FDRSM allows a simple estimation of the 8-year risk for developing diabetes as follows: fasting blood glucose levels between 5.4 and 6.9 mmol/L (10 points); BMIs of 30.0 or greater (5 points); HDL-C levels less than 0.9 mmol/L in men and 1.2 mmol/L in women (5 points) positive parental histories (3 points); triglyceride levels greater than 1.7 mmol/L (3 points); BMIs of 25.0 to 29.9 (2 points) and elevated blood pressure greater than 130/85 mm Hg (2 points) (23).

Identifying patients with high risks for diabetes by using a validated scoring model, such as the FDRSM, enables physicians to implement prevention interventions in those patients and, as a result, halt or delay the onset of the disease. It may also help in realizing cost avoidance for the healthcare system.

Methods

We obtained data from the CPCSSN database. We extracted the records of 571 631 patients from the CPCSSN 2013Q3 database. The average age of patients in the database is 44.78 years and approximately 56% are women. There are 666 485 records of diseases in the database because some patients are diagnosed with more than 1 disease and have multiple records as a result. Approximately 9.4% of patients were diagnosed with diabetes. The average BMIs of patients, based on the most recent record for each patient, was 26.54, with a standard deviation of 7.37. All records for blood pressure, HDL, triglycerides and fasting blood glucose are based on the most recent laboratory results for each patient. All laboratory results in the CPCSSN database were recorded in mmol/L. Table 1 shows the characteristics of the patient population in the CPCSSN database and the most recent data for each patient.

Our study sample included patients who had laboratory results for all variables that were addressed by the FDRSM between 2002 and 2005. Like the FDRSM, we included all middle-aged patients (from 45 to 64 years of age). We developed the research cohorts as follows: we extracted the information of patients with the most recent laboratory and examination results in 2002, including age, sex, BMI, fasting blood glucose levels, HDL, triglycerides and blood pressure and then calculated the risk scores of these patients based on the FDRSM. The parental histories of diabetes were not available in the database, so we excluded this variable. We tracked these patients for 8 years, from January 1, 2003, to December 31, 2010, to find out whether they had been diagnosed with diabetes. We repeated the procedure for the years 2003 to 2005 and merged all of the information to develop the final research sample.

We used the area under the receiver operating characteristic curve (AROC) or C statistic as a tool for estimating the discriminant capability of the FDRSM in Canadian primary care patients (24–26). To assess the within-study scoring model validity, we used a jackknife procedure. We took 10 random samples of 90% of the participants and found the AROC for each sample. We also compared the means of overall diabetes risk scores based on the FDRSM for a number of groups of patients with different diseases in order to