Annals of Epidemiology 40 (2019) 21–27

Contents lists available at ScienceDirect

Annals of Epidemiology

Original article

Contextualizing tuberculosis risk in time and space: comparing time-restricted genotypic case clusters and geospatial clusters to evaluate the relative contribution of recent transmission to incidence of TB using nine years of case data from Michigan, USA

Grace A. Noppert, PhD [a, b, c, *], Zhenhua Yang, MD, PhD [a], Philippa Clarke, PhD [a, d], Peter Davidson, PhD [e], Wen Ye, PhD [f], Mark L. Wilson, ScD [a]

[a] Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor
[b] Center for the Study of Aging and Human Development, Duke University, Durham, NC
[c] Carolina Population Center, University of North Carolina, Chapel Hil
[d] Institute for Social Research, University of Michigan, Ann Arbor
[e] Michigan Department of Health and Human Services, Lansing, MI
[f] Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor

ARTICLE INFO

Article history:
Received 19 April 2019
Accepted 2 October 2019
Available online 17 October 2019

Keywords:
Tuberculosis
Neighborhood
Social disparities
Transmission

ABSTRACT

Purpose: Novel approaches must address the underlying factors sustaining the tuberculosis (TB) epidemic in the United States, specifically what maintains new *Mycobacterium tuberculosis* (*Mtb*) transmission.
Methods: Culture-confirmed TB cases reported to the Michigan Department of Health and Human Services (2004–2012) were analyzed for time-restricted genotypic and/or geospatial clustering. Cases with both types of clustering were used as a proxy for recent, local transmission. Modified, multivariate Poisson regression models were fit to estimate this prevalence in relation to various individual- and neighborhood-level demographic and socio-economic variables.
Results: Those individuals that were spatially clustered were 1.7 times as likely to also be time-restricted genotypically clustered. The prevalence of recent, local transmission was higher among U.S.-born cases, males, and non-Hispanic blacks. Moreover, people living in neighborhoods in the highest poverty quartile had 13.8 times the prevalence of recent, local transmission compared with those in the lowest poverty neighborhoods.
Conclusions: Our results suggest geographic areas with high concentration of TB cases are likely driven by ongoing transmission, rather than enclaves of individuals who have reactivated a case of latent TB. Furthermore, efforts to continue reducing *Mtb* transmission in the United States, and other low-incidence settings, must better identify community-level sources of risk, manifested through the complex social interactions among people and their environments.

Background

Beginning in 2013, the United States experienced stagnation in the decades-long decline in tuberculosis (TB) incidence, currently at ~3 per 100,000 persons per year [1,2]. This is disconcerting in light of the new World Health Organization goal of TB elimination in low-incidence countries such as the United States [3]. It suggests that new approaches are needed to understand and modify the factors that sustain transmission in the United States, and specifically the role of recent transmission.

TB is caused by infection with *Mycobacterium tuberculosis* (*Mtb*), transmitted primarily via airborne droplets [4]. An estimated 90%–

Disclosures: The study was approved by the Institutional Review Boards of the University of Michigan (HUM 00082884) and the Michigan Department of Health and Human Services (201509-07-EA).

Availability of data and material: The data that support the findings of this study are available from the Michigan Department of Health and Human Services but restrictions apply to the availability of these data, which were used under a data use agreement for the present study, and so are not publicly available. These data may be available through request to the Michigan Department of Health and Human Services.

Conflicts of interest: The authors declare that they have no competing interests.

* Corresponding author. Carolina Population Center, University of North Carolina, 123 West Franklin Street, Chapel Hil, NC 27516.
E-mail address: gnop@email.unc.edu (G.A. Noppert).

https://doi.org/10.1016/j.annepidem.2019.10.001
1047-2797/

95% of people who are infected with *Mtb* produce immune responses that force the pathogen into latency, resulting in latent TB infection (LTBI) [4]. Individuals with LTBI may later progress to active TB disease, a process that is highly dependent on individual risk factors, the most noteworthy being older age, malnutrition, immunosuppression (especially coinfection with HIV or diabetes) [5], stress, and other factors affecting host immunity [4,6]. Alternatively, active TB disease may arise shortly after initial infection if the initial immune response is inadequate. TB that appears in this manner is considered to reflect a recent transmission event, requiring a different public health strategy for control. Accordingly, a key feature of TB control strategies involves estimating the extent and context of recent transmission in a given population.

As molecular and epidemiologic tools have advanced, so has our ability to distinguish recent transmission from activation of LTBI. Many present studies have relied on molecular markers that use *Mtb* genotypic similarity as an indicator of recent transmission [7–9]. In addition, studies of people in the United States and in other developed countries have also used common social links with index cases to classify recent transmission [10]. Yet, there are inherent limitations to these methods that involve the discriminatory power of current genotyping technology and of establishing epidemiological links among cases. To address these challenges, TB researchers are increasingly developing more integrative approaches to better classify recent transmission [11–14]. One such approach, for example, is the use of whole-genome sequencing (WGS), in addition to data on epidemiologic links, to more accurately estimate chains of recent TB transmission [15].

Traditional paradigms of infectious disease epidemiology emphasize the importance of spatial patterns among individuals and populations in affecting transmission risk, particularly for those diseases in which close contact is necessary for pathogen spread. Accordingly, measures of physical proximity between people have been used in many analyses as an indicator of local transmission [11,12,16,17]. However, physical proximity may not always indicate local transmission given the unique natural history of TB. Individuals can be exposed and infected with *Mtb* long before they develop active TB disease, as the infection can exist in a latent state. Thus, it is plausible that there are concentrations of individuals with LTBI that do not represent ongoing recent transmission. Accordingly, it is imperative to infer whether clusters of individuals with TB disease in physical space correspond to recent, local transmission.

With a TB incidence rate of 1.3 per 100,000 persons per year in 2015 [18], the state of Michigan is a relatively low-burden state within the United States. Yet, burden is not evenly distributed, and there are persistent disparities, both at the individual and neighborhood levels [19,20]. Our study aims were two-fold. First, we developed a method of inferring recent, local *Mtb* transmission by integrating information on *Mtb* genotype, time, and location. Second, using nine years of TB case data from Michigan, we applied this method of inferring recent, local *Mtb* transmission to enhance understanding of how individual- and neighborhood-level social factors impact TB disparities, particularly among cases of recent, local transmission.

Methods

Data sources

There were 1800 TB cases reported by the Michigan Department of Health and Human Services from January 1, 2004, to December 31, 2012. Of the 1800 reported cases, 1390 (77%) were culture-confirmed, allowing for *Mtb* isolate genotyping. Our study sample initially comprised all TB cases reported in Michigan from 2004 through 2012 for which genotyping and residential address data were available. Cases that lacked both spoligotype and 12-locus-MIRU-VNTR results, and/or address information were excluded. The final study sample of 1235 cases represented 69% of all 1800 cases recorded by Michigan Department of Health and Human Services, and 89% of the 1390 culture-confirmed cases.

Demographic characteristics of each case were drawn from the Report of a Verified Case of TB form [21], developed by the U.S. Centers for Disease Control and Prevention and used for routine TB surveillance in Michigan. TB case addresses were geocoded using ESRI ArcGIS (v10.4) and then linked to U.S. census block groups. Block group–level characteristics were derived from the American Community Survey (ACS), 2012, 5-year estimates [22]. We used the linkage to the ACS data to assess neighborhood density and neighborhood poverty. Additional information on the creation of both of these variables is included in the supplementary material.

Cluster definitions

In this study, we classified individuals as belonging to one of four types of clusters (including nonclustered) based on a combination of genotype (spoligotype and 12-locus MIRU/VNTR), diagnostic year, and spatial location data (Fig. 1). Individuals in the study sample were classified as being time-restricted genotypically clustered TB cases only, geospatially clustered TB cases only, both time-restricted genotypically and geospatially clustered TB cases, or neither time-restricted genotypically nor geospatially clustered TB cases. Time-restricted genotypic clusters of TB cases were those sharing an identical genotype by spoligotype and 12-locus MIRU/VNTR, as well as diagnosis dates within one year of each other, which is similar to previous studies [14,23]. Genotypically nonclustered TB cases were those not sharing an identical genotype or having diagnosis dates more than one year apart.

Geospatial clusters of TB cases were identified through an algorithm based on Kulldorff's scan statistic [24] using the standard SaTScan (ver. 9.4.3) software, in which various parameters are specified by the investigator. This analysis "scans" the total study region in user-defined circular subareas to detect those areas with a greater-than-expected number of cases per population. Distances between cases were based on the geographic centroid of each census block group. We used the discrete Poisson model for purely spatial clusters, with a spatial cluster maximum radius of 1.0 miles. The "expected" case counts were determined based on the Poisson assumption that the expected number of cases is proportional to the population size of a given area. Both case and population counts per census block group were considered to account for population differences among block groups. Sensitivity analyses were also performed to evaluate whether inferences regarding geospatial clustering would change with different sizes of the circle radius (see Supplementary Table 2).

There were 47 unique spatial clusters of TB cases with an average radius distance of 0.64 miles. The cluster sizes ranged from three to 37 cases. Eight of the clusters likely reflect cases within the same household. Additional information on the spatial clusters of cases is given in Supplementary Table 3.

Using these genotypic clustering and geospatial clustering results, we then defined a case of recent, local transmission as any case that was both genotypically and geospatially clustered with at least one other case. Previous studies have shown that the use of additional data points can improve the accuracy of identifying recent transmission. Notably, using a "gold-standard," field-based epidemiology investigation, France et al. [13] demonstrated that by defining recent transmission through integration of several data types (i.e., genotype, geographic distance, time of diagnosis, source case infectiousness), the accuracy of recent transmission

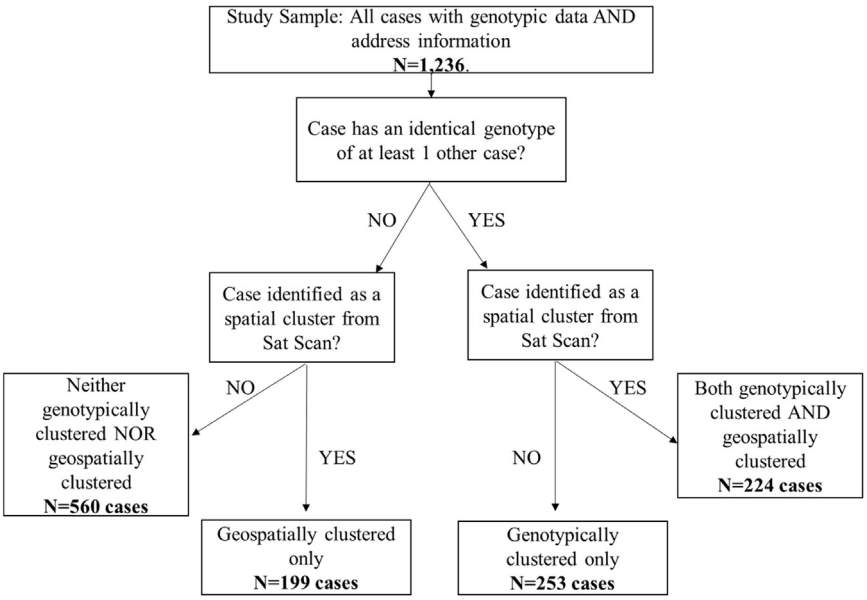


Fig. 1. Flowchart showing the assignment of clustering definitions in the study sample.

identification was improved compared with that based solely on genotype or space. By the same logic, our approach of analyzing multiple data types to classifying recent-local transmission should also improve the accuracy of inferences regarding recent transmission.

Statistical analyses

The primary outcome of this study was being classified as a case of recent, local transmission (vs. any of the three individual cluster classifications). Again, individuals were classified as a case of recent, local transmission if they were both time-restricted genotypically clustered and spatially clustered. We first examined differences in the demographic characteristics of TB cases by type of cluster using a χ^2 test. Regression analyses were then used to estimate the prevalence of recent, local transmission by a series of individual- and neighborhood-level social factors. Specifically, we used modified, multivariable Poisson regression models with generalized estimating equations (GEE) and a compound symmetry correlation structure and log link. This approach was necessary because the prevalence of recent, local transmission was greater than 10% in the study sample, thus precluding traditional logistic regression in which the odds ratio would overestimate the prevalence ratio. These Poisson regression models were adjusted for individual- and neighborhood-level sociodemographic factors. We used GEE to account for nesting of cases within census block groups [25]. Modified Poisson models with GEE allow for accurate parameter estimation and robust variance estimates by accounting for the correlation existing among cases in the same block group. The association of individual-level factors (race/ethnicity, sex, nativity, and age) with the prevalence of recent, local transmission was evaluated (model 1). Model 2 includes the covariates in model 1 in addition to neighborhood-level factors (neighborhood density and neighborhood poverty).

The prevalence ratio of recent, local transmission and corresponding 95% confidence interval (CI) were calculated based on the above Poisson models. We used a two-tailed alpha level equal to 0.05 to determine significance. Statistical analyses were carried out in SAS (ver. 9.4).

The study was approved by the Institutional Review Boards of the University of Michigan and the Michigan Department of Health and Human Services.

Results

Study population

The demographic composition of the 1235 TB cases included in the study sample is described in Supplementary Table 1. The average annual incidence rate of TB for the state of Michigan during the study period was 1.49 per 100,000 person years (ranging from 1.02 to 2.01). Using the ACS, the cumulative incidence for the total 9-year study period was 16.2 per 100,000 person, ranging from 0 to 6250 per 100,000 persons by census block group. The block groups with incidence rates greater than the average were most concentrated in the Metro Detroit area.

Prevalence of clustering

We then examined the distribution of each cluster type in the sample (see Supplementary Fig. 1). A total of 253 cases (20%) were classified as solely time-restricted genotypically clustered, meaning they shared an identical genotype with at least one other case in the sample and were diagnosed within one year of that case. Another 199 cases (16%) were classified as solely geospatially clustered, which meant that their census block group at the time of diagnosis was within a 1-mile radius of another case's block group in the study sample. Yet another 224 cases (18%) were classified as cases of recent, local transmission (both time-restricted genotypically and geospatially clustered). The remaining 559 cases (45%) were nonclustered. As compared with nongenotypically clustered cases, those that were time-restricted genotypically clustered were nearly twice as likely to also be spatially clustered (prevalence ratio = 1.79, $P < .001$).

The distribution of cluster type showed differing patterns by key covariate categories (Table 1). A higher proportion of male cases was classified as recent, local transmission compared with female cases. Over half of the non-Hispanic white, Asian, and Hispanic cases were classified as nonclustered, whereas the largest

Table 1
Comparison of types of clustering across study year and key sociodemographic characteristics

Sociodemographic factor	Nonclustered (neither geospatially nor genotypically clustered) N (%)	Geospatially clustered only N (%)	Genotypically clustered only N (%)	Recent, local transmission (geospatially and genotypically clustered) N (%)	*P*-value
Total sampled	559 (45)	199 (16)	253 (20)	224 (18)	
Gender					
Male	318 (43)	108 (15)	160 (22)	148 (20)	**.02**
Female	241 (48)	91 (18)	93 (19)	75 (15)	
Race/Ethnicity					
Non-Hispanic white	182 (64)	29 (10)	55 (19)	19 (7)	**<.0001**
Non-Hispanic black	124 (25)	95 (19)	104 (21)	173 (35)	
Asian	179 (59)	47 (15)	60 (20)	18 (6)	
Hispanic	62 (51)	23 (19)	25 (20)	12 (10)	
Other	12 (43)	5 (18)	9 (32)	2 (7)	
Nativity					
U.S.-born	235 (35)	91 (13)	166 (24)	188 (28)	**<.0001**
Foreign-born	324 (59)	108 (20)	86 (16)	35 (6)	

P-value corresponds to a χ^2 test examining whether the distribution of cluster type differs by each sociodemographic variable. The proportions reflect row proportions. Example: Males are more likely to be nonclustered (43%) compared with geospatially clustered alone (15%).
P values below .05 are bolded.

proportion of non-Hispanic black cases were classified as recent, local transmission. Nearly 60% of the foreign-born cases were classified as nonclustered compared with 35% of the U.S.-born cases.

Prevalence of local transmission

We developed and compared models of recent, local transmission in relation to various individual- and neighborhood-level covariates (Table 2). Using only individual-level characteristics (model 1), race/ethnicity, nativity, and age were all significant independent predictors of recent, local transmission. After controlling for all other covariates, the prevalence of recent, local transmission in non-Hispanic blacks was 4 times greater than that of non-Hispanic whites. Hispanics and Asians were not statistically significantly different from non-Hispanic whites in prevalence of recent, local transmission, although a greater risk was observed for both Hispanics (2.5 times) and Asians (1.9 times), as compared with the non-Hispanic whites (Table 2).

The prevalence of recent, local transmission among U.S.-born individuals was 3.5 times that of foreign-born persons ($P = .0001$), after controlling for all other covariates (Table 2). Furthermore, the prevalence of recent, local transmission was lower in younger age groups than that in the 18- to 64-year-old group. People aged less than 18 years had a 69% reduced prevalence of recent, local transmission, which was significantly lower than that in the 18- to 64-year-old group.

In model 2, we added neighborhood density and neighborhood poverty to the analysis. Using both individual- and neighborhood-level covariates, results showed that nativity, age, and neighborhood-level poverty were significant predictors of recent, local transmission (Table 2). The prevalence of recent, local transmission among U.S.-born was 2.14 times that of foreign-born ($P = .004$), controlling for all other covariates. Those who were younger than 18 years had a 63% lower prevalence of local transmission than the 18- to 64-year-old group, which was statistically significant.

Individuals residing in neighborhoods in the highest quartile of poverty had 13.8 times the prevalence of recent, local transmission compared with those in the lowest poverty neighborhoods (Q4 vs. Q1) ($P < .0001$), controlling for all other covariates (Table 2). Moreover, there was a monotonic relationship of increasing neighborhood poverty and increasing prevalence of recent, local transmission.

Sensitivity analyses

Several sensitivity analyses were performed to further interrogate our findings. First, we tested whether time-restricted genotypic clusters were more likely to be spatial clusters. Results showed that those individuals who were classified as time-restricted genotypic clusters were more likely to be spatial clusters (Supplemental Table 4) (prevalence ratio: 1.71; 95% CI: 1.48, 1.97).

We also further examined the group of individuals who were time-restricted genotypically clustered, but not spatially clustered. On average, those genotypic clusters that were time-restricted (but not spatial clusters) were less likely to be non-Hispanic black and U.S.-born. These individuals were also more likely to be older (65+ years) and live in less-dense and lower-poverty neighborhoods. Finally, they were less likely to report homelessness in the past 12 months, injection drug use, HIV-positive status, and pulmonary TB. Full results of this additional analysis are reported in the supplement (Supplemental Table 5).

We also examined the distribution of race/ethnicity and nativity by neighborhood-level covariates. That analysis showed significant differences in the distribution of both race and nativity by level of neighborhood poverty. Full results are provided in the supplement (Supplemental Table 6).

Discussion

Using Michigan TB surveillance data from 2004 through 2012, we applied an integrative approach for classifying recent, local *Mtb* transmission that accounted for bacterium genotype, geographic location of cases, and time to examine sociodemographic disparities in disease. There were two key findings from these analyses. First, our more complex classification of recent transmission highlighted ongoing disparities in TB risk, and particularly the need to address *Mtb* transmission in high-poverty neighborhoods. Second, as compared with previous studies, we developed more robust inferences concerning risk factors associated with recent, local transmission by using an integrative approach involving multiple data types. We believe that the more conservative spatial and temporal parameters we used are well-suited to infer recent transmission in urban environments, where the prevalence of *Mtb* infection (i.e., LTBI) is unknown.

One aim of our study was to develop a method that explores how integrating multiple types of data could improve identification of ongoing transmission in an urban community setting. Previous

Table 2
Estimates of the adjusted prevalence ratio of recent, local transmission by neighborhood-level and individual-level factors

Factor	Model 1 PR (95%)	*P*	Model 2 PR (95%)	*P*
Race/Ethnicity				
Non-Hispanic black	3.97 (1.80, 8.73)	**<.0001**	1.77 (0.97, 3.22)	.0722
Asian	1.87 (0.64, 5.45)		1.66 (0.72, 3.83)	
Hispanic	2.47 (0.87, 7.04)		1.25 (0.59, 2.64)	
Non-Hispanic white	Ref.		Ref.	
Gender				
Male	1.19 (0.86, 1.64)	.3000	1.12 (0.87, 1.45)	.3811
Female	Ref.		Ref.	
Nativity				
U.S.	3.54 (1.87, 6.72)	**.0001**	2.14 (1.27, 3.60)	**.0044**
Foreign-B	Ref.		Ref.	
Age category				
<18 y	0.31 (0.12, 0.71)	**.0013**	0.37 (0.18, 0.76)	**<.0001**
65 + y	0.73 (0.37, 1.43)		0.84 (0.46, 1.52)	
18–64 y	Ref.		Ref.	
Neighborhood density				
Q4			1.28 (0.86, 1.89)	.5300
Q3			1.09 (0.72, 1.65)	
Q2			1.08 (0.70, 1.67)	
Q1			Ref.	
Neighborhood poverty*				
Q4			13.78 (4.76, 39.87)	**<.0001**
Q3			8.07 (2.82, 23.10)	
Q2			3.25 (1.07, 9.93)	
Q1			Ref.	

Results based on modified Poisson models with GEE. *P*-values reflect the type 3 analysis effects.
Model 1 considers only individual-level demographic characteristics. In model 2, neighborhood density and poverty are added as covariates.
P values below .05 are bolded.
* Higher quartiles reflect greater poverty or higher density.

studies have demonstrated that the use of additional data sources to classify recent transmission may produce inferences that are closer to the "gold standard", that is shoe-leather epidemiological contact investigations [11,13]. Using this logic, we sought to make use of data on *Mtb* genotype, time of diagnosis, and spatial location of cases to provide a more robust proxy of recent transmission. We tested whether time-restricted genotypically clustered cases were more likely to be spatially clustered than nongenotypically clustered cases, and applied sensitivity analysis to assess the robustness of our working definition of spatially clustered cases. The results indicated that time-restricted genotypically clustered cases were nearly twice as likely to be spatially clustered as nongenotypically clustered cases. The value of this approach is that two complementary types of information were combined in hopes of increasing the accuracy of classifying recent transmission. Combining multiple lines of data improves our ability to understand or infer recent transmission, and thereby improves policies for prevention.

Several recent reports have documented the value of integrative approaches to classifying recent *Mtb* transmission. Moonan et al. [12] used a method that first defined genotypically matched cases, and then identified geographic areas with higher than expected prevalence of genotypic clustering, which they defined as geospatial clustering. However, residence was represented by zip code, which would have missed finer-resolution spatial patterns. The report by France et al. [13] that established the plausible-source case approach was a rigorous assessment of several different methods of classifying recent transmission. Notably, they compared their approach to the gold standard of epidemiologic field evaluation by optimizing sensitivity and specificity given varying geographic distances and disease prevalences. However, that study did not consider whether the geographic distance that is relevant for predicting transmission of *Mtb* may differ depending on other contextual features. For example, in places where much of the TB burden is concentrated in urban areas, a 10-mile radius may be too large a distance to recognize meaningful geospatial clusters of individuals with a shared environment conducive to transmitting *Mtb*. In comparison, our approach is likely more conservative using smaller geographic distances to determine geospatial clusters which may more accurately reflect the shared spatial environment of individuals leading to ongoing transmission of *Mtb*. Moreover, our analysis extends beyond these previous studies by performing both multivariable individual-level and neighborhood-level risk factor analyses.

Notably, using WGS, Guthrie et al. [15] found the estimated transmission rate among pediatric cases was considerably lower than using MIRU-VNTR alone. WGS is a major development in TB molecular epidemiology and may represent the new standard for incorporating genomic data into studies of recent transmission of TB.

In sensitivity analyses, we examined the 20% of the sample that were time-restricted genotypically clustered, but not spatially clustered. Based on our findings, we believe these likely represent cases linked by remote transmission. For example, the transmission event may have happened sometime in the past, individuals dispersed and then reactivated at a later time point. However, given the unique natural history of TB infection and the complexity of TB pathogenesis, as well as the limitations in the discriminatory power of the genotyping methods, it is difficult to determine with certainty when transmission occurred.

Nearly half of cases in our study were neither geographically nor genotypically clustered, likely due to reactivation of LTBI. Notably, more foreign-born (59%) than U.S.-born cases were classified as nonclustered, which was confirmed in multivariable analyses. This finding is consistent with the hypothesis that foreign-born persons are more likely to be infected with *Mtb* in their country of origin, and then progress to active TB disease sometime after their arrival in the United States [26–28]. Other studies have reported similar findings regarding foreign-born populations [11,12,16]. One implication may be that TB control efforts should continue to focus on reducing activation of LTBI among foreign-born populations.

One of the key findings of our study was the importance of social context in sustaining *Mtb* transmission in the community. Markers of both individual- and neighborhood-level social context are critical, and interact with one another. One of our previous studies reported that U.S.-born, non-Hispanic blacks were the highest risk racial/ethnic group for TB incidence [19], and that the association with race was largely explained by markers of neighborhood poverty [20]. However, that report did not incorporate geospatial data. In the present analysis, we tested whether this trend held true when using a more nuanced measure of recent transmission that simultaneously also accounted for spatial context. Indeed, our findings have affirmed the importance of the neighborhood social environment in determining recent, local transmission.

Despite the strengths in our approach to classifying recent, local transmission, there are several limitations. First, the residential address may not be accurate for all cases. In addition, while the use of GEE can account for correlations between individuals in the same block group, there could also be some residual dependency among blocks group captured within the same SaTScan spatial cluster. This could result in biased standard errors—smaller than what would be expected if such dependency were corrected.

Although a more specific designation of recent, local transmission was devised, we are still not able to infer the direction of pathogen transmission. In addition, while we believe that adding a time restriction to the genotypic cluster definition improved accuracy, some misclassification may still have occurred, particularly for those cases diagnosed at the beginning or end of the study period. There were also 535 individuals excluded from the study sample because of either missing genotypic information or address information. Notably, among the 535 missing, 16% were among individuals aged less than 18 years. Pediatric TB is widely considered a useful proxy for recent transmission as children (particularly those aged less than 5 years) have necessarily only been infected for short time [29–32]. Thus, excluding those individuals likely leads to an underestimate of the prevalence of recent transmission in our sample.

Finally, recent studies have shed light on the limits of current genotyping methods in inferring recent transmission. Studies find that transmission clusters involving the Beijing family are particularly poorly resolved using MIRU-VNTR typing [33,34]. The Beijing genotype is the major component of the L2 lineage [35]. In the present study, 60% of cases were from the Euro-American lineage (L4) and only 14% from the East Asian lineage (L2). Thus, we believe that based on genotyping alone, the MIRU-VNTR may be performing better than in other studies which have higher proportions of the L2 lineage. However, more importantly, these studies further highlight the necessity of approaches such as the current one. Incorporating multiple sources of data to infer recent transmission is critical, especially in contexts where WGS was not available.

Conclusion

As the United States moves toward the laudable goal of TB elimination, it is imperative to understand which factors contribute, and in which ways, to recent *Mtb* transmission. Our findings highlight the critical role that neighborhood social context continues to play in creating and/or maintaining disparities in TB incidence in the United States. Improved TB control will require a deeper understanding of how complex social interactions occurring among individuals and as part of their environments are driving transmission.

Acknowledgment

The authors thank their collaborators at the TB Control Program at the Michigan Department of Health and Human Services, United States for their contribution to the collection of the data used for this study.

Dr. Noppert received partial support from the University of Michigan Rackham Graduate School, United States and from the following grants through the National Institutes of Health: National Institute on Aging, United States (grant number T32-AG000029–40 and K99AG062749-01A1 and Eunice Kennedy Shriver National Institute of Child Health and Human Development, United States (grant number T32 HD091058).

Authors' contributions: G.N. participated in the design of the study, carried out the data analysis, interpreted the data, and drafted and revised the manuscript. Z.Y. assisted with interpretation of study results and revising of the manuscript. P.C. assisted with interpretation of study results and revising of the manuscript. W.Y. assisted with the data analysis and revising of the manuscript. P.D. coordinated the data collection and assisted with revising the manuscript. M.W. conceived and directed the study, assisted with interpretation of study results, and revising of the manuscript. All authors read and approved the final manuscript.

References

[1] Salinas JL. Leveling of tuberculosis incidence—United States, 2013–2015. MMWR Morb Mortal Wkly Rep 2016;65:273–8.
[2] Talwar A, Tsang C, Price S, Pratt R, Walker W, Schmit K, et al. Tuberculosis — United States. MMWR Morb Mortal Wkly Rep 2018;2019(68):257–62.
[3] Lönnroth K, Migliori GB, Abubakar I, D'Ambrosio L, De Vries G, Diel R, et al. Towards tuberculosis elimination: an action framework for low-incidence countries. Eur Respir J 2015;45(4):928–52.
[4] Nelson KE, Williams CM, Graham NM, Balows A. Infectious disease epidemiology theory & practice. Massachusetts: Elsevier; 2001.
[5] Dye C, Williams BG. The population dynamics and control of tuberculosis. Science 2010;328(5980):856–61.
[6] Russell DG. Who puts the tubercle in tuberculosis? Nat Rev Microbiol 2007;5(1):39.
[7] Rodwell TC, Kapasi AJ, Barnes RF, Moser KS. Factors associated with genotype clustering of Mycobacterium tuberculosis isolates in an ethnically diverse region of southern California, United States. Infect Genet Evol 2012;12(8):1917–25.
[8] Curtis A, Ridzon R, Novick L, Driscoll J, Blair D, Oxtoby M, et al. Analysis of Mycobacterium tuberculosis transmission patterns in a homeless shelter outbreak. Int J Tuberc Lung Dis 2000;4(4):308–13.
[9] Fenner L, Gagneux S, Helbling P, Battegay M, Rieder HL, Pfyffer GE, et al. Mycobacterium tuberculosis transmission in a country with low tuberculosis incidence: role of immigration and HIV infection. J Clin Microbiol 2012;50(2):388–95.
[10] Fitzpatrick LK, Hardacker JA, Heirendt W, Agerton T, Streicher A, Melnyk H, et al. A preventable outbreak of tuberculosis investigated through an intricate social network. Clin Infect Dis 2001;33(11):1801–6.
[11] Yuen CM, Kammerer JS, Marks K, Navin TR, France AM. Recent transmission of tuberculosis—United States, 2011–2014. PLoS One 2016;11(4):e0153728.
[12] Moonan PK, Ghosh S, Oeltmann JE, Kammerer JS, Cowan LS, Navin TR. Using genotyping and geospatial scanning to estimate recent Mycobacterium tuberculosis transmission, United States. Emerg Infect Dis 2012;18(3):458.
[13] France AM, Grant J, Kammerer JS, Navin TR. A field-validated approach using surveillance and genotyping data to estimate tuberculosis attributable to recent transmission in the United States. Am J Epidemiol 2015;182(9):799–807.
[14] Berzkalns A, Bates J, Ye W, Mukasa L, France AM, Patil N, et al. The road to tuberculosis (Mycobacterium tuberculosis) elimination in Arkansas; a re-examination of risk groups. PLoS One 2014;9(3):e90664.
[15] Guthrie JL, Delli Pizzi A, Roth D, Kong C, Jorgensen D, Rodrigues M, et al. Genotyping and whole-genome sequencing to identify tuberculosis transmission to pediatric patients in British Columbia, Canada, 2005–2014. J Infect Dis 2018;218(7):1155–63.
[16] Moonan PK, Bayona M, Quitugua TN, Oppong J, Dunbar D, Jost KC, et al. Using GIS technology to identify areas of tuberculosis transmission and incidence. Int J Health Geogr 2004;3(1):23.
[17] Prussing C, Castillo-Salgado C, Baruch N, Cronin WA. Geo-epidemiologic and molecular characterization to identify social, cultural, and economic factors where targeted tuberculosis control activities can reduce incidence in Maryland, 2004–2010. Public Health Rep 2013;128(6_suppl3):104–14.
[18] Michigan TB Epidemiology 2018. Lansing, MI: Michigan Department of Health and Human Services; 2019.
[19] Noppert GA, Wilson ML, Clarke P, Ye W, Davidson P, Yang Z. Race and nativity are major determinants of tuberculosis in the US: evidence of health

disparities in tuberculosis incidence in Michigan, 2004–2012. BMC Public Health 2017;17(1):538.
[20] Noppert GA, Yang Z, Clarke P, Ye W, Davidson P, Wilson ML. Individual-and neighborhood-level contextual factors are associated with Mycobacterium tuberculosis transmission: genotypic clustering of cases in Michigan, 2004–2012. Ann Epidemiol 2017;27(6):371–376.e5.
[21] CDC|TB| Report of a Verified Case of Tuberculosis Form (RVCT), Available at https://www.cdc.gov/tb/programs/rvct.
[22] American Community Survey (ACS), Five-Year Estimates. Generated by Grace Noppert; using Social Explorer. U.S. Census Bureau; 2016.
[23] France AM, Cave MD, Bates JH, Foxman B, Chu T, Yang Z. What's driving the decline in tuberculosis in Arkansas? A molecular epidemiologic analysis of tuberculosis trends in a rural, low-incidence population, 1997–2003. Am J Epidemiol 2007;166(6):662–71.
[24] Kulldorff M. A spatial scan statistic. Commun Stat-Theory Methods 1997;26(6):1481–96.
[25] Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol 2004;159(7):702–6.
[26] Nuzzo JB, Golub JE, Chaulk P, Shah M. Postarrival tuberculosis screening of high-risk immigrants at a local health department. Am J Public Health 2015;105(7):1432–8.
[27] Ricks PM, Cain KP, Oeltmann JE, Kammerer JS, Moonan PK. Estimating the burden of tuberculosis among foreign-born persons acquired prior to entering the US, 2005–2009. PLoS One 2011;6(11):e27405.
[28] Stennis N, Trieu L, Perri B, Anderson J, Mushtaq M, Ahuja S. Disparities in tuberculosis burden among South Asians living in New York City, 2001–2010. Am J Public Health 2015;105(5):922–9.
[29] Parrinello CM, Crossa A, Harris T. Seasonality of tuberculosis in New York City, 1990–2007. Int J Tuberc Lung Dis 2012;16(1):32–7.
[30] Ruwende J, Sanchez-Padilla E, Maguire H, Carless J, Mandal S, Shingadia D. Recent trends in tuberculosis in children in London. J Public Health 2010;33(2):175–81.
[31] Marquez L, Feske ML, Teeter LD, Musser JM, Graviss EA. Pediatric tuberculosis: the litmus test for tuberculosis control. Pediatr Infect Dis J 2012;31(11):1144.
[32] Centers for Disease Control and Prevention. TB in Children 2014. https://www.cdc.gov/tb/topic/populations/tbinchildren/default.htm. [Accessed 9 January 2019].
[33] Koster KJ, Largen A, Foster JT, Drees KP, Qian L, Desmond E, et al. Genomic sequencing is required for identification of tuberculosis transmission in Hawaii. BMC Infect Dis 2018;18(1):608.
[34] Wyllie DH, Davidson JA, Smith EG, Rathod P, Crook DW, Peto TE, et al. A quantitative evaluation of MIRU-VNTR typing against whole-genome sequencing for identifying Mycobacterium tuberculosis transmission: a prospective observational cohort study. EBioMedicine 2018;34:122–30.
[35] Shitikov E, Kolchenko S, Mokrousov I, Bespyatykh J, Ischenko D, Ilina E, et al. Evolutionary pathway analysis and unified classification of East Asian lineage of Mycobacterium tuberculosis. Sci Rep 2017;7(1):9227.

Supplementary Material

Creation of neighborhood indices

TB case addresses were geocoded using ESRI ArcGIS (v10.4) and then linked to U.S. census block groups. Block group–level characteristics derived from the American Community Survey (ACS), 2012 5-year estimates [21]. We used the linkage to the ACS data to create an index of neighborhood poverty.

We conducted a factor analysis on a set of seven demographic variables at the block group level. From the factor analysis, we chose variables which had factor loadings greater than 0.4 to include in the final index. These variables included percent of the block group that was black, percent of the block group with less than a high school education, percent of the block group unemployed, percent of the block group utilizing public assistance, percent of all properties in the block group that were vacant properties, and percent of the block group with an income-to-poverty level ratio below 2. Neighborhood poverty was calculated as the mean of each of the six items at the block group level (alpha reliability = 0.82). With regard to the income-to-poverty ratio, the U.S. Census defines an income-to-poverty ratio below 2 as being "poor or struggling." Thus, we used 2.0 as our cut-point to determine the proportion of the census block group with an income-to-poverty ratio below 2.0.

Supplementary Table 1
Comparison of the study sample ($n = 1235$) with those excluded from the study sample ($n = 565$), Michigan 2004–2012

Sociodemographic factor	Study sample ($n = 1235$)		Excluded from study sample ($n = 565$)		*P*-value
	N	%	N	%	
Race/ethnicity					
Non-Hispanic white	285	23	154	27	.06
Non-Hispanic black	496	40	210	37	
Asian	304	25	119	21	
Other	28	2	10	2	
Hispanic	122	10	72	13	
Age (y)					
<18	35	3	89	16	**<.0001**
18–64	890	72	373	66	
65+	310	25	103	18	
Sex					
Male	734	59	313	55	.21
Female	500	40	252	45	
Missing	1	<0.1	0		
Nativity					
Foreign-born	553	45	245	43	.16
U.S.-born	680	55	316	56	
Missing	2	<0.2	4	<1	
Site of disease					
Pulmonary	855	69	333	59	**<.0001**
Extrapulmonary	276	22	187	33	
Both	102	8	41	7	
Missing	2	<0.2	4	<1	

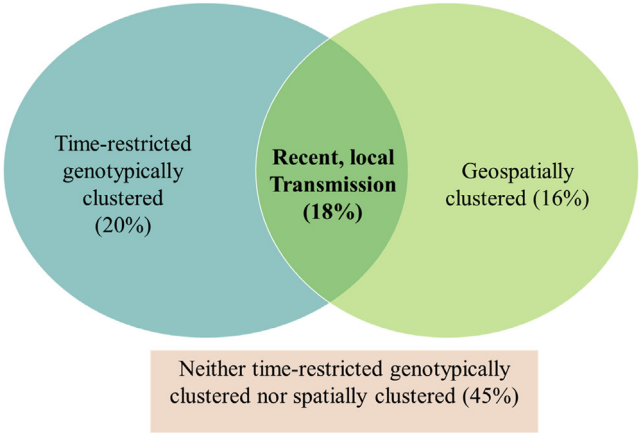
Classifications of race/ethnicity, age, sex, nativity, and site of disease were defined base on the Report of Verified Case of TB form developed by the Centers for Disease Control and Prevention.
P values less than .05 are bolded.

Supplementary Table 2
Sensitivity analyses to show the results based on different radii size of the circle

Radius of circle	Nonclustered (neither geospatially nor genotypically clustered)	Geospatially clustered only	Genotypically clustered only	Recent, local transmission (geospatially and genotypically clustered)
	N (%)	N (%)	N (%)	N (%)
0.5 mile	599 (49)	159 (13)	314 (25)	163 (13)
1 mile (our study)	559 (45)	199 (16)	253 (20)	224 (18)
1.5 miles	504 (41)	254 (21)	227 (18)	250 (20)
2 miles	512 (41)	246 (20)	216 (17)	261 (21)
5 miles	427 (35)	331 (27)	183 (15)	294 (24)
10 miles (AM France)	329 (27)	429 (35)	133 (11)	344 (28)

Supplementary Table 3
Detailed information on the 1-mile spatial clusters

Cluster number	Number of cases	Radius (km)
1	37	1.44
2	11	0
3	20	1.49
4	20	1.56
5	16	1.58
6	13	1.02
7	16	1.28
8	13	1.28
9	13	1.35
10	7	0.81
11	14	1.53
12	11	1.36
13	15	1.5
14	9	1.5
15	11	1.54
16	12	1.6
17	6	0.78
18	7	0.98
19	10	1.4
20	7	0.82
21	9	1.57
22	5	0
23	11	1.6
24	5	0.56
25	9	1.58
26	8	1.58
27	7	1.33
28	8	1.14
29	4	0.34
30	7	1.38
31	4	0
32	6	1.11
33	9	1.56
34	3	0
35	10	1.35
36	6	1.47
37	5	0.79
38	3	0
39	5	0.81
40	5	1.36
41	3	0
42	3	0
43	5	0.93
44	5	1.39
45	3	0.47
46	3	0
47	4	0.77
	Mean	1.019 (km) 0.634 (mi)



Supplementary Fig. 1. Venn diagram showing the relationships among the different types of clusters.

Supplementary Table 4
Testing whether those in a spatial cluster were more likely to be time-restricted genotypic clustered as well

Spatial cluster	Time restricted genotypic cluster	
	=0	=1
=0	559	253
=1	199	224

P value of χ^2 test <.0001.

Model results

Those that were spatially clustered had 1.7 (95% CI: 1.48, 1.97) times the odds of being time-restricted genotypically clustered compared with those not in a spatial cluster ($P < .0001$).

Supplementary Table 5
Comparison of demographic characteristics of those individuals who were time-restricted genotypically clustered only to those that were both time-restricted genotypically clustered and spatially clustered

Sociodemographic factor	Time-restricted genotypically clustered only N (%)	Geospatially and genotypically clustered N (%)	*P*-value
Total sampled	253 (20)	224 (18)	
Gender			
Male	160 (22)	148 (20)	.48
Female	93 (19)	75 (15)	
Race/ethnicity			
Non-Hispanic white	55 (19)	19 (7)	**<.0001**
Non-Hispanic black	104 (21)	173 (35)	
Asian	60 (20)	18 (6)	
Hispanic	25 (20)	12 (10)	
Other	9 (32)	2 (7)	
Nativity			
U.S.-born	166 (24)	188 (28)	**<.0001**
Foreign-born	86 (16)	35 (6)	
Age category			
<18 y	10 (4)	8 (4)	**.03**
18–64 y	197 (78)	194 (87)	
65 + y	46 (18)	22 (10)	
Neighborhood density			
Q4	71 (28)	81 (36)	**.0005**
Q3	60 (24)	56 (25)	
Q2	53 (21)	60 (27)	
Q1	69 (27)	27 (12)	
Neighborhood poverty			
Q4	55 (22)	138 (62)	**<.0001**
Q3	71 (28)	64 (29)	
Q2	61 (24)	18 (8)	
Q1	66 (26)	4 (2)	
Homeless in the last 12 mo	18 (7)	29 (13)	**.03**
Contact of an active TB case	22 (9)	31 (14)	.07
Noninjection drug use	29 (11)	39 (17)	.17
Injection drug use	7 (3)	20 (9)	**.01**
HIV positive	17 (7)	23 (10)	**.02**
Extrapulmonary TB	68 (27)	42 (19)	**.04**

P-value corresponds to a χ^2 test examining whether the distribution of cluster type differs by each sociodemographic variable.
P values less than .05 are bolded.

Supplementary Table 6
Comparison of neighborhood density and neighborhood SES by race and nativity

Sociodemographic factor	Non-Hispanic white %	Non-Hispanic black %	Asian %	Hispanic %	Other race %	U.S.-born %	Foreign-born %
Neighborhood density							
Q4	18	32	18	33	21	26	24
Q3	22	26	25	25	32	25	25
Q2	21	27	30	16	14	25	25
Q1	39	15	28	26	32	24	26
Neighborhood poverty							
Q4	7	53	2	15	11	41	5
Q3	20	30	17	38	18	26	24
Q2	33	14	32	34	36	17	35
Q1	40	3	50	14	36	15	37

Higher quartiles of neighborhood poverty are indicative of increased poverty. χ^2 test of differences by density: density*race: $P < .0001$; density*nativity: $P = .73$; χ^2 test of differences by neighborhood poverty: neighborhood poverty*race: $P < .0001$; neighborhood poverty *nativity: $P < .0001$.