BMJ Open

Protocol for a population-based molecular epidemiology study of tuberculosis transmission in a high HIV-burden setting: the Botswana Kopanyo study

N M Zetola,[1] C Modongo,[2,3] P K Moonan,[4] E Click,[4] J E Oeltmann,[4] J Shepherd,[5] A Finlay[4]

For numbered affiliations see end of article.

Correspondence to
Dr NM Zetola;
nzetola@gmail.com

ABSTRACT

Introduction: *Mycobacterium tuberculosis* (Mtb) is transmitted from person to person via airborne droplet nuclei. At the community level, Mtb transmission depends on the exposure venue, infectiousness of the tuberculosis (TB) index case and the susceptibility of the index case's social network. People living with HIV infection are at high risk of TB, yet the factors associated with TB transmission within communities with high rates of TB and HIV are largely undocumented. The primary aim of the Kopanyo study is to better understand the demographic, clinical, social and geospatial factors associated with TB and multidrug-resistant TB transmission in 2 communities in Botswana, a country where 60% of all patients with TB are also infected with HIV. This manuscript describes the methods used in the Kopanyo study.

Methods and analysis: The study will be conducted in greater Gaborone, which has high rates of HIV and a mobile population; and in Ghanzi, a rural community with lower prevalence of HIV infection and home to the native San population. Kopanyo aims to enrol all persons diagnosed with TB during a 4-year study period. From each participant, sputum will be cultured, and for all Mtb isolates, molecular genotyping (24-locus mycobacterial interspersed repetitive units-variable number of tandem repeats) will be performed. Patients with matching genotype results will be considered members of a genotype cluster, a proxy for recent transmission. Demographic, behavioural, clinical and social information will be collected by interview. Participant residence, work place, healthcare facilities visited and social gathering venues will be geocoded. We will assess relationships between these factors and cluster involvement to better plan interventions for reducing TB transmission.

Ethics: Ethical approval from the Independent Review Boards at the University of Pennsylvania, US Centers for Disease Control and Prevention, Botswana Ministry of Health and University of Botswana has been obtained.

Strengths and limitations of this study

Over the past two decades, the use of molecular methods coupled with classical epidemiological approaches has allowed researchers to detect previously undocumented tuberculosis (TB) transmission events and direct evidence for exogenous reinfection among people with recurrent disease. However, biased and underpowered study designs, as well as the use of lower resolution genotyping techniques, have limited our interpretation of these studies. To our knowledge, the Kopanyo study will be one of the largest prospective population-based molecular epidemiological studies to assess factors associated with TB transmission within entire communities with high rates of HIV and TB.

INTRODUCTION

Tuberculosis (TB) continues to be a major contributor to morbidity and mortality globally, particularly among persons infected with HIV.[1–3] TB in low-resource countries accounts for the majority of the world's burden of the disease and increased global migration threatens TB control efforts in low-burden, high-resource countries.[3,4] The emergence of community-acquired multidrug-resistant TB (MDR-TB) strains emphasises the critical need for improving TB control and prevention strategies.[5–9]

The WHO estimates that Botswana has one of the highest TB notification rates in the world.[10] The high rate of TB is closely linked to the high HIV seroprevalence in Botswana, where ~61% of patients with TB are also infected with HIV (hereafter referred to as TB/HIV).[11] TB remains the leading cause of death among people living with HIV/AIDS in Botswana and worldwide.

The design of effective TB prevention strategies depends on our understanding of TB transmission in communities. However, most interventions to prevent ongoing TB transmission are based on data from the pre-HIV and pre-MDR era, or from areas with low TB and HIV prevalence. It is known that extensive and prolonged exposure to someone with infectious TB disease increases the risk of TB transmission,[12 13] but little is known about the role of casual or brief encounters, particularly among persons with HIV-related immunosuppression.

Genotyping *Mycobacterium Tuberculosis* (Mtb) isolates enables the visualisation of select genetic loci.[14] Isolates that share common genetic loci are referred to as genotypically clustered, a potential marker for recent TB transmission.[14 15] TB genotype clustering, when combined with epidemiological links between patients with TB, has improved our understanding of TB transmission.[14 15] However, many TB molecular epidemiological studies have been limited by some of the following methodological issues:

1. Limited enrolment of people with TB in a given community. Most molecular studies, particularly in high TB incidence settings, have been based on convenience samples drawn from available clinical isolates of Mtb.[16–18] Moreover, passive (self-referral) TB index case identification, or the omission of cases, has the potential to introduce bias into a study,[16 19] because failure to include all cases into a TB genotyping study can lead to unsolvable or biased investigations of transmission chains because a given cluster with missing cases cannot be fully reconstructed.
2. Limited temporospatial data for determination of potential sites of transmission.
3. Limited molecular epidemiological analyses from settings with high HIV prevalence. Therefore, the effect of this key clinical and epidemiological factor on TB transmission dynamics is not known.
4. Limited information regarding contacts and secondary cases exposed to index cases. Retrospective clinical data used in many studies do not include a comprehensive list of people exposed to index cases and physical locations potentially involved in a transmission event.

STUDY GOALS
Kopanyo is the Setswana word meaning 'to link'. Using TB genotyping, geospatial analysis and patient interviews, we plan to link patients together to better understand TB transmission networks and locations of transmission. Our overarching goal is to better understand transmission dynamics of TB within a country characterised by high rates of TB and HIV, while addressing what we perceive to be the major methodological limitations of prior studies.

STUDY METHODS
Setting
We will conduct a population-based survey of all persons diagnosed with TB in Gaborone and Ghanzi, Botswana,

from September 2012 to March 2016. Botswana is a politically and economically stable country with a strong national HIV antiretroviral programme and a universal healthcare system. Half of the population of ~1.8 million people is concentrated in urban areas. Kopanyo study staff will recruit participants from two sites: (1) the capital city, Gaborone (population 231 592) and (2) the rural town of Ghanzi (population 44 100). These two sites differ with regard to HIV prevalence (23% in Gaborone, 16% in Ghanzi). The population of Gaborone is primarily Tswana, the dominant ethnic group in Botswana, whereas the population of Ghanzi is primarily indigenous San (figure 1). Gaborone, the largest urban centre and its surrounding suburbs, contain nearly 20% of Botswana's population. Gaborone has one of the highest prevalences of HIV infection in the world,[20] and ~70% of persons with TB in Gaborone are also infected with HIV.[11] Ghanzi is in rural Western Botswana, where the vast majority of the population lives on large farms and often in crowded houses. There is little or no migration within the area. Over the past two decades, the TB notification rate (defined as the number of new and relapse TB cases notified per year) in Ghanzi has consistently been the highest in the country and one of the highest in the world, at 722 per 100 000 population per year.[11 20]

Sample size calculation and sample selection
To decrease the potential bias introduced by an analysis of only a sample of the population, we aim to enrol the entire population of patients with TB disease diagnosed during a 4-year period. Thus, sample size calculation or selection was not applicable to our design.

Participant recruitment
Management of TB is fully integrated into the primary healthcare system, and services are provided through district hospitals, clinics and health posts. Trained healthcare workers and lay community volunteers provide medical care, social support and directly observed therapy (DOT) for patients with TB. Participants will be recruited from TB clinics, DOT centres and HIV care clinics (details regarding inclusion and exclusion criteria are provided in online supplementary information S1). Participants will be identified and enrolled at the time of their TB diagnosis or during the initiation of TB treatment (figures 2 and 3). Close contacts of TB index cases identified through routine programmatic contact investigation activities will be prospectively enrolled and linked to the index case. This will allow for easy establishment of epidemiological links should they be diagnosed with TB in the future and consequently become study participants.

Determination of coverage completeness
Understanding completeness of case ascertainment in this study is crucial for valid interpretation and generalisation of the results.[16] To do this, we will:
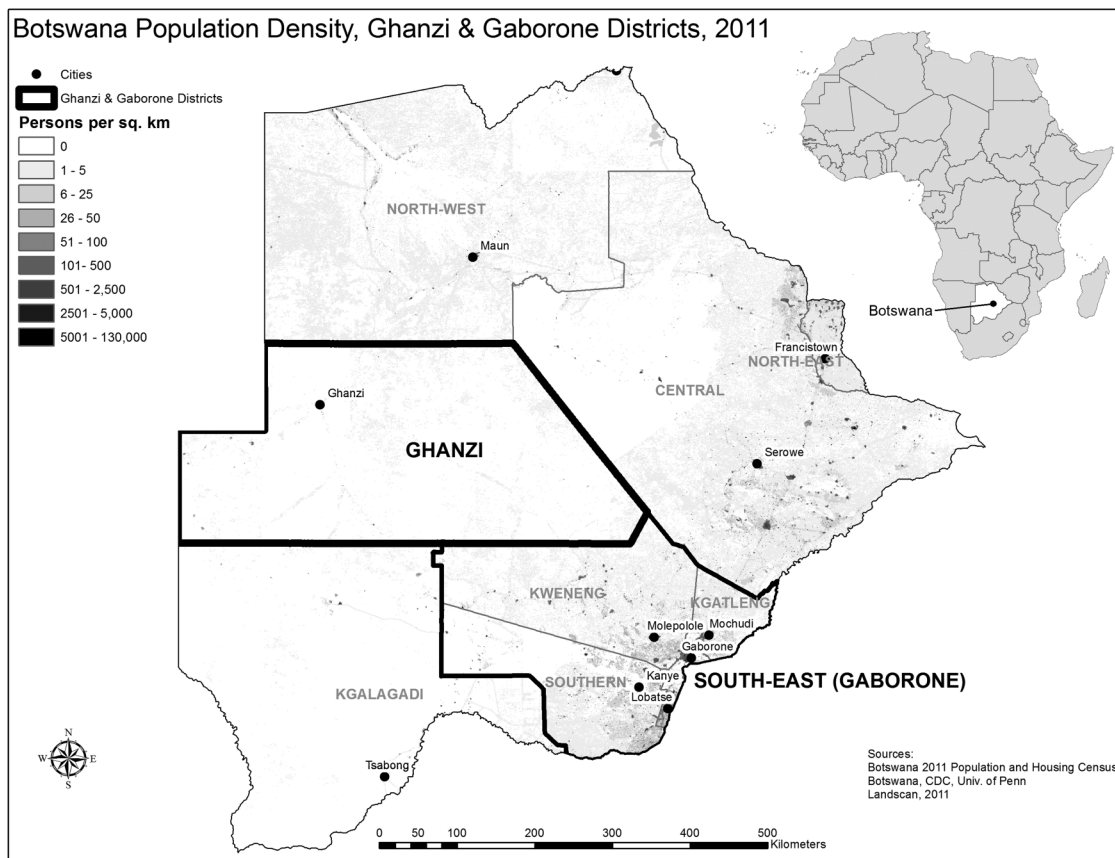
Figure 1 Districts and cities selected for this tuberculosis transmission study.
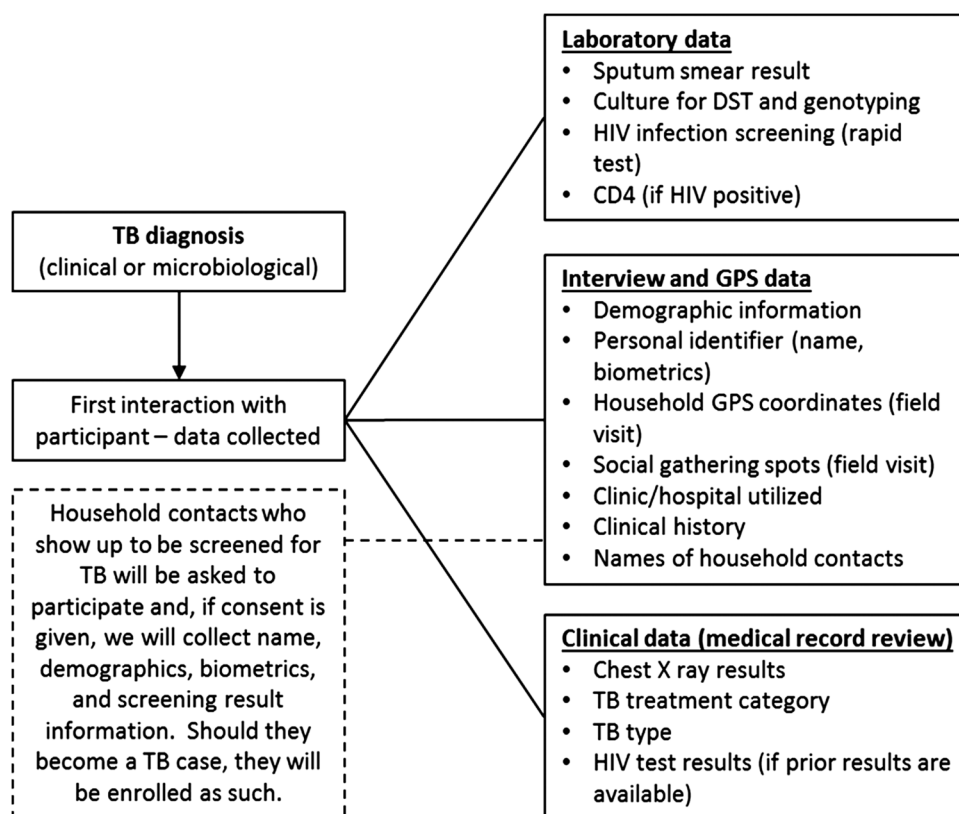

Figure 2 Overview of the study procedures and data collection. GPS, Global Positioning System; TB, tuberculosis.
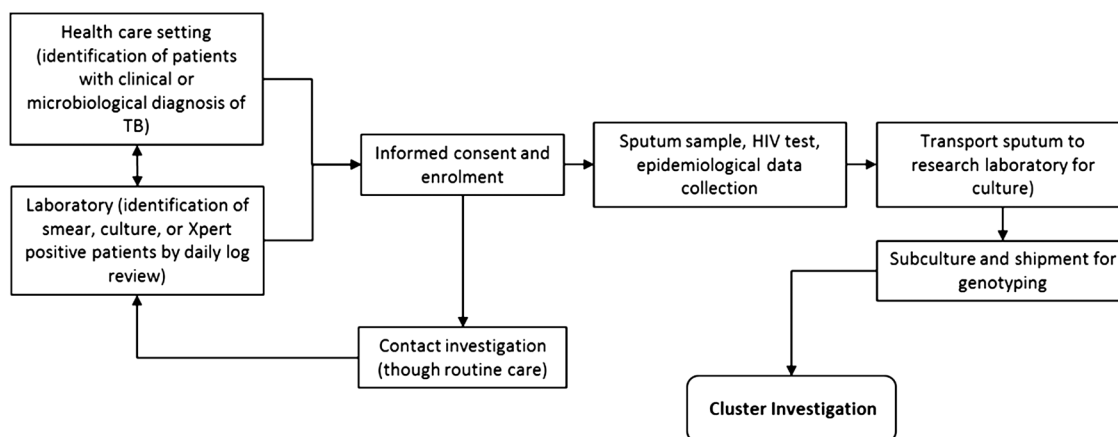
Figure 3 Simplified algorithm for participant identification, enrolment and study procedures. TB, tuberculosis.

A. Access the electronic records and database of the Botswana National TB Program to determine the number of patients diagnosed and treated in each city (TB reporting is mandatory in Botswana).

B. Use the official Botswana National TB Treatment Register to identify patients who are enrolled for treatment, including persons with sputum smear-negative disease. This will be performed through periodic on-site visits to all clinics.

C. Use clinic logs of samples submitted for microscopy or cultures (direct verification through on-site clinic visits), register logs of samples received and peripheral laboratories (through full-time staff based at those facilities) and through periodic assessment of the central data register at the National TB Reference Laboratory.

Data collection

At the time of enrolment, a sputum sample will be obtained from all participants (detailed information regarding data elements is provided in online supplementary information S2). The sputum will be cultured and, if positive, the isolate will be genotyped using 24-locus mycobacterial interspersed repetitive units-variable number of tandem repeats (MIRU-VNTR, see online supplementary information S3 for details).[21] [22] All cultured isolates will also be tested for susceptibility to first-line antituberculosis drugs. Isolates found to be resistant to rifampin or isoniazid will be tested for susceptibility to second-line drugs. Additionally, each participant will be interviewed with a standardised questionnaire to obtain demographic, social, behavioural and clinical information. If needed, additional or confirmatory clinical information will be collected from paper and electronic medical records, and TB registers. HIV testing results will be collected for all enrolled participants. If a patient does not have a documented HIV test result in the 3 months prior to enrolment, HIV testing will be performed according to Botswana HIV/AIDS policy guidelines.[11]

For all participants, the locations of residences (plot numbers), work place, healthcare facilities visited and places of social gathering, such as churches and bars, will be collected and translated into global positioning system coordinates.[23] Approximate dates and frequency for visiting each social venue will also be recorded. If home plot numbers are not available, addresses will be physically located by the study staff and Global Positioning System (GPS) coordinates will be collected using handheld devices (refer to online supplementary information S4 for details regarding geocoding procedures).

Assessing TB transmission

For more than two decades, molecular characterisation of Mtb, or TB genotyping, when combined with epidemiological data, has been used as a proxy for TB transmission. The proportion of cases that belong to a genotype cluster is influenced by the discriminatory power of the genotyping approached used. Currently, the most commonly used genotyping approaches include insertion sequence 6110 (IS6110) restriction fragment length polymorphism (RFLP), spoligotyping and MIRU-VNTR.[14] Relative to IS6110-RFLP, MIRU-VNTR typing does not have to exclude isolates with a low IS6110 copy number, has a faster turnaround time and it is high throughput. Both MIRU-VNTR and spoligotyping allow for easier interlaboratory comparisons.[24] However, MIRU-VNTR has significantly higher discriminatory power than spoligotyping and equal to that of the IS6110-RFLP.[14] Although less discriminatory than whole genome sequencing,[25] the advantage of 24-loci MIRU-VNTR has increased its use for TB molecular epidemiological studies and it is currently considered the gold standard for large population-based TB transmission studies.[18]

In this study, genotype clusters will be defined as two or more TB cases with Mtb isolates with the same 24-locus MIRU-VNTR results. Non-clustered TB cases will have isolates that have non-matching, or unique,

24-locus MIRU-VNTR results. Among genotype clustered cases, patient interview, symptom onset and treatment dates, and geospatial data will be used to further categorise putative transmission events into one or more of the following types[14 23 26 27]:

A. Household transmission—two or more patients with matching genotypes residing in the same house.
B. Workplace transmission—two or more patients with matching genotypes working in the same venue.
C. School transmission—two or more patients with matching genotypes that reported congregating in the same school, college, university or other educational venue.
D. Social setting transmission—two or more patients with matching genotypes that reported congregating in the same social setting (bar, church, etc).
E. Healthcare facility transmission—two or more patients with matching genotypes who reported visiting a healthcare facility (clinic or hospital) at the same time.
F. Residence/neighbourhood transmission—two or more patients with matching genotypes residing in close geographic proximity (eg, spatial scanning methodology or by politically defined areas, such as neighbourhoods or districts).
G. Social network transmission—two or more patients with matching genotypes linked by naming each other as someone they spent time with when they were potentially infectious or when the person they spent time with was potentially infectious.
H. Suspected transmission of unknown origin—two or more patients with matching genotypes who are not members of any of the above defined cluster types.

It is possible that multiple transmission event types can occur for an individual participant. This study aims to quantify the amount of TB attributed to each of the aforementioned transmission event types to guide public health action to interrupt ongoing TB transmission in the community. These categories will not be considered mutually exclusive. The putative infectious period for index cases will be 12 months prior to their TB diagnosis. For select clusters, we will conduct intensive field or cluster investigations to further identify the people exposed to participants during their infectious periods. The resultant lists of persons known to have been exposed will then be prioritised for TB screening and evaluation. These intensive field investigations are intended to lead to early identification of people with undiagnosed TB disease or provision of TB preventive therapy for people exposed and infected. We acknowledge the potential bias that can be introduced by exclusion of prisoners from our study. Exclusion of prisoners might preclude the identification of TB cases resulting from community-based transmission events who were later imprisoned for the remaining duration of the study. However, given that our study extends for a 4-year period, it allows enrolment of most TB cases imprisoned for shorter periods. Further, our procedures include extensive data collection regarding prior time and place of imprisonment, potentially allowing the identification of TB transmission events occurring in those settings.

Cluster investigations

In collaboration with the Government of Botswana, we will select several TB genotype clusters based on public health importance (eg, MDR-TB transmission, healthcare-associated transmission, multijurisdictional transmission or suspected transmission of unknown origin) to intensively investigate. Study participants within selected clusters will be reinterviewed. The reinterviews will focus on social history, behaviours (such as excessive alcohol use, illicit drug use, tobacco smoking, exposure to persons known to have TB, sexual partners), identification of places where they work and spend their leisure time, which healthcare facilities they have visited, foreign travel and incarceration history. Responses during these interviews will be used as prompts for subsequent interviews among other participants in the cluster. Venues and neighbourhoods with high proportions of epidemiologically or genotypically linked cases of TB can be prioritised for intensified case-finding activities.

Geospatial clustering

Spatial scan statistics can be used to identify locations with a high concentration of disease, including TB.[23 27 28] SaTScan V.9.1.0 will be employed to identify geographic areas with a larger-than-expected rate of discrete TB genotype clusters using all other culture-positive cases of TB counted during the study as the background rate.[23 29] In brief, all enrolled participants with TB disease and genotype results will be aggregated by genotype according to residential address. Thus, each genotype will be the unit of analysis and scanned separately. Then, applying a purely spatial perspective, the number of TB transmission events in an area, assuming a Poisson distribution, will be estimated by SaTScan generated circular zones of various sizes up to a maximum radius of 1 km. A log-likelihood ratio will be calculated for each zone in comparison with all possible zones, with the maximum likelihood ratio representing the zone most likely to identify spatial clustering for each genotype. A Monte Carlo simulation with 999 repetitions will be used to determine the distribution of the scan statistic under the null hypothesis of spatial randomness; significant spatial clusters will be chosen with an α of $p < 0.05$. No duplicative case counting will occur.

ANALYTIC PLAN AND REPORTING OF RESULTS

First, we will determine whether each patient belongs to a genotype-defined cluster or not. Since we expect that case detection will be high, we will assume that nonclustered patients had either imported their infection from an area outside the study, or that the infection that

led to disease did not happen recently. The primary analyses to be performed are:

1. Describe the diversity of Mtb genotypes and lineages in Botswana, and after stratification by geographic region (Ghanzi vs Gaborone) and/or HIV status.
2. Determine the proportion of TB cases by genotype cluster types (ie, healthcare facility-based cluster, household cluster, etc).
3. Identify geographic areas with high concentrations of TB disease and/or larger-than-expected rates of clustering.
4. Using cases with unique genotypes as a referent population, determine demographic, behavioural and clinical risk factors associated with clustering.

The results of this study will provide information required to identify currently unknown social networks, locations and geographic areas of recent and perhaps ongoing TB transmission as well as the identification of TB outbreaks in communities with varying prevalence of HIV infection. We will use this information to identify neighbourhoods, healthcare facilities and social networks associated with increased levels of transmission that could benefit from targeted interventions. We will strictly adhere to recently published standards for reporting of molecular epidemiology for infectious diseases.[15 18]

ETHICS AND HUMAN SUBJECT PROTECTION

Written informed consent will be obtained from all participants aged ≥18 years or their parents or legal guardians if <18 years. Assent for participation will be obtained from all minors able to provide it prior to obtaining consent from their legal guardians. Consent and assent for participation in the study will be obtained in the participant's native language.

Anticipated benefits to participants

Benefits to the participant will include a very comprehensive assessment of his/her health needs and/or problems interacting with the healthcare system. The antituberculosis drug susceptibility results of all Mtb isolates will be made available to the treating clinician to inform the clinical management of disease for any patient diagnosed, as part of this study. Current national guidelines indicate first-line drug susceptibility testing to be performed on all culture-positive samples. Mtb culture is routinely indicated only for children (<12 years of age) and for patients considered at risk for MDR TB. All clinical management decisions will be deferred to the treating physician.

Risks and methods to minimise risks

The most important social and psychological risk will be having a breach in confidentiality and/or one's TB or HIV status inadvertently disclosed. Every effort will be made to ensure the confidentiality of the participants enrolled in this study. At the time of the interview, the

participant's name and contact information will be linked with an identification number. Thereafter, the participant will only be identified by this number. No personal names and no contact information will be recorded in study data forms nor in the study database.

Monitoring of adverse incidents

At the close of each interview (including interviews for in-depth cluster investigation), we will assess the occurrence of any adverse events that were the result of participating in the Kopanyo study. We will record and report all adverse incidents in accordance with the Institutional Review Boards (IRBs) that approved the study.

Author affiliations
[1]Department of Radiation Oncology, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[2]Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[3]Botswana-UPenn Partnership, Department of Medicine, University of Pennsylvania Gaborone, Gaborone, Botswana
[4]Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
[5]Division of Infectious Diseases, Department of Medicine, Yale University, New Heaven, Connecticut, USA

Disclaimer 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/the Agency for Toxic Substances and Disease Registry.

Competing interests None declared.

Patient consent Obtained.

Ethics approval Institutional Review Boards (IRBs) of the University of Pennsylvania, the US Centers for Disease Control and Prevention, Botswana Ministry of Health, and The University of Botswana and Princess Marina Hospital in Gaborone, Botswana.

REFERENCES

1. Kamerbeek J, Schouls L, Kolk A, *et al*. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. *J Clin Microbiol* 1997;35:907–14.
2. World Health Organization. *Global tuberculosis control: surveillance, planning, financing*. Geneva: World Health Organization, 2014. http://www.who.int/tb/publications/global_report/en/
3. World Health Organization. *TB/HIV fact sheet 2009*. Geneva: World Health Organization, 2009. http://www.who.int/tb/challenges/hiv/factsheet_hivtb_2009.pdf
4. World Health Organization. *Interim policy on collaborative TB/HIV activities*. Geneva: World Health Organization, 2004; WHO/HTM/TB/2004.330. http://www.who.int/hiv/pub/tb/tbhiv/en/index.html
5. Roeger LI, Feng Z, Castillo-Chavez C. Modeling TB and HIV co-infections. *Math Biosci Eng* 2009;6:815–37.
6. Sharomi O, Podder CN, Gumel AB, *et al*. Mathematical analysis of the transmission dynamics of HIV/TB coinfection in the presence of treatment. *Math Biosci Eng* 2008;5:145–74.
7. Dowdy DW, Chaisson RE, Maartens G, *et al*. Impact of enhanced tuberculosis diagnosis in South Africa: a mathematical model of expanded culture and drug susceptibility testing. *Proc Natl Acad Sci USA* 2008;105:11293–8.
8. Bock NN, Jensen PA, Miller B, *et al*. Tuberculosis infection control in resource-limited settings in the era of expanding HIV care and treatment. *J Infect Dis* 2007;196(Suppl 1):S108–13.
9. Gomes MG, White LJ, Medley GF. Infection, reinfection, and vaccination under suboptimal immune protection: epidemiological perspectives. *J Theor Biol* 2004;228:539–49.
10. World Health Organization. *Global tuberculosis control: surveillance, planning, financing*. Geneva: World Health Organization, 2008. http://www.who.int/tb/publications/global_report/en/
11. Botswana Tuberculosis National Programme. Botswana Tuberculosis National Programme Report, 2008. Botswana Ministry of Health, 2009.
12. Ma N, Zalwango S, Malone LL, *et al*, Tuberculosis Research Unit (TBRU). Clinical and epidemiological characteristics of individuals resistant to M. tuberculosis infection in a longitudinal TB household contact study in Kampala, Uganda. *BMC Infect Dis* 2014;14:352.
13. Sloot R, Schim van der Loeff MF, Kouw PM, *et al*. Risk of tuberculosis after recent exposure. A 10-year follow-up study of contacts in Amsterdam. *Am J Respir Crit Care Med* 2014;190:1044–52.
14. Barnes PF, Cave MD. Molecular epidemiology of tuberculosis. *N Engl J Med* 2003;349:1149–56.
15. Field N, Cohen T, Struelens MJ, *et al*. Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE statement. *Lancet Infect Dis* 2014;14:341–52.
16. Murray M. Sampling bias in the molecular epidemiology of tuberculosis. *Emerging Infect Dis* 2002;8:363–9.
17. Borgdorff MW, van den Hof S, Kalisvaart N, *et al*. Influence of sampling on clustering and associations with risk factors in the molecular epidemiology of tuberculosis. *Am J Epidemiol* 2011;174:243–51.
18. Mears J, Abubakar I, Cohen T, *et al*. Effect of study design and setting on tuberculosis clustering estimates using mycobacterial interspersed repetitive units-variable number tandem repeats (MIRU-VNTR): a systematic review. *BMJ Open* 2015;5:e005636.
19. Plazzotta G, Cohen T, Colijn C. Magnitude and sources of bias in the detection of mixed strain M. tuberculosis infection. *J Theor Biol* 2015;368:67–73.
20. The National AIDS Coordinating Agency (NACA). Botswana AIDS Impact Survey (BAIS) IV. Botswana Ministry of Health, 2014.
21. Supply P, Allix C, Lesjean S, *et al*. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. *J Clin Microbiol* 2006;44:4498–510.
22. Supply P, Lesjean S, Savine E, *et al*. Automated high-throughput genotyping for study of global epidemiology of Mycobacterium tuberculosis based on mycobacterial interspersed repetitive units. *J Clin Microbiol* 2001;39:3563–71.
23. Moonan PK, Ghosh S, Oeltmann JE, *et al*. Using genotyping and geospatial scanning to estimate recent mycobacterium tuberculosis transmission, United States. *Emerging Infect Dis* 2012;18:458–65.
24. Varma-Basil M, Kumar S, Arora J, *et al*. Comparison of spoligotyping, mycobacterial interspersed repetitive units typing and IS6110-RFLP in a study of genotypic diversity of Mycobacterium tuberculosis in Delhi, North India. *Mem Inst Oswaldo Cruz* 2011;106:524–35.
25. Guerra-Assunção JA, Crampin AC, Houben RM, *et al*. Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. *Elife* 2015;4.
26. Middelkoop K, Bekker LG, Mathema B, *et al*. Molecular epidemiology of Mycobacterium tuberculosis in a South African community with high HIV prevalence. *J Infect Dis* 2009;200:1207–11.
27. Moonan PK, Teeter LD, Salcedo K, *et al*. Transmission of multidrug-resistant tuberculosis in the USA: a cross-sectional study. *Lancet Infect Dis* 2013;13:777–84.
28. Moonan PK, Bayona M, Quitugua TN, *et al*. Using GIS technology to identify areas of tuberculosis transmission and incidence. *Int J Health Geogr* 2004;3:23.
29. Touray K, Adetifa IM, Jallow A, *et al*. Spatial analysis of tuberculosis in an urban west African setting: is there evidence of clustering? *Trop Med Int Health* 2010;15:664–72.