**ABSTRACT**

**Introduction** Tuberculosis (TB) transmission is difficult to measure, and its drivers are not well understood. The effectiveness of infection control measures at healthcare clinics and the most appropriate intervention strategies to interrupt transmission are unclear. We propose a novel approach using clinical, environmental and position-tracking data to study the risk of TB transmission at primary care clinics in TB and HIV high burden settings in sub-Saharan Africa.

**Methods and analysis** We describe a novel and rapid study design to assess risk factors for airborne TB transmission at primary care clinics in high-burden settings. The study protocol combines a range of different measurements. We will collect anonymous data on the number of patients, waiting times and patient movements using video sensors. Also, we will collect acoustic sound recordings to determine the frequency and intensity of coughing. Environmental data will include indoor carbon dioxide levels (CO2 in parts per million) and relative humidity. We will also extract routinely collected clinical data from the clinic records. The number of *Mycobacterium tuberculosis* particles in the air will be ascertained from dried filter units using highly sensitive digital droplet PCR. We will calculate rebreathed air volume based on people density and CO2 levels and develop a mathematical model to estimate the risk of TB transmission. The mathematical model can then be used to estimate the effect of possible interventions such as separating patient flows or improving ventilation in reducing transmission. The feasibility of our approach was recently demonstrated in a pilot study in a primary care clinic in Cape Town, South Africa.

**Ethics and dissemination** The study was approved by the University of Cape Town (HREC/REF no. 228/2019), the City of Cape Town (ID-8139) and the Ethics Committee of the Canton Bern (2019-02131), Switzerland. The results will be disseminated in international peer-reviewed journals.

**INTRODUCTION**

Tuberculosis (TB), caused by the bacterium *Mycobacterium tuberculosis* (MtB), remains a major global public health problem, particularly in the context of HIV and drug resistance. Sub-Saharan Africa is one of the most heavily burdened regions globally, although control measures have been in place since the beginning of the 20th century. Over a century of investment in TB control has reduced TB mortality, but effective strategies are urgently needed to reduce TB transmission. Drivers of the TB epidemic in sub-Saharan Africa are HIV-infection and the resulting immunodeficiency (the strongest risk factor), delayed diagnosis and treatment as well as undetected and untreated cases of TB or drug-resistant TB. These factors allow patients with infectious TB to transmit MtB to the community.
There are still many gaps in our knowledge on TB transmission such as the factors and locations associated with the risk of transmission, the effectiveness of infection control measures at clinics in high-burden settings and the most appropriate intervention strategies to interrupt transmission.5–7

For TB transmission to occur, infected individuals must expel Mtb bacilli from their respiratory tract, and an uninfected individual must inhale aerosols containing live bacilli to become infected. Transmission of Mtb is difficult to measure due to the lack of an in vitro test assay. The preferred approach is to measure presumptive transmission resulting in secondary cases as determined by molecular/genomic epidemiology. TB transmission has traditionally been investigated using contact tracing, analyses of geo-temporal clustering and molecular typing.8 9 However, molecular methods require resource-intensive culturing of strains and measure only transmission resulting in secondary cases. Furthermore, contact tracing is difficult to implement in resource-limited settings and may not be an effective control strategy in endemic areas where casual contacts are increasingly recognised as an important contributor to transmission.10 New approaches are therefore urgently needed.

Mtb is carried in airborne particles (called infectious droplets), which are generated when people with TB cough, sneeze or shout.5 Indoor carbon dioxide (CO₂) levels can be used to assess the amount of exhaled air in a room and the amount of rebreathed air.11–13 Humidity is associated with Mtb survival in the air.14 Viable Mtb particles have been captured from contaminated air.11 15–17 We describe a unique study design to assess risk factors for airborne TB transmission in primary care clinics in high TB/HIV-burden settings. The approach combines a range of relevant measurements, including data on patient and infrastructure, movements of patients through the facility, coughing, environmental indoor CO₂ levels and concentration of Mtb particles in the air.

We hypothesise that (1) exposure to Mtb particles at the clinic can be estimated by studying the patient flow and CO₂ levels; (2) the number of individuals present, the rebreathed fraction and the frequency of coughs influence exposure to Mtb; (3) clinics with small and crowded waiting rooms, low ventilation and suboptimal patient separation have an increased risk of TB transmission.

**METHODS**

**Study design**

Longitudinal study with data collection at the levels of the patients, the clinic and the environment. Data collection will take place for over 4 weeks. Figure 1 shows the floor plan of an exemplary primary clinic with the planned study activities, including the recording of patient movements, coughing, CO₂ levels and Mtb particles in the air.

Study setting and study population

The study will take place at several antiretroviral therapy (ART) or primary care clinics in countries participating in the International Epidemiology Database to Evaluate AIDS (IeDEA) in Southern Africa collaboration with a high TB and HIV burden.18 These clinics are located in urban and peri-urban communities with predominantly young and black African residents. TB and HIV are both prevalent in these communities.

**Patient-level variables**

*Video sensor data*

A person-tracking sensor system developed by Xovis (Zollikofen, Switzerland and Cambridge, Massachusetts, USA; see www.xovis.com) will be used to monitor the clinic attendees’ movements. The data will be used to calculate waiting times, the number of people in different locations and the average distance between people, and to identify highly frequented areas. Several sensors will be installed to cover the clinic area, calibrated and validated. Sensors with overlapping ranges will be combined for the seamless coverage of people’s movements over large areas (figure 2). The raw data consist of the person’s height, time, date and the position (x–y coordinates) for each unique individual during the duration of their stay within the clinic. The data are captured every 0.25 s. The raw data are then parsed by a python script to calculate the height, total movement and observation time at different locations, and to visualise hotspots.

The sensors have four levels of privacy. In our study, privacy will be set to level 2, which means that the data are a fully anonymised stream of the coordinates of...
moving dots. These dots will be linked probabilistically to the anonymised clinical data thus excluding any risk of re-identification. Further, the images of tracked individuals taken by the sensors will not be stored (see also the data privacy and security statement by Xovis).19 20

Patient and clinic data
We will extract clinical data from the electronic patient registry for all patients who visited the clinics during the 4 weeks of data collection. The data will include date and time of registration, age, sex, height (used for linkage with video sensor data), HIV status, presumptive TB (‘TB suspects’), mode of TB diagnosis (sputum smear microscopy, Xpert MTB/RIF, mycobacterium culture), TB diagnostic test results, date of TB treatment start and current anti-TB treatment regimen. We will not extract any personal data, such as names or social security numbers. The clinic-level data will include information such as the setting (urban or rural), level of care, the number of patients seen each year, availability of adult and paediatric care, TB control measures (natural ventilation, use of masks, separation of patients with TB or coughers) and the floor plan of the clinics. The data will be collected using the web-based REDCap Data Entry System (www.redcap.org).21

Definitions
Presumptive TB refers to a patient who presents with clinical symptoms or signs suggestive of TB. A bacteriologically confirmed TB case is a person for whom a biological specimen is positive by smear, microscopy, culture or rapid diagnostic such as Xpert MTB/RIF or line probe assay.22

Environmental variables
CO₂, relative humidity and temperature monitoring
We will use CO₂ monitors (Digital CO₂ Monitor Carbon Dioxide Metre XE-2000, XEAST, Guangdong, China) which include the COZIR-A sensor for ambient CO₂ levels of 0%–1% (Gas Sensing Solutions, Cumbernauld, Scotland). The monitors will record indoor CO₂ concentrations (in parts per million (ppm)), temperature and relative humidity at minute intervals. Data from the five monitors are stored and exported in serial digital format.15 16 At each clinic we will install five monitors to cover the most visited spaces and a control area outside the clinic (figure 1). The monitors auto-calibrate over time to a standard minimum value of 400 ppm. This value is close to the average monthly outdoor carbon dioxide concentration measured at Cape Point, South Africa (405.48 ppm in December 2018).23 Monitors will be run for 1 week before the start of the study to allow them to calibrate and to confirm that they delivered comparable data.

Cough monitoring
We will install a microphone (RØDE NT-USB, Sydney, Australia) near the ceiling to continuously record the sounds in the waiting room (figure 1). CoughSense, a deep learning cough detection algorithm based on MXNet, an open-source deep-learning software framework, was developed to classify audio signals as coughing or other sounds.24 The algorithm uses spectrograms extracted from the raw audio for classification. The model was trained and tested using multiple audio recordings obtained through clinical and ambulatory deployments. We will identify audio records with cough sounds and calculate the cough frequency, intensity and duration. The cough data will be linked with the CO₂ data and video sensor data by time and date. Finally, we will probabilistically link the video sensor data (ie, the moving dots) with the clinical data by the height of the patient, time and date of the visit.

Bio-aerosol sampling and molecular detection
We will collect Mtb particles from the air using mobile bio-aerosol sampling devices (Dry Filter Unit (DFU) 1000, Lockheed Martin Integrated Systems, Gaithersburg, Maryland, USA). The DFU 1000 is a portable biological air sampler. Ambient air is drawn through 1 µm polyesterfelt filters at a rate of ~1000 L/min via an electrical blower. The DFU is a useful collection system for low concentration aerosols, capturing particles of 1 µm or larger, and allowing easy access for the retrieval of filters. One DFU will be placed in the waiting room and the other in the TB treatment room (figure 1). Each DFU collects air for 7 hours (two periods of 3.5 hours) every day onto two filters. Due to logistical and human resource limitations, we will be able to change filters only two times per day. Therefore, we decided to run the mobile bio-aerosol sampling devices during the busiest time at the clinic, between 07:00 and 14:00. All other data will be collected from 07:00 to 16:00.

As previously described,17 duplicate filters from each sampling session will be transferred to 50 mL Falcon tubes and vortexed in sterile phosphate buffered saline (with 0.05% Tween 80). Following centrifugation at 3750 rpm for 15 min, filters will be removed and the pellet subjected to DNA extraction. DNA from Mtb cells will...
be extracted using an in-house lysis buffer with subsequent pelleting (centrifugation 13 000 rpm for 10 min) of DNA and resuspension in 50 µl of Tris-EDTA buffer (10 mM Trist, 1 mM EDTA, pH 8.0). Given that droplet digital PCR (ddPCR) is relatively robust against inhibitors, no further DNA purification will be required. The primer/probe combinations and reaction conditions for Mtb-specific ddPCR have been described. Samples with known amounts of purified Mtb DNA (0.01 ng and 0.001 ng) will be included to serve as positive and nuclease-free water as negative control. The data generated from the ddPCR reaction will be analysed via the Umbrella pipeline (25) using wells with a minimum of 10 000 droplets.

### Statistical analyses

We will describe the data and examine associations between sources of data and then use results to parameterise a mathematical model (Table 1). The statistical analysis will quantify the joint association between the clinical and environmental variables and the number of Mtb particles measured by the mobile aerosol sampling. To this end, we will use Poisson regression, with the number of Mtb particles by the period of time as the dependent variable, considered as a proxy measure for the risk of TB transmission. As this variable is measured by periods of 3.5 hours, the other variables will be aggregated over the same periods. We will consider all combinations of independent variables, including second-order interactions, and compare the model using standard model selection methods. We will use variable selection methods in a Bayesian framework, including the deviance information criterion and the leave-one-out information criterion. We will operationalise the variables as follows:

1. **Patient data**: Numbers and characteristics of patients consulting the clinic overall will be summarised using descriptive statistics.
2. **Video sensor data**: Raw data about individuals’ movements will be transformed into waiting times until medical consultation, number of individuals in the different locations, highly frequented areas in the clinic and the average distance between patients (ie, clustering). We will link each tracked individual to the clinical data collected from the clinic’s database (ie, TB/HIV diagnosis), using the order of arrival and time of registration.
3. **CO₂ data**: Measurements of CO₂ concentration at the different locations in the clinic (waiting room, registration desk, TB treatment room), together with estimates in outdoor air, will be used to estimate the proportion

<table>
<thead>
<tr>
<th>Table 1 Description of the measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data source</strong></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>CO₂ monitor</td>
</tr>
<tr>
<td>Relative humidity</td>
</tr>
<tr>
<td>Temperature</td>
</tr>
<tr>
<td>Cough recording</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Mobile aerosol sampling</td>
</tr>
<tr>
<td>Video sensor</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Patient charts</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

dB, decibel; Mtb, *Mycobacterium tuberculosis*; n, number; ppm, parts per million; s, second; TB, tuberculosis.
of air in the different locations that was expired by individuals, the rebreathed fraction. We will use a modified Wells-Riley model appropriate for non-steady states conditions of ventilation and number of individuals to describe and calculate the shared rebreathed air in the different locations in either litre/minute or litre/day and the air exchange (litre/hour per person). Table 2 describes the parameters to calculate the shared rebreathed air and the air exchange. The proportion of rebreathed air (f) will be calculated from the excess CO2 measured indoors, divided by the exhaled CO2 (Ca):

\[ f = \frac{(C-Co)}{Ca} \]  (1)

(see table 2 for definitions of parameters in this and the following equations).

Before we can calculate the rebreathed air volume (RAV) we will need the rebreathed proportion from other people (fo). Therefore, we need to know the number of people (n) present at each time point at the clinic:

\[ fo = f \times \frac{(n-1)}{n} \]  (2)

Finally, we can calculate the RAV for each minute by multiplying fo and the minute respiratory volume (8 L/min, (p)) as:

\[ \text{Rebreathed air volume (RAV)} = (pfo) \]  (3)

In the next step, we aim to calculate the ventilation rate (air exchange). The indoor CO2 generated rate is the product of the average volume of gas exhaled per person (0.13 L/s per person, (V)) and the CO2 concentration of the exhaled air:

\[ \text{Indoor CO2 generation rate in l/s per person (G)} = V \times Ca \]  (4)

The ventilation rate is expressed as:

\[ \text{Ventilation rate in l/s per person (Q)} = \frac{G}{C-Co} \]  (5)

To calculate the air exchange per hour (Equation 6) we need to know the volume of air in a given location.

4. Relative humidity and temperature: We will describe changes in relative humidity and temperature over the day.

5. Cough sounds: Frequency, intensity, duration of recorded coughs per time period.

6. Detection of Mtb in bio-aerrosol sampling: The number of Mtb genome copies present in each sample.

### Mathematical modelling

A mathematical model (figure 3) will integrate all sources of data to model the risk of TB transmission and identify key drivers of transmission. We hypothesise that the number of individuals present, the rebreathed fraction and the frequency of coughs will have the greatest influence on the risk. As we do not observe transmission events, we will use yu, the number of Mtb genomes counted by the bio-aerosol sampling for each 6 hour period u, as the dependent variable. The model will describe yu as a Poisson process with time-dependent intensity λ(t) (also called a Cox process). This intensity can be interpreted as a proxy for the (unobserved) risk of TB transmission. We will model λ(t) using 10 min time periods t within u.

![Figure 3 Model structure. Mtb, Mycobacterium tuberculosis; TB, tuberculosis](image)

Table 2 Description of the variables to calculate the shared rebreathed air volume and air exchange as well as the parameters to construct the mathematical transmission model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Observed CO2 concentration in the indoor air per minute</td>
<td>Observed</td>
</tr>
<tr>
<td>Co</td>
<td>CO2 concentration in the outdoor air per minute</td>
<td>400–420 ppm</td>
</tr>
<tr>
<td>Ca</td>
<td>CO2 concentration in the exhaled air</td>
<td>38 000–40 000 ppm</td>
</tr>
<tr>
<td>f</td>
<td>Proportion of rebreathed air</td>
<td>Equation</td>
</tr>
<tr>
<td>n</td>
<td>Number of people recorded at the location</td>
<td>Observed</td>
</tr>
<tr>
<td>fo</td>
<td>Rebreathed proportion from other people</td>
<td>Equation</td>
</tr>
<tr>
<td>p</td>
<td>Minute respiratory volume</td>
<td>8 L/min</td>
</tr>
<tr>
<td>RAV</td>
<td>Rebreathed air volume</td>
<td>Equation</td>
</tr>
<tr>
<td>V</td>
<td>Average volume of gas exhaled per person</td>
<td>0.13 L/s per person</td>
</tr>
<tr>
<td>G</td>
<td>Indoor CO2 generation rate (L/s per person)</td>
<td>Equation</td>
</tr>
<tr>
<td>Q</td>
<td>Ventilation rate (L/s per person)</td>
<td>Equation</td>
</tr>
<tr>
<td>vol</td>
<td>Volume of the room</td>
<td>Calculated</td>
</tr>
</tbody>
</table>
the number of coughs. The model will be estimated in a Bayesian framework using Stan, a probabilistic programming language.\textsuperscript{23}

We will use several metrics to assess the goodness of fit of the model. We will then add more complexity to the model by integrating other independent variables such as the characteristics of the patients, shared rebreathed air (CO\textsubscript{2} monitors),\textsuperscript{12 13 15 16} patient flow (clustering of individuals, movements of patients) and other potential drivers (intensity/frequency/duration of coughs).\textsuperscript{20 30}

Having developed the model of the risk of TB transmission, we will evaluate the possible effect of interventions such as limiting the number of patients in the waiting room, separating coughers or increasing ventilation. All analyses will be performed in R (V.3.6.0) or Stata (V.15.1, Stata Corporation, Texas, USA).

**Reporting**
The results from this study will be reported following the recommendations of the Strengthening the Reporting of Observational Studies in Epidemiology statement.\textsuperscript{34 35}

**Pilot study**
We conducted a pilot study in Cape Town (South Africa) to examine the feasibility of our approach (figure 1). The pilot study took place at a primary care clinic, which incorporates HIV counselling and testing, and a TB clinic for the diagnosis and management of TB. The clinic is situated in Masiphumelele, a large settlement of formal and semiformal housing in Cape Town, which has previously been described.\textsuperscript{35 36} The clinic is open on workdays from Monday till Friday from 07:00 till 16:00. We studied clinic activities for over 4 weeks on workdays between July 25 and August 23 2019.

Data collection was successful overall, however, we experienced several power cuts during the pilot study. An important lesson learnt is that power banks for laptops and WiFi routers are needed, as well as Uninterruptible Power Supply (UPS) and access to a generator as back-up for the DFU and CO\textsubscript{2} monitors.

**Patient and public involvement**
We discussed the aims and study design with local clinic staff, colleagues at the University of Cape Town and public health specialists early on in the planning phase and developed the specific objectives and data collection procedures in collaboration with them. Patients were not involved in the design, recruitment or conduct of the study. We will make the results of this study available to the participating clinics and the public health authorities.

**DISCUSSION**
TB control is particularly relevant for sub-Saharan Africa, which carries a disproportionately large portion of the global burden of both TB and HIV. There is an urgent need to understand the drivers of TB transmission to reduce TB incidence using new intervention approaches.\textsuperscript{3} A better understanding of transmission, coupled with a rapid test system are likely to contribute to improving TB control in clinic settings. This project will provide new insights into the complex TB transmission framework at a primary care clinic in an endemic setting using clinical data, CO\textsubscript{2} levels, cough analyses, video tracking and Mtb particles sampled from the air. Although the large-scale implementation and evaluation of interventions are beyond the scope of this study, the results will generate new hypotheses and opportunities for public health intervention studies (eg, randomised controlled or cluster-randomised trials).

Importantly, with mathematical models based on real-life data, we can evaluate the likely effect of interventions thus improving intervention studies and inform help with logistical and infrastructural planning of primary care clinics to reduce the transmission risk. Having established feasibility in one clinic in South Africa, we are planning to collect data in several countries in sub-Saharan Africa. The results from this broader study have the potential to inform national and international guidelines to reduce TB transmission at healthcare centres.

**Strengths and limitations**
This is a novel and rapid approach to studying TB transmission combining a wide range of different measurements, which goes beyond the traditional methods such as contact tracing, geo-temporal clustering or molecular genotyping. It will lead to a comprehensive transmission model to measure the effects of various interventions in the clinic setting, paving the way for future studies. Study limitations pertain to the need for stable electrical power and WiFi for over 24 hours for data collection at the primary care clinic, which is often an issue in low-income and middle-income countries. We will address this problem by using UPS power stabilisers with access to generator back-up power.

**ETHICS AND DISSEMINATION**
The University of Cape Town Faculty of Health Sciences Human Research Ethics Committee (HREC/REF: 228/2019), the City of Cape Town (Project ID 8139) and the Ethics Committee of the Canton of Bern (2019-02131), Switzerland approved the pilot study. Expansion of the study will include other clinics in the greater Cape Town metropolitan area, but also clinics in Zambia, Zimbabwe or Malawi, including urban and rural sites to ensure representative and generalisable results. Separate ethical approval will be sought from the relevant ethics committees or institutional review boards.

The results will be disseminated in international peer-reviewed journals and presented at national and international conferences.

**Author affiliations**
\textsuperscript{1}Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland
Acknowledgements We would like to thank Mbali Mohlamayone for setting up the bio-aerosol sample collection system, Zeneat Hoonen and Ronnelt Seldon for developing the laboratory protocols. We are also grateful to the clinical manager, the staff members at the clinic and all patients who contributed to this study. Finally, we would like to thank the City of Cape Town for the support they gave us for the use of one of their clinic facilities. Research findings and recommendations do not represent an official view of the City of Cape Town.

Contributors Overall concept: KZ, CM, MB, RW, ME, LF. Medical informatics: SB, JR. Cough sound concept: XL, MS, CM, RW. Video sensor concept: KZ, JR, SB, ME, LF. Patient data concept: CM, KM, RW. Laboratory work concept: CM, ASK, KM, DW, RW. Modelling work concept: KZ, JR, SB, ME, LF. Pilot study coordination: CM, KM, RW, KZ, MB, ME, LF. Wrote the first draft of the paper, which was reviewed by all authors and revised on the basis of the comments received by co-authors. All authors approved the final version of the manuscript.

Funding The roll-out of this study is supported by a grant from the Swiss National Science Foundation (grant no. CR53X-1_190781). KZ and LF are supported by the National Institute of Allergy and Infectious Diseases (NAII) through grant no. U01-AI059924-05, ME by special project funding (grant no. 174281) from the Swiss National Science Foundation.

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of changes were made. See: http://creativecommons.org/licenses/by/4.0/

ORCID iDs Kathrin Zürcher http://orcid.org/0000-0002-7915-3194 Marie Ballif http://orcid.org/0000-0003-3133-3011 Anastasia Sideris Koch http://orcid.org/0000-0002-5897-4196 Keren Middelkoop http://orcid.org/0000-0001-9922-4263 Matthias Egger http://orcid.org/0000-0001-7462-5132 Lukas Fenner http://orcid.org/0000-0003-3309-4835

REFERENCES


