

Spatio-temporal analysis of tuberculous infection risk among clients of a homeless shelter during an outbreak

A. Crisan,* H. Y. Wong,[†] J. C. Johnston,[‡] P. Tang,[§] C. Colijn,[¶] M. Otterstatter,** L. Hiscoe,** R. Parker,** S. L. Pollock,** J. L. Gardy**[#]

*Communicable Disease Prevention and Control Services, British Columbia Centre for Disease Control, Vancouver, [†]Department of Microbiology & Immunology, University of British Columbia, Vancouver, [‡]Clinical Prevention Services, British Columbia Centre for Disease Control, Vancouver, [§]British Columbia Public Health Microbiology and Reference Laboratory, Vancouver, British Columbia, Canada; [¶]Department of Mathematics, Imperial College London, London, UK; [#]School of Population and Public Health, University of British Columbia, Vancouver, **Interior Health Authority, Kelowna, British Columbia, Canada

SUMMARY

SETTING: British Columbia (BC) has a low incidence of tuberculosis (TB), with the burden of endogenously acquired disease concentrated among vulnerable populations, including the homeless. In May 2008, a TB outbreak began in a BC homeless shelter, with a single index case seeding multiple secondary cases within the shelter.

OBJECTIVE: To use nightly shelter records to quantify the risk of latent tuberculous infection (LTBI) among shelter clients as a function of their sleeping distance from and duration of exposure to the index case.

DESIGN: Distance and duration of exposure were visualised and assessed using logistic regression with LTBI status as outcome. We used a novel machine learning approach to establish exposure thresholds that

optimally separated infected and non-infected individuals.

RESULTS: Of 161 exposed shelter clients, 58 had a recorded outcome of infected ($n=39$) or non-infected ($n=19$). Only duration of exposure to the index was associated with increased odds of infection (OR 1.26); stays of ≥ 5 nights put shelter clients at higher odds of infection (OR 4.97).

CONCLUSION: The unique data set and analytical approach suggested that, in a shelter environment, long-term clients are at highest risk of LTBI and should be prioritised for screening during an outbreak investigation.

KEY WORDS: transmission; indoor air; exposure; infectiousness

IN LOW-INCIDENCE REGIONS such as North America, endemic tuberculosis (TB) transmission is concentrated in at-risk populations, including the homeless. As infection typically requires prolonged exposure to an infectious case,^{1,2} the crowded sleeping quarters of homeless shelters place clients, particularly those with comorbidities,³ at increased risk of acquiring TB. Several reported TB outbreaks have been linked to shelters or non-traditional sleeping arrangements,^{4–7} but the influence of duration of exposure and distance from an infectious source in these environments has yet to be examined. Understanding these spatio-temporal influences on TB transmission may inform the design of indoor environments, and guide screening and outbreak investigation procedures in other congregate settings.

In May 2008, an outbreak of TB began in a homeless shelter in Kelowna, British Columbia, Canada. The index case, diagnosed with acid-fast bacilli (AFB) 4+ pulmonary TB, used the shelter for

12 nights, seeding a large first wave of both active cases and latent tuberculous infection (LTBI). By January 2015, the outbreak had grown to include 52 active cases, with 2369 community members identified as potentially exposed and investigated using the tuberculin skin test (TST). By combining a unique data set of the names and bed locations of shelter clients with the TST screening database, we were able to examine the influence of duration of exposure and distance from the index case on risk of LTBI among shelter clients.

STUDY POPULATION AND METHODS

The shelter environment

The shelter is a two-storey building purpose-built in 1985–1986. The dormitory occupies the second storey, and comprises a large central room measuring 209 m² (19 m × 11 m) and a second room measuring 24 m² (3 m × 8 m), containing 32 and five bunk beds,

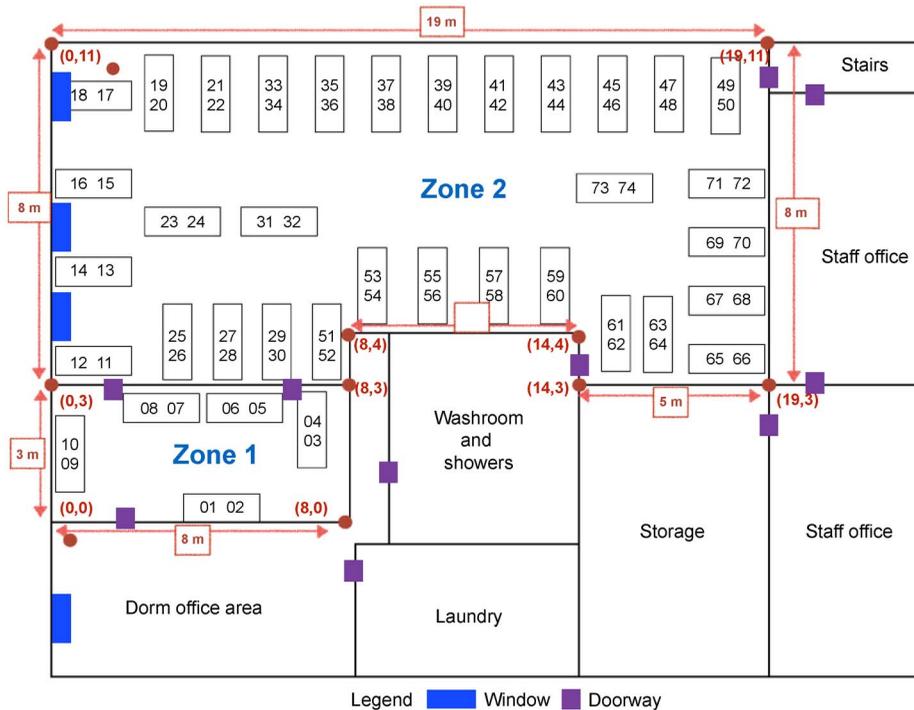


Figure 1 Shelter layout at the time of the index case's visits. Measurements of the two sleeping rooms (Zones 1 and 2) are indicated by red arrows and the corresponding x, y coordinates. Rectangles indicate bunk beds, which are numbered by the shelter and linked to nightly roll call data; even numbers indicate top bunks. Blue boxes mark the location of awning windows opening to the street; purple boxes indicate doorways. This image can be viewed online in colour at <http://www.ingentaconnect.com/content/ijutld/ijltd/2015/00000019/00000009/art00007>

respectively (Figure 1). The ventilation system is forced air, with two 10 000 BTU (British thermal unit) rooftop units serving the dormitory. At the time of the outbreak, two commercial-grade air purifiers were also in place. Four awning-style windows are located along one wall of the second storey, three of which open onto the larger room.

Bed map and roll call data

Each night, incoming clients are assigned a numbered bunk; the bed number and client name are recorded on roll calls. As part of the outbreak investigation, the shelter provided the public health unit with the rolls for the 12 nights during which the index case used the shelter, along with a scaled map of the numbered bed locations as the shelter was configured in 2008. From these data, we identified individuals exposed to the index case at the shelter, their sleeping distance from the index case and the number of nights spent in the shelter with the index.

Client TB status definitions

We cross-referenced individuals exposed to the index case using the community TST screening database, assigning shelter clients to one of three outcomes: 'infected' (a diagnosis of active TB or a positive TST result with no prior history of TB disease and no prior positive TST), 'non-infected' (at least one negative

TST >2 months after exposure to the index case), and 'unknown' (clients who were not screened, received a TST but did not have it read, had a prior positive TST, had prior LTBI, or were bacille Calmette-Guérin vaccinated).

Evaluation of time and distance variables

For each client, we calculated the total time exposed to the index case (T) as the cumulative, non-consecutive number of days during which a client slept in the shelter when the index was present. We used the shelter layout to calculate the distance of each client's bed relative to the index case's bed (D) for each day of the index's visits (Appendix).^{*} We quantified a client's sleeping distance from the index in three ways: D_{sum} – the sum of the nightly distances, D_{avg} – the average nightly distance and D_{min} – the minimum distance. We also normalised D_{sum} by T to create a combined measure, D_T – the distance-to-exposure time ratio.

Odds ratios (ORs) were calculated using univariable logistic regression with the infected/non-infected response variable and predictor variables T , D_{sum} , D_{avg} , D_{min} and D_T . All tests for significance were two-

^{*} The appendix is available in the online version of this article, at <http://www.ingentaconnect.com/content/ijutld/ijltd/2015/00000019/00000009/art00007>

Table 1 Distribution of shelter clients' tuberculosis screening results*

Outcome	Total	Both visits	First only	Second only
Active	8	2	3	3
Latent	31	13	7	11
Non-infected	19	3	1	15
Prior infection	3	1	1	1
Unknown: not in database	99	2	52	45
Unknown: TST not read	1	0	0	1

* A total of 161 clients were exposed to the index case during their stay in the shelter. The index case made two visits to the shelter: the first for 7 nights over an 11-night window and the second 4 months later for 5 nights over a 10-night window.

TST = tuberculin skin test.

sided and were considered significant if $P < 0.05$. Clients with unknown TB status were not included in the analysis. Recognising that some positive TSTs in exposed individuals likely represented a baseline level of LTBI unrelated to the current outbreak, we also examined ORs under multiple scenarios of baseline LTBI rates (Appendix). Predicted numbers of infections were calculated using the Wells-Riley model (Appendix).

Establishing a threshold for exposure time

A threshold was established using recursive partitioning, with a maximum tree depth of 1 using *R* software v3.0.2 and *rpart* v4.1 (R Computing, Vienna, Austria). The recursive partitioning took as input exposure time T and established a threshold that optimally separated infected and non-infected clients (Appendix). Sensitivity and specificity were used to evaluate the separability of each threshold. Bootstrapped optimism penalisation was used to adjust raw performance measures and provide an estimate of internal validation.⁸

Ethics approval for the study was received by the University of British Columbia, Vancouver, BC, Canada.

RESULTS

Reach of community screening among exposed clients

The index case stayed in the shelter a total of 12 days on two non-consecutive occasions: Visit 1, 7 nights' stay over an 11-night window; and Visit 2, approximately 4 months after Visit 1, with 5 nights' stay over a 10-day window. Overall, 161 shelter clients were exposed to the index: 64 during Visit 1, 76 during Visit 2 and 21 during both visits (Table 1).

Although an intensive TST-based screening programme was implemented at the shelter and in the community following the diagnosis of the index case, the highly mobile client base meant that 62% of the exposed clients ($n = 99$) did not present for screening. Of all 161 clients, we found that those with longer shelter use, approximated by time of exposure to the

index, had higher odds of presenting for screening (OR 1.38, $P < 0.001$). Clients using the shelter exclusively during Visit 2 were also more likely to have been screened (Table 1, $\chi^2 32.5$, $P < 0.001$), likely due to the index being diagnosed shortly after this visit, facilitating follow-up.

Community screening among shelter clients

Three clients were not screened due to a reported prior history of TB. Of those exposed clients with a screening result and no prior history ($n = 58$), 39 (67%) had tuberculous infection. These cases included eight active TB cases definitively linked to the index case through genomic analysis.⁹ Of the 31 LTBI cases, we assumed that all were infected by the index in the analyses presented hereafter; scenarios with different baseline LTBI rates are described in the Appendix. The Wells-Riley equation¹⁰ predicted 11–12 new infections, substantially fewer than the 39 observed.

Qualitative visualisation of shelter sleeping locations

To qualitatively explore the impact of proximity on infection, we visualised the shelter clients' TB status and their sleeping locations during the index's visits (Figure 2 and Appendix). We observed that infection among individuals sleeping near the index case was more common in Visit 1 vs. Visit 2. During Visit 1, both the index's sleeping location and those of other clients were stable; however, during Visit 2 the index tended to spend one or two nights in the shelter, followed by nights elsewhere, and as a result, slept in four different locations. To further explore the relationship between these spatio-temporal variables and infection, we next performed a quantitative analysis of infection risk.

Exposure time, but not distance, was associated with an increased risk of infection

We explored whether exposure time or distance was associated with an increased risk of tuberculous infection. We independently evaluated cumulative days of exposure T , cumulative distance from the index D_{sum} , average distance from the index D_{avg} , minimum distance from the index D_{min} and the distance-to-exposure time ratio D_T using univariable logistic regression with infected or non-infected as outcome (Table 2). Only the time variable, T , had a significant effect on the probability of infection, with an OR of 1.26. We explored the effect of different baseline LTBI levels on this OR (Appendix) and determined that the finding of increased risk was observable and significant up to and including a baseline LTBI prevalence of 16%.

Threshold of risk of tuberculous infection

Given the observation that only cumulative exposure days T was linked to an increased risk of tuberculous

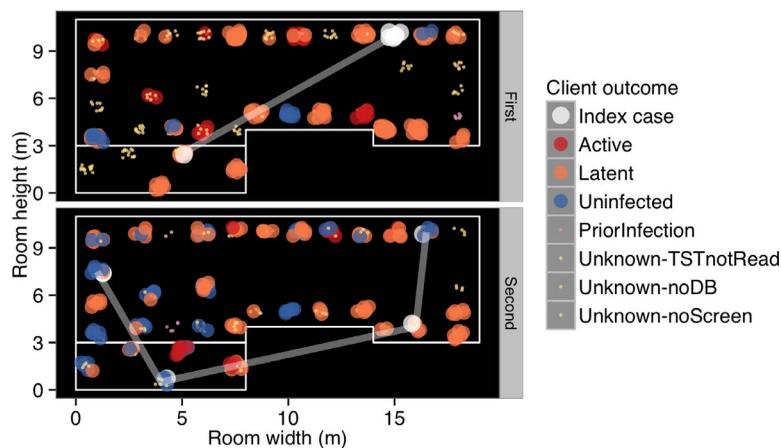


Figure 2 Client TB status according to index case stay. Dots indicate where a specific client slept each night and are coloured according to TB status. The index case is shown in white, and his movement is emphasised with a white line. All of the days from the index's first (top panel) and second (bottom panel) visits are shown superimposed on top of each other, with a transparency and position jitter applied to each dot to show the day-to-day variability of clientele occupying that bunk. TB = tuberculosis. This image can be viewed online in colour at <http://www.ingentaconnect.com/content/ijuatld/ijtld/2015/00000019/00000009/art00007>

infection, we next explored whether or not a threshold exposure time exists that maximally separates infected from non-infected clients. Using a recursive partitioning algorithm on T , we found that a threshold of 5 days' shelter use optimally separates infected and non-infected clients (Figure 3). The sensitivity and specificity of the 5-day threshold, with penalisation for over-fitting, were respectively 63% and 57%, indicating that this threshold is not a definitive boundary of increased risk of infection. However, given the alternative of considering every shelter client at risk—a generalisation that has a sensitivity of 100% and specificity of 0%—even this non-definitive boundary offers useful discriminatory power.

Re-applying the logistic regression with a dichotomised exposure time variable, i.e., classifying clients as either above or below the 5-day threshold, increased the OR to 4.97 (95% confidence interval [CI] 1.58–17.0); a P value is not reported as the exposure threshold was generated and evaluated on the same data.

Table 2 ORs of active tuberculosis and latent tuberculous infection, according to time exposed to the index case and sleeping distance from the index case

Variable	OR (95%CI)	P value
T	1.26 (1.05–1.58)	0.02
D_{sum}	1.01 (1.00–1.03)	0.08
D_{avg}	0.95 (0.85–1.03)	0.33
D_{min}	0.95 (0.86–1.03)	0.21
D_T	0.95 (0.85–1.05)	0.33

OR = odds ratio; CI = confidence interval; T = cumulative nights exposed to the infectious index; D_{sum} = sum of the nightly sleeping distance from the index; D_{avg} = average nightly distance from the index; D_{min} = minimum distance from the index; D_T = distance-to-exposure time ratio.

DISCUSSION

In this observational study, we leveraged a unique data set—the detailed record of the dates during which an infectious individual resided at a homeless shelter and his nightly sleeping locations—to retrospectively examine the impact of exposure duration and sleeping distance on the risk of tuberculous infection in other shelter clients. By linking the shelter data to a community screening database, we combined traditional statistical and epidemiological approaches with data mining techniques into a novel method to explore the data set. Although our study conclusions are limited due to the small number of exposed individuals with a recorded outcome, we believe the method presented here may be useful for investigating outbreaks in other congregate settings, particularly larger facilities with electronic record keeping systems.

Our investigation suggested that cumulative time exposed to the index case was associated with higher odds of tuberculous infection, in agreement with recent work,¹¹ and our observation proved robust to varying levels of baseline LTBI prevalence. More specifically, we determined that clients spending ≥ 5 nights in this shelter with the index were at 4.97-fold higher odds of tuberculous infection than clients exposed for ≤ 4 nights—an increase in risk greater than the 2.74-fold risk ratio of LTBI estimated in other congregate settings, largely prisons, in high-income settings.¹² Although the sensitivity and specificity indicate our 5-day threshold is not a definitive boundary, reflecting our study's small size and opportunistic but ultimately incomplete data set, this threshold may be refined through future investi-

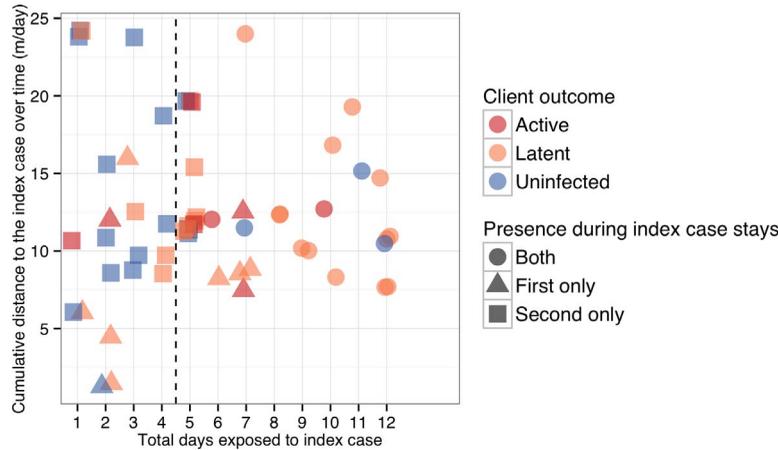


Figure 3 Visualisation of the cumulative exposure threshold. Clients are plotted according to their cumulative exposure time T and their distance-to-exposure time ratio D_T . The dotted line indicates the maximally separating exposure time threshold determined via machine learning that best separates infected and non-infected clients. TST = tuberculin skin test; DB = database. This image can be viewed online in colour at <http://www.ingentaconnect.com/content/iatld/ijtld/2015/00000019/00000009/art00007>

gations in shelter environments. In the interim, we suggest that the threshold might be used in screening programmes to prioritise TB-exposed clients for testing; we believe that our statistical and computational approach to interrogating our data set might be useful in other TB outbreak scenarios in congregate settings.

We did not find that sleeping proximity was significantly associated with risk of infection. This is likely due to the nature of the shelter environment itself: although clients had structured sleeping arrangements, they also moved about the facility freely and made use of a communal eating space downstairs. Thus, transmission from the index case need not have occurred only in the dormitory and would not necessarily be captured by sleeping distance. However, a client's number of nights in the shelter—a surrogate for total exposure time in the dormitory and meal centre settings—should capture these other interactions. Our observations also suggest that in short-term congregate settings with high client mobility, such as shelters and jails, concentric circle-based contact tracing approaches focusing only on the beds immediately surrounding an infectious case may be missing infected contacts, and that exposure time should be considered.

The importance of exposure time is also demonstrated by the discrepancy between the number of infections predicted by the Wells-Riley model ($n = 11$ – 12) and the number actually observed ($n = 39$). Doubling the exposure time in the model from 8 to 16 hours to account for exposure outside of sleeping hours would increase the predicted number of infected contacts to 22. To generate the observed 39 new infections—likely an underestimation of the true number of new LTBI cases—would further require

almost doubling the number of infectious quanta per hour to 23, twice the reported value for an office-based outbreak of TB and 18 times the rate reported for the average TB patient.¹³ The discrepancy between the Wells-Riley model predictions and the observed number of infected clients reflects the complexity of establishing appropriate model parameters, particularly infectious quanta per hour. Work by Fennelly et al. showed that AFB4+ pulmonary TB patients produced a highly variable number of aerosolised bacilli,¹⁴ with patients producing 19 colony-forming units (cfu) on average, but some patients expelling hundreds of infectious bacilli.

Although bed spacing was not found to be significantly associated with infection in this study, anecdotal evidence from a second shelter in the outbreak community, at which fewer beds were placed in a larger room and at which germicidal ultraviolet (UV) treatment was installed during the outbreak, suggests that these environmental controls may have impacted transmission—the second shelter was linked to only two secondary cases of TB arising from one infected client.⁹ Design elements such as minimum air changes, UV germicidal irradiation and ventilation of areas other than sleeping quarters have previously been suggested to reduce TB transmission in shelters.¹⁵ Unfortunately, the lack of a bed map and roll call data from the second shelter precludes a quantitative evaluation. The contribution of distance to TB transmission in congregate settings should continue to be explored, particularly in large facilities with greater client densities, and bed map and roll call data similar to that collected here may prove an important tool in understanding the impact of facility design on TB transmission.

CONCLUSIONS

We report a novel analysis of the risk of infection in a shelter-based TB outbreak, leveraging a unique data set of mapped sleeping locations. We demonstrate that threshold exposure times optimally separating infected and non-infected shelter clients can be calculated using a machine learning approach. Our results suggest that in an outbreak involving homeless clients, individuals with a history of long-term shelter use should be prioritised for screening, as they are at a greater risk of infection than infrequent shelter clients; however, due to the size and nature of our retrospective study, it is necessary to further validate the 5-day threshold we calculated in other congregate settings.

Acknowledgements

The authors would like to thank D McKay, P Hasselback, M Fillion and the Interior Health TB Outbreak Management Team for their role in investigating and managing the outbreak; K Heinitz for details of the shelter environment; R Balshaw for the statistics conversations; and the British Columbia Centre for Disease Control Foundation for Population and Public Health (Vancouver, BC, Canada) for funding support.

Conflicts of interest: none declared.

References

- Riley R L. Airborne pulmonary tuberculosis. *Bacteriol Rev* 1961; 25: 243–248.
- Riley R L. Aerial dissemination of pulmonary tuberculosis. *Am Rev Tuberc* 1957; 76: 931–941.
- Marais B J, Lönnroth K, Lawn S D, et al. Tuberculosis comorbidity with communicable and non-communicable diseases: integrating health services and control efforts. *Lancet Infect Dis* 2013; 13: 436–448.
- Gardy J L, Johnston J C, Ho Sui S J, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 2011; 364: 730–739.
- Khan K, Rea E, McDermaid C, et al. Active tuberculosis among homeless persons, Toronto, Ontario, Canada, 1998–2007. *Emerg Infect Dis* 2011; 17: 357–365.
- Curtis A B, Ridzon R, Novick F, et al. Analysis of *Mycobacterium tuberculosis* transmission patterns in a homeless shelter outbreak. *Int J Tuberc Lung Dis* 2000; 4: 308–313.
- McElroy P D, Southwick K L, Fortenberry E R, et al. Outbreak of tuberculosis among homeless persons coinfecting with human immunodeficiency virus. *Clin Infect Dis* 2003; 36: 1305–1312.
- Harrell F E, Lee K L, Mark D B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15: 361–387.
- Didelot X, Gardy J, Colijn C. Bayesian inference of infectious disease transmission from whole-genome sequence data. *Mol Biol Evol* 2014; 31: 1869–1879.
- Riley E C, Murphy G, Riley R L. Airborne spread of measles in a suburban elementary school. *Am J Epidemiol* 1978; 107: 421–432.
- Kanamori H, Kimura R, Weber D J, et al. Lessons learned from earthquake-related tuberculosis exposures in a community shelter, Japan, 2011. *Am J Infect Control* 2014; 42: 246–248.
- World Health Organization. WHO policy on TB infection control in health-care facilities, congregate settings, and households. WHO/HTM/TB/2009.419. Geneva, Switzerland: WHO, 2009.
- Nardell E A, Keegan J, Cheney S A, Etkind S C. Airborne infection. Theoretical limits of protection achievable by building ventilation. *Am Rev Respir Dis* 1991; 144: 302–306.
- Fennelly K P, Jones-López E C, Ayakaka I, et al. Variability of infectious aerosols produced during coughing by patients with pulmonary tuberculosis. *Am J Respir Crit Care Med* 2012; 186: 450–457.
- Toronto Public Health. Environmental control best practices: guidelines to reduce TB transmission in homeless shelters and drop-in centres. Toronto, ON, Canada: Toronto Public Health, 2007: p 83.

APPENDIX

ESTIMATING DISTANCES BETWEEN SHELTER CLIENTS

The shelter comprises a large central room measuring 209 m² (19 m × 11 m) and a secondary room of 24 m² (3 m × 8 m), containing respectively 32 and 5 bunk beds, for a total of 74 individual beds. The exact spacing between beds during the index case's visits was not available; we therefore estimated inter-bed spacing to be 0.5 m and 1.5 m by assuming bed dimensions of a standard twin bed rounded to the nearest meter (height 2 m, width 1 m), and using the bed map provided by the shelter as a guide (Figure A.1). Top and bottom bunks were assigned values of 1 and 0, respectively.

The distance between the index case and a specific shelter client was calculated using the standard distance equation for points in 3-D space:

$$d = \sqrt{(x_1 - x_{ref})^2 + (y_1 - y_{ref})^2 + (z_1 - z_{ref})^2},$$

where x_{ref} , y_{ref} , and z_{ref} refer to the co-ordinates of the index case on a given night and x_1 , y_1 , and z_1 are the co-ordinates of another client on the same night. Given that the shelter facility has two zones, we assumed that walls impeded transmission and assigned a distance of 24 m—the maximal distance between any two points in the dormitory—to clients outside the index case's sleeping zone.

EXPLORING THE EFFECT OF VARYING BASELINE LTBI RATES ON THE CALCULATED RISK OF INFECTION

Taking into account that some positive tuberculin skin test (TST) results in exposed individuals likely represented a baseline level of latent tuberculous infection (LTBI) unrelated to the current outbreak, we examined how odds ratios (ORs) varied under multiple scenarios of baseline LTBI rates. For this analysis, individuals in the active TB and non-infected outcome groups remained the same; our previous genomic work on this outbreak¹ indicated that active TB cases could be linked via their genomic sequence to the index case. The 31 individuals in the LTBI group were handled as follows: we performed a sensitivity analysis that varied the number of LTBI clients whose positive TST was assumed to result from a prior, non-outbreak infection from 1 to 31, evaluating all possible case label permutations for each step. For example, if we assume that one client had baseline LTBI unrelated to the outbreak, there are 31 (31 choose 1) possible label permutations; if we assumed two clients had baseline infections, there are 465 (31 choose 2) possible label permutations. We limited the maximum number of label permutations

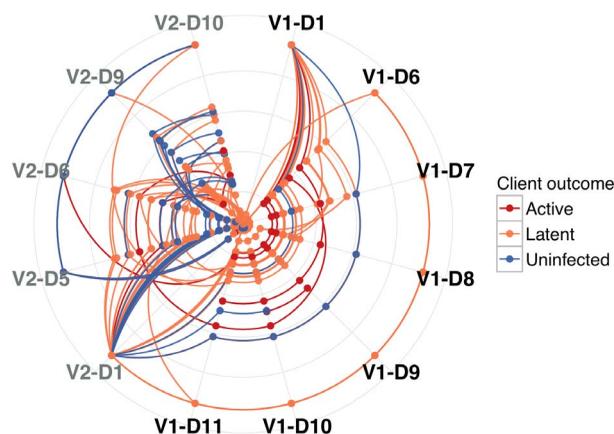


Figure A.1 Visualisation of sleeping locations and sleeping distance from the index case over time. The centre of the plot is the location of the index case; dots correspond to other shelter clients coloured by tuberculosis status. Distance from the centre of the plot reflects a client's sleeping distance from the index case on a given day. Days are labelled based on the index client visit: days of the first (V1) and second (V2) visits are shown in black and grey, respectively, and numbered according to the day (i.e., V1-D1 = Visit 1 Day 1, etc.). Consecutive numbers indicate consecutive days in the shelter.

evaluated to 10 000. For each label permutation, we performed univariable regression for each of the distance and exposure time variables and recorded whether or not the result was significant and the OR.

As with the analysis described in the main text, none of the distance variables gave statistically significant results; the results presented below are for the exposure time T variable only. In the main text, the OR, assuming all LTBI cases were infected by the index, is 1.26 (95% confidence interval [CI] 1.05–1.58, $P = 0.02$). As the number of baseline LTBI cases increases from 1 to 31, the point estimates and ranges of OR decrease in a linear fashion (Figure A.2). Similarly, as the number of baseline LTBI cases increases, the percentage of permutations returning a statistically significant OR also decreases.

Taken together, exposure time confers an increased risk of infection (OR lower bound >1.0) at a largely statistically significant level ($P < 0.05$ in 50% or more runs) when we assume between zero and three clients with baseline LTBI. Assuming four or five clients had LTBI unrelated to the outbreak is also associated, albeit not as significantly, with increased risk.

Lashley reported a 16% baseline LTBI in a North American inner city homeless shelter (46 LTBI cases among 282 screened shelter clients).² Given that our outbreak community is quite different from that in Baltimore, MD, USA, with a smaller under-housed population, better access to health care and lower tuberculosis (TB) pre-outbreak rates (1/100 000 vs. $\sim 4/100\,000$ in Baltimore), we can use Lashley's data to set an upper bound that five of our 31 cases could

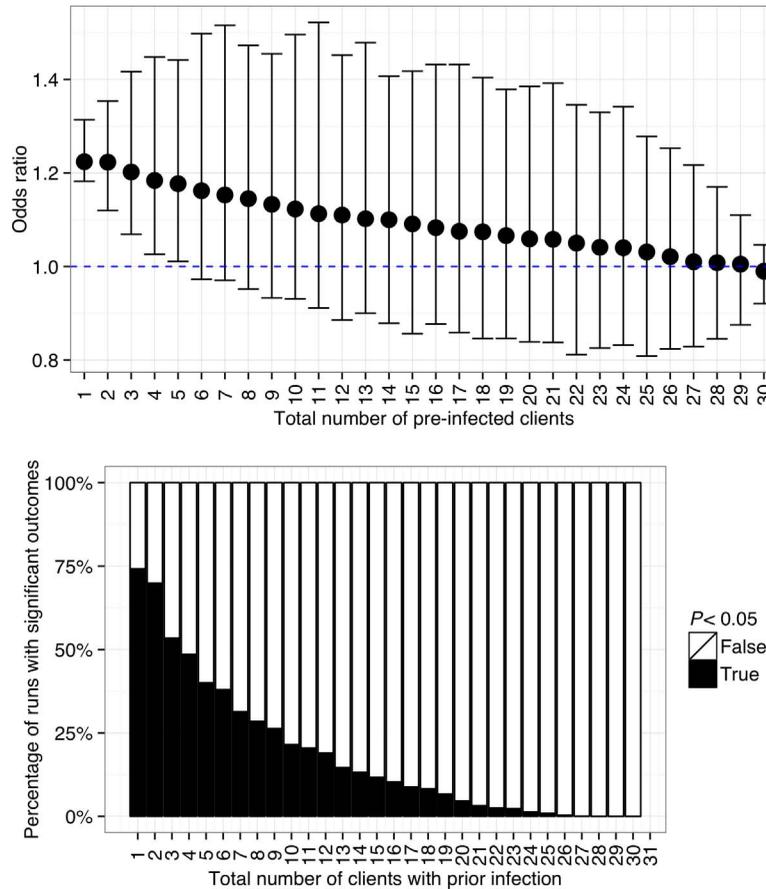


Figure A.2 Number of baseline LTBI cases and OR point estimates and ranges.

have been a baseline LTBI case. We believe that the actual rate of prior TST positivity is likely <10%, or three individuals, in our exposed-infected cohort. Under all three scenarios—all cases infected by the index (main text), Lashley’s 16% baseline (26/31 cases infected by the index) or our 10% baseline (28/31 cases infected by the index)—we find ORs with lower bounds >1 and generally significant *P* values, indicating that our method is appropriately sensitive to underlying variations in the baseline LTBI-positive population.

PREDICTING THE NUMBER OF INFECTED CLIENTS WITH THE WELLS-RILEY MODEL

We used the Wells-Riley equation to predict the expected number of new infections *N* on each night the index case stayed in the shelter:³

$$N = S \left(1 - e^{-Iqt/\Omega} \right),$$

where

- *S* (susceptibles) = 9 or 63 (number of other clients in sleeping zone 1 and 2, respectively)
- *I* (infectors) = 1 (one index case)
- *q* (infectious quanta/h) = 12.7 (value established in Nardell et al.⁴)

- *p* (pulmonary ventilation rate) = 0.36 m³/hour (resting ventilation rate of 6 l/min)
- *t* (exposure time) = 8 h (overnight stay)
- *Q*, the absolute ventilation of the room (m³/h), calculated as *NV*, where *N* = air changes per hour (estimated at 4) and *V* = room volume in m³ (using the sleeping zone areas and assuming a ceiling height of 3 m).

Because the shelter had two zones, both of which were used by the index case, we calculated the number of new cases on each night for the zone in which the index slept. We summed the estimated number of new infections per night over the 12 nights to obtain a total estimate of predicted new infections arising from the index case.

ESTABLISHING AND EVALUATING A CUMULATIVE EXPOSURE THRESHOLD

A threshold to separate infected and non-infected clients based on exposure duration or proximity was established using recursive partitioning with a maximum tree depth of one using *R* software v3.0.2 and the *rpart* package v4.1 (R Computing, Vienna, Austria). Clients with unknown TB status were excluded from the analysis. The recursive partitioning

takes, in this case, a single continuous variable (i.e., exposure time, distance) as input and constructs a decision tree. We limited this tree to only one level, also known as a decision stump;⁵ the single split established by the recursive partitioning establishes the threshold for the input variable.

We wished to evaluate the efficacy of this threshold within our data set. To do this, we used sensitivity and specificity to evaluate whether the chosen threshold could discriminate between infected and non-infected clients. As we did not have a separate data set to validate this threshold and the available data set was too small to partition into training and testing sets, we used bootstrapped optimism penalisation to adjust the raw sensitivity and specificity measures.⁵ The adjusted measures provide a reasonably accurate estimate of the threshold's discriminability in a data set with similar characteristics; this is known as 'internal validation'.⁶ Splitting a data set into training and testing also provides a measure of internal validation, but given the relatively small sample size

of our data set this split would be both ineffective and wasteful of the limited data available. Bootstrapped optimism penalisation is an effective alternative.

References

- 1 Didelot X, Gardy J, Colijn C. Bayesian inference of infectious disease transmission from whole-genome sequence data. *Mol Biol Evol* 2014; 31: 1869–1879.
- 2 Lashley M. A targeted testing program for tuberculosis control and prevention among Baltimore city's homeless population. *Public Health Nurs* 2007; 24: 34–39.
- 3 Riley E C, Murphy G, Riley R L. Airborne spread of measles in a suburban elementary school. *Am J Epidemiol* 1978; 107: 421–432.
- 4 Nardell E A, Keegan J, Cheney S A, Etkind S C. Airborne infection: theoretical limits of protection achievable by building ventilation. *Am Rev Respir Dis* 1991; 144: 302–306.
- 5 Viola P, Jones M. Robust real-time face detection. *Int J Comput Vis* 2004; 57: 137–154.
- 6 Harrell F E, Lee K L, Mark D B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15: 361–387.

RESUME

CONTEXTE : La Colombie Britannique (BC) a une faible incidence de tuberculose (TB) dont la part de maladie acquise sur place est concentrée dans les populations vulnérables, notamment les sans-abri. En mai 2008, une flambée épidémique de TB a débuté dans un centre d'accueil de sans-abris de BC avec un cas index unique qui a provoqué de nombreux cas secondaires au sein du centre d'accueil.

OBJECTIF : Recourir aux registres d'accueil de nuit pour quantifier le risque d'infection tuberculeuse parmi les clients des centres d'accueil en fonction de leur distance de couchage vis-à-vis du cas index et de la durée de leur exposition à ce cas.

SCHEMA : La distance et la durée d'exposition ont été visualisées et évaluées grâce à une régression logistique dont le résultat a été le statut de l'infection. Nous avons utilisé une nouvelle approche d'apprentissage-machine

pour établir les seuils d'exposition séparant les personnes infectées des personnes non infectées.

RÉSULTATS : Sur 161 clients du centre exposés, 58 ont eu un devenir documenté d'infection ($n = 39$) ou de non infection ($n = 19$). Seule la durée d'exposition au cas index a été associée à un risque accru d'infection (OR 1,26) ; le fait de séjourner 5 nuits ou davantage a augmenté les risques des clients d'être contaminés (OR 4,97).

CONCLUSIONS : Cet ensemble de données unique et cette approche analytique ont suggéré que dans un environnement de centre d'accueil de sans-abri, les clients restant longtemps courent un risque plus élevé d'infection tuberculeuse et devraient être prioritaires lors du dépistage organisé dans le cadre de l'investigation d'une flambée épidémique.

RESUMEN

MARCO DE REFERENCIA: Columbia Británica presenta una baja incidencia de tuberculosis (TB) y su carga de morbilidad por enfermedad endógena se concentra en las poblaciones vulnerables, como las personas sin domicilio. En mayo del 2008, se declaró un brote epidémico de TB en un albergue para personas indigentes y un caso inicial único dio lugar a múltiples casos secundarios en la misma institución.

OBJETIVO: Utilizar los registros de los albergues nocturnos con el propósito de cuantificar el riesgo de infección tuberculosa en sus usuarios, en función de la distancia física del caso inicial y la duración de la exposición.

MÉTODO: La distancia al caso inicial y la duración de la exposición se visualizaron y se evaluaron mediante un análisis de regresión logística, en el cual el criterio de valoración fue la situación con respecto a la infección tuberculosa. Se utilizó una estrategia de aprendizaje

automático con el fin de establecer umbrales de exposición que distinguieran de manera óptima las personas infectadas y las personas sin infección.

RESULTADOS: De los 161 usuarios expuestos en el albergue, en 58 existía un registro del criterio de valoración, como infectado ($n = 39$) o sin infección ($n = 19$). Solo la duración de la exposición al caso inicial se asoció con una mayor probabilidad de infección (OR 1,26); las estancias de ≥ 5 noches aumentaban la probabilidad de infección de los usuarios del albergue (OR 4,97).

CONCLUSIÓN: Una base de datos única y la estrategia analítica indicaron que en el entorno de un albergue, los usuarios de larga duración presentan el riesgo más alto de contraer la infección tuberculosa y por lo tanto deben ser prioritarios al aplicar las medidas de detección sistemática durante la investigación de un brote epidémico.