HHS Public Access

Author manuscript

Am J Prev Med. Author manuscript; available in PMC 2022 August 01.

Published in final edited form as:

Am J Prev Med. 2021 August ; 61(2): 201–208. doi:10.1016/j.amepre.2021.02.006.

Tuberculosis Genotype Clusters and Transmission in the U.S., 2009–2018

Jonathan M. Wortham, MD,

Rongxia Li, PhD,

Sandy Althomsons, MA, MHS,

Steve Kammerer, MBA,

Maryam B. Haddad, PhD, MPH,

Krista M. Powell, MD, MPH

Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, Atlanta, Georgia

Abstract

Introduction: In the U.S., universal genotyping of culture-confirmed tuberculosis cases facilitates cluster detection. Early recognition of the small clusters more likely to become outbreaks can help prioritize public health resources for immediate interventions.

Methods: This study used national surveillance data reported during 2009–2018 to describe incident clusters (≥3 tuberculosis cases with matching genotypes not previously reported in the same county); data were analyzed during 2020. Cox proportional hazards regression models were used to examine patient characteristics associated with clusters doubling in size to ≥6 cases.

Results: During 2009–2018, a total of 1,516 incident clusters (comprising 6,577 cases) occurred in 47 U.S. states; 231 clusters had ≥6 cases. Clusters of ≥6 cases disproportionately included patients who used substances, had recently experienced homelessness, were incarcerated, were U.S.-born, or self-identified as of American Indian or Alaska Native race or of Black race. A median of 54 months elapsed between the first and third cases in clusters that remained at 3–5 cases, compared with 9.5 months in clusters that grew to ≥6 cases. Longer time between first and third cases and presence of ≥1 patient aged ≥65 years among the first 3 cases predicted lower hazard for accumulating ≥6 cases.

Conclusions: Clusters accumulating ≥3 cases within a year should be prioritized for intervention. Effective responses strategies should include plans for targeted outreach to U.S.-born individuals, incarcerated people, those experiencing homelessness, people using substances, and individuals self-identifying as of American Indian or Alaska Native race or of Black race.

INTRODUCTION

Despite low tuberculosis (TB) incidence in the U.S., outbreaks continue to cause preventable illnesses and deaths.[1] Traditionally, outbreak detection has depended on contact

Address correspondence to: Jonathan Wortham, MD, Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, 1600 Clifton Rd. NE, MS E-10, Atlanta GA 30333. vij5@cdc.gov.

investigations that are triggered by identification of infectious TB cases, consisting of focused efforts to identify people exposed and evaluate them for TB disease and infection.[2] Unfortunately, contact investigations are imperfect despite the best efforts of public health investigators because, willfully or not, patients sometimes do not name all of their contacts.[1] Nonetheless, routine analyses of universal genotyping data can supplement contact investigations and other outbreak detection and response efforts, based on the principle that *Mycobacterium tuberculosis* isolates from cases related by recent transmission will share a genotype.[3] To that end, since 2004, the Centers for Disease Control and Prevention (CDC) has provided universal genotyping of at ≥1 *M. tuberculosis* isolate for each culture-confirmed TB case in the U.S., allowing for detection of genotype-matched cases clustered in space and time that might represent TB outbreaks.[1,3]

Prompt cluster detection facilitates timely response to prevent and control outbreaks and reduce TB-associated morbidity. Early recognition of the small clusters that are more likely to turn into large outbreaks can help prioritize public health resources for immediate interventions.[4] CDC's predictive analysis of genotype-clustered cases during 2004–2010 identified homelessness, excess alcohol or illicit drug use, or incarceration among any case in a cluster of 3 cases within 5.3 months as associated with the cluster becoming an outbreak; outbreaks were defined as ≥6 cases with matching spacer oligonucleotide typing (spoligotype) and 12-locus mycobacterial interspersed repetitive units variable number of tandem repeats analysis (MIRU-VNTR) patterns that were confirmed by local epidemiologic investigation to represent outbreaks.[5] Additionally, knowledge of case characteristics associated with outbreaks could prompt critical review and optimization of TB control and prevention strategies focused toward certain demographic and social groups. After that last analysis, TB incidence in the United States continued to decline, and routine genotyping methods expanded from 12- to 24-locus MIRU-VNTR to improve discriminatory power.[2,6,7] Using these more recent, higher resolution data, this study aims to identify factors associated with growth of genotype clusters from 3 to 6 cases, which might be early signs of outbreaks.

METHODS

Study Sample

Since 1993, the National Tuberculosis Surveillance System (NTSS) has collected demographic, clinical, and laboratory characteristics of all TB cases reported in the U.S.[8] NTSS captures also captures whether patients experienced homelessness or used substances (excess alcohol or any illicit drugs) in the year before TB diagnosis, whether patients were incarcerated at diagnosis, and their counties of residence. CDC's National TB Genotyping Service used spoligotyping and 12-locus MIRU-VNTR analyses during 2004–2008. In January 2009, MIRU-VNTR analysis expanded to include 12 additional loci (i.e., 24-locus MIRU-VNTR).

This study included all TB cases with genotyping results reported from the 50 U.S. states and District of Columbia to NTSS during 2009–2018. Because these data are used for public health practice and obtained during routine TB surveillance, CDC deemed this analysis not to be human subjects research and, therefore, did not require IRB review.

Measures

This analysis focused on incident clusters, defined as ≥3 genotype-matched cases preceded by a 24-month period during which no cases with matching spoligotype and 24-locus MIRU-VNTR patterns (or matching spoligotype and 12-locus MIRU-VNTR during any part of the 24-month period before 2009) were reported in that county. The date used for each case in this analysis was the earliest of the following 3 dates: the date of treatment initiation, the date the case was verified and counted by the state TB controller, or the date of the first clinical specimen from which drug-susceptibility results were available.

Statistical Analysis

Chi-square tests were used to compare the frequency distributions of demographic and clinical characteristics of patients in incident clusters of 3–5 cases with characteristics of patients in clusters with ≥6 cases.

After examination of the patient-level characteristics, cluster-level characteristics were examined; clusters were classified as having a given demographic or clinical characteristic if ≥1 patient among the first 3 case patients had the given demographic or clinical characteristic (e.g., those involving ≥1 person experiencing homelessness compared with those that had none). Then, Cox proportional hazards regression models were used to assess associations between these cluster-level characteristics and the hazard of accumulating ≥6 cases; incident clusters that did not accumulate 6 cases before December 31, 2018 were censored. The Cox model facilitated investigation of associations between demographic and clinical factors among ≥1 of the first 3 cases and the hazard of clusters doubling in size from 3 to 6 cases, provided for censoring of clusters that never met the outcome of interest (i.e., ≥6 cases), and accounted for variable periods of time between the third case and censoring. Associations were quantified between demographic and clinical factors and risk per unit time (i.e., hazard rate) of accumulating ≥6 cases; hazard ratios (HRs) and 95% CIs were calculated that compared the hazard of clusters accumulating ≥6 cases given any of following characteristics among ≥1 of the initial 3 cases: 5 age groups, race/ethnicity, experiencing homelessness, diagnosis while incarcerated, U.S.-born, substance use, sputum smear positivity, presence of radiograph findings consistent with pulmonary TB, and residence in a predominately urban versus rural county (or county-equivalent area) to the hazard of accumulating ≥6 cases among clusters that did not have ≥1 patient with the given characteristic (e.g., clusters with ≥1 patient aged 5–14 years compared with clusters with no patients aged 5–14 years).

The Cox model also facilitated simultaneous examination of multiple demographic and clinical covariates. Adjusted HRs (AHRs) and 95% CIs were estimated by accounting for time (in months) between the diagnoses of the first and third case in the cluster and the presence of any of the following covariates among the initial 3 cases that had p-values <0.20 in bivariate models: age 5–14 years, age 45–64 years, age ≥65 years, American Indian or Alaska Native race, Black race, homelessness within past year, U.S. birth, excess drug or alcohol use, and urban county as defined by the U.S. Department of Agriculture.[9] The appropriateness of the proportional hazards assumption for each of these variables

was verified by calculating Schoenfeld residuals, examining log–log curves, and assessing models with time as an interaction variable.

Because the variable representing time between the first and third case did not satisfy the proportional hazards assumption for the entire data set, stratified analyses were conducted calculating AHRs for risk factors associated with the hazard of cluster growth within the 24 months following the third case in the cluster. The proportional hazards assumptions were met for the strata consisting only of incident clusters and follow-up periods of ≤24 months after the third case. Because the highest risk for developing TB disease is thought to be during the 24-month period following infection, subsequent analyses used this model for identifying risk factors among incident clusters that would benefit from immediate intervention (i.e., compared with a model of follow-up periods >24 months after the third case).[10]

RESULTS

During 2009–2018, a total of 97,921 TB cases were reported to NTSS (Figure 1). Of these, 26,412 (27%) cases were excluded because no genotyping results were available; about 3 quarters (n=19,959) were clinically diagnosed cases without laboratory confirmation. Of the remaining 71,509 genotyped TB cases, investigators excluded 51,012 with a genotype unique to a particular jurisdiction, 6,934 that matched only 1 other case in a jurisdiction, and 6,986 that were in pre-existing clusters (i.e., there was no 24-month period without a genotype-matched case; therefore, the first case was not identifiable). These exclusions left 6,577 cases in 1,516 incident clusters for analysis; 2,091 (32%) of these cases were in 231 incident clusters of ≥6 cases.

Most (68%) cases in incident clusters occurred in patients aged 25–64 years (Table 1). Although the overall number of non-Hispanic American Indian or Alaska Native patients in clusters was small, their proportion in clusters with ≥6 cases was 4 times that of clusters with 3–5 cases (8% vs 2%, p<0.0001). More patients in ≥6 case clusters were non-Hispanic Black (Black) compared with 3–5 case clusters (33% vs 26%, p<0.0001). Homelessness was more common among patients in clusters with ≥6 cases compared with 3–5 case clusters (12% vs 9%, p<0.00001). More patients in ≥6 case clusters were U.S.-born compared with 3–5 case clusters (66% vs 48%, p<0.01), and HIV coinfection was similar in ≥6 case clusters (7% vs 6%, p=0.06). Other characteristics were similar across incident clusters. Approximately a third of patients in incident clusters had radiographic evidence of cavitary disease; for most patients in incident TB clusters, symptomatic presentations to health care prompted their TB diagnostic evaluations.

Of the 1,516 incident clusters, 534 had ≤24 months between the third and sixth case or December 31, 2018, the end of the study period. The median time between the third and sixth cases was 11.7 (IQR=6.7–17.1) months in the 110 clusters that became ≥6 cases. Similarly, the median time between the third case and censoring was 12.3 (IQR=7.0–18.1) months among the 424 clusters that accumulated only 3–5 cases. However, among 424 incident clusters that accumulated only 3–5 cases (i.e., did not become ≥6 cases), the median time between the first and third cases was 54 (IQR=31.3–75.8) months, compared with a

median time of 9.5 (IQR=3.7–20.4) months between the first and third cases among 110 clusters that did accumulate ≥6 cases. Table 2 shows that a longer time interval between the first and third case was associated with lower hazard of a cluster becoming ≥6 cases in both the crude (HR=0.94 per month, 95% CI=0.93, 0.95) and adjusted (AHR=0.95 per month, 95% CI=0.93, 0.96) models. In other words, each additional month between the first and third case was associated with a 5%–7% reduction in the hazard of becoming ≥6 cases in the unadjusted model and a 4%–7% reduction in the adjusted model.

Incident clusters with ≥1 patient aged ≥65 years among the first 3 cases had lower hazard of becoming ≥6 cases in both the crude (HR=0.4, 95% CI=0.2, 0.6) and adjusted (AHR=0.4, 95% CI=0.2–0.7) models. Having ≥1 patient with Black race (HR=1.5, 95% CI=1.0, 2.1) or ≥1 patient with American Indian/Alaska Native race (HR=3.8, 95% CI=2.0, 7.2) predicted higher hazard of cluster growth to or ≥6 cases in unadjusted models. Nonetheless, neither race nor ethnicity of first 3 cases were statistically significant predictors of cluster growth in the multivariable model (Table 2).

DISCUSSION

County-based incident clusters of ≥6 TB cases disproportionately included patients who used substances, had recently experienced homelessness, were incarcerated at diagnosis, were U.S.-born, or self-identified as of American Indian or Alaska Native or Black race. However, after accounting for the simultaneous effects of several demographic and clinical factors, only 2 factors among the first 3 were statistically significant predictors of hazard of growth to ≥6 cases. First, each additional month between the first and third cluster case was associated with approximately a 5% lower hazard rate of cluster growth. Second, diagnosis of ≥1 of the first 3 cases in patients aged ≥65 years was associated with an approximately 60% lower hazard of growth to ≥6 cases.

County-based incident genotype clusters with ≥3 cases were infrequent, and even fewer doubled in size to ≥6 cases. However, incident clusters of ≥3 cases occurred in nearly every state during the study period. This finding is consistent with other data that suggest recent *M. tuberculosis* transmission represents a relatively small but persistent fraction of TB disease in the U.S., perhaps 10%–15% of genotyped cases.[11] Though relatively infrequent, *M. tuberculosis* transmission likely affects distinct groups that might differ from a county's typical TB caseload of primarily non–U.S. born patients and could require different TB control strategies. Geospatially concentrated clusters of cases with matching *M. tuberculosis* genotypes often involve U.S.-born patients, representing pockets of *M. tuberculosis* transmission in the U.S.[6,11]

Approximately 4%–6% of TB cases annually in the U.S. occur among patients who are or have recently experienced homelessness; in this analysis, 16% of cases in clusters with ≥6 cases, and 9% among incident clusters with 3–5 cases, occurred among such individuals.[6] These findings are consistent with others' observations that *M. tuberculosis* transmission among people experiencing homelessness remains a problem in the U.S., perhaps further demonstrating that traditional TB control strategies are less effective among those with unstable housing.[1,13–16] Traditional TB control measures that depend on patients being able

to name close contacts often have limited utility in large congregate settings (e.g., homeless overnight facilities) where there is a greater opportunity for transmission among people who are less likely to know identifying information about those with whom they have had shared airspace.[1,13,16,17] This association between homelessness and large clusters could also reflect delayed diagnoses, because individuals experiencing homelessness often have limited access to the healthcare system, leading to delayed TB diagnoses, longer periods of symptomatic, contagious disease, and more *M. tuberculosis* transmission.[18]

Excess alcohol use and illicit drug use were also common among patients in incident clusters of both sizes. Both characteristics have been associated with infectiousness and poor treatment outcomes among TB patients; these characteristics could also be proxies for people with increased contact with others (e.g., shared substance use venues), lack of healthcare access, and, perhaps, delayed diagnosis.[20] For example, patients might be unwilling or unable to name contacts encountered at substance use venues. To stop *M. tuberculosis* transmission in these social networks, public health programs must be able to establish trust and gain access to contacts at risk, which is often challenging.

Clusters involving individuals who self-identified as being American Indian or Alaska Native or Black race also had higher hazard of becoming clusters with ≥6 cases. Clusters involving American Indians and Alaska Natives were rare; most clusters fitting this description occurred in communities with substantial geographic and systemic barriers to health care. Anecdotally, the clustered TB cases among American Indians and Alaska Natives often resulted from active case finding efforts in isolated communities where transmission had already occurred; therefore, cases were diagnosed in short succession and these clusters accumulated 6 cases quickly. Anecdotally, among Black individuals, many of the clustered diagnoses occurred in people with other social risk factors and the time between cases was often short. Adjustment for these social risk factors and time interval between first and third cases made the observed associations between race and accumulation of ≥6 cases disappear. Conversely, ≥1 cases among people aged ≥65 years predicted clusters that stayed 3–5 cases; this is likely because TB among older patients disproportionately represents reactivation of latent TB infection acquired in the remote past.

Limitations

This analysis has limitations. First, this approach might have overestimated sizes of some clusters, because matching genotype patterns were defined by spoligotype and 24-locus MIRU-VNTR (i.e., conventional but incomplete genetic characterization of *M. tuberculosis* isolates). On the other hand, cluster size might have been underestimated for some, because genotyping can only be performed for culture-confirmed cases—often excluding young children with TB, for example, for whom culture confirmation is less common.[21] Additionally, clusters were defined using county (or county equivalent) borders, but *M. tuberculosis* transmission often involves patients in multiple jurisdictions. Nonetheless, this analysis should still be useful because public health decision making and service delivery often occur at the county level. Although there is no standard definition for clusters or outbreaks, given the relative rarity of clusters, doubling of cluster size (i.e., growing from 3 to 6 cases) should prompt additional consideration and potential public health

action in any jurisdiction. The cluster definition also facilitated comparisons with previous work in the U.S. as well as similar work in other countries.[5,22,23] Other unmeasured case characteristics, such as social determinants (e.g., poverty, household size, access to health care) and occupation, could have been factors in *M. tuberculosis* transmission and ultimately cluster size. One potential explanation behind the finding that 3 cases in rapid succession was associated with cluster growth is that further transmission had already occurred (e.g., prolonged infectious period of an initial case) by the time public health authorities were able to implement effective countermeasures. To examine these important social determinants and the promptness or adequacy of TB control and prevention measures would require a different kind of study design; these key factors are not part of national TB surveillance and unable to be included in this analysis. Finally, though this study examined patient-level characteristics associated with cluster growth, information from epidemiologic investigations was not available. Without these data, systematically determining where cluster-associated transmission occurred and the relative contributions of transmission in different settings (e.g., household versus non-household congregate settings) was not possible.

CONCLUSIONS

In recent years, approximately two thirds of TB cases reported in the U.S. have occurred in non–U.S. born patients.[6] Despite this, 48% of patients in clusters of ≥3 cases were U.S.-born, and >60% of patients in clusters of ≥6 cases were U.S.-born. Additionally, clusters disproportionately included patients who used substances, had recently experienced homelessness, were incarcerated, were U.S.-born, or self-identified as of American Indian or Alaska Native race or of Black race. Because incident clusters occurred in virtually every state, each state should continue implementing strategies to identify, investigate, and intervene to stop *M. tuberculosis* transmission in groups that might have different demographic profiles from the typical TB caseload of non–U.S. born individuals.[1,24] Despite differences in demographics between patients in clusters of ≥6 cases compared with either clusters of 3–5 cases or TB cases overall, the best predictor for cluster growth from 3 to ≥6 cases seems to be the occurrence of 3 matching TB cases in rapid succession (median=9.5 months), suggesting the importance of ensuring public health capacity to intervene when these incident clusters are detected.

ACKNOWLEDGMENTS

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

No financial disclosures were reported by the authors of this paper.

REFERENCES

1. Mindra G, Wortham JM, Haddad MB, Powell KM. Tuberculosis outbreaks in the United States, 2009–2015. Pub Health Rep. 2017;132(2):157–163. 10.1177/0033354916688270. [PubMed: 28147211]
2. National Tuberculosis Controllers Association, CDC. Guidelines for the investigation of contacts of persons with infectious tuberculosis. MMWR Recomm Rep. 2005;54(RR-15):1–47.

3. Ghosh S, Moonan PK, Cowan L, Grant J, Kammerer JS, Navin TR. Tuberculosis genotyping information management system: enhancing tuberculosis surveillance in the United States. Infect Genet Evol. 2012;12(4):782–788. 10.1016/j.meegid.2011.10.013. [PubMed: 22044522]
4. CDC. Prioritizing Tuberculosis Genotype Clusters for Further Investigation and Public Health Action. Atlanta, GA: HHS, CDC. https://www.cdc.gov/tb/programs/genotyping/Prioritizing_Tuberculosis_Genotype_Clusters_August2017.pdf. Published 2017. Accessed July 2, 2020.
5. Althomsons SP, Kammerer JS, Shang N, Navin TR. Using routinely reported tuberculosis genotyping and surveillance data to predict tuberculosis outbreaks. PLoS One. 2012;7(11):e48754. 10.1371/journal.pone.0048754. [PubMed: 23144956]
6. CDC. Reported tuberculosis in the United States, 2018. Atlanta, GA: HHS, CDC; 2019. https://www.cdc.gov/tb/statistics/reports/2018/table1.htm. Published 2019. Accessed July 2, 2020.
7. Teeter LD, Kammerer JS, Ghosh S, et al. Evaluation of 24-locus MIRU-VNTR genotyping in Mycobacterium tuberculosis cluster investigations in four jurisdictions in the United States, 2006–2010. Tuberculosis (Edinb). 2017;106:9–15. 10.1016/j.tube.2017.05.003. [PubMed: 28802410]
8. CDC. Tuberculosis surveillance data training—report of verified case of tuberculosis. Washington, DC: HHS, CDC; 2009.
9. U.S. Department of Agriculture, Economic Research Service. 2013 Rural–Urban Continuum Codes. Washington DC: U.S. Department of Agriculture. https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/. Published 2013. Accessed May 19, 2019.
10. CDC and American Thoracic Society. Targeted tuberculin testing and treatment of latent tuberculosis infection. MMWR Recomm Rep. 2000;49(RR-6):1–51.
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. 10.1371/journal.pone.0153728. [PubMed: 27082644]
12. Kammerer JS, Shang N, Althomsons SP, Haddad MB, Grant J, Navin TR. Using statistical methods and genotyping to detect tuberculosis outbreaks. Int J Health Geogr. 2013;12:15. 10.1186/1476-072x-12-15. [PubMed: 23497235]
13. Munn MS, Duchin JS, Kay M, Pecha M, Thibault CS, Narita M. Analysis of risk factors for tuberculous infection following exposure at a homeless shelter. Int J Tuberc Lung Dis. 2015;19(5):570–575. 10.5588/ijtld.14.0648. [PubMed: 25868026]
14. Lofy KH, McElroy PD, Lake L, et al. Outbreak of tuberculosis in a homeless population involving multiple sites of transmission. Int J Tuberc Lung Dis. 2006;10(6):683–689. [PubMed: 16776457]
15. CDC. Tuberculosis outbreak associated with a homeless shelter—Kane County, Illinois, 2007–2011. MMWR Morb Mortal Wkly Rep. 2012;61(11):186–189. [PubMed: 22437912]
16. Powell KM, VanderEnde DS, Holland DP, et al. Outbreak of drug-resistant Mycobacterium tuberculosis among homeless people in Atlanta, Georgia, 2008–2015. Public Health Rep. 2017;132(2):231–240. 10.1177/0033354917694008. [PubMed: 28257261]
17. CDC. Notes from the field: tuberculosis cluster associated with homelessness—Duval County, Florida, 2004–2012. MMWR Morb Mortal Wkly Rep. 2012;61(28):539–540. [PubMed: 22810268]
18. Wille SM, Kemp KA, Greenfield BL, Wallis ML. Barriers to healthcare for American Indians experiencing homelessness. J Soc Distress Homeless. 2017;26(1):1–8. 10.1080/10530789.2016.1265211. [PubMed: 29375241]
19. Volkmann T, Moonan PK, Miramontes R, Oeltmann JE. Excess alcohol use and death among tuberculosis patients in the United States, 1997–2012. J Tuberc Res. 2016;4(1):18–22. 10.4236/jtr.2016.41003. [PubMed: 27054144]
20. Oeltmann JE, Kammerer JS, Pevzner ES. Tuberculosis and substance abuse in the United States, 1997–2006. Arch Intern Med. 2009;169(2):189–197. 10.1001/archinternmed.2008.535. [PubMed: 19171816]
21. Winston CA, Menzies HJ. Pediatric and adolescent tuberculosis in the United States, 2008–2010. Pediatrics. 2012;130(6):e1425–e1432. 10.1542/peds.2012-1057. [PubMed: 23184110]

22. Hamblion EL, Le Menach A, Anderson LF,et al. Recent TB transmission, clustering and predictors of large clusters in London, 2010–2012: results from first 3 years of universal MIRU-VNTR strain typing. Thorax. 2016;71(8):749–756. 10.1136/thoraxjnl-2014-206608. [PubMed: 27417280]

23. Izumi K, Murase Y, Uchimura K, et al. Transmission of tuberculosis and predictors of large clusters within three years in an urban setting in Tokyo, Japan: a population-based molecular epidemiological study. BMJ Open. 2019;9(5):e029295. 10.1136/bmjopen-2019-029295.

24. Advisory Council for the Elimination of Tuberculosis, CDC. Essential components of a tuberculosis prevention and control program. MMWR Recomm Rep. 1995;44(RR-11):1–16.

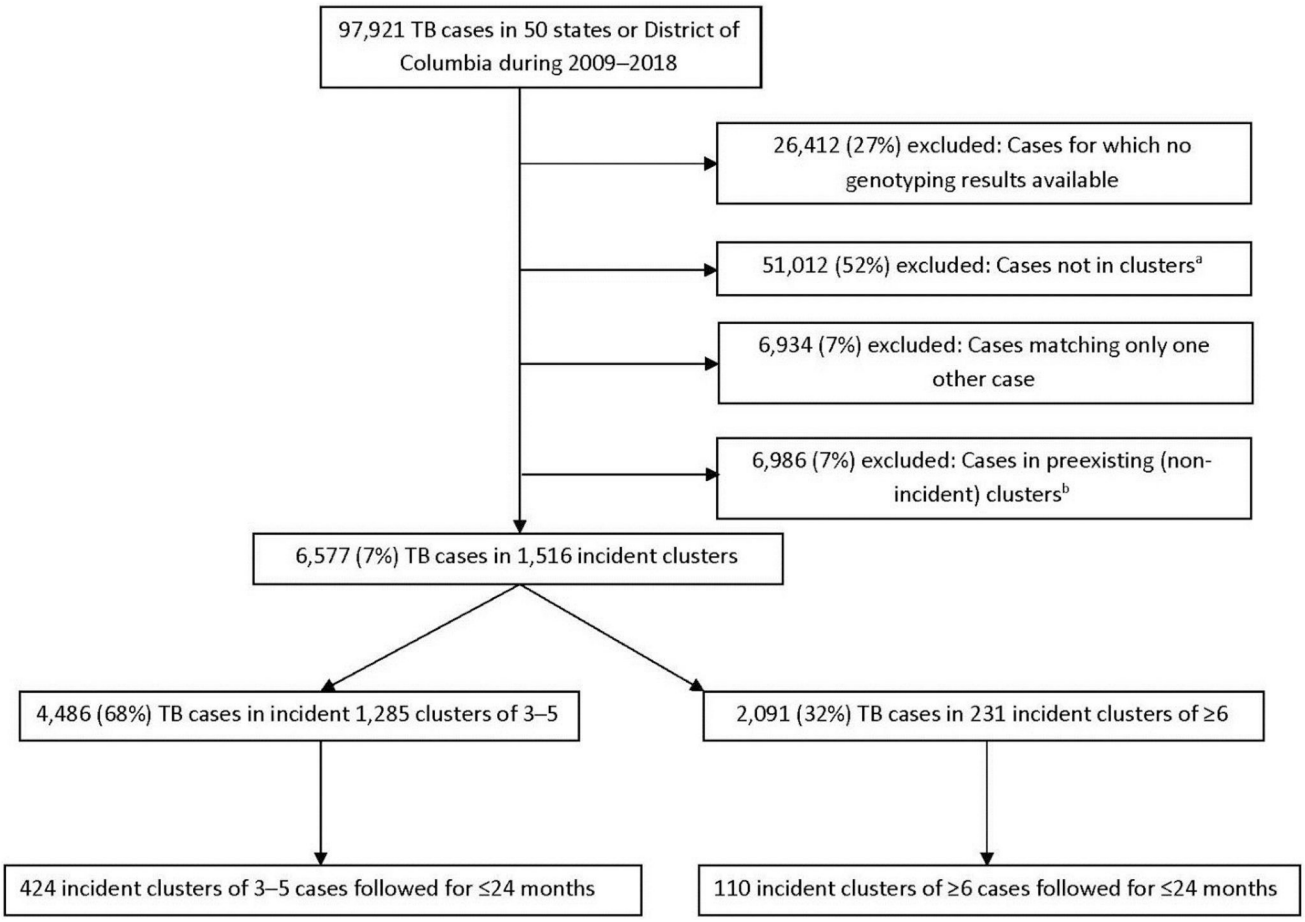

Figure 1.

Inclusion and exclusion algorithm for the primary analysis.

[a]TB cases for which only one case with that genotype was reported in the same county during 2009–2018.

[b]TB clusters were considered pre-existing if there were any cases in the county with matching spoligotype and 24-locus MIRU-VNTR during the 24 months preceding the first case (or matching spoligotype and 12-locus MIRU-VNTR for cases genotyped before 2009).

TB, tuberculosis; MIRU-VNTR, mycobacterial interspersed repetitive units variable number of tandem repeats.

Table 1.

Demographic, Clinical, and Laboratory Characteristics of Tuberculosis Cases in Incident, County-based Genotype Clusters, 2009–2018

Patient characteristics	**All cases in incident clusters N=6,577 patients n (%)**	**3–5 case clusters N=4,486 patients n (%)**	**≥6 case clusters N=2,091 patients n (%)**	***p*-value[f]**
Age group				**<0.0001**
0–4 years	190 (3)	124 (3)	66 (3)	
5–14 years	124 (2)	75 (2)	49 (2)	
15–24 years	952 (15)	619 (14)	333 (16)	
25–44 years	2,251 (34)	1,480 (33)	771 (37)	
45–64 years	2,226 (34)	1,550 (35)	676 (32)	
≥65 years	833 (13)	637 (14)	196 (9)	
Race/ethnicity				**<0.0001**
Non-Hispanic American Indian/Alaska Native	264 (4)	87 (2)	177 (8)	
Non-Hispanic Asian	1,137 (17)	848 (19)	289 (14)	
Non-Hispanic Black	1,866 (28)	1,169 (26)	697 (33)	
Non-Hispanic Native Hawaiian/other Pacific Islander	2,284 (35)	63 (1)	43 (2)	
Non-Hispanic White	869 (13)	628 (14)	241 (12)	
Hispanic or Latino	2,284 (34)	1,661 (37)	623 (30)	
Experienced homelessness[a]	647 (9)	391 (9)	256 (12)	**<0.0001**
Incarcerated[b]	358 (5)	207 (5)	151 (7)	**<0.0001**
U.S.-born[c]	3,529 (54)	2,147 (48)	1,382 (66)	**<0.0001**
HIV infected	420 (7)	269 (6)	151 (7)	0.09
Excess alcohol or drug use	1,841 (28)	1,159 (26)	682 (33)	**<0.001**
Reason evaluated for TB				0.09
Abnormal chest radiograph	1,214 (19)	845 (19)	369 (18)	
Contact investigation	613 (9)	364 (8)	249 (12)	
Incidental lab result	507 (8)	347 (8)	160 (8)	
Targeted testing	266 (4)	142 (3)	124 (6)	
Symptoms of TB disease	3,768 (58)	2,617 (59)	1,151 (55)	
Pulmonary disease	5,905 (90)	4,011 (89)	1,894 (91)	0.15
Sputum-smear positive[d]	3,661 (62)	2,496 (62)	1,165 (62)	0.42
Radiographic findings[e]				
Lung abnormalities	5,390 (91)	3,686 (92)	1,704 (90)	0.51
Cavitary lesions in lungs	1,982 (37)	1,355 (37)	627 (37)	0.46

Note: Boldface indicates statistical significance ($p<0.05$).

[a] Within year before TB diagnosis.

[b] At time of TB diagnosis.

[c] A person born in 1 of the 50 states or the District of Columbia, or a person born outside the U.S. to at least one parent who was a U.S. citizen.

[d] Among patients with pulmonary disease.

[e] Among patients with pulmonary disease who also had chest radiograph results.

[f] *p*-values based on χ^2 test of proportions. Comparisons were between proportions of cases in clusters of 3–5 case and ≥6 case clusters.

TB, tuberculosis.

Am J Prev Med. Author manuscript; available in PMC 2022 August 01.

Table 2.

Proportion of Tuberculosis Clusters with Selected Demographic, Clinical, and Laboratory Characteristics Among First 3 Cases—U.S., 2009–2018

Cluster-level characteristics (≥1 of first 3 patients)	**n=424 clusters of 3–5 cases followed for ≤24 months following the 3rd case n (%)**	**n=110 clusters of ≥6 cases followed for ≤24 months following the 3rd case n (%)**	**HR (95% CI)[d]**	**Adjusted HR (95% CI)[d,e]**
Median time between 1st and 3rd case[a]	54 months	9.5 months	**0.94 (0.93, 0.95)** [f]	**0.95 (0.93, 0.96)** [e, f]
Age group				
0–4 years	30 (7)	13 (12)	1.4 (0.8, 2.6)	—
5–14 years	15 (3)	7 (6)	**2.3 (1.1, 5.0)**	1.2 (0.7, 1.9)
15–24 years	134 (32)	40 (36)	1.2 (0.8, 1.8)	—
25–44 years	276 (65)	82 (75)	1.3 (0.9, 2.1)	—
45–64 years	291 (69)	68 (62)	0.7 (0.5, 1.0)	0.7 (0.5, 1.1)
≥65 years	153 (36)	14 (13)	**0.4 (0.2, 0.6)**	**0.3 (0.2, 0.6)**
Race/ethnicity				
Non-Hispanic American Indian/Alaska Native	7 (1)	11 (10)	**3.8 (2.0, 7.2)**	2.0 (0.9, 4.8)
Non-Hispanic Asian	106 (25)	21 (19)	0.8 (0.5, 1.3)	—
Non-Hispanic Black	156 (37)	55 (50)	**1.5 (1.0, 2.1)**	0.8 (0.5, 1.3)
Non-Hispanic Native Hawaiian/other Pacific Islander	7 (2)	3 (3)	2.1 (0.7, 6.8)	—
Non-Hispanic White	103 (24)	19 (17)	0.7 (0.4, 1.2)	—
Hispanic or Latino	189 (45)	44 (40)	0.9 (0.6, 1.3)	—
Experienced homelessness[b]	88 (21)	39 (35)	**1.5 (1.0, 2.1)**	1.2 (0.8, 2.0)
Incarcerated[c]	43 (10)	16 (15)	1.3 (0.7, 2.2)	—
U.S.-born	271 (64)	91 (83)	**2.3 (1.4, 3.8)**	1.6 (0.9, 2.9)
HIV infection	72 (17)	25 (23)	1.2 (0.8, 1.9)	—
Excess alcohol or drug use	113 (27)	39 (35)	1.3 (0.9, 1.9)	1.0 (0.6, 1.6)
Sputum-smear positive[g]	—	—	1.1 (0.5, 2.5)	—
Radiographic findings[h]				
Lung abnormalities	422 (99)	109 (99)	0.2 (0.1, 1.7)	—
Cavitary lesions in lungs	273 (64)	84 (76)	1.4 (0.9, 2.2)	—

Cluster-level characteristics (≥1 of first 3 patients)	n=424 clusters of 3–5 cases followed for ≤24 months following the 3rd case n (%)	n=110 clusters of ≥6 cases followed for ≤24 months following the 3rd case n (%)	HR (95% CI)[d]	Adjusted HR (95% CI)[d,e]
County setting[i]				
Urban, metro county[j]	393 (93)	97 (88)	**0.5 (0.3, 1.0)**	0.9 (0.5, 1.9)
Urban or rural county adjacent to metro county[k]	23 (5)	8 (6)	1.4 (0.7, 2.9)	—
Urban, non-metro county[l]	8 (2)	3 (3)	1.8 (0.6, 5.7)	—
Rural county[m,n]	0 (0)	2 (2)	*n*	*n*

Note: Boldface indicates statistical significance ($p<0.05$).

[a] Median number of months between the first and third cluster cases.

[b] Within year before TB diagnosis.

[c] At time of TB diagnosis.

[d] Hazard ratio of growth to ≥6 cases with ≥1 case of the given demographic compared with clusters with 0 cases with the given demographic.

[e] Adjusted hazard ratios (aHR) and 95% CIs for growth to ≥6 cases were calculated by adjusting for time (in months) between the diagnoses of the first and third case in the cluster and the presence of ≥1 case of any of the following among the initial 3 cases: age 5–14 years, age 45–64 years, age ≥65 years, American Indian or Alaska Native race, Black race, homelessness within past year, U.S. birth, excess drug or alcohol use, and urban county as defined by the U.S. Department of Agriculture, Economic Research Service.

[f] Time was calculated as a continuous variable; therefore, for each additional month between the first and third cases, the hazard ratio was lower by the factor specified. In other words, each additional month between the first and third case was associated with a 5%–7% reduction in the hazard of becoming ≥6 cases in the unadjusted model and a 4%–7% reduction in the adjusted model.

[g] Almost all clusters had at least one pulmonary case (99% of 3–5 case clusters and all clusters with ≥6 cases).

[h] Among cases with chest radiograph results.

[i] As classified by the U.S. Department of Agriculture, Economic Research Service. Accessed at https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/ in May 2019.

[j] As classified by the U.S. Department of Agriculture, Economic Research Service. Accessed at https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/ in May 2019. Indicates rural–urban continuum codes 1–3, which represent counties in metro areas with populations ≥250,000.

[k] As classified by the U.S. Department of Agriculture, Economic Research Service. Accessed at https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/ in May 2019. Indicates rural–urban continuum codes 4 and 6, which represent counties with urban populations adjacent to metro areas with populations of ≥2,500.

[l] As classified by the U.S. Department of Agriculture, Economic Research Service. Accessed at https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/ in May 2019. Indicates rural–urban continuum code 8, which represents completely rural counties and counties with urban populations <2,500 that are adjacent to metro areas.

[m] As classified by the U.S. Department of Agriculture, Economic Research Service. Accessed at https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/ in May 2019. Indicates rural–urban continuum code 9, which represents completely rural counties and counties with urban populations <2,500 that are not adjacent to metro areas.

[n] Although $p<0.05$ for the comparison (Fisher's exact) between the number of clusters of 3–5 cases and ≥6 cases, this was not included in the final model due to a 0 cell size.

TB, tuberculosis.