

Aiding the practice of tuberculosis control: a decision support model to predict transmission

Hiroshi Mamiya*, Kevin Schwartzman, Aman Verma, Christian Jauvin, Marcel Behr and David Buckeridge

McGill University, Montreal, QC, Canada

Objective

To develop and validate a prediction model that estimates the probability of a newly diagnosed tuberculosis (TB) case being involved in ongoing chain of transmission, based on the case's clinical and sociodemographic attributes available at the time of diagnosis.

Introduction

A new TB case can be classified as: (1) a source case for transmission leading to other, secondary active TB cases; (2) a secondary case, resulting from recent transmission; or (3) an isolated case, uninvolved in recent transmission (i.e., neither source nor recipient). Source and secondary cases require more intense intervention due to their involvement in a chain of transmission; thus, accurate and rapid classification of new patients should help public health personnel to effectively prioritize control activities. However, the currently accepted method for classification, DNA fingerprint analysis, takes many weeks to produce the results (1); therefore, public health personnel often solely rely on their intuition to identify the case who is most likely to be involved in transmission. Various clinical and sociodemographic features are known to be associated with TB transmission (2). By using these readily available data at the time of diagnosis, it is possible to rapidly estimate the probabilities of the case being source, secondary and isolated.

Methods

A multinomial logistic regression model was developed based on the information of 1552 TB cases reported on the island of Montreal between 1996 and 2007. The predictor variables were age, sex, area of residence in Montreal, country of origin, presence of cavitory lesion in chest X-ray, sputum smear result, HIV infection, illicit drug use, living in apartment, history of TB diagnosis, being alcoholic, and coughing. DNA fingerprint analysis was used as the reference standard to define the dependent variable of the model. The data were multiply imputed, and model selection was performed by Bayesian Model Averaging. Cross-validation was performed on each of the imputed datasets to measure the predictive performance of the model using the area under the receiver operating curve (AUC).

Results

A total of 1552 cases comprising 107(6.9%) source cases, 207(13.4%) secondary cases and 1238 (79.8%) isolated cases were available to train the model. The AUCs of the model to discriminate source, secondary and isolated case are shown in Table 1. HIV infection (odds ratio [OD] 3.72, 95% confidence interval [CI] 2.10; 6.59) and the interaction of living in apartment and presence of cavitory lesion (OD 3.19, 95% CI 1.71; 5.96) were found to be significant predictors of being a source, while being Canadian-born (OD 2.87, 95% CI 1.94; 4.26)

Table 1. Discriminative performance of the prediction model

Outcome to be predicted	AUC (95% CI)
Source	0.62 (0.57; 0.67)
Secondary	0.64 (0.61; 0.66)
Isolated	0.65 (0.65; 0.66)

Abbreviations: CI, confidence interval, AUC, area under curve

and being Haitian-born (OD 3.09, 95% CI 1.99; 4.78) were significant predictors of being a secondary case.

Conclusions

Performance of the prediction model was promising as it was significantly better than random prediction (i.e., the AUCs were higher than 0.5). Small proportions of source and secondary cases in the available data may have limited performance. However, the model can be an effective decision support tool if its ability to identify a case likely to be involved in transmission is superior to the intuition of public health officials. Thus, further evaluation of the model in the context of TB control program should be conducted. If effective, the model would be particularly useful when incidence of TB increases in a resource limited setting, in which efficient prioritization of investigation is desired. Overall, the current study has important implications in promoting the approach of evidence-based practice in control of TB.

Keywords

Tuberculosis; transmission; prediction model; public health; decision support

Acknowledgments

The authors thank Ms. Fiona Macintosh and Mr. Carmine Rossi for providing the results of DNA fingerprint analysis and Ms. Kimberly Kotar for maintaining the tuberculosis database.

References

1. Nguyen LN, Gilbert GL, Marks GB. Molecular epidemiology of tuberculosis and recent developments in understanding the epidemiology of tuberculosis. *Respirology*. 2004;9:313-9.
2. Fok A, Numata Y, Schulzer M, FitzGerald MJ. Risk factors for clustering of tuberculosis cases: a systematic review of population-based molecular epidemiology studies. *Int J Tuberc Lung Dis*. 2008;12:480-92.

*Hiroshi Mamiya

E-mail: hiroshi.mamiya@mail.mcgill.ca