OJPHI

ISDS 2013 Conference Abstracts

ISDS INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

A Probabilistic Case-finding Algorithm for Chronic Disease Surveillance

Stephanie Brien*[1], Luke Mondor[1], Nancy Mayo[2] and David Buckeridge[1]

[1]McGill University, Montreal, QC, Canada; [2]McGill University Health Center, Montreal, QC, Canada

Objective

To develop and validate a multivariable probabilistic algorithm for detecting cases of diabetes mellitus (DM) using clinical and demographic data.

Introduction

There is a clear need for improved surveillance of chronic diseases to guide public health practice and policy. Chronic disease surveillance has tended to use administrative data, due to the need to link encounters for an individual over time and to have complete capture of all encounters. Case-detection algorithms generally combine variables found in the data using Boolean operators (i.e., AND, OR, NOT). For example, a commonly used algorithm for DM surveillance requires a patient to have one hospitalization or two physician visits within two years with a diagnostic code for DM. While this approach to defining case-detection algorithms is straightforward, it has limitations. For example, if more than simple combinations of one or two variables are used, then it becomes unwieldy to represent the algorithm and it can be difficult to identity how different variables in the definition contribute to detection accuracy.

A multivariable probabilistic case-detection algorithm can address these problems and facilitate exploration of how the multiple variables available from different data sources might improve case-detection accuracy[1]. In this research, we develop an approach for probabilistic multivariable case-detection and apply the method to a cohort of older adults with known DM status to demonstrate and evaluate the method.

Methods

The cohort was drawn from residents of the region of Sherbrooke, Quebec. Eligible subjects were ≥65 years of age between 2002 and 2007, lived in Sherbrooke for at least three consecutive years, were insured under the provincial prescription drug plan, and had at least one laboratory result processed at a hospital or community laboratory in the region. Individuals with elevated plasma glucose, 2-hour plasma glucose or glycated hemoglobin tests were considered to have DM.

Data on predictors of DM were obtained from Quebec's administrative databases, which hold linked information of physician visits, hospitalizations, and drug prescriptions. We considered the following predictors: age at baseline, male sex and at least one diabetes-related hospitalization (1H), physician visit (1P) or dispensed medication (1Rx). We also considered two physician visits within two years (2P2Y). Individuals with International Classification of Diseases 9th Revision (ICD-9) or ICD-10 diagnostic codes, 250.x or E10.x-E14.x, respectively, were considered to have a DM-related service. Oral hypoglycemic agents were used to identify medications dispensed for diabetes.

We created probabilistic case-detection algorithms by sequentially adding predictors to a logistic regression model. We internally validated the models using a 2:1 split sample approach. Models were compared using measures of goodness of fit (e.g., AIC), discrimination and accuracy, area under the receiver operating curve (AUC), and sensitivity, specificity, PPV and NPV calculated at the optimal threshold of the receiver operating curve.

Results

A total of 9,507 people ≥65 years of age were included in the study. Two thousand fifty-three (21.6%) had DM. All models performed well, with sensitivity specificity, and AUC equal to or greater than 0.75, 0.90 and 0.85, respectively (Table 1). Dispensed medications improved model performance and accuracy for all combinations. Adding sex and age slightly increased the discriminative performance of the models however they did not improve model accuracy. Model #4, incorporating all predictors of DM, performed better than all other models.

Conclusions

We used a multivariable probabilistic case-detection model, incorporating male sex, age, at least one hospitalization, physician visit and drug dispensed for diabetes, to develop an accurate case ascertainment algorithm for diabetes with excellent performance. In future research, we will attempt to further increase the accuracy and performance of the diabetes case ascertainment algorithm by including other variables in the model, such as socioeconomic status and risk factors, complications and comorbidities of DM.

Table 1. Accuracy, Goodness of Fit and Discriminative Performance of Models

Model	Sensitivity	Specificity	PPV	NPV	AIC	AUC (95% CI)
1) 1H+1P	0.81	0.90	0.70	0.95	3545.7	0.87 (0.86, 0.89)
2) 1H+2P2Y	0.75	0.95	0.81	0.93	3519.8	0.85 (0.84, 0.87)
3) 1H+1P+1Rx	0.83	0.90	0.70	0.95	3246.2	0.89 (0.87, 0.90)
4) 1H+1P+1Rx+male sex+age	0.83	0.90	0.70	0.95	3226.6	0.91 (0.89, 0.92)
5) 1H+2P2Y+1Rx	0.77	0.95	0.80	0.94	3304.4	0.87 (0.85, 0.88)
6) 1H+2P2Y+1Rx+male sex+age	0.77	0.95	0.80	0.94	3280.5	0.87 (0.85, 0.88)

Keywords

Diabetes Mellitus; Case-detection; Algorithm; Administrative data

Acknowledgments

This research was supported by the CIHR.

References

1. Van Walraven C, Austin PC, Manuel D, Knoll G, Jennings A and Forster AJ. The usefulness of administrative databases for identifying disease cohorts is increased with a multivariate model. J Clin Epidemiol. 2010;63:1332–41.

*Stephanie Brien
E-mail: brien.steph@gmail.com