HHS Public Access

Author manuscript

Am J Epidemiol. Author manuscript; available in PMC 2018 June 12.

Published in final edited form as:

Am J Epidemiol. 2015 November 01; 182(9): 799–807. doi:10.1093/aje/kwv121.

A Field-Validated Approach Using Surveillance and Genotyping Data to Estimate Tuberculosis Attributable to Recent Transmission in the United States

Anne Marie France[*],

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

Juliana Grant,

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

J. Steve Kammerer, and

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

Thomas R. Navin

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

Abstract

Tuberculosis genotyping data are frequently used to estimate the proportion of tuberculosis cases in a population that are attributable to recent transmission (RT). Multiple factors influence genotype-based estimatesof RTand limit the comparison of estimates over time and across geographic units. Additionally, methods used for these estimates have not been validated against field-based epidemiologic assessments of RT. Here we describe a novel genotype-based approach to estimation of RT based on the identification of plausible-source cases, which facilitates systematic comparisons over time and across geographic areas. We compared this and other genotype-based RT estimation approaches with the gold standard of field-based assessment of RT based on epidemiologic investigation in Arkansas, Maryland, and Massachusetts during 1996–2000. We calculated the sensitivity and specificity of each approach for epidemiologic evidence of RT and calculated the accuracy of each approach across a range of hypothetical RT prevalence rates plausible for the United States. The sensitivity, specificity, and accuracy of genotype-based RT estimates varied by approach. At an RT prevalence of 10%, accuracy ranged from 88.5% for state-based clustering to 94.4% with our novel approach. Our novel, field-validated approach allows for systematic assessments over time and across public health jurisdictions of varying geographic size, with an established level of accuracy.

Keywords

disease transmission; genotyping; molecular epidemiology; recent transmission; tuberculosis

[*]Correspondence to Dr. Anne Marie France, Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Mail Stop E-10, Atlanta, GA 30329 (afrance@cdc.gov).

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

Conflict of interest: none declared.

Tuberculosis (TB) disease may result from infection acquired recently (recent transmission (RT)) or from infection acquired many years in the past (1). TB disease attributable to RT represents active transmission in a population, which can be interrupted by prompt public health intervention (2). Measuring TB attributable to RT provides a tool with which public health programs can focus limited public health resources and track progress in TB control efforts to interrupt transmission (3–6).

TB attributable to RT cannot be distinguished clinically from TB resulting from the reactivation of remotely acquired infection (7); for this reason, measuring RT requires a knowledge of the transmission connections (epidemiologic links) between cases. Field-based epidemiologic investigation to identify the source of a given case (a source case–secondary case pair) represents the gold standard for determining which cases are likely to be due to RT. However, these source-case investigations are extremely resource-intensive and are rarely conducted. Contact investigations, which differ from source-case investigations in that they are used to identify persons who have been exposed to a person with contagious TB, are an essential component of TB control (2) and represent another source of information on transmission connections that can be used to understand RT in a population (8). However, these investigations are also resource-intensive, and the data are often not collected systematically. Additionally, con-tact investigations frequently miss connections between cases, particularly in hard-to-reach populations, such as persons who abuse chemical substances and persons experiencing home-lessness (9–12).

TB genotyping is frequently used to estimate the proportion of TB cases in a population that are attributable to RT (5, 8, 13–21), as it provides a tool with which to infer transmission connections between cases. Genotype-based methods for estimating RT rely on the fundamental assumption that cases related by RT will have identical genotypes (typically referred to as clustered cases), while cases resulting from the reactivation of remotely acquired infection will have unique genotypes in the population (nonclustered cases). While genotype data alone suggest transmission, they do not always correlate with RT (22). Nevertheless, genotype-based estimates of RT have many advantages: They are not dependent on field data, they can be made on a large scale with data that are routinely available in many jurisdictions, and they can be easily automated and systematically applied over time and across geographic units. Genotype-based methods do require that a high proportion of culture-positive TB cases in the population are genotyped (23); however, in the United States, where genotyping is routine for all culture-positive cases, this is not an obstacle (24, 25).

While genotype-based estimates of RT have been generated often, the methods used for these estimates have not been standardized. As a result, estimates have varied dramatically, even across populations that would be expected to have similar rates of RT (15–19, 21, 26). Multiple factors affect genotype-based RT estimates, including the geographic unit over which clustering is defined (20), the time window of the study (26), and whether or not source cases are taken into account (27). Because RT estimates are sensitive to the geographic unit and time window over which clustering is defined, comparisons of RT across jurisdictions of different geographic sizes and over time are problematic. A

systematic approach to address both of these factors is essential to the use of RT estimation approaches for programmatic purposes, such as tracking estimates of RT over time across jurisdictions of different sizes. At the same time, estimation approaches that do not exclude likely source cases (e.g., by excluding the first case in a genotype cluster) can substantially overestimate the proportion of cases that are attributable to RT, as they estimate the proportion of cases simply involved in RT, as opposed to cases attributable to RT.

Regardless of the methods used, genotype-based estimates of RT have not been validated against field-based epidemiologic assessments of RT. As a result, the sensitivity and specificity of previous RT estimates based on genotyping are unknown, and it is not clear which approaches are the most accurate approaches for estimating RT.

In this paper, we validate previously published genotype-based approaches to RT estimation against the gold standard of field-based assessments of RT based on the epidemiologic investigation and calculate the sensitivity and specificity of each for epidemiologic evidence of RT using various geographic definitions. Additionally, we propose a novel approach for estimating TB attributable to RT that overcomes many of the limitations of prior approaches, facilitates systematic comparisons of RT across jurisdictions of different geographic sizes and over time, and accounts for source cases. We validated this approach, which we call the plausible-source case approach, against field-based assessments of RT derived from epidemiologic investigation, and we calculated the sensitivity and specificity of our estimate for epidemiologic evidence of RT. We compared our proposed approach with previously published approaches in order to identify the most accurate approach to estimation of RT for the US population.

Methods

Study population

The National TB Genotyping and Surveillance Network (NTGSN) was established as part of a population-based study conducted at 7 US sites during 1996–2000 (28, 29). As part of this study, isolates from cases identified at the participating sites were genotyped using IS*6110* restriction fragment length polymorphism (RFLP), and epidemiologic field investigations were conducted to identify potential transmission relationships between cases. Epidemiologic field investigations included contact investigation for all cases. At 3 study sites, researchers also conducted detailed epidemiologic investigations of cases with the same genotype pattern (cluster investigations) for cases reported during 1998–2000. Details on the original NTGSN study have been previously reported (28, 29).

We included a subset of the NTGSN study population in our analysis: culture-positive cases who had an isolate with ≥6 RFLP bands from participating states that conducted the most rigorous epidemiologic field investigations (including both contact investigation and cluster investigation): Arkansas, Maryland, and Massachusetts. We restricted our analysis to isolates with ≥6 RFLP bands because the discrimination of this method is poor among isolates with fewer than 6 bands (30) and because discrimination among isolates with ≥6 RFLP bands has been described as being similar to discrimination using the combination of spoligotyping and 24-locus mycobacterial interspersed repetitive unit–variable-number tandem repeat

(MIRU-VNTR) analysis (31), the genotyping methods currently used in the United States. Information on clinical and demographic characteristics, risk factors, and zip code of residence was available from routinely collected surveillance data for all cases.

Field-based gold standard for RT

We defined the gold standard for RT using data collected through the comprehensive epidemiologic field investigations conducted as part of the NTGSN study. Potential transmission relationships between cases (epidemiologic links) were investigated through both contact investigation and cluster investigation. Epidemiologic links included potential transmission relationships identified within a household as well as nonhousehold relationships (e.g., coworkers, friends, and residents of congregate settings such as jails, homeless shelters, and long-term care facilities). When epidemiologic links were identified, TB control staff in the field determined the direction of transmission where possible, describing each identified source case–secondary case pair.

While the NTGSN investigation identified epidemiologic links and the direction of transmission between cases where possible, field investigators did not apply criteria for classifying transmission links as recent or not. Therefore, we applied criteria to the NTGSN field data to classify each case into one of 3 field-based gold-standard RT categories: 1) field evidence of RT, 2) no field evidence of RT, and 3) possible RT. We defined RT as transmission occurring within 2 years before the diagnosis of a reported case, consistent with the time period when risk of progression from infection to active disease is highest (32, 33) and when the opportunity for public health intervention is greatest. Practically, we implemented this definition by requiring that an identified source case be diagnosed within 2 years before the diagnosis date of the putative secondary case, or, because source cases are sometimes identified after a secondary case, that the source case be diagnosed at any time following the secondary case. Since the diagnosis date is not reported in routinely collected surveillance data, we used the earliest of 3 dates as a proxy for diagnosis date: the date on which a patient specimen was collected for drug susceptibility testing, the date on which TB treatment was initiated, or the date on which the patient was counted as having a verified case of TB.

We categorized each case into one of the 3 field-based gold-standard RT categories using the following definitions:

- *Field evidence of RT*: a case of TB disease with an identified source case (i.e., an epidemiologic link between 2 cases and a known direction of transmission), where the source case was diagnosed during the period between 2 years before the given case's diagnosis date and any time following the given case's diagnosis date and the cases had matching RFLP patterns. Matching RFLP patterns were required because genotyping data often disprove suspected transmission links between cases (34).
- *No field evidence of RT*: case had no identified epidemiologic link to another case.

- *Possible RT*: a case with an identified epidemiologic link to another case but no known direction of transmission (i.e., TB control staff in the field could not determine which case was the source case and which was the secondary case), or a case with an identified source case for which the RFLP patterns did not match those of the identified source case.

In order to allow each case the same time interval for a source case to be identified (2 years prior), we only categorized cases for which the full time interval could be evaluated. Therefore, only cases reported during January 1, 1998–September 30, 2000 were classified into one of the 3 field-based gold-standard RT categories. The χ^2 test was used to compare demographic and risk factors across the 3 categories.

Genotype-based RT estimates

We used 4 approaches to estimate RT for the cases reported during January 1, 1998–September 30, 2000; each of these approaches utilized only routinely collected genotyping and surveillance data and was applied independently of the field-based gold standard. Three of the approaches, referred to here as geographic approaches, were based on geography and genotyping alone. The geographic units in these approaches have been previously used as the basis for published estimates of RT (14, 19–21). The fourth approach, the plausible-source case approach, is a novel approach we propose that is based on identifying a plausible-source case of transmission based on clinical and demographic factors, the time interval between cases, geography, and genotyping data. Each of these 4 approaches was applied to TB cases to estimate cases attributable to RT, and then each approach was compared with the field-based gold standard.

Geographic approaches

State-based clustering: Cases with an RFLP pattern that matched that of at least 1 other case in the same state during January 1, 1998–September 30, 2000 were classified as RT.

County-based clustering: Cases with an RFLP pattern that matched that of at least 1 other case in the same county during January 1, 1998–September 30, 2000 were classified as RT.

SaTScan-based clustering: SaTScan is a software tool, developed by Dr. Martin Kulldorff (Harvard Medical School, Boston, Massachusetts) in conjunction with Information Management Services, Inc. (Calverton, Maryland), that analyzes data using spatial scanning statistics (35) and has been used to estimate TB RT by identifying geographic areas with larger-than-expected rates of genotype clustering (20). TB cases in a statistically significant SaTScan cluster ($P < 0.05$) with at least 1 other case during the period January 1, 1998–September 30, 2000 were classified as RT. SaTScan version 8.0.2 was run using a circular geography that is not restricted by administrative boundaries (e.g., state or county) and a maximum search radius of 50 km (31.3 miles), and case location was determined as the latitude and longitude of the geographic center of the case's zip code of residence.

Novel approach: the plausible-source case approach—We evaluated each case reported during January 1, 1998– September 30, 2000 by comparing it with cases reported in the full study population (January1,1996–December 31, 2000), to determine whether a

plausible-source case could be identified; if a plausible-source case was identified, the given case was classified as attributable to RT. Comparison with cases in the longer time window was necessary in order to evaluate plausible-source cases over the same time unit for each case. A plausible-source case was defined as a case that 1) involved a respiratory form of TB (pulmonary, laryngeal, pleural, or miliary) in a patient over 4 years of age and 2) was diagnosed within 2 years prior to the given case. For cases that could not themselves be classified as source cases (e.g., a case without a respiratory form of TB or in a patient ≤4 years of age), we also allowed for a plausible-source case to be diagnosed up to 3 months following a given case (an arbitrarily selected time frame in which case-finding activities such as contact investigation or source-case investigation could reasonably be expected to occur). The plausible-source case must have resided within the specified geographic distance from the given case (calculated as the distance between the centroids of the cases' reported zip codes of residence or, where the zip code was missing or invalid, the centroids of the cases' reported cities of residence) and must have had the same RFLP pattern.

We included pleural TB as a respiratory form of disease because of the high probability that a case reported as pleural also has pulmonary involvement (36–38).

If any plausible-source case was identified, the related case was classified as attributableto RT. However, cases diagnosed in foreign-born persons less than 100 days after their entry into the United States were never classified as attributable to RT, even if a plausible-source case was identified; this criterion was included because of the likelihood that cases diagnosed in foreign-born persons during this time period represent infection acquired outside of the United States (39).

We implemented the plausible-source case approach using a range of geographic distance thresholds between cases, from 0 miles to 500 miles (800 km).

Comparison of genotype-based RT estimates with field-based gold standard

For each of the 4 genotype-based estimation approaches, we calculated the sensitivity and specificity of the estimate for true RT, as defined by the field-based gold standard. Using these sensitivity and specificity calculations, we evaluated the accuracy of the method across a range of hypothetical prevalence rates of RT plausible for the US population or subpopulations within the United States.

Calculation of sensitivity and specificity—Sensitivity was calculated as the percentage of “field evidence of RT” cases that were classified as RT by the given method. Specificity was calculated as the percentage of cases with “no field evidence of RT” that were classified as not RT by a given method. Cases of “possible RT” (i.e., cases with unclear evidence of RT) were not included in sensitivity and specificity calculations.

Calculation of accuracy—Using the sensitivity and specificity calculated for each method, accuracy was evaluated across a range of hypothetical RT prevalence values using the following formula:

$$\text{Accuracy} = (\text{sensitivity} \times \text{prevalence}) + [\text{specificity} \times (1 - \text{prevalence})] .$$

For the RT algorithm, sensitivity, specificity, and accuracy were calculated for each geographic distance threshold considered.

Sensitivity analysis—To evaluate the impact of our decision to exclude cases in the "possible RT" field-based gold-standard RT category from our sensitivity and specificity calculations, we conducted a sensitivity analysis in which we included these cases as having "field evidence of RT" (see Web Tables 1 and 2, available at http://aje.oxfordjournals.org/) and an additional sensitivity analysis in which we included these cases as having "no field evidence of RT" (Web Tables 3 and 4).

To evaluate the impact of our criterion that cases in foreign-born persons diagnosed less than 100 days after entry into the United States never be classified as attributable to RT, even if a plausible-source case was identified by our plausible-source case approach, we conducted a sensitivity analysis in which we removed this criterion (such that any case for which a plausible-source case was identified would be classified as attributable to RT, regardless of country of origin or time since arrival in the United States) (Web Table 5).

To evaluate the impact of our definition of RT as transmission between cases diagnosed within a 2-year period, we conducted a sensitivity analysis in which we defined RT as transmission between cases diagnosed within a 1-year period (Web Appendix, Web Tables 6 and 7).

Stratified analysis—We stratified sensitivity, specificity, and accuracy calculations according to cases' nativity (US-born (Web Tables 8 and 9) or foreign-born (Web Tables 10 and 11)).

We conducted analyses using SAS, version 9.3 (SAS Institute, Inc., Cary, North Carolina), and SaTScan, version 8.0.2 (Dr. Martin Kulldorff and Information Management Services, Inc.).

Results

Study population

During 1996–2000, a total of 2,935 cases of culture-positive TB were reported in Arkansas, Maryland, and Massachusetts; IS*6110* RFLP genotyping results were available for 2,842 (96.8%). Of these, 2,198 cases (77.3%) had an isolate with ≥6 RFLP bands and were included in our full study population. Of the 2,198 cases in the study population, 1,188 (54%) were reported between January 1, 1998, and September 30, 2000, and therefore could be classified into one of the 3 field-based gold-standard RT categories for sensitivity and specificity calculations. Zip code was missing or invalid for 69 (3.1%) of the 2,198 cases; for 68 (98.6%) of these cases, city of residence was available and was used for geographic distance calculations in the plausible-source case approach.

Of the 2,198 cases in the study population, 1,862 (84.7%) had a respiratory form of TB. This included 88 cases with pleural TB but no other respiratory site reported.

Field-based gold-standard RT classifications

Of the 1,188 cases that could be classified into one of the 3 field-based gold-standard RT categories, 72 (6.1%) met the definition for field evidence of RT, 103 (8.7%) met the definition of possible RT, and 1,013 (85.3%) had no field evidence of RT. Cases with field evidence of RT were more frequently aged ≤4 years, were more frequently US-born, and more frequently had reported excess alcohol abuse or injecting drug use in the year prior to diagnosis (Table 1).

Genotype-based RT estimates

The percentage of cases estimated to be due to RT by the genotype-based methods varied by approach. Of the geographic approaches considered, the percentage estimated to be due to RT ranged from a low of 12.7% using SaTScan-based clustering to a high of 22.3% using state-based clustering (Figure 1). Using the plausible-source case approach, the percentage of cases estimated to be due to RT ranged by geographic threshold considered, from a low of 5.6% using a 1-mile (1.6-km) threshold to a high of 17.8% using a 500-mile (800-km) threshold (Figure 2).

Comparison of genotype-based RT estimates with the field-based gold standard

The sensitivity and specificity of estimates varied by approach. Among the geography-based approaches, state-based clustering yielded the highest sensitivity but the lowest specificity (Table 2), while the sensitivity and specificity of SaTScan and county-based clustering were similar. The accuracy of each approach varied across the hypothetical prevalence rates of RT that were considered. The accuracy of state-based clustering was lowest when the prevalence of RT was less than or equal to 25% but was the most accurate of the geography-based methods considered when the prevalence of RT was 30% or more. At prevalence rates less than or equal to 25%, SaTScan was the most accurate of the geography-based methods considered.

The sensitivity and specificity of the plausible-source case approach varied by the distance threshold (Table 3). Increasing the distance threshold across which plausible-source cases could be identified increased the sensitivity but decreased specificity. For RT prevalence ≤30%, the accuracy was maximized at 10 miles (16 km).

For the range of RT prevalence rates most plausible for the United States (≤30%), the accuracy of the plausible-source case approach with a distance threshold of 10 miles (16 km) was slightly higher than that of other methods considered, ranging from 93.2% for an RT prevalence of 30% to 94.4% for an RT prevalence of 10%. Using this approach, the estimated percentage of cases attributable to RT in the study population was 11.4%.

Discussion

In this paper we have presented, to our knowledge, the first systematic evaluation of approaches to estimation of TB attributable to RT, validating estimates against field-based epidemiologic data. We observed a range of estimates of TB attributable to RT according to the method used, illustrating the importance of having a validated and standardized approach to estimation.

While there is no perfect gold standard for identifying TB attributable to RT, the population-based data used as our gold standard were unique in their strength, including findings from field-based cluster investigations as well as contact investigations. Cluster investigations often identify transmission events missed by contact investigation alone, including transmission in difficult-to-reach populations (9) and transmission that crosses geographic boundaries (40). Case-patients with field evidence of RT disproportionately reported characteristics that have previously been associated with RT and outbreaks, including US birth and substance abuse (41).

While differences in accuracy between the methods we considered were relatively small (ranging from 89.2% to 94.1% when the true RT prevalence was 15%), the difference in prevalence estimates generated by the methods was substantial (ranging from 22.3% for state-based clustering to 11.4% for the plausible-source case method with a 10-mile (16-km) threshold), indicating that small differences in accuracy have important consequences for population-based prevalence estimates. In low-incidence populations like that of the United States, where the expected prevalence of TB attributable to RT is probably less than 30%, a high specificity is more important than high sensitivity in the accuracy of an RT measure. Our plausible-source case approach, with a threshold of 10 miles, yielded a slightly more accurate RT estimate than other methods for the range of prevalence rates that are plausible for the United States.

SaTScan has been proposed as an alternative to methods that limit clustering definitions to a single jurisdiction, such as the state- or county-based methods (20). While SaTScan does have the potential to capture transmission that crosses jurisdictional boundaries, geographic units over which clustering is defined are not consistent across geographic areas (35), because the search radius can be varied up to the maximum radius in order to capture the highest concentration of cases with a given genotype. SaTScan, as well as the state-and county-based methods, cannot easily be applied to assess trends over time.

The plausible-source case approach we propose overcomes many limitations of previous approaches to estimation of RT. It identifies potential transmission across geographic boundaries, using a consistent geographic distance that allows for comparison across jurisdictions of different geographic size. It evaluates the same time frame for a plausible-source case for all cases, regardless of when during the study period the case was diagnosed. Unlike the $n - 1$ method (8), the commonly used approach to account for source cases (15, 16, 18, 19, 21), our plausible-source case approach does not assume that the first case in the cluster is the source case for every other case in the cluster. Rather, it considers plausible-

source cases for each case individually, requiring that a plausible-source case have characteristics associated with infectiousness.

The plausible-source case approach provides a standard method that allows for comparability across time and geographic jurisdictions. At the same time, it allows users the flexibility to choose a different threshold that is best suited to their population and application. We chose a threshold that maximizes the accuracy of RT estimates in the range of RT prevalence expected in the United States (≤30%). The appropriate threshold will probably differ in different populations, depending on the true prevalence of RT.

Our study had a number of limitations. First, our validation was limited by an imperfect gold standard. Despite rigorous epidemiologic field investigations, some transmission links were likely to have been missed; even the best investigations can miss links, particularly in hard-to-reach populations (9). Second, because genotyping data were a component of both gold-standard criteria and genotype-based estimation methods, sensitivity and specificity calculations were probably inflated. Third, the US National Tuberculosis Genotyping Service currently uses 24-locus MIRU-VNTR and spoligotyping to genotype TB, not IS6110 RFLP, the method used in this study. However, evidence suggests that the resolution of 24-locus MIRU-VNTR is similar to that of IS6110 RFLP (31); therefore, it is reasonable that the findings obtained here are applicable to current genotyping methods. Finally, our study included only 3 US states, which are not representative of the United States as a whole. RT estimates for the study population therefore cannot be generalized to the full US population, and movement of populations might be different in areas not considered. The 10-mile (16-km) threshold might not be ideal in all settings within the United States.

We have identified an approach for estimating TB attributable to RT that overcomes many limitations of previous approaches and have validated it against field-based epidemiologic data. While it is unlikely that any estimate will be completely accurate, this validated approach allows for systematic assessments over time and across jurisdictions of varying geographic size, with an established level of accuracy. Our plausible-source case approach, with a threshold of 10 miles, yields RT estimates with a high level of accuracy for the range of prevalence rates plausible for the United States, and it facilitates comparisons over time and across public health jurisdictions of varying geographic size. To our knowledge, this represents the first validated approach to estimation of TB attributable to RT.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the investigators in the National Tuberculosis Genotyping and Surveillance Network (NTGSN) and the health officials in Arkansas, Maryland, and Massachusetts who supported the NTGSN's epidemiologic field investigations in those states.

References

1. Frieden TR, Sterling TR, Munsiff SS, et al. Tuberculosis. Lancet. 2003; 362(9387):887–899. [PubMed: 13678977]
2. National Tuberculosis Controllers Association; Centers for Disease Control and Prevention. Guidelines for the investigation of contacts of persons with infectious tuberculosis. Recommendations from the National Tuberculosis Controllers Association and CDC. MMWR Recomm Rep. 2005; 54(RR-15):1–47.
3. Berzkalns A, Bates J, Ye W, et al. The road to tuberculosis (Mycobacterium tuberculosis) elimination in Arkansas; a re-examination of risk groups. PLoS One. 2014; 9(3):e90664. [PubMed: 24618839]
4. Borgdorff MW, van der Werf MJ, de Haas PE, et al. Tuberculosis elimination in the Netherlands. Emerg Infect Dis. 2005; 11(4):597–602. [PubMed: 15829200]
5. France AM, Cave MD, Bates JH, et al. 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):662–671. [PubMed: 17625223]
6. Moonan PK, Oppong J, Sahbazian B, et al. What is the outcome of targeted tuberculosis screening based on universal genotyping and location? Am J Respir Crit Care Med. 2006; 174(5):599–604. [PubMed: 16728707]
7. Geng E, Kreiswirth B, Burzynski J, et al. Clinical and radiographic correlates of primary and reactivation tuberculosis: a molecular epidemiology study. JAMA. 2005; 293(22):2740–2745. [PubMed: 15941803]
8. Small PM, Hopewell PC, Singh SP, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N Engl J Med. 1994; 330(24): 1703–1709. [PubMed: 7910661]
9. Malakmadze N, González IM, Oemig T, et al. Unsuspected recent transmission of tuberculosis among high-risk groups: implications of universal tuberculosis genotyping in its detection. Clin Infect Dis. 2005; 40(3):366–373. [PubMed: 15668858]
10. Reichler MR, Reves R, Bur S, et al. Evaluation of investigations conducted to detect and prevent transmission of tuberculosis. JAMA. 2002; 287(8):991–995. [PubMed: 11866646]
11. Yun LW, Reves RR, Reichler MR, et al. Outcomes of contact investigation among homeless persons with infectious tuberculosis. Int J Tuberc Lung Dis. 2003; 7(12 suppl 3):S405–S411. [PubMed: 14677830]
12. McNabb SJ, Kammerer JS, Hickey AC, et al. Added epidemiologic value to tuberculosis prevention and control of the investigation of clustered genotypes of Mycobacterium tuberculosis isolates. Am J Epidemiol. 2004; 160(6):589–597. [PubMed: 15353420]
13. Alland D, Kalkut GE, Moss AR, et al. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. N Engl J Med. 1994; 330(24): 1710–1716. [PubMed: 7993412]
14. Kempf MC, Dunlap NE, Lok KH, et al. Long-term molecular analysis of tuberculosis strains in Alabama, a state characterized by a largely indigenous, low-risk population. J Clin Microbiol. 2005; 43(2):870–878. [PubMed: 15695694]
15. Durmaz R, Zozio T, Gunal S, et al. Population-based molecular epidemiological study of tuberculosis in Malatya, Turkey. J Clin Microbiol. 2007; 45(12):4027–4035. [PubMed: 17928426]
16. Tazi L, Reintjes R, Bañuls AL. Tuberculosis transmission in a high incidence area: a retrospective molecular epidemiological study of Mycobacterium tuberculosis in Casablanca, Morocco. Infect Genet Evol. 2007; 7(5):636–644. [PubMed: 17689298]
17. Iñigo J, Arce A, Palenque E, et al. Decreased tuberculosis incidence and declining clustered case rates, Madrid. Emerg Infect Dis. 2008; 14(10):1641–1643. [PubMed: 18826835]
18. Love J, Sonnenberg P, Glynn JR, et al. Molecular epidemiology of tuberculosis in England, 1998. Int J Tuberc Lung Dis. 2009; 13(2):201–207. [PubMed: 19146748]
19. Vanhomwegen J, Kwara A, Martin M, et al. Impact of immigration on the molecular epidemiology of tuberculosis in Rhode Island. J Clin Microbiol. 2011; 49(3):834–844. [PubMed: 21159930]

20. Moonan PK, Ghosh S, Oeltmann JE, et al. Using genotyping and geospatial scanning to estimate recent Mycobacterium tuberculosis transmission, United States. Emerg Infect Dis. 2012; 18(3): 458–465. [PubMed: 22377473]
21. Rodwell TC, Kapasi AJ, Barnes RF, et al. 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):1917–1925. [PubMed: 22982156]
22. Braden CR, Templeton GL, Cave MD, et al. Interpretation of restriction fragment length polymorphism analysis of Mycobacterium tuberculosis isolates from a state with a large rural population. J Infect Dis. 1997; 175(6):1446–1452. [PubMed: 9180185]
23. Glynn JR, Vynnycky E, Fine PE. Influence of sampling on estimates of clustering and recent transmission of Mycobacterium tuberculosis derived from DNA fingerprinting techniques. Am J Epidemiol. 1999; 149(4):366–371. [PubMed: 10025480]
24. Centers for Disease Control and Prevention. Reported Tuberculosis in the United States 2012. Atlanta, GA: Centers for Disease Control and Prevention; 2013.
25. Ghosh S, Moonan PK, Cowan L, et al. Tuberculosis genotyping information management system: enhancing tuberculosis surveillance in the United States. Infect Genet Evol. 2012; 12(4):782–788. [PubMed: 22044522]
26. Vynnycky E, Nagelkerke N, Borgdorff MW, et al. The effect of age and study duration on the relationship between ‘clustering’ of DNA fingerprint patterns and the proportion of tuberculosis disease attributable to recent transmission. Epidemiol Infect. 2001; 126(1):43–62. [PubMed: 11293682]
27. Murray M, Alland D. Methodological problems in the molecular epidemiology of tuberculosis. Am J Epidemiol. 2002; 155(6):565–571. [PubMed: 11882530]
28. Ellis BA, Crawford JT, Braden CR, et al. Molecular epidemiology of tuberculosis in a sentinel surveillance population. Emerg Infect Dis. 2002; 8(11):1197–1209. [PubMed: 12453343]
29. Crawford JT, Braden CR, Schable BA, et al. National Tuberculosis Genotyping and Surveillance Network: design and methods. Emerg Infect Dis. 2002; 8(11):1192–1196. [PubMed: 12453342]
30. Chaves F, Yang Z, el Hajj H, et al. Usefulness of the secondary probe pTBN12 in DNA fingerprinting of *Mycobacterium tuberculosis*. J Clin Microbiol. 1996; 34(5):1118–1123. [PubMed: 8727887]
31. Allix-Béguec C, Fauville-Dufaux M, Supply P. Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol. 2008; 46(4):1398–1406. [PubMed: 18234864]
32. Sutherland I. Recent studies in the epidemiology of tuberculosis, based on the risk of being infected with tubercle bacilli. Adv Tuberc Res. 1976; 19:1–63. [PubMed: 823803]
33. Comstock GW, Livesay VT, Woolpert SF. The prognosis of a positive tuberculin reaction in childhood and adolescence. Am J Epidemiol. 1974; 99(2):131–138. [PubMed: 4810628]
34. Bennett DE, Onorato IM, Ellis BA, et al. DNA fingerprinting of Mycobacterium tuberculosis isolates from epidemiologically linked case pairs. Emerg Infect Dis. 2002; 8(11):1224–1229. [PubMed: 12453346]
35. Kulldorff M. A spatial scan statistic. Commun Stat Theory Methods. 1997; 26(6):1481–1496.
36. Seiscento M, Vargas FS, Bombarda S, et al. Pulmonary involvement in pleural tuberculosis: how often does it mean disease activity? Respir Med. 2011; 105(7):1079–1083. [PubMed: 21392956]
37. Kim HJ, Lee HJ, Kwon SY, et al. The prevalence of pulmonary parenchymal tuberculosis in patients with tuberculous pleuritis. Chest. 2006; 129(5):1253–1258. [PubMed: 16685016]
38. Antonangelo L, Vargas FS, Puka J, et al. Pleural tuberculosis: is radiological evidence of pulmonary-associated disease related to the exacerbation of the inflammatory response? Clinics (Sao Paulo). 2012; 67(11):1259–1263. [PubMed: 23184200]
39. Cain KP, Benoit SR, Winston CA, et al. Tuberculosis among foreign-born persons in the United States. JAMA. 2008; 300(4):405–412. [PubMed: 18647983]
40. McNabb SJN, Braden CR, Navin TR. DNA fingerprinting of Mycobacterium tuberculosis: lessons learned and implications for the future. Emerg Infect Dis. 2002; 8(11):1314–1319. [PubMed: 12453363]

41. Mitruka K, Oeltmann JE, Ijaz K, et al. Tuberculosis outbreak investigations in the United States, 2002–2008. Emerg Infect Dis. 2011; 17(3):425–431. [PubMed: 21392433]

Abbreviations

MIRU-VNTR	mycobacterial interspersed repetitive unit–variable-number tandem repeat
NTGSN	National TB Genotyping and Surveillance Network
RFLP	restriction fragment length polymorphism
RT	recent transmission
TB	tuberculosis

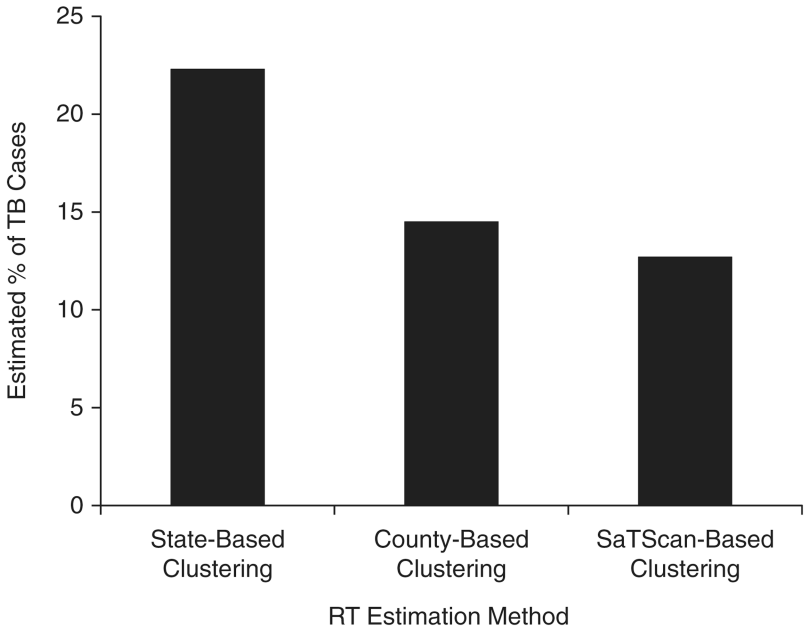

Figure 1.
Estimated percentage of tuberculosis (TB) cases attributable to recent transmission as determined using state-based clustering, county-based clustering, and SaTScan-based (35) clustering, Arkansas, Maryland, and Massachusetts, January 1, 1998–September 30, 2000.

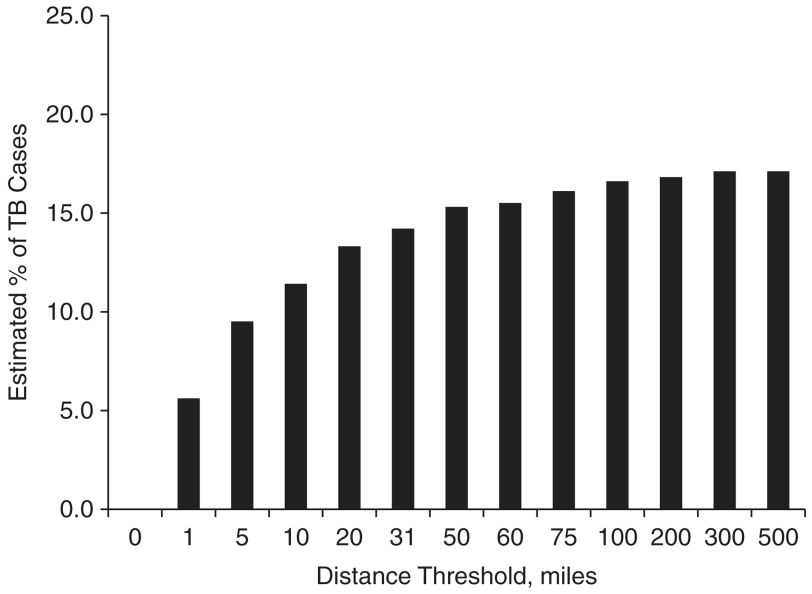

Figure 2.

Estimated percentage of tuberculosis (TB) cases attributable to recent transmission as determined using the novel plausible-source case approach, with distance thresholds ranging from 0 to 500 miles (800 km), Arkansas, Maryland, and Massachusetts, January 1, 1998–September 30, 2000. 1 mile = 1.6 km.

Table 1

Distribution of Patient Characteristics Among Cases With Field Evidence of Recent Transmission (RT) (*n* = 72), Patients With Evidence of Possible RT (*n* = 103), and Patients With No Field Evidence of RT (*n* = 1,013), Arkansas, Maryland, and Massachusetts, January 1, 1998–September 30, 2000

Characteristic	Evidence of RT (*n* = 72)		Possible RT (*n* = 103)		No Evidence of RT (*n* = 1,013)		*P* Value
	No.	Row %	No.	Row %	No.	Row %	
Sex							0.0420
Female	22	4.5	35	7.2	431	88.3	
Male	50	7.1	68	9.7	582	83.1	
Age, years							<0.0001
≤4	4	66.7	0	0.0	2	33.3	
5–14	0	0.0	1	9.1	10	90.9	
15–24	16	11.8	17	12.5	103	75.7	
25–44	30	7.4	47	11.6	327	80.9	
45–64	18	6.0	29	9.7	252	84.3	
≥65	4	1.2	9	2.7	319	96.1	
Nativity							<0.0001
US-born	60	9.9	60	9.9	488	80.3	
Foreign-born	11	1.9	43	7.5	523	90.6	
Unknown	1	33.3	0	0.0	2	66.7	
Race/ethnicity[a]							0.0006
White	17	4.6	29	7.8	327	87.7	
Black	42	9.7	48	11.1	341	79.1	
Hispanic	7	5.1	13	9.5	117	85.4	
American Indian/Alaska Native	0	0.0	0	0.0	1	100.0	
Asian or Pacific Islander	5	5.8	13	5.4	224	92.6	
Unknown	1	25.0	0	0.0	3	75.0	
Homeless							0.0950
Yes	5	12.2	7	17.1	29	70.7	
No	67	5.9	96	8.4	976	85.7	
Unknown	0	0.0	0	0.0	5	100.0	

Characteristic	Evidence of RT (*n* = 72)		Possible RT (*n* = 103)		No Evidence of RT (*n* = 1,013)		*P* Value
	No.	Row %	No.	Row %	No.	Row %	
Excess alcohol use							<0.0001
Yes	13	10.5	25	20.2	86	69.4	
No	54	5.5	70	7.2	853	87.3	
Unknown	5	5.8	8	9.3	73	84.9	
Injecting drug use							0.0013
Yes	6	22.2	5	18.5	16	59.3	
No	63	5.8	93	8.6	932	85.7	
Unknown	3	4.1	5	6.9	65	89.0	

Abbreviation: RT, recent transmission.

[a]Persons of Hispanic ethnicity might be of any race; non-Hispanic persons were categorized as Asian or Pacific Islander, black, white, American Indian/Alaska Native, or unknown.

Table 2

Sensitivity and Specificity of Tuberculosis Recent Transmission (RT) Estimates Using State-Based, SaTScan-Based, and County-Based Clustering Methods and Accuracy of Estimates Across a Range of Plausible Hypothetical RT Prevalence Rates (*n* = 1,085), Arkansas, Maryland, and Massachusetts, January 1, 1998–September 30, 2000

RT Estimation Method	No. of True-Positive Cases	No. of False-Positive Cases	No. of True-Negative Cases	No. of False-Negative Cases	Sensitivity, %	Specificity, %	Accuracy of Estimate by Hypothetical RT Prevalence, %					
							10%	15%	20%	25%	30%	50%
State-based clustering	72	129	884	0	100.0	87.3	88.5	89.2	89.8	90.5	91.1	93.6
SaTScan-based clustering[a]	58	53	960	14	80.6	94.8	93.3	92.6	91.9	91.2	90.5	87.7
County-based clustering	58	68	945	14	80.6	93.3	92.0	91.4	90.7	90.1	89.5	86.9

Abbreviation: RT, recent transmission.

[a]SaTScan was developed by Dr. Martin Kulldorff (Harvard Medical School, Boston, Massachusetts) in conjunction with Information Management Services, Inc. (Calverton, Maryland) (35).

Table 3

Sensitivity and Specificity of the Plausible-Source Case Approach to Estimation of Tuberculosis Recent Transmission (RT) Across Distances Varying From 0 to 500 Miles (800 km) and Accuracy of Estimates Across a Range of Plausible Hypothetical RT Prevalence Rates (*n* = 1,085), Arkansas, Maryland, and Massachusetts, January 1, 1998–September 30, 2000

Distance, miles (km)	No. of True-Positive Cases	No. of False-Positive Cases	No. of True-Negative Cases	No. of False-Negative Cases	Sensitivity, %	Specificity, %	Accuracy of Estimate by Hypothetical RT Prevalence, %					
							10%	15%	20%	25%	30%	50%
0 (0)	0	0	1,013	72	0.00	100.0	90.0	85.0	80.0	75.0	70.0	50.0
1 (1.6)	39	19	994	33	54.2	98.1	93.7	91.5	89.3	87.1	84.9	76.2
5 (8)	58	36	977	14	80.6	96.5	94.9	94.1	93.3	92.5	91.7	88.5
10 (16)	64	50	963	8	88.9	95.1	94.4	94.1	93.8	93.5	93.2	92.0
20 (32)	65	65	948	7	90.3	93.6	93.3	93.1	92.9	92.8	92.6	91.9
31.3 (50)	65	73	940	7	90.3	92.8	92.5	92.4	92.3	92.2	92.0	91.5
50 (80)	68	82	931	4	94.4	91.9	92.2	92.3	92.4	92.5	92.7	93.2
60 (96)	68	84	929	4	94.4	91.7	92.0	92.1	92.3	92.4	92.5	93.1
75 (120)	69	89	924	3	95.8	91.2	91.7	91.9	92.1	92.4	92.6	93.5
100 (160)	69	93	920	3	95.8	90.8	91.3	91.6	91.8	92.1	92.3	93.3
200 (320)	70	94	919	2	97.2	90.7	91.4	91.7	92.0	92.4	92.7	94.0
300 (480)	70	97	916	2	97.2	90.4	91.1	91.4	91.8	92.1	92.5	93.8
500 (800)	70	105	908	2	97.2	89.6	90.4	90.8	91.2	91.5	91.9	93.4

Abbreviation: RT, recent transmission.