Networks and tuberculosis: an undetected community outbreak involving public places

A.S. Klovdahl\textsuperscript{a,\ast*}, E.A. Graviss\textsuperscript{b}, A. Yaganchoost\textsuperscript{c}, M.W. Ross\textsuperscript{c}, A. Wanger\textsuperscript{d}, G.J. Adams\textsuperscript{e}, J.M. Musser\textsuperscript{b,1}

\textsuperscript{a}Department of Sociology, The Faculties, Australian National University, Canberra 0200, Australia
\textsuperscript{b}Institute for the Study of Human Bacterial Pathogenesis, Department of Pathology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
\textsuperscript{c}School of Public Health, University of Texas, Texas Medical Center, Houston, TX 77225, USA
\textsuperscript{d}Department of Pathology, University of Texas Medical School, Houston, TX 77225, USA
\textsuperscript{e}Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Abstract

After decades of decline in developed countries, there was a resurgence of tuberculosis in the mid-1980s accompanied by increased recognition that this infectious disease has long remained a major public health problem at the global level. New methods from molecular biology, in particular DNA 'fingerprinting' (of \textit{Mycobacterium tuberculosis}), made it clear that current transmission and recent infection (in contrast to reactivation of earlier, latent infection) were much more significant than previously believed. Studies of tuberculosis outbreaks using these new tools pointed to complex networks through which infection was spreading and highlighted the need for new approaches to outbreak investigation and disease control. In the study reported here a new approach — combining methods from molecular biology, epidemiology and network analysis — was used to examine an outbreak of tuberculosis in Houston, Texas. Initial investigation using conventional strategies revealed few contacts among 37 patients with identical (six-band) DNA (16S/16O-based) fingerprints but subsequent research uncovered over 40 places (including many gay bars) to which patients in this outbreak could be linked. Network methods were used to reconstruct an outbreak network and to quantify the relative importance (here, 'betweenness' centrality) of different actors (persons and places) playing a role in the outbreak. The multidisciplinary work provides the basis for a new approach to outbreak investigation and disease control. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Mycobacterium tuberculosis; DNA fingerprinting; Outbreak investigation; Network analysis; Outbreak network; Place-finding

Introduction

Increasing incidence of tuberculosis in the 1980s pushed efforts to control this infectious disease higher on public health agendas. In developed nations tuberculosis disease had been declining for many decades, in some areas for more than a century (McDougall, 1949; Lowell, Edwards & Palmer, 1969; McKeown, 1976). In view of this, there were expectations among experts that it could be eliminated in the foreseeable future, initially in developed nations and later in the rest of the world (Comstock, 1975; Mason, 1986).

At the turn of the century, however, tuberculosis remains a public health problem of global proportions. At least 1.7 billion people are infected, approximately...
one-third of the population of the world, about eight
million new cases emerge each year and roughly three
million people die annually of tuberculosis (Kochi, 1991;
Ravignione, Snider & Kochi, 1995).

TB 101

Understanding the transmission of pathogens affecting
humans and the development of disease that results — both
required for effective disease control — poses research
problems that transcend disciplinary boundaries. At the
time, an essential starting point is available knowledge
about the relevant biology and epidemiology.

Tuberculosis is typically caused by Mycobacterium
tuberculosis, a bacterium. When healthy persons are
infected with M. tuberculosis in an estimated 90-95% of
cases their immune systems deal effectively with it; their
infections are kept in check and these (latent or
dormant) infections do not lead to disease (Comstock
and O’Brien, 1991). In fact, one result of infection is
usually increased immune system response to mycobac-
teria, reflected in a ‘positive’ reaction to a tuberculin skin
test (e.g. Mantoux test; Snider, 1982). On the one
hand, this augmented response reduces the likelihood of
reinfection; on the other, tuberculin skin tests have been
very useful in tuberculosis control. That is, a negative
test for a person who has been in contact with an
infectious patient followed by a positive result, i.e. a
‘conversion’, has long been used as an indicator of
transmission between an infected and a susceptible

Persons with latent tuberculosis infection are not
infectious and thus cannot transmit their infection to
others. In a small proportion of otherwise healthy adults
infected with M. tuberculosis (roughly 5-10%) infection
progresses to active tuberculosis disease, within a short
time, or later as a result of endogenous reactivation. The
disease can affect virtually any part of the body (Slavin,
Walsh & Pollack, 1980; Hopewell, 1994) but ordinarily
only larynx involving the respiratory tract (e.g.
lungs, larynx) can become infectious. Talking, coughing,
sneezing and singing are mechanisms by which secre-
tions containing bacteria in human respiratory passages
are aerosolized and expelled into the surrounding air as
droplets (Mims, Dimmock, Nash & Stephen, 1995).
These droplets, too heavy to travel far, after evaporation
become ‘droplet nuclei’, which are able to circulate
widely in confined spaces (Wells, 1934). Airborne
droplet nuclei containing viable tubercle bacilli may be
inhaled by others and cause exogenous infection (Riley,
Although "[the principal risk behavior for acquiring
TB infection is breathing]" (Bloom & Murray, 1992),
there has been a widespread view that tuberculosis is not
highly contagious. Evidence cumulated over decades
suggested that close contact with an infectious person
over an extended period of time in a confined space was
usually required for transmission (Riley, 1957; Loudon,

An important characteristic of tuberculosis is that
most cases — if diagnosed early — can be treated
successfully. However, organisms resistant to anti-
tuberculosis drugs were observed shortly after the
introduction of the first of these antibiotics (streptomycin)
and concern about drug resistance has increased
(BMRC, 1948; Riedel, 1993b; Frieden et al., 1996; Ridzon
et al., 1997; Taylor, Bergmanie-Sweat & Suarez,
1999). As incomplete treatment can lead to the
emergence of drug resistant organisms, directly observed
therapy (DOT) to improve patient adherence to
prescribed antibiotics has become an important element
in tuberculosis control (Sumartojo, 1993; Bayer &

Treatment assumes that cases of active tuberculosis
disease are found. Case-finding has been a key
component of tuberculosis control programs and con-
tact tracing an important tool for active case-finding for
many years (Riedel, 1993a; Etkind, 1993). In low-
prevalence (e.g. developed) countries, with relatively few
persons with infectious tuberculosis disease in the larger
community, contact tracing has been guided by the
’stone-in-pond’ or concentric circle principle (Veen,
1992; Etkind, 1993). That is, contacts are ranked by
their likely degree of exposure and those perceived most
at risk targeted first. In the past this has meant that close
typically, within-household) contacts of patients were
the earliest to be examined, then — resources permitting
— casual contacts.

The development of molecular biological strain-
typing methods, for example restriction fragment length
polymorphism (RFLP) fingerprinting of M. tuberculo-
sis, has had a profound impact on understanding of
tuberculosis epidemiology. Basic research on the M.
tuberculosis genome discovered DNA probes (e.g.
Insertion Sequence IS6110) that produced good genetic
markers and made it possible to differentiate between
strains.

Studies carried out using DNA fingerprint methodology
have provided evidence that much more tuberculosis
transmission was taking place than had been believed,
recently, outside households, and often involving
persons for whom no close contact to another case
could be found. To illustrate, in the Canton of Bern
(Switzerland) 163 isolates were fingerprinted (IS6110-
based RFLP). 50 patients were found to be potentially
linked and 23 with the same pattern (‘RFLP-pattern ’)
were used as the starting point for an outbreak
investigation (Genewein et al., 1993). Three main
groupings of patients with active tuberculosis disease
were then identified. The first consisted of homeless
persons, drug addicts and alcoholics who used shelters and other services provided for persons on the margins of society. The second centered on a restaurant, consisting of the restaurant owner, a waitress, a patron with a link to one of the drug addicts, and three others who frequented the restaurant. The third consisted of two patients who could not be linked to the other groups. One worked in a nursing home and one was a homeless alcoholic. This study demonstrated the usefulness of fingerprint methods in outbreak investigation and also called attention to the potential for tuberculosis to spread from groups at highest risk to others. As the authors observed, "there is a definite 'spillover' to the general population" (Genewein et al., 1993).

In another study on-going transmission of multidrug resistant M. tuberculosis was observed by Bifani and colleagues (Bifani et al., 1996). In New York City 206 patients were found to have the same fingerprint (IS6110, 17-band). This strain — strain W — was resistant to four of five first-line tuberculosis antibiotics, and in some cases to second-line drugs as well. An additional 67 tuberculosis patients with similar or closely related fingerprints were subsequently identified. Geographically, the 273 patients with this strain resided in New York, Colorado, Florida, Georgia and Nevada.

The limitations of conventional public health methods became clear as a result of studies such as these. Although individual patients often were identified by traditional surveillance procedures at some point in an outbreak, and some small clusters of disease found, typically the magnitude of the larger outbreaks of which they were part was not realized. The application of DNA fingerprint methods, thus, exposed a serious weakness in tuberculosis control programs. For example, one path-breaking (IS6110) study was submitted "a population-based study using conventional and molecular methods" (Small et al., 1994). The study underscored the problem: cutting edge methods developed in molecular biology were being used in conjunction with methods of outbreak investigation unchanged for decades. In this San Francisco study only 10% of the patients with tuberculosis disease who were linked by common fingerprints (IS6110) had been identified by conventional contact investigation and the authors pointed to the need to develop "novel approaches to contact tracing" (Small et al., 1994). In the Swiss study cited above, moreover, the authors noted the difficulty in trying to observe the 'complex transmission networks' in the outbreak they had uncovered with DNA fingerprints (Genewein et al., 1993). Commenting on this study, Day and colleagues pointed to the potential power of using social network methods in conjunction with DNA fingerprinting to study tuberculosis outbreaks (Day, Jison, Ward & Weber, 1993).

The Houston Tuberculosis Initiative

The Houston Tuberculosis Initiative was undertaken to explore a new approach to outbreak investigation and tuberculosis control based on combining methods from molecular biology, epidemiology and network analysis.

The work focussed on Houston (Texas), the fourth largest city in the United States (1990: 1.6 million), located in Harris County, the third most populous county in the country (1990, 2.8 million). In terms of racial diversity, the county was eleventh in the country, comprised of about 54% whites, 22% Hispanics, 19% blacks and 4% Asians or Pacific Islanders. In addition, Houston — about an eight hour drive from the Mexican-American border — was located within a major corridor for illegal immigrants travelling from south of the border, i.e. from higher tuberculosis prevalence countries, to other parts of North America.

Texas itself has long been a high prevalence state (Lowell et al., 1969) and case rates for Houston in the 1990s were three to four times the US national average (CDC, 1994). In 1991 there were 609 new cases of tuberculosis disease in Houston: in 1992, 758 new cases. This represented an increase of roughly 25%, the largest for any American city. By 1997 there were 780 new cases, an increase of almost 30% over the 1991 figures. In terms of rates, in 1991 the rate case for Houston was 34.5/100,000, over three times the US national average (10.4/100,000); in 1992 it was 40.4/100,000, roughly four times the average (10.5/100,000). This placed Houston among the US metropolitan areas with the highest case rates (CDC, 1994).

Part of the explanation for tuberculosis in Houston in the 1990s was the amount of human immunodeficiency virus (HIV) infection and the number of acquired immunodeficiency syndrome (AIDS) cases. HIV-induced reduction in immune system effectiveness can contribute to the development of tuberculosis disease, which in turn has been a criterion for AIDS since 1993 (CDC, 1992; Shafer & Edlin, 1996). By March 1994 over 10,000 (10,231) AIDS cases had been reported in Houston and at least 1000 new cases were being notified annually. These figures put Houston among the top five or six American cities in AIDS cases.

Network concepts

A network is a set of nodes connected together by links of one kind or another. A social network is a set of persons connected together by social relationships. Depending on the focus of a study, the links between these persons may be friendships, kinship relations, work relationships, relationships with neighbors, sexual relationships and so on (Mitchell, 1969; Scott, 1992).
Social networks channel flows of various kinds through populations, for example, information, influence and infection. Studies of the flow of infections through social networks have concentrated on pathogens transmitted in the course of sexual contact. Early work has included studies of patients with AIDS in the initial 'patient zero' cluster (Klovda halm, 1985), the geographical dissemination of HIV infection (Haraldsdottir, Gupta & Anderson, 1992), the possibility of its heterosexual spread (Klovda halm, Potterat, Woodhouse, Muth, Muth & Darrow, 1994; Rothenberg, Potterat, Woodhouse, Muth, Darrow & Klovda halm, 1998a), and the flow of 'traditional' sexually transmitted infections, such as gonorrhea and syphilis (Ghani et al., 1996; Rothenberg et al., 1998b; Day, Ward, Ison, Bell & Weber, 1998).

Social networks are generally conceptualized as graphs or digraphs. A mathematical graph is a set of points (vertices) connected by lines (edges). The lines in a graph are undirected ($l_{ij} = l_{ji}$) (Wilson, 1985). A digraph (i.e., a directed graph) is a set of points connected by directed lines. In a digraph, if a line $l_{ij}$ is present ($l_{ij} \neq 0$), a line $l_{ji}$ may ($l_{ji} = 1$) or may not ($l_{ji} = 0$) be (Harary, Norman & Cartwright, 1965). Conceptualizing networks as graphs (or digraphs) makes relevant powerful mathematical and statistical tools (Wasserman & Faust, 1994).

The points in the mathematical graph representing a network through which HIV may be spreading typically have been all persons. In studying HIV/AIDS, moreover, the definition of a personal (risk) relationship has been sufficiently elastic to incorporate needle-sharing by injecting drug users as well as sexual relationships and thus encompass the most important (horizontal) ways in which HIV infection was being transmitted (Neaigus et al., 1994; Curtis, Friedman, Neaigus, Jose, Goldstein & Idefonso, 1995; Latkin, Mandell, Vladhov, Oziemkowska & Celentano, 1996; Friedman et al., 1997).

Tuberculosis is more complicated. Outbreaks have been associated with places and personal relationships linking infected persons to possible sources of infection — if they exist — may not be ascertainable (Allos et al., 1996; Kenyon, Valway, Ihe, Onorato & Castro, 1996; Sepkowitz, 1996; Valway et al., 1998). In consequence, it is useful to think in terms of 'outbreak networks'.

An outbreak network is a set of actors together with the links that connect them in an infectious disease outbreak. The actors may be persons, places and other 'objects' (animate or inanimate) that 'play a role' in the transmission of an infectious agent. The links are epidemiologically-relevant interactions among the persons, places and any other actors involved. In outbreaks involving only persons, an outbreak network reduces to a social network in which an epidemiologically-relevant subset of links among persons is considered.

A tuberculosis outbreak network, thus, can consist of persons, places and the relevant links connecting them. For simplicity, it is assumed here that the graph describing an outbreak network consists of 'equivalent' points, persons or places, infectious or capable of being infected. By the same token, a place would be infected infectious in a different sense than a person. For example, a household could be considered infected infectious if any person(s) associated with it had active infectious tuberculosis disease. There may be attributes of a place, of course, such as physical features or social characteristics, that potentiate its role in an outbreak.

The concept of 'centrality' is a useful starting point for using network methods to study outbreak networks. Students of social networks have devoted a good deal of attention to centrality and its measurement (Wasserman & Faust, 1994). The original focus of this work was on communication networks, in particular on 'gatekeepers'. persons whose location in a network enabled them to exert control over the flow of information (Freeman, 1979). Location in a network can also affect the flow of infection (Klovda halm, 1985; Rothenberg et al., 1998a).

Everything else being constant, more centrally-located actors in a network should have a higher probability of being infected, and — if infected — of infecting others.

Methods

The Houston Tuberculosis Initiative attempted to locate all new (incidence) cases of tuberculosis disease diagnosed in Houston (Harris County) during a five-year period, to culture isolates of M. tuberculosis from these patients, and to fingerprint them using standardized methods (e.g. IS6110-based RFLP procedures).

Information about new cases was obtained from the Houston Tuberculosis Control Office, public clinics, private physicians and area hospitals. This entailed ongoing communication with the appropriate persons and organizations, and included regular reviews of hospital records for suspected new cases. Information on new cases was also available from four diagnostic laboratories responsible for about 90% of sputum and or culture tests in Houston, and from about 40 other private laboratories that carried out the remainder.

The methodology used to fingerprint isolates has been described in detail elsewhere (van Embden et al., 1993; Small & van Embden, 1994). After fingerprints were obtained, they were analyzed and archived in a computerized database (BioImage Whole Band Analysis, version 3.2; BioImage, Ann Arbor, Michigan, USA) and pattern matching software used to identify both matching and similar fingerprints (prints). For similar prints, gels were rerun to unambiguously match strains, with visual confirmation.
Patients with newly diagnosed tuberculosis were approached by members of the research team, the study was described and a request to participate made. If informed consent was obtained, patients were interviewed using the 'Houston Mycobacteria Active Surveillance Form' (Baylor Interview). In this patients were asked about their demographic characteristics, living situations, modes of transport, travel, social contacts and places frequented. Additionally, they were asked to provide tobacco, alcohol and drug use, sexual and medical histories. For consenting patients, clinical records were reviewed for symptoms (e.g. cough), dates of symptom onset, specimen details and dates, diseases, treatments, adherence to medication and outcomes, including death. Where possible, information on patients who had died (or left the area) was sought from proxy persons. Members of the research team met weekly with field staff to maximize the likelihood of locating prospective study participants. A payment of U$10 was offered to each participant.

Although transmission of infection is inherently a directed process, e.g. from an infected person to a susceptible person, to simplify the analysis here it was assumed that the links were non-directional. This was equivalent to saying that although transmission from one actor to another was indicated, the evidence considered was not sufficient to specify in which direction. Data were analyzed using SAS (version 6.11) (SAS Institute Inc., 1994), and Ucinet 5.0 (Borgatti, Everett & Freeman 1999). Network diagrams were created using Microsoft Access (version 2.0) and Graphviz (version 1.4; Koutsofios and North, 1993).

The relative importance of the persons and places identified in the outbreak study was quantified using the classic betweenness measure of centrality for points in a graph. Expressed formally, the measure of overall betweenness centrality for each point was

$$\beta(p_i) = \Sigma \beta_{ij}(p_i)$$

which — for each point (p_i) on any shortest path from between points i and j — sums the partial betweenness scores (\(\beta_{ij}(p_i)\)). For graphs \(\beta_{ij}\) is summed for each \(p_i\) over every pair of points \(i < j \land i \neq j\) (Freeman, 1979).

The 'Print 4' cluster: an undetected outbreak

Early in the project, clusters of tuberculosis patients in Houston with identical DNA fingerprints were observed. These clusters provided evidence against endogenous reactivation of latent infection acquired in the past; they pointed to recently acquired infection and thus to recent transmission. The organism associated with one of these clusters was labeled the 'Print 4' strain. This cluster was then made the focus of an investigation using resources not normally available to public health departments, such as ethnographic fieldwork, laboratory and computer software resources.

In Fig. 1, a schematic representation of 14 DNA (IS6110 RFLP) fingerprints with the Print 4 banding pattern is shown. Lanes 1 and 14 are from control isolates. The other 12 (lanes 2-13) are examples of prints from Print 4 patients. As can be seen, this was a 'six-band' strain; it had six copies of the Insertion Sequence IS6110.

Examined here are 37 cases, occurring between 1993 and 1996, which were identified at the time of this analysis with the Print 4 strain. All 37 of these patients were diagnosed with active tuberculosis disease (standard case definition), about 68% (25/37) with pulmonary disease, 22% (8/37) with extra-pulmonary disease and 10% (3/37) with both. Of the 37, 36 were male; one was female. More detailed information was available for 34 patients, in some cases through a proxy person. Of these, 47% (16/34) were between 25 and 34 years old when diagnosed, 38% (13/34) between 35 and 44, and 15% (5/34) 45 or older. About three-quarters (25/34) were white, the remainder black (3/34) and Hispanic (6/34).

Four of the Print 4 patients were married or cohabiting, six were separated, divorced or widowed, and 24 (70%) were never married. About 80% (27/34) had 12 years or more of formal education; roughly 25% (9/34) were employed at time of diagnosis. Moreover, about 20% (7/34) identified themselves as heterosexual, 70% (24/34) as gay and 10% (3/34) as bisexual. In addition, about 70% (24/34) answered 'yes' when asked if they had ever had a positive HIV test and about 30% (10/34) 'no'.

With 37 patients, each patient theoretically could be in contact with up to 36 other (Print 4) patients, so there was a possible maximum of 666 links connecting them \(6(n-1)/2 = 37(36)/2\). In other words, there could be a link (e.g. personal contact) between patient no. 1 and patient no. 2, between patient no. 1 and patient no. 3, ..., between patient no. 2 and patient no. 3, ..., and, between patient no. 36 and patient no. 37. Although this mathematically possible maximum would be unlikely to be reached, early studies of AIDS patients indicated quite large numbers of sexual contacts. For example, about 75% of (29) men in the 'patient zero' cluster of AIDS patients reported 50 or more male sexual partners in the year before onset of symptoms (Auerbach, Detroy, Jaffe & Curran, 1984).

In Houston, however, when contacts of these 37 Print 4 patients were examined, only 12 person-to-person (case-case) links were found, from both contact investigation and ethnographic fieldwork. The index patient could not be linked to any other patient; half of the patients (51%, 19/37) had no known link to another patient using information from proxy persons as well.

A. S. Klevdalh et al. Social Science and Medicine 52 (2001) 681-694
In Fig. 2, the distribution of cases over time, using date of diagnosis of tuberculosis disease (morbidity date) to place patients chronologically, is presented together with the contacts found to link patients. Twenty-nine of the 37 cases are shown. Clearly, the information on personal contacts was not sufficient to understand how this outbreak was occurring.

After further investigation in collaboration with the (City of Houston) Health Department, it emerged that places rather than persons were the key. Virtually all patients for whom data were available (33/34) reported having frequented bars, taverns, restaurants or similar establishments (e.g. cafes, ‘ice houses’). And, 32 of the 37 patients (86%) could be linked to one or more specifically-named places of this kind. In fact, the fieldwork yielded a list of over 40 bars and other places frequented — during the period studied — by patients diagnosed with active disease caused by the Print 4 strain of *M. tuberculosis*. Importantly, the significance of places in this outbreak was supported by a case-control study (Yaganehdoost et al., 1999).

Of the (44) public places described by these patients, 50% (22/44) were gay (and/or lesbian) bars (or dance clubs). Twelve of these (22) gay/lesbian bars were considered ‘public’ bars as well, i.e. they were frequented by members of the public not known to be gay or lesbian. In one instance a bar was a ‘gay bar’ one night during the week, and a ‘public’ bar the rest of the time. Further, 11 of the 44 places reported were not identifiable as gay bars, including two ‘topless’ bars. There was not adequate information for another nine places to allow characterization with any confidence, including two local bars thought to have a predominantly Hispanic clientele, i.e. possible ‘Latin’ bars.

Most of these places, as would be expected, were in Houston or ‘Greater Houston’ (Harris County), but two were in Dallas (Texas), about five hours away by automobile, one was in San Diego (California), one San Francisco (California), one Las Vegas (Nevada) and one New York City (New York).

A next step

At this point the outbreak might have been considered ‘explained’. To have done so, however, would have meant ignoring data obtained in the outbreak investigation (as in many others) that could be used to reconstruct an outbreak network. It also would have
The overall pattern of case–place links represented — in effect — an hypothesis about transmission of tuberculosis infection in this outbreak. That is, 29 (about 80%) of the Print 4 patients were linked together in a single connected region of an outbreak network, directly, or indirectly though other patients or places. As patients were often linked to more than one place (most often a bar) multiple opportunities had been available for exposure to infection.

**Centrality in a tuberculosis outbreak**

Although attempting to depict all of the links found between patients and places would have rendered Fig. 3
unintelligible, all of the relevant data were used to calculate centrality scores. These data consisted of 37 patients with identical fingerprints, 13 places (e.g., bars) that linked two or more patients, and over 100 (non-directional) links among patients, and between patients and places (where $l_{ij} = 1$ and represents one link).

The centrality scores (rounded) for the 10 most important actors in this outbreak are shown in Table 1. Higher scores indicate greater centrality. Clearly, places were critical for understanding this outbreak. Six of the top 10 most central actors were places, in particular, gay bars. Each served food and drinks; each was in the gay public bar category referred to earlier. The most central place, Place 30, was named by 13 patients, though for simplicity only some of its links to these patients are shown (in Fig. 3). This bar was known as a ‘pickup joint’, i.e. a bar popular with clients in search of casual sex or ‘one night stands’.

The most second central of these bars, Place 20, was also named by 13 patients as a place frequented by them. Place 20 was described as a popular place for socializing. Of the remaining four central places, Place 12 was named by 12 patients, Place 24 by 13, Place 26 by 12 and Place 33 by seven patients.

Four patients were in the top 10 in terms of their importance in this outbreak. All were white males. Two were between 25 and 34 years old, one between 35 and 44 and one between 55 and 64. Only one was employed. And, three of the four were known to be HIV positive. In terms of contacts, one patient was reported to have had contact with nine other (Print 4) patients or places in the outbreak, one with seven and two with six patients or places.
Table 1
Importance of persons and places in the Print 4 outbreak

<table>
<thead>
<tr>
<th>Rank</th>
<th>Case place</th>
<th>Centrality score (betweenness)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Place 30</td>
<td>148</td>
</tr>
<tr>
<td>2</td>
<td>Place 20</td>
<td>146</td>
</tr>
<tr>
<td>3</td>
<td>Place 12</td>
<td>108</td>
</tr>
<tr>
<td>4</td>
<td>Place 24</td>
<td>105</td>
</tr>
<tr>
<td>5</td>
<td>Place 26</td>
<td>75</td>
</tr>
<tr>
<td>6</td>
<td>Case 14</td>
<td>67</td>
</tr>
<tr>
<td>7</td>
<td>Case 28</td>
<td>58</td>
</tr>
<tr>
<td>8</td>
<td>Case 21</td>
<td>57</td>
</tr>
<tr>
<td>9</td>
<td>Place 33</td>
<td>56</td>
</tr>
<tr>
<td>10</td>
<td>Case 17</td>
<td>44</td>
</tr>
</tbody>
</table>

*The centrality scores have been rounded to the nearest integer. Calculations were based on all relevant data (patients, 37; places, 13; non-directional links, 112) although not all are shown in the figures.

Discussion

The patients in the Print 4 cluster were not identified as being part of a single outbreak by conventional public health case-finding and contact tracing procedures. DNA (IS6110-based) fingerprint methods made possible the recognition that a large outbreak was occurring. Ethnographic fieldwork uncovered additional links among persons and found many links between persons and places.

The failure of traditional public health methods to recognize this ongoing outbreak was not unusual. The use of fingerprint methods has led to the discovery of a number of outbreaks of tuberculosis disease undetected by conventional approaches. In part this demonstrates the power of new molecular methods; in part, it reflects the limited resources and outdated tools currently available for outbreak investigation. Indeed, it would have been remarkable if any public health department using traditional methods had been able to identify these cases as part of a single on-going outbreak.

The main difficulty in understanding this outbreak initially was the limited number of close personal contacts identified. Each of the 37 patients could potentially have had contact with up to 36 other (Print 4) patients, but the number of personal connections identified was less than 2% of the number possible (12,666). As the outbreak developed over several years, and some patients died during this period, the number of person-to-person links actually possible would have been less than the theoretical maximum. For example, a patient who became part of the outbreak toward the end of the period may have had no opportunity to be in contact with a patient who died early in the outbreak. Nevertheless, the number of links found among Print 4 patients — using both contact investigation and ethnographic fieldwork — was still quite small.

Places, principally gay bars, were found to be critical to the Print 4 outbreak. Although many bars were described as gay bars, they were frequented — to varying degrees — by persons who did not characterize themselves as gay or lesbian. In other words, the bars could be perceived by some who patronized them during the outbreak as ‘gay’ bars and by others as ‘neighborhood’ bars, or just ‘bars’.

The importance of places in tuberculosis transmission has been documented in a number of other studies. Moreover, this was not the first time bars or taverns have been associated with tuberculosis. For example, studies carried out earlier reported barmen had high rates of tuberculosis mortality (Logan and Benjamin, 1957). And, in a small town outbreak that began in 1989 (in Maine), 10% (3/30) of ‘tavern regulars and employees’ in contact with an infectious case were believed to have been infected by him (Alls et al., 1996). Further, Kline, Hedemark and Davies (1995) documented (IS6110) an outbreak of tuberculosis among the patrons of a bar in Minneapolis several years ago. In Houston, the evidence suggested that a significant proportion of over 40 places — including many bars — may have been associated with the transmission of a single strain of *M. tuberculosis*.

Regarding other places, Barnes et al. (1997) identified eight clusters in a population of 162 isolates from patients in central Los Angeles. Although patients within each cluster (ranging from two to 43 patients) had similar fingerprints (IS6110 and pTBN12), only two patients named each other as contacts. Homeless shelters were found to be the connecting links and the authors pointed to the importance of extra-household locations in tuberculosis transmission.

In Houston it was possible to find evidence supporting recent transmission of a strain of *M. tuberculosis* in a community outbreak involving a large number of places and — using network methods — to make an initial attempt to quantify the role they played in the outbreak. Places were as important as persons. Although it appeared unlikely that the outbreak could have been sustained without the places identified, neither could it have occurred in the absence of infectious persons. Both persons and places were among the top 10 most central actors in the outbreak. Thus, effective tuberculosis control must focus on both persons and places.

The attention given to different kinds of actors in tuberculosis outbreaks, however, is likely to vary from outbreak to outbreak. The appropriate relative person place emphasis would depend on characteristics of an outbreak and the populations at risk. For example, recent work (Sreevatsan et al., 1997) suggested that *M. tuberculosis* strains can be placed into one of three genotypic groups (group 1, group 2, group 3) and that
strains in groups 1 and 2 have been associated more often with larger outbreaks. Print 4 was a group 2 organism, and the outbreak associated with it was one of 46 large clusters (of 47) in Houston and New York caused by group 1 or group 2 organisms (Sreevatsan et al., 1997). This suggested that place-finding may be indicated — and most cost-effective — when group 1 or group 2 organisms are identified. In other words, place-finding — used when indicated by evidence that a more easily transmitted and or more virulent strain is involved — may help to prevent small numbers of cases from developing into large outbreaks.

The outbreak network diagram presented in Fig. 3 represents an hypothesis about the spread of a pathogen. The plausibility of the hypothesis depends on the evidence available to support it. Accordingly, it should be noted that a schematic representation of the DNA fingerprints (e.g. Fig. 1) can convey an unwarranted sense of precision. That is, there are a number of possible sources of error in the DNA fingerprinting procedures. For present purposes, suffice it to note that previous research has provided strong evidence that DNA fingerprints of M. tuberculosis meet the requirements for good genetic markers, and that the cultures used here were retested when questions arose (Eisenach, 1994). In this analysis, moreover, only patients with identical six-band (IS6110) prints were included (Yeh et al., 1998; Burman et al., 1997). In addition, IS6110profiling was supported by other strain differentiation methods (Yaganehdoost et al., 1999).

Another possible source of error was variation in individual rates of progression from infection to disease. As morbidity rates were used to locate cases in time (Figs. 2 and 3), this invariably meant that some individuals would be 'misplaced' chronologically. However, the measure of centrality used was based on the assumption of non-directional links between network nodes (persons/places) and representing the outbreak network in this manner (i.e. as a mathematical graph) reduced the likelihood of some errors attributable to patient 'placement'. By the same token, only one measure of importance was used and other measures may have provided additional (or better) information (Stephenson & Zelen, 1989; Altmann, 1993; Rothenberg, Potterat, Woodhouse, Darrow, Muth & Klov Dahl, 1993; Bell, Atkinson & Carlson, 1999). Although beyond the present scope, sensitivity analyses that consider possible periods-of-infectiousness errors can be useful (Behr et al., 1999).

Other factors may have affected measurement of the relative importance of places in this outbreak. On the one hand, the prominence of places such as bars could have been overstated to the extent that patients found it easier to remember places where they had spent time rather than persons with whom they were in contact. Thus, place-associated transmission did not necessarily occur within the confines of the places found to be important in this outbreak.

On the other hand, the importance of places may have been underestimated. To illustrate, in Fig. 3 Place 30 was represented twice, as 'Place 30 (1)' and as 'Place 30 (2)'. As many of the patients had compromised immune systems (HIV/AIDS, with lowered CD4 counts), tuberculosis infection was more likely to progress to tuberculosis disease, and more rapidly (Shafer & Edlin, 1996). It appeared likely, then, that cases such as no. 26 (Fig. 3), diagnosed in 1996, were more likely to have had their tuberculosis-associated contact with Place 30 closer to 1996 than to 1993 when this bar originally appeared in the outbreak. If so, Place 30 would be on additional shortest paths between the index case and (for example) case no. 36 (which were not included in this analysis) and thus score higher in centrality. In short, the measures of importance used here did not take into account that some places were associated with this outbreak over a period of years.

Another factor supports the possibility that the relative importance of places in this outbreak may have been underestimated. As noted previously, M. tuberculosis infection in about 90-95% of healthy persons never progresses to disease. Here, almost three-quarters of the (37) patients were not 'healthy'; they were persons with HIV/AIDS. As tuberculosis transmission is more difficult to document in the absence of disease, other (unobserved) transmission of the Print 4 strain associated with the places identified was likely to have occurred, to healthy (non-HIV/AIDS) persons. In some of these, disease may develop in the future.

Overall, the importance of the places identified in this outbreak appeared to be at odds with much previous research, which found tuberculosis transmission largely confined to households. There are a number of possible reasons for this. For example, the lifestyles of those studied here may have been significantly different from those in earlier outbreaks. That is, the Print 4 patients were mostly men, 25-44 years old, almost three-quarters of whom had never married and identified themselves as gay or bisexual. They may have lived in smaller households and spent more time socializing in public places such as bars or restaurants. Equally, lifestyles in the larger population (e.g. of Houston) may have changed since earlier work; more meals may have been taken outside the household during the period studied, in restaurants or other establishments that serve food.

Whatever the complete explanation, persons with infectious tuberculosis disease expose others outside their households more often than has been appreciated. In the absence of methods for strain differentiation, observation of more tuberculosis among household and other close contacts was likely to have been influenced by limited knowledge about how outbreaks were actually occurring. This does not reflect adversely on
the many high quality studies carried out over the past 100 years, it merely serves as a reminder of the inherent difficulties involved in studying *M. tuberculosis* and its transmission.

Conclusions and implications

As the incidence of tuberculosis disease has declined in more prosperous areas of the world, increasingly case-finding and contact tracing have been directed primarily at household members and other close contacts. DNA fingerprinting of *M. tuberculosis* strains, however, has made it clear that significant transmission has been occurring in circumstances not previously recognized as important. In consequence, programs to reduce tuberculosis incidence that concentrate almost exclusively on within-household and other close contacts require serious reconsideration.

In this research, DNA fingerprinting procedures, contact tracing, fieldwork, network concepts and methods were used to study the likely course of an outbreak of tuberculosis that occurred over a number of years. When place-finding was added to conventional case-finding methods, the case-place links identified made it possible to develop a plausible explanation for this outbreak.

Put succinctly, it had been possible to eliminate place-associated transmission early in this outbreak, over 90% of the cases caused by the Print 4 organism may have been prevented.

Several implications of this work stand out. Firstly, there is a need to examine critically the outbreak investigation methods currently employed and to evaluate their effectiveness in identifying persons exposed to tuberculosis. Relevant to this need is a large body of social network research aimed at developing methods for eliciting valid and reliable information about personal contacts (de Sola Pool & Kochen, 1978; Bernard, Johnsen, Killworth, McCarty, Shelley & Robinson, 1990; Brewer, 1999).

Secondly, in outbreaks where places as well as persons may be involved, i.e. places other than private households, outbreak investigations may be more effective if not only 'person oriented' (case-finding) but also 'place oriented' (place-finding). Importantly, there is no evidence that procedures currently used to find persons who have been in epidemiologically-relevant contact with infectious cases are effective (or efficient) for locating epidemiologically-significant places.

Thirdly, developing valid and reliable measures of the importance of actors (or classes of actors) in an outbreak network could help to improve disease control. For example, calculating quantitative measures of the importance periodically could help public health departments identify key persons and places in outbreaks, thereby assisting in directing resources more effectively. For persons identified as important, the range of relevant tools for tuberculosis control would include tuberculin skin testing, chest radiography, chemoprophylaxis, chemotherapy, directly observed therapy (DOT) and — in recalcitrant cases of non-adherence and/or drug resistance — isolation. For places identified as important, the range of possible interventions could include site-specific educational programs designed to increase tuberculosis awareness, more extensive tuberculin skin testing, possibly mobile clinics, as well as programs for ascertainment of compliance with relevant public health regulations, such as those relating to ventilation in public places. Clearly, places identified as important in an outbreak are of interest not only because of their possible direct involvement in infectious agent transmission. Places are also potentially significant as prospective locations for more cost-effective public health interventions aimed at preventing future transmission, even when (paradoxically) actual transmission may be known to be occurring elsewhere.

To end on a somewhat positive note, resources for tuberculosis control were augmented during the 1990s and helped to reduce case numbers, in Houston and elsewhere in the US (CDC, 1998). Yet, in view of the magnitude of the overall problem, especially at the global level, there remains a pressing need for new, more effective strategies for tuberculosis control. The approach explored here can help to provide a basis for these. This multidisciplinary approach can also contribute to improved outbreak investigation and disease control more generally.

Acknowledgements

Grant support — NIDA, NIH (DA-09238, J.M. Musser, Principal Investigator). The authors express their appreciation to members of the City of Houston Health Department for their cooperation and to Jim Frewin for his database expertise.

References


van Embden, J. D. A., Cave, M. D., Crawford, J. T., Dale, J. W., Eisenach, K. D., Giequelin, B., Hermans, P. W. M., Martin, C., McAdam, R., Shinnick, T. M., & Small, P. M.


