Assessing the risk of cardiovascular disease due to latent tuberculosis infection among foreign-born Canadians in British Columbia, Canada, 1985-2019: a machine learning approach

Md. Belal Hossain *, Dr. James C. Johnston, Dr. Hubert Wong, Dr. Mohsen Sadatsafavi, Dr. Mohammad Ehsanul Karim

*Correspondence to: Md. Belal Hossain; Email: belal.hossain@ubc.ca

Summary of the poster

Background

Latent tuberculosis infection (LTBI) could increase the risk of cardiovascular disease (CVD) [1], which is responsible for one-third of all global deaths [2]. However, evidence is limited, with no evidence from low tuberculosis incidence settings like Canada, where tuberculosis is primarily concentrated among immigrants. In this study, we aimed to explore the long-term risk of CVD associated with LTBI among foreign-born Canadians in British Columbia, Canada.

Study cohort

We developed a retrospective cohort of people born outside Canada and immigrating to BC between 1985 and 2019. The cohort consists of ~1.4 million foreign-born individuals. The cohort was developed using an existing linkage between Immigration, Refugees and Citizenship Canada (IRCC) and Population Data BC databases with the following databases: Vital Statistics database (deaths in BC) [3], tuberculosis registry (tuberculosis-related information) [4], Medical Services Plan (enrolment/cancellation dates, ICD-9 diagnostic codes) [5], Hospital Discharge Abstract Database (ICD-9/10 codes for inpatients and day surgeries) [6], PharmaNet (drugs dispensed from BC community pharmacies) [7], Cancer Agency database (all malignancies) [8], HIV registry (HIV test results) [9], Renal database (chronic kidney disease, dialysis and transplant data) [10], and Statistics Canada Census (neighbourhood income quintiles) [11].

Participants

The participants for this study were those permanent residents who tested for Interferon-Gamma Release Assay (IGRA) during their residency in BC (N = 9,858). We excluded those participants diagnosed with LTBI before 1985, who developed CVD before LTBI, missing or unknown IGRA test results, and had a history of tuberculosis. The final analytic sample was 8,781. The Median follow-up time since immigration was 15 years (IQR: 10-23 years).

Index date or time zero

The date of the first IGRA test.

Exposure variable

Participants were classified as having LTBI if tuberculosis antigen was ≥ 0.35 IU/ml [12].

Outcome variable(s)

The primary outcome variable was time from the index date to death or diagnosis with composite CVD events of ischemic heart disease, cerebrovascular disease (stroke), or peripheral vascular disease. The secondary outcomes were time from the index date to death or diagnosis with myocardial infarction, hemorrhagic stroke, ischemic stroke, and peripheral vascular disease. The CVD events were identified from linked administrative data using validated algorithms proposed by Tonelli et al. [13], using ICD-9 and ICD-10 codes [13,14].

Confounders

Age, sex, income, tobacco use, alcohol use, substance use, hypertension, diabetes, and chronic kidney disease were the known confounders for the LTBI–CVD relationship [15–20]. The other potential confounders were education, WHO region of birth, immigration class, marital status, obesity, HIV/AIDS, dyslipidemia, organ transplant, and TNF-alpha [15,16,18,20,21]. Variables were defined in a one-year covariate assessment window prior to the index date.

Statistical analysis

Handling missing data: We have a high percentage of missing values for tobacco use (~60%) and income (~1%). We used multiple imputation to impute those missing values by considering the missing at random assumption. Since we were dealing with a time-to-event outcome, predictors used to build the imputation model included all confounders, CVD event, and the Nelson–Aalen estimator of CVD event [22]. We imputed five datasets with five iterations.

Regression analysis with Cox LASSO: We first aimed to use Cox proportional hazards model to explore the relationship between LTBI and CVD. However, with a rare CVD event (~2.5% in our data and <1% for specific CVD events), the Cox model would likely overfit, and neither the full model (model with all confounders) nor backward elimination (reduced model) is recommended in rare event scenarios [23]. Therefore, we used the penalized version of the Cox model (i.e., Cox LASSO) that can deal with multicollinearity between confounders, reduce the dimensions of the potential confounders, and deal with overfitting [24]. Keeping in mind that we must adjust for the

true confounders (listed above), we used the variable selection only for the potential confounders (listed above). In other words, we forced the models always to include the LTBI exposure and true confounders and make the variable selection only for the potential confounders. We used 5-fold cross-validation to choose the hyperparameters of the models and considered 200 bootstrap replicates to obtain the 95% CI [25].

Propensity score analysis: As an alternative confounding adjustment tool and as a sensitivity of the findings with Cox LASSO, we used inverse probability weighting and 1:1 propensity score matching analyses. First, we estimated the propensity scores (i.e., probability of LTBI) using a 5-fold cross-validated super learner to reduce model-misspecification [26], with logistic, LASSO, random forest, and gradient boosting as the candidate learners. We included all confounders (true and potential) as covariates in the propensity score model. Stabilized weights were calculated to prevent extreme weights and increase precision [27]. We observed the standardized mean difference (SMD) of less than 0.1 on the inverse probability weighted data, meaning that a good covariate balancing was achieved on the pseudo-population [28]. Then the association between LTBI and CVD was explored by fitting the Cox proportional hazards model on the pseudo-population created by re-weighting each participant's contribution by the calculated stabilized inverse probability weights. Second, we used 1:1 nearest neighbor matching to match an LTBI exposed with LTBI unexposed within the caliper of 0.2 times the standard deviation of the logit of propensity score [29]. Similar to the inverse probability weighting, we observed a good covariate balancing on the matched data (SMD $\leq$ 0.1). We then fitted the Cox proportional hazards model on the matched data. The robust sandwich standard error was used to estimate the 95% CI [30].

References

1. Huaman MA, Henson D, Ticona E, Sterling TR, Garvy BA. Tuberculosis and cardiovascular disease: linking the epidemics. Trop Dis Travel Med Vaccines. 2015;1(1):1–7.
2. World Health Organization. Cardiovascular diseases (CVDs) [Internet]. 2021 [cited 2022 Jan 30]. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
3. British Columbia Ministry of Health [creator]. Vital Events Deaths. V2. Population Data BC [publisher]. Data Extract. MOH(2015) [Internet]. 2017 [cited 2021 Oct 22]. Available from: https://www.popdata.bc.ca/data
4. BC Centre for Disease Control [creator]. BC Provincial TB Registry (BCCDC-iPHIS). Population Data BC [publisher]. Data Extract. BCCDC (2014) [Internet]. 2015 [cited 2021

Oct 22]. Available from: http://www.popdata.bc.ca/data

5. British Columbia Ministry of Health [creator]. Consolidation File (MSP Registration & Premium Billing). V2. Population Data BC [publisher]. Data Extract. MOH(2015). [Internet]. 2017 [cited 2021 Oct 22]. Available from: https://www.popdata.bc.ca/data

6. Canadian Institute for Health Information [creator]. Discharge Abstract Database (Hospital Separations). V2. Population Data BC [publisher]. Data Extract. MOH(2015) [Internet]. 2017 [cited 2021 Oct 22]. Available from: https://www.popdata.bc.ca/data

7. BC Ministry of Health [creator]. PharmaNet. BC Ministry of Health [publisher]. Data Extract. Data Stewardship Committee (2014) [Internet]. 2015. Available from: http://www.popdata.bc.ca/data

8. BC Cancer. BC Cancer Registry Data. V2. Population Data BC. Data Extract. BC Cancer(2015) [Internet]. 2017 [cited 2021 Oct 22]. Available from: http://www.popdata.bc.ca/data

9. BC Centre for Disease Control [creator]. BC Provincial HIV/AIDS Surveillance Database (BCCDC-HAISIS). Population Data BC [publisher]. Data Extract. BCCDC (2014) [Internet]. 2015 [cited 2021 Oct 22]. Available from: http://www.popdata.bc.ca/data

10. BC Renal Agency [creator]. BC Provincial Renal Agency Database (PROMIS). Population Data BC [publisher]. Data Extract. PROMIS(2014) [Internet]. 2015 [cited 2021 Oct 22]. Available from: http://www.popdata.bc.ca/data

11. Statistics Canada. Statistics Canada Income Band Data. Catalogue Number: 13C0016. V2. Population Data BC. Data Extract. Population Data BC(2015) [Internet]. 2009 [cited 2021 Oct 22]. Available from: https://www.popdata.bc.ca/data

12. Campbell JR, Pease C, Daley P, Pai M, Menzies D. Chapter 4: Diagnosis of tuberculosis infection. Can J Respir Crit Care, Sleep Med. 2022;6(sup1):49–65.

13. Tonelli M, Wiebe N, Fortin M, Guthrie B, Hemmelgarn BR, James MT, Klarenbach SW, Lewanczuk R, Manns BJ, Ronksley P. Methods for identifying 30 chronic conditions: application to administrative data. BMC Med Inform Decis Mak. 2016;15(31):1–11.

14. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C, Saunders LD, Beck CA, Feasby TE, Ghali WA. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005;43(11):1130–9.

15. Huaman MA, Ticona E, Miranda G, Kryscio RJ, Mugruza R, Aranda E, Rondan PL, Henson D, Ticona C, Sterling TR. The relationship between latent tuberculosis infection and acute myocardial infarction. Clin Infect Dis. 2018;66(6):886–92.

16. Al Khoufi EA. Association between latent tuberculosis and ischemic heart disease: a hospital-based cross-sectional study from Saudi Arabia. Pan Afr Med J. 2021;38.

17. Faksri K, Reechaipichitkul W, Pimrin W, Bourpoern J, Prompinij S. Transmission and risk factors for latent tuberculosis infections among index case-matched household contacts. Southeast Asian J Trop Med Public Heal. 2015;46(3):486.

18. Kiazyk S, Ball TB. Latent tuberculosis infection. Canada Commun Dis Rep. 2017;43(3):62–6.

19. Yap P, Tan KHX, Lim WY, Barkham T, Tan LWL, Mark I, Chen C, Wang YT, Chee CBE. Prevalence of and risk factors associated with latent tuberculosis in Singapore: a cross-sectional survey. Int J Infect Dis. 2018;72:55–62.

20. Hasanain AFA, El-Maghraby KM, Zayed AAH, Nafee AMA, Abdel-Aal SM, Bakkar SM. Latent tuberculosis infection among patients with coronary artery stenosis: A case–Control study. Int J Mycobacteriology. 2018;7(2):143.

21. Huaman MA, De Cecco CN, Bittencourt MS, Ticona E, Kityo C, Ballena I, Nalukwago S, Nazzinda R, Ticona C, Azañero R. Latent Tuberculosis Infection and Subclinical Coronary Atherosclerosis in Peru and Uganda. Clin Infect Dis. 2021;

22. White IR, Royston P. Imputing missing covariate values for the Cox model. Stat Med. 2009;28(15):1982–98.

23. Ojeda FM, Mueller C, Boernigen D, Tregouet D-A, Schillert A, Heinig M, Zeller T, Schnabel RB. Comparison of cox model methods in a low-dimensional setting with few events. Genomics Proteomics Bioinformatics. 2016;14(4):235–43.

24. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med. 1997;16(4):385–95.

25. Chatterjee A, Lahiri SN. Bootstrapping lasso estimators. J Am Stat Assoc. 2011;106(494):608–25.

26. Pirracchio R, Petersen ML, Van Der Laan M. Improving propensity score estimators' robustness to model misspecification using super learner. Am J Epidemiol. 2015;181(2):108–19.

27. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Heal. 2006;60(7):578–86.

28. Zhang Z, Kim HJ, Lonjon G, Zhu Y. Balance diagnostics after propensity score matching. Ann Transl Med. 2019;7(1):16.

29. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharm Stat. 2011;10(2):150–61.

30. Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. Stat Med. 2013;32(16):2837–49.