Published in final edited form as:
Int J Tuberc Lung Dis. 2011 July ; 15(7): 899–905. doi:10.5588/ijtld.10.0556.

Local epidemic history as a predictor of tuberculosis incidence in Saskatchewan Aboriginal communities

Caitlin Pepperell[1,*], **Alicia H Chang**[1,*], **Wendy Wobeser**[2], **Julie Parsonnet**[1], and **Vernon H. Hoeppner**[3]

[1]School of Medicine, Division of Infectious Diseases, Stanford University, Stanford, California, United States of America

[2]Department of Medicine, Division of Infectious Diseases, Queen's University, Kingston, Ontario, Canada

[3]Department of Medicine, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

SUMMARY

Background—Average tuberculosis (TB) incidence rates are high in Canadian Aboriginal communities, but there is significant variability within this group.

Objective—To determine whether local history of post-contact TB epidemics is predictive of contemporary epidemiology among Aboriginal communities in Saskatchewan, Canada.

Methods—TB incidence, age-specific morbidity patterns, and rates of clustering of TB genotypes from 1986 to 2004 were compared between two groups of communities: Group 1, in which post-contact epidemics of TB were established around 1870, and Group 2, in which they were delayed until after 1920. Concomitant effects of socioeconomic and geographic variables were explored with multivariate models.

Results—Group 2 communities were characterized by higher annual incidence of TB (median 431/100 000 versus 38/100 000). In multivariate models that included socioeconomic and geographic variables, historical grouping remained a significant independent predictor of community incidence of TB. Clustering of TB genotypes was associated with Group 2 (OR 8.7, 95%CI 3.3–22.7) and age 10–34y (OR 2.5, 95%CI 1.1–5.7).

Conclusions—TB transmission dynamics can vary significantly as a function of a population's historical experience with TB. Populations at different stages along the epidemic trajectory may be amenable to different types of interventions.

Keywords

Native American; Health Status Disparity; Epidemic; Tuberculosis; Canada

*Equal contributions

Description of author contributions

CP conceived of the study, following observations made by VHH. VHH and CP collected and archived data and materials. TB genotyping was performed by CP and WW. Statistical analyses were designed and performed by AC, with input from CP and JP. CP drafted the paper, with input from AC. All authors revised the manuscript.

Final published version of manuscript

See References cited section[26] for citation of final published version of the manuscript

INTRODUCTION

Aboriginal Canadians remain at high risk for tuberculosis (TB), with recent incidence rates approximately five times the national average and as high as 20 times the rate among other Canadian-born populations[1]. There is, however, significant geographic variability in incidence and prevalence of tuberculosis among different Aboriginal populations in Canada[1]. Rates of tuberculosis among Aboriginal Canadians in the province of Saskatchewan are relatively high: estimated incidence rate 62/100 000/year (2007[2]) and prevalence of latent infection 47–61% among adults > 30 years (reserve communities, 1991–8[3]). We have observed significant, unexplained variability in rates of TB among Aboriginal communities within Saskatchewan: range of mean annual incidence 0-930/100 000/year, 1986–2004. Socioeconomic status is known to affect the risk of many illnesses, and could offer a potential explanation for the observed variability in rates of TB. A previous study of several individual- and area-level socioeconomic indicators in this population failed to find any consistent association with risk of tuberculosis[4]. Household crowding has been shown to potentiate transmission of tuberculosis in other settings[5–8], and another study found that high levels of household crowding increased the risk of 2 or more cases of tuberculosis in Canadian Aboriginal communities[9].

Early observers of tuberculosis epidemiology noted that even without effective interventions, rates of disease in a population will eventually peak, and then gradually fall[10–12]. The time scale for this process – centuries - is long in comparison with many other infectious diseases. Mathematical models of tuberculosis have demonstrated that this decline may occur simply as a result of intrinsic transmission dynamics, without having to invoke changes in the host or environment[13]. Using archival data, we have previously delineated the timing of TB epidemic initiation among Aboriginal communities of Saskatchewan, Canada[14] and classified these communities into two groups. **Group 1** communities experienced a relatively early shift to epidemic TB (1870s), whereas the shift to epidemic TB occurred later (>1920) in **Group 2** communities. The aim of this study was to determine whether historical classification (remote onset epidemic TB, **Group 1**, versus recent onset epidemic TB, **Group 2**) was predictive of the contemporary epidemiology of TB among Saskatchewan Aboriginal communities. A secondary aim was to determine whether differences between historical groups could be accounted for by known socioeconomic and geographic predictors of TB incidence.

METHODS

Data sources

This is a retrospective cohort study of clinical data from incident cases of TB, as well as genotyping data from associated isolates of *Mycobacterium tuberculosis* (*M.tb*). Data and isolates have been systematically collected and archived in database format by the TB control division of Saskatchewan, Canada since 1986; clinical data are described in detail in the supplemental online material (SOM). Inclusion criteria for clinical data and *M.tb* strains analyzed in this study were: residence in one of 67 Aboriginal communities (First Nations reserves and Métis communities, as defined by Census Canada) and diagnosis date between 1986 and 2004. The study was approved by the institutional review boards of Stanford University and the University of Saskatchewan.

Population registry data were obtained from Saskatchewan Health registrations. These data are described in detail in the SOM. Socioeconomic variables analyzed in this study include: measures of household crowding, education, unemployment and income (described in detail in SOM); all of these data were obtained from census reports (community profiles, 1996 and 2001). In addition to socioeconomic variables, we analyzed geographic remoteness data for

the communities included in this study. We used the Indian and Northern Affairs (INAC) "geographic zone" designation[15] for the communities.

Outcome measures

We compared two primary outcomes: incidence of tuberculosis, and risk of bacterial 'clustering'. Tuberculosis incidence was calculated by dividing the mean number of incident tuberculosis diagnoses per year that occurred from 1986 to 2004 by the mean number of persons per year living in the communities during the same time period. In addition, age-specific incidence was calculated for the following five age categories: 0–4, 5–9, 10–34, 35–64, and ≥65. These categories correspond to previously described patterns of age-specific tuberculosis morbidity, which have been observed to shift over time as a result of the cohort effect[16]. *M.tb* isolates were typed by restriction fragment length polymorphism analysis (RFLP), as described previously[14]. Cases were defined as clustered if the isolated *M.tb* strain shared an identical RFLP band pattern with that from another case and the cases occurred within two years of each other[17].

Definition of predictor variable

Historical community classification was our main predictor of interest. Classification procedures have been described previously[14]. In brief, 65 communities were divided on the basis of archival data into two groups: **Group 1**, in which epidemics of tuberculosis were initiated prior to 1920, and **Group 2**, in which epidemics were delayed until after 1920. The cutoff between the two groups (1920) is empirical and reflects different historical events linked with these epidemics. Two communities did not clearly fit either category and were excluded from the main analyses.

Statistical analysis

Group 1 and **2** communities were examined for differences in age distribution, community size, geographic remoteness, crowding, education, unemployment and income using t-, Wilcoxon and chi-square tests of association. Comparisons of tuberculosis incidence between the historical groups were evaluated using analysis of covariance (ANCOVA) to control for differences in remoteness, crowding, and education (variables found to be predictive of TB incidence in single factor analyses). Similarly, we used multivariate analysis of covariance (MANCOVA) to compare five age-specific TB incidences between the historical groups. Income data were available for 55 of the 65 communities and analyses of this smaller cohort were conducted separately. We tested for the presence of interactions in all models and addressed multiplicity in comparisons using the Tukey method.

For the evaluation of risk of tuberculosis case clustering based on historical grouping, differences in characteristics between groups were evaluated using Wilcoxon and chi-square tests of association. Characteristics that were significantly different among the study groups were tested in univariate logistic regression with clustering as the outcome. In order to not miss true associations that might be masked by confounding, terms that approached significance ($p \leq 0.10$) were tested in multivariate logistic regression. Statistics were performed using SAS version 9.13 (Cary, N. Carolina), with significance level set at 0.05 for all analyses.

RESULTS

Characteristics of **Group 1** and **Group 2** populations are shown in Table 1 and Figure 1a. Community size, age and sex composition were similar in the two groups. **Group 2** (recent epidemic onset) communities were more remote, had higher levels of household crowding,

lower levels of high school completion among adults, and higher levels of unemployment than **Group 1** communities. Household income was higher in **Group 2** communities.

Over the study interval (1986–2004), there were 228 cases of tuberculosis diagnosed in **Group 1** communities and 1243 cases in **Group 2** communities (see Table 2). Children < 5 years accounted for the largest proportion of cases, in both groups (see Figure 1b). Children ≥ 5y, adolescents and young adults < 35y accounted for a larger proportion of cases in **Group 2** versus **Group 1** communities (50% VS 32%, Figure 1b); there was a statistically significant difference in median age and age composition of **Group 1** versus **Group 2** TB cases ($p<0.001$ for both). There were four documented cases of reinfection in the **Group 2** population (i.e. distinct *M.tb* RFLP types isolated from the same individual on two occasions at least two years apart). There were no documented cases of reinfection in the **Group 1** population. **Group 2** cases were more likely to have a history of BCG vaccination, while **Group 1** cases were more likely to have a medical risk factor for TB identified.

Compared with **Group 1**, **Group 2** had higher median annual incidence of TB and a higher proportion of *M.tb* isolates clustered by RFLP analysis ($p<0.001$ for both, Table 3). **Group 2** communities also had a higher incidence of TB across the five age categories; the difference was not statistically significant in those over 65y (Table 3).

Predictors of higher TB incidence in single factor analyses were historical **Group 2**, geographic remoteness, increased household crowding and lower educational attainment among adults (Table 4). However, in multi-factor ANCOVA, historical grouping and geographic zone were the only variables that remained significant (Table 4). In multi-factor analysis of age-specific incidences of TB, historical group was a significant predictor in all categories except those older than 65y, remoteness was predictive in all except those < 5y and > 65y, and household crowding was significant only among those age < 5y (Table S2, SOM).

Among the 55 communities with income data, higher household income was associated with higher TB incidence in single factor ANCOVA ($p=0.002$, R-square 0.17). However, this association disappeared when income was added to the multi-factor analysis ($p=0.98$), and again only historical grouping ($p=0.002$) and remoteness ($p<0.001$) remained significant.

Univariate analysis of *M.tb* genotyping data showed several variables associated with increased risk of clustering (Table 5). These included community features (**Group 2**, remoteness, increased household crowding, lower education and higher unemployment) and individual characteristics (age and absence of a medical risk factor). However, in multivariate analysis, only residence in a **Group 2** community (OR 8.7) and age 10–34y (OR 2.5) were associated with a higher risk of clustering, whereas relapse/reinfection was associated with a lower risk of clustering (OR 0.4).

DISCUSSION

Our results demonstrate a strong influence of local historical dynamics on epidemiology of tuberculosis among Aboriginal populations of Saskatchewan, Canada. A simple system of historical epidemic classification was a powerful, and independent, predictor of tuberculosis incidence, age-specific patterns of tuberculosis morbidity, and clustering of *M.tb* genotypes. **Group 1** communities, where epidemic forms of tuberculosis were established in the 1870s[14], are now characterized by relatively low incidence of tuberculosis (mean 38/100 000/year between 1986 and 2004); incidence is, however, higher than contemporaneous estimates for non-Aboriginal Canadians (1.3/100 000/y in 2000[18]). Cases of TB in **Group 1** communities skew to older age categories (> 35y), are more likely to be associated with a medical risk factor, and are less likely to be associated with a clustered bacterial genotype.

Descriptive studies of epidemics presumed to be in decline (e.g. over recent decades in the United States[19, 20] and Canada[21]) and modeling studies[13] suggest that these features would predominate late versus early in an epidemic.

Epidemic forms of tuberculosis were delayed in **Group 2** populations until after 1920[14]. These populations are now characterized by high rates of TB (median annual incidence 431/100 000/y). Children and young adults <35 years predominated among TB cases from these communities. This age distribution of cases is consistent with theoretical predictions about an early stage epidemic[16, 22]. Reinfection with a new strain of *M.tuberculosis* was observed in the **Group 2** population, as it has been in other high incidence settings[23].

Relative to **Group 1, Group 2** communities had higher levels of household crowding, lower levels of adult educational attainment, and higher levels of unemployment. However, in statistical analyses that controlled for these variables, historical grouping remained a significant predictor of TB incidence. We observed an unexpected pattern of association between TB incidence and income, with higher income predictive of higher TB incidence in univariate analysis. This observation is likely due to confounding, as income was higher in **Group 2** communities and the association disappeared when historical grouping and zone were added to the analysis.

For both **Group 1** and **Group 2**, children < 5 years accounted for the highest proportion of TB cases. This may be due to active screening for TB among First Nations children living on reserve (see SOM for details). Active screening extends through the sixth grade; relatively lower rates in children 5–9y could reflect the efficacy of early case finding efforts. Alternatively, this observation could reflect basic patterns of susceptibility or transmission that vary with age: a similar nadir in rates of TB among school aged children has been observed in many populations[16]. Household crowding was predictive of age-specific incidence for children < 5y (and no other category), possibly reflecting the importance of household transmission for the youngest age group. Children < 10y were not statistically associated with clustering of *M.tb* genotypes; the sample size was probably too small to detect such an association, as few of the cases among young children were associated with a positive culture for *M.tb* (88/768 cases, age <10y).

The authors of a previous study of TB among Canadian First Nations identified an association between geographic remoteness and community risk of TB; given that it may be more difficult to deliver health care services to remote regions, they hypothesized that higher rates of tuberculosis in some communities were the result of lack of access to health care services[9]. We observed an association between geographic remoteness and **Group 2.** However, we do not believe that relative lack of access to health services is a complete explanation for high rates of TB in **Group 2** communities. In our study, historical grouping remained a strong predictor of TB incidence after accounting for differences in geographic remoteness. Furthermore, the signature of provincial TB control policies (active screening of young children and more diagnoses in this group) was evident to the same extent in **Group 1** and **Group 2**. A history of BCG vaccination was also more common among **Group 2** cases (Table 2), suggesting that, if anything, TB control measures were applied more widely in **Group 2**. Finally, TB incidence in **Group 2** trended downward over the study interval (see Figure S1). We speculate that geographic remoteness is one of the factors related to delayed onset of epidemic TB in **Group 2** communities. Inaccessibility of these populations could have resulted in different historical patterns of exposure to *M.tb* and/or historical delays in the ecological antecedents of epidemic TB.

The proportion of clustered TB genotypes was significantly higher in **Group 2** versus **Group 1** communities (87% VS 31%) and clustered genotype was strongly associated with

residence in a **Group 2** community (OR=8.7 in the multivariate model). Clustering of genotypes is usually interpreted as evidence of rapid progression to disease among recently infected individuals[24]. Theoretical predictions[13, 25] and descriptive epidemiological studies[21] suggest that this type of "fast" tuberculosis should predominate early in the epidemic cycle[14].

Although we are struck by the consistency of our findings with observations from other populations and theoretical predictions, there are limitations that bear mentioning. This was a study of a single region. It is possible that some features of the populations we studied preclude generalization of our findings to other settings. The clinical data analyzed here were not collected specifically for this study. Although attempts were made to systematically collect the data, there may be missing data that have biased our results. Finally, it is possible that there are unexplored explanatory variables and residual confounding that account for the differences we observed between historical groups.

Conclusions

We have delineated differences in tuberculosis epidemiology within a high risk population that correspond with distinct epidemic trajectories over the preceding century. Our findings suggest that underlying dynamics of disease transmission differ along the epidemic trajectory. Tailoring TB control efforts to the epidemic stage in the host population has the potential to improve disease eradication efforts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Edward Burns (professor emeritus, State University of New York at Binghamton) for thoughtful discussions of the material presented here. We are also grateful to staff of the TB control division of Saskatchewan, Canada for their efforts archiving data and materials used in this study.

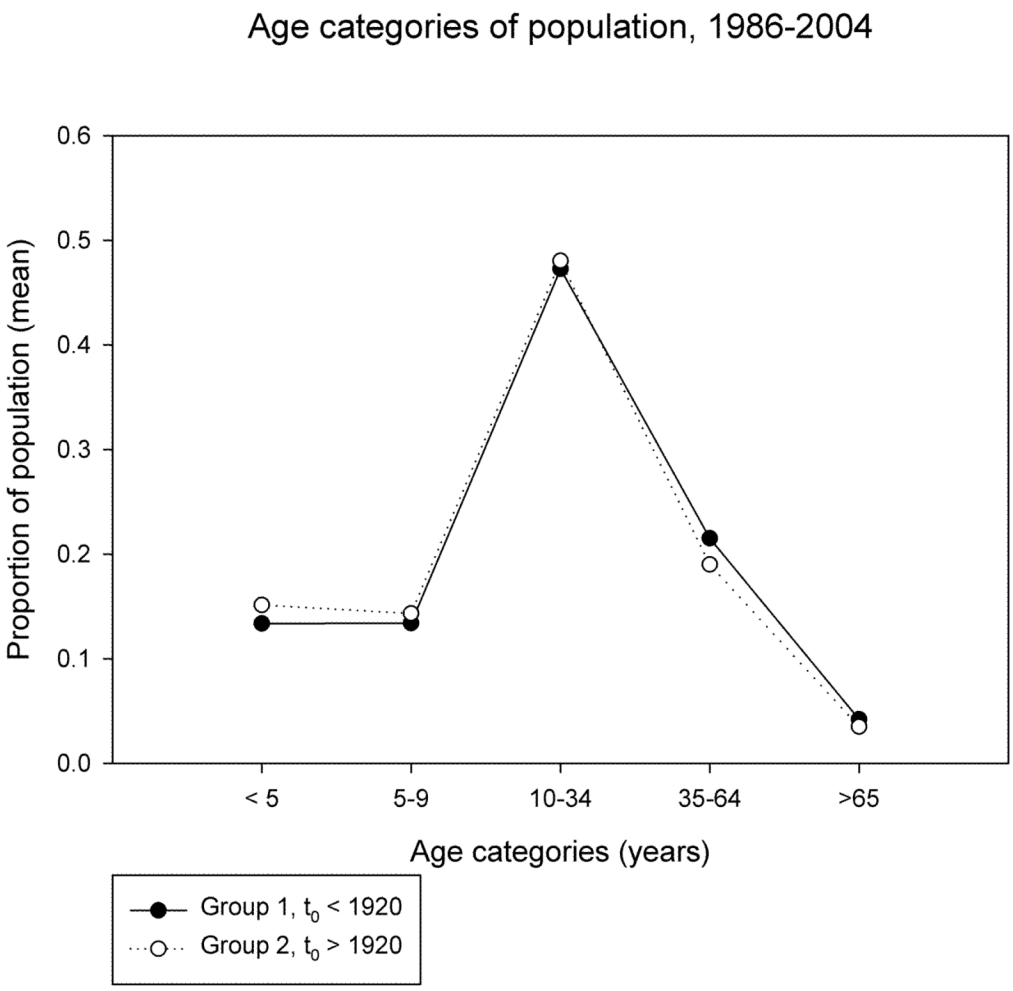
Funding

This work was supported by the National Institute of Allergy and Infectious Diseases at the National Institutes of Health (1 K08 AI067458-01A1 to C.P.) and the Stanford Center on Longevity (fellowship to A.C.).

References

1. Long, R. Health Mo. Her Majesty the Queen in Right of Canada. 2007. Canadian Tuberculosis Standards.
2. Ellis, E. Canada PHAo. Tuberculosis in Canada 2007. Ottawa: Public Health Agency of Canada; 2007.
3. Clark, M.; Riben, P. Canada H. Tuberculosis in First Nations Communities, 1999. Ottawa: 1999.
4. Ward, HA. Risk Factors in the Progression from Tuberculosis Infection to Disease. Saskatoon: University of Saskatchewan; 2004.
5. Bhatti N, Law MR, Morris JK, Halliday R, Moore-Gillon J. Increasing incidence of tuberculosis in England and Wales: a study of the likely causes. BMJ. 1995 Apr 15; 310(6985):967–9. [PubMed: 7728031]
6. Mangtani P, Jolley DJ, Watson JM, Rodrigues LC. Socioeconomic deprivation and notification rates for tuberculosis in London during 1982–91. BMJ. 1995 Apr 15; 310(6985):963–6. [PubMed: 7728030]
7. Hawker JI, Bakhshi SS, Ali S, Farrington CP. Ecological analysis of ethnic differences in relation between tuberculosis and poverty. BMJ. 1999 Oct 16; 319(7216):1031–4. [PubMed: 10521193]

8. Cantwell MF, McKenna MT, McCray E, Onorato IM. Tuberculosis and race/ethnicity in the United States: impact of socioeconomic status. Am J Respir Crit Care Med. 1998 Apr; 157(4 Pt 1):1016–20. [PubMed: 9563713]
9. Clark M, Riben P, Nowgesic E. The association of housing density, isolation and tuberculosis in Canadian First Nations communities. Int J Epidemiol. 2002 Oct; 31(5):940–5. [PubMed: 12435764]
10. Grigg ER. The arcana of tuberculosis with a brief epidemiologic history of the disease in the U.S.A. IV. Am Rev Tuberc. 1958 Oct; 78(4):583–603. [PubMed: 13583421]
11. Grigg ER. The arcana of tuberculosis; with a brief epidemiologic history of the disease in the U.S.A. III. Am Rev Tuberc. 1958 Sep; 78(3):426–53. contd. [PubMed: 13571601]
12. Grigg ER. The arcana of tuberculosis with a brief epidemiologic history of the disease in the U.S.A. Am Rev Tuberc. 1958 Aug; 78(2):151–72. contd. [PubMed: 13559649]
13. Blower SM, McLean AR, Porco TC, Small PM, Hopewell PC, Sanchez MA, et al. The intrinsic transmission dynamics of tuberculosis epidemics. Nat Med. 1995 Aug; 1(8):815–21. [PubMed: 7585186]
14. Pepperell C, Hoeppner VH, Lipatov M, Wobeser W, Schoolnik GK, Feldman MW. Bacterial genetic signatures of human social phenomena among M. tuberculosis from an Aboriginal Canadian population. Mol Biol Evol. 2010 Feb; 27(2):427–40. [PubMed: 19861642]
15. INAC. First Nation Profiles. [12/30/10]; Available from: www.ainc-inac.gc.ca
16. Frost WH. The age selection of mortality from tuberculosis in successive decades. Am J Hyg. 1939; 30:91–6.
17. Glynn JR, Bauer J, de Boer AS, Borgdorff MW, Fine PE, Godfrey-Faussett P, et al. Interpreting DNA fingerprint clusters of Mycobacterium tuberculosis. European Concerted Action on Molecular Epidemiology and Control of Tuberculosis. Int J Tuberc Lung Dis. 1999 Dec; 3(12): 1055–60. [PubMed: 10599007]
18. Ellis, E. Canada H. Tuberculosis in Canada 2000. Ottawa: 2000.
19. CDC. Services UDoHaH. Reported Tuberculosis in the United States, 2008. Atlanta: CDC; 2009.
20. Cronin WA, Golub JE, Lathan MJ, Mukasa LN, Hooper N, Razeq JH, et al. Molecular epidemiology of tuberculosis in a low- to moderate-incidence state: are contact investigations enough? Emerg Infect Dis. 2002 Nov; 8(11):1271–9. [PubMed: 12453355]
21. Grzybowski S. Tuberculosis A Look at the World Situation. Chest. 1983; 84(6):756–61. [PubMed: 6641310]
22. Vynnycky E, Fine PE. Lifetime risks, incubation period, and serial interval of tuberculosis. Am J Epidemiol. 2000 Aug 1; 152(3):247–63. [PubMed: 10933272]
23. van Rie A, Warren R, Richardson M, Victor TC, Gie RP, Enarson DA, et al. Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. N Engl J Med. 1999 Oct 14; 341(16):1174–9. [PubMed: 10519895]
24. Riley, LW. Molecular Epidemiology of Infectious Diseases Principles and Practices. Washington, D.C.: ASM Press; 2004.
25. Porco TC, Blower SM. Quantifying the intrinsic transmission dynamics of tuberculosis. Theor Popul Biol. 1998 Oct; 54(2):117–32. [PubMed: 9733654]
26. Pepperell C, Chang AH, Wobeser W, Parsonnet J, Hoeppner VH. Local epidemic history as a predictor of tuberculosis incidence in Saskatchewan Aboriginal communities. The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease. 2011 Jul; 15(7):899–905. [PubMed: 21682962]



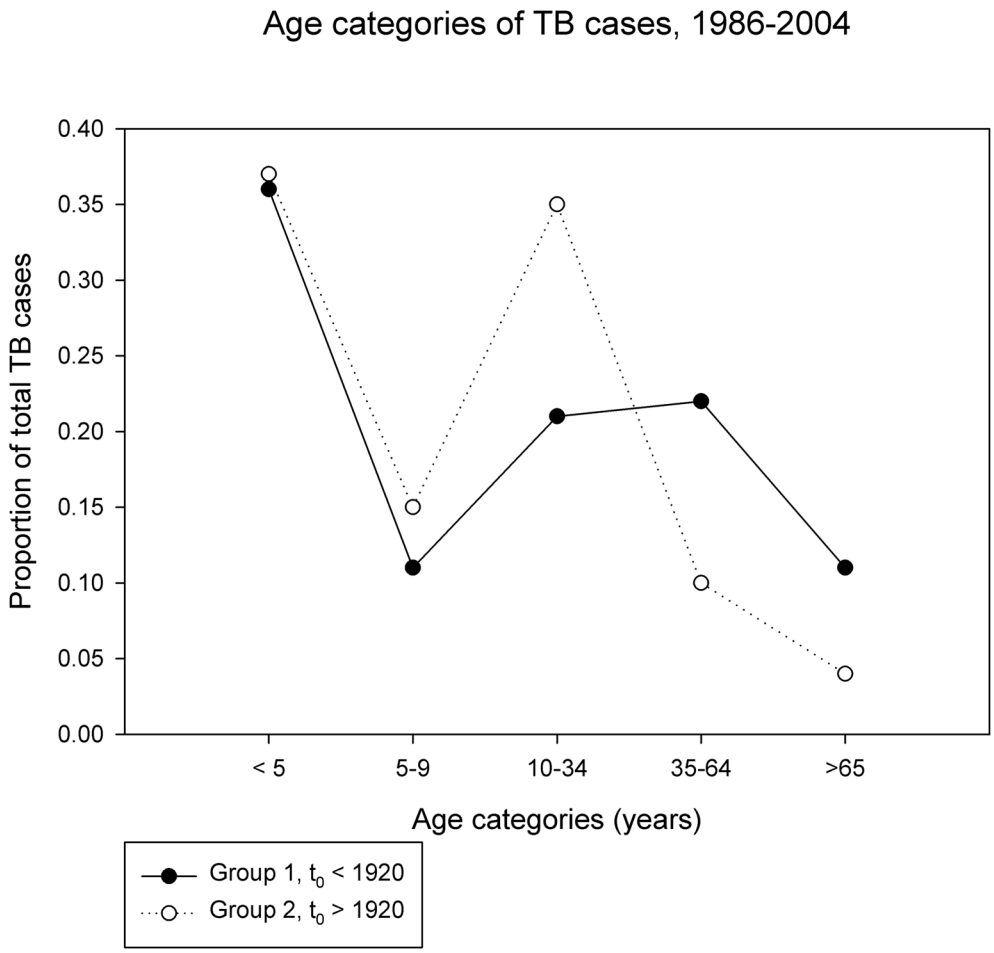


Figure 1. Age structure of populations

a. Average census totals for five age categories described in the text (see *Methods* and SOM) were calculated for each community, over the study period (1986–2004). The proportion of total population attributable to each category is shown on the Y axis. Filled circles are average values from communities where historical TB epidemics occurred prior to 1920, open circles are from communities where the epidemic started after 1920.

b. Proportion of total TB cases (1986–2004) attributable to different age categories is shown on the Y axis.

Table 1

Demographic, geographic and socioeconomic descriptors of **Group 1** and **Group 2** communities

	Group 1 Epidemic t_0 < 1920	**Group 2 Epidemic t_0 > 1920**	**P value[a]**
Number of communities	**45**	**20**	
	value (standard deviation)		
Average community size (1986–2004)[b]	598 (370)	887 (912)	0.48
Average number, aged < 5 years	78 (54)	141 (146)	0.62
5–9	78 (51)	133 (137)	
10–34	276 (172)	447 (467)	
35–64	126 (70)	177 (197)	
≥ 65	25 (14)	33 (38)	
% female (1986–2004)[c]	48.5 (2)	49.1 (1)	0.13
Average % of population in Group > 350 km from service center or without year-round road access to service center (1986–2004)[d]	0	28 (2)	< 0.001
Average % households > 1 person per room (1996, 2001)[e]	17 (11)	31 (12)	0.006
Average % adults in community with < high school education (2001)[f]	44 (13)	58 (13)	< 0.001
Average rate of unemployment (1996, 2001)[g]	29 (7)	32 (4)	0.02
Median annual household income (2001)[h]	20,121 (3,294)	23,575 (5,853)	0.03

[a] Wilcoxon rank-sum or chi-square test

[b] Average size of communities within each group over the study interval (1986–2004). Source: Saskatchewan Health registrations. Age category data are summarized from data reported in 5-year bins. Age category averages do not sum to total averages because age category data were not available for all years for which totals were available. Details in SOM.

[c] Mean % females within group

[d] Between 1986 and 2004, average percentage of population within the group that lived in the two most remote Indian and Northern Affairs (INAC) geographic categories; INAC classification is by distance from nearest service center.

[e] Average percentage of community households with > 1 person per room (Census 1996 and 2001). Room is defined as an enclosed indoor space, exclusive of bathrooms.

[f] Average percent of community aged 20–64y with less than a high school education, 2001 (Census).

[g] Average rate of unemployment for communities within each group (Census 1996 and 2001).

[h] Median household income for each community (Census 2001). Result shown is the average of communities within each historical group. Data missing (suppressed in census reports) for 7 communities in Group 1, and 3 communities in Group 2. Currency is the Canadian dollar.

Table 2

Characteristics of TB cases, **Group 1** and **Group 2**, 1986–2004

	Group 1 Epidemic t_0 < 1920	Group 2 Epidemic t_0 > 1920	P value[a]
Total number of TB cases	228	1243	
Age category, number (%)			<0.001
<5 years	82 (36)	456 (37)	
5–9	26 (11)	184 (15)	
10–34	47 (21)	434 (35)	
35–64	49 (22)	122 (10)	
≥65	24 (11)	47 (4)	
Median age (range)	11 (4mo-84y)	9 (2mo-88y)	<0.001
Male, number (%)	124 (54)	649 (52)	0.48
Disease category, number (%)[b]			0.21
pulmonary	164 (80)	931 (85)	
extra-pulmonary	29 (14)	133 (12)	
disseminated	11 (5)	36 (3)	
Cavitary pneumonia (%)[c]	13/206 (6)	90/1117 (8)	0.92
Relapse or reinfection (%)[d]	13 (6)	100 (8)	0.22
Documented reinfection (new strain)[e]	0	4	
History of LTBI treatment (%)[f]	5 (2)	26 (2)	0.95
History of BCG (%)	108 (47)	700 (56)	0.01
Sputum smear positive (%)[g]	44/104 (42)	231/513 (45)	0.44
Culture positive (%)[h]	81/180 (45)	475/1042 (46)	0.88
Medical risk factor identified (%)[i]	19 (8)	9 (0.7)	<0.001
Identified by screening (%)[j]	99/227 (44)	580/1240 (47)	0.31

[a] Wilcoxon rank-sum or chi-square test. Unless otherwise indicated, the denominator for the proportions is the total number of TB cases in each group. Detailed descriptions of variables are in *Methods* and SOM.

[b] Disease category data (pulmonary, extra-pulmonary, or disseminated) were available from 204 of the Group 1 cases and 1100 of the Group 2 cases. Percentages reported are proportion of cases from which disease classification data were available.

[c] Cavity identified on chest X ray/cases for which chest X ray results were available.

[d] History of discrete tuberculosis disease episode a minimum of 2 years prior to the current diagnosis

[e] Bacteria with ≥ 2 distinct RFLP types isolated from a single individual during distinct disease episodes. RFLP types differed by ≥5 bands. Denominator is the total number of cases within the group for which RFLP typing data were available.

[f] History of treatment for latent tuberculosis infection

[g] Number of sputum smear specimens with acid fast bacilli/number of sputum smear examinations recorded

[h] Number of patient specimens with *M.tuberculosis* identified in culture/number of tests recorded (≤1 specimen/case)

[i] Risk factor for tuberculosis identified in medical history. Risk factors included HIV infection, malignancy, renal failure, diabetes mellitus, and treatment with corticosteroids. No data were recorded regarding use of tobacco, alcohol or other recreational drugs.

[j] Case identified during through active surveillance, such as contact tracing investigations, tuberculin skin testing surveys, or population-based screens for tuberculosis disease. Denominator is number of cases for which these data were available.

Table 3

TB incidence and *M.tb* clustering, **Groups 1** and **2**, 1986–2004

	Group 1 Epidemic t_0 < 1920	**Group 2 Epidemic t_0 > 1920**	**P value[a]**
Median incidence (IQR)[b]	38 (15–61)	431 (52–554)	<0.001
Age < 5y	63 (0–162)	621 (117–1215)	<0.001
5–9y	0 (0–66)	341 (25–676)	<0.001
10–34y	19 (0–29)	119 (32–418)	<0.001
35–64y	29 (0–75)	140 (58–400)	0.001
>65y	0 (0–177)	180 (0–851)	0.06
Clustered genotype (%)[c]	19/61 (31)	334/386 (87)	<0.001

[a] Wilcoxon rank-sum or chi-square test

[b] Median incidence/100,000 population/year for Group 1 and 2 communities, interquartile range in brackets, adjusted for multiple comparisons.

[c] Number of *M.tb* genotypes identified as clustered by IS6110 RFLP fingerprinting/total number of bacterial isolates genotyped

Table 4

Predictors of community incidence of TB 1986–2004: Analysis of Covariance

ANCOVA	Independent variable[a]	F statistic[b]	P value	R-square[c]
Single factor	historical group[d]	44.94	<0.001	0.42
	zone[e]	27.86	<0.001	0.58
	crowding[f]	22.04	<0.001	0.27
	education[g]	20.33	<0.001	0.25
	unemployment[h]	1.97	0.17	0.031
Multi-factor	historical group	13.29	<0.001	0.71
	zone	13.19	<0.001	
	crowding	2.02	0.16	
	education	0	0.97	

[a] Independent (predictor) variable; dependent variable is community incidence of tuberculosis

[b] Mean square for factor (predictor) divided by mean square for error.

[c] Proportion of variation in dependent variable (community incidence of TB, 1986–2004) accounted for by the independent variable (single factor) or set of independent variables (multi-factor).

[d] Group 1 (epidemic $t_0 < 1920$) or Group 2 (epidemic $t_0 > 1920$)

[e] Indian and Northern Affairs (INAC) geographic categories; INAC classification is by [road] distance from nearest service center.

[f] Percentage of households within the community reporting > 1 person per room occupancy. Source: Census 1996 and 2001.

[g] Percent of the community aged 20–64y with less than a high school education, 2001 (Census).

[h] Average community rate of unemployment (Census 1996 and 2001).

Table 5

Logistic regression analysis of *M.tb* clustering among TB cases, 1986–2004

Category	OR (univariate)[a]	95% CI	P value	OR (multivariate)	95% CI	P value
Group 1 t_0 < 1920	1.0 (Ref)			1.0 (Ref)		
Group 2 t_0 > 1920	14.2	7.7–26.3	<0.001	8.7	3.3–22.7	<0.001
Crowding [c]	1.8	1.4–2.2	<0.001	1.2	0.9–1.8	0.26
Unemployment[d]	2.3	1.6–3.5	<0.001	1.3	0.6–2.7	0.48
Education[e]	2.4	1.4–2.1	<0.001	0.8	0.6–1.2	0.27
Zone 1[f]	1.0 (Ref)			1.0 (Ref)		
Zone 2	2.2	0.6–8.3	0.26	0.5	0.1–2.7	0.45
Zone 3	18.1	4–78	<0.001	2.9	0.4–23.1	0.32
Zone 4	7.1	1.7–29.3	0.006	1	0.2–6.1	0.99
Age < 5	4.3	1.5–12	0.006	1.5	0.4–5	0.5
Age 5–9	3.3	1–11.1	0.057	1.4	0.4–5.6	0.62
Age 10–34	4	2–7.9	<0.001	2.5	1.1 –5.7	0.034
Age 35–64	1.33	0.65–2.73	0.43	1.4	0.6–3.4	0.43
Age ≥ 65	1.0 (Ref)			1.0 (Ref)		
No medical risk factor[g]	5.5	2.33–13	<0.001	0.9	0.3–2.8	0.82
BCG vaccinated	1.17	0.73–1.89	0.52			
Relapse or reinfection[h]	0.6	0.32–1.1	0.10	0.4	0.2–0.9	0.025

[a] Odds ratio for clustering of bacterial isolate by RFLP analysis

[b] Variables approaching statistical significance (p≤0.10) in univariate analysis were included in the multivariate analysis so as not to miss associations hidden by confounding variables.

[c] Percentage of households within the community of residence with > 1 person per room (% >1ppr). Odds ratio for clustering is for 10% increase in this value.

[d] Rate of unemployment in community of residence (Census 1996 and 2001). Odds ratio for clustering is for 10% increase in unemployment rate.

[e] Percentage of adults in community of residence with less than a high school education (Census 2001). Odds ratio for clustering is for 10% increase in percentage of adults with less than a high school education.

[f] Geographic zone of community of residence. Zone defined by Indian and Northern Affairs Canada according to distance by road to the closest service center. Remoteness increases from zone 1 to zone 4.

[g] No known risk factor for tuberculosis. Risk factors included HIV infection, malignancy, renal failure, diabetes mellitus, and treatment with corticosteroids. No data were recorded regarding use of tobacco, alcohol or other recreational drugs.

[h] History of discrete tuberculosis disease episode a minimum of 2 years prior to the current diagnosis