Diabetes Risk Assessment – Application and Validation of a Diabetes Screening Score Applied to the National Health and Wellness Survey

Paul C. Langley, Ph.D., Adjunct Professor, College of Pharmacy, University of Minnesota; Heejung Bang, Ph.D., Associate Professor, Weill Medical College, Cornell University; Hillary Gross, Kantar Health; Kathy Annunziata, Kantar Health

Abstract

OBJECTIVES: To apply a risk assessment algorithm to a national US population sample (i) to determine the prevalence of at-risk pre-diabetic and undiagnosed diabetic participants and (ii) to validate the role of the risk factors included in the scoring algorithm in the prediction of diagnosed type 2 diabetes.

METHODS: Data from the 2009 National Health and Wellness Survey (NHWS), a national survey of the US population, were used to apply an adapted version of a recently developed scoring algorithm/prediction model to identify patients at risk for screening for pre-diabetes and undiagnosed diabetes. The algorithm combines six risk factors to generate total risk scores: age, gender, family history of hypertension, high blood pressure, obesity, and physical activity. Logistic regression analysis was then used to fit the prediction model (risk logistic regression) in order to assess the statistical significance of the identified risk factors. Other validation measures were also computed.

RESULTS: Among those 18 years of age and over in the NHWS sample (n=74,474) with a confirmed diagnosis of diabetes, a 27.3% of participants were estimated to be at elevated risk for diabetes and a further 16.5% for pre-diabetes. The fitted prediction model indicated the following variables were strongest risk factors in the NHWS sample: (i) age 60 years and over (odds ratio 6.90); (ii) morbid obesity (odds ratio 6.11); (iii) a family history of diabetes (odds ratio 4.70); and (iv) being obese (odds ratio 3.04). The sensitivity was 23.2%, specificity 97.8%, positive predictive value 57.3% and negative predictive value 91.0%, with area under the ROC curve of 0.847.

CONCLUSIONS: This study seems to provide strong support for the new scoring algorithm to identify both pre-diabetic and diabetic at-risk populations, although some modifications may be necessary for the majority of existing databases and further validation among diverse populations in various real world settings are warranted.

Introduction

In a recent paper, Bang et al. (2009) developed a new diabetes screening score for undiagnosed diabetes in a multicentric adult US population. The objective was to improve existing algorithms for diabetes risk-scoring by using recent NHANES data (1999-2004) and to develop an algorithm that could easily be incorporated into software for use in primary health care settings.

The algorithm proposed by Bang et al. (2009) comprises questions on age, sex, family history of diabetes, history of hypertension, obesity and physical activity. Scores range from -1 to 9 with a designated cut-off point to designate individuals at low risk for undiagnosed diabetes.

The algorithm was tested by Bang et al. (2009) to determine the prevalence of at-risk pre-diabetic and undiagnosed diabetic participants and to validate the role of the risk factors included in the scoring algorithm in the prediction of diagnosed type 2 diabetes.

Methods

Data from the 2009 National Health and Wellness Survey (NHWS), an internet-based national survey of the US adult population, were used to assess the role of the algorithm in at risk for screening for pre-diabetes and undiagnosed diabetes and to validate the algorithm.

For the analysis, the NHWS sample excluded those respondents who indicated they had a confirmed diagnosis of diabetes. The prevalence estimates were generated directly from the NHWS data set by application of the scoring algorithm.

There was a minor modification to the activity measure compared to the Bang et al (2009) definition to meet NHWS activity requirements. The algorithm was validated by fitting the logistic regression model with the six risk factors identified by Bang et al. (2009) to generate the scoring algorithm. Validation involved (i) assessing the magnitude and significance of the model parameters (e.g., odds ratios) and (ii) generating measures of sensitivity, specificity, positive predictive value, negative predictive value and ROC analysis.

Results

Application of the risk score algorithm to the adult US population identified a total of 27.3% of the adult population as being ‘at risk’ for a diagnosis of diabetes (Table 1). A further 16.5% were identified as ‘at risk’ for a pre-diabetes diagnosis. Among those who reported they had a confirmed diagnosis of diabetes, the risk score identified 77.4% as ‘at risk’ and a further 12.2% as ‘pre-diabetes’ risk.

Logistic regression results are presented in Table 2. The fitted prediction model indicated the following variables were strongest risk factors in the NHWS sample: (i) age 60 years and over (odds ratio 6.90); (ii) morbid obesity (odds ratio 6.11); (iii) a family history of diabetes (odds ratio 4.70); and (iv) being obese (odds ratio 3.04). Physical activity was not significant.

The sensitivity was 23.2%, specificity 97.8%, positive predictive value 57.3% and negative predictive value 91.0%, with area under the ROC curve (AUC) of 0.847 (Table 3).

Discussion

The NHWS has been used on a number of occasions to validate patient-reported outcome (PRO) instruments. This is the first time it has been used to apply and validate a scoring algorithm to assess populations at potential risk in the US population.

The results from this first application of the algorithm suggest that among those without a confirmed diagnosis of diabetes, a possible 27.3% of the adult population (an estimated 40.5 million) are at risk for a diagnosis of diabetes and a further 16.5% (an estimated 30.3 million) at risk of a pre-diabetes diagnosis.

The modelled logistic regression results support the validity of the scoring algorithm. With the exception of physical activity (which is generally difficult to measure accurately), all other characteristics identified by Bang et al (2009) entered with acceptable odds ratios. The validation measures (e.g., combination of sensitivity and specificity, AUC) were also acceptable.

Conclusions

This study seems to provide strong support for the new scoring algorithm to identify both pre-diabetic and diabetic at-risk populations, although some modifications may be necessary for the majority of existing databases and further validation among diverse populations in various real world settings are warranted.

References