1   **TITLE PAGE**   Journal: *J Infect Dis* (major article)

2

3   Estimating tuberculosis transmission risks in a primary care clinic in
4   South Africa: modelling of environmental and clinical data

5

6   Kathrin Zürcher [1], Julien Riou [1], Carl Morrow [2,3], Marie Ballif [1], Anastasia Koch [4], Simon
7   Bertschinger [1,5], Digby F. Warner [2,4], Keren Middelkoop [2,3], Robin Wood [2,3],
8   Matthias Egger [1,6,7], Lukas Fenner [1]*

9

10   1) Institute of Social and Preventive Medicine, University of Bern, Switzerland
11   2) Institute of Infectious Disease and Molecular Medicine, University of Cape Town, University
12   of Cape Town, South Africa
13   3) Desmond Tutu HIV Centre, Department of Medicine, University of Cape Town, University
14   of Cape Town, South Africa
15   4) SAMRC/NHLS/UCT Molecular Mycobacteriology Research Unit & DST/NRF Centre of
16   Excellence for Biomedical TB Research, Department of Pathology, Faculty of Health Sciences,
17   University of Cape Town, South Africa
18   5) Medical Informatics, Bern University of Applied Sciences, Bern, Switzerland
19   6) Centre for Infectious Disease Epidemiology & Research, School of Public Health & Family
20   Medicine, University of Cape Town, South Africa
21   7) Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

22

23   *** Correspondence:**
24   Professor Lukas Fenner, MD
25   Institute of Social and Preventive Medicine, University of Bern (ISPM)
26   Mittelstrasse 43, 3012 Bern, Switzerland
27   Phone: +41 31 631 35 18; email: lukas.fenner@ispm.unibe.ch

28

29   **Key words:** Mycobacterium tuberculosis, Tuberculosis, Transmission, CO_2

30

31   **Running head:** Estimating tuberculosis transmission

32

33   **Summary:** By modelling epidemiological, clinical and environmental data at a
34   primary care clinic in South Africa, we identified young adults and relative humidity as
35   potentially important factors for tuberculosis transmission. This approach should be
36   used to estimate transmission and evaluate interventions

37

43

44   **Conflict of interests:** All authors declare that they have no conflicts of interest.

45

46   **Word count, inserts, and additional files:**

47   Abstract 200 words, main text 3,497 words
48   Inserts: 2 tables, 5 figures, 50 references
49   Supplemental information: 3 tables, 2 figures

ABSTRACT

Background: Congregate settings, such as healthcare clinics, may play an essential role in *Mycobacterium tuberculosis (Mtb)* transmission. Using patient and environmental data, we studied transmission at a primary care clinic in South Africa.

Methods: We collected patient movements, cough frequency, and clinical data, and measured indoor carbon dioxide (CO_2) levels, relative humidity, and *Mtb* genomes in the air. We used negative binomial regression model to investigate associations.

Results: We analyzed 978 unique patients who contributed 14,795 data points. The median patient age was 33 years ([IQR] 26-41), 757 (77.4%) were female. Overall, median CO_2 levels were 564ppm (IQR 495-646), highest in the morning. Median number of coughs/day was 466 (368-503), overall median *Mtb*-DNA-copies/μL/day 4.2 (IQR 1.2-9.5). We found an increased presence of *Mtb*-DNA in the air of 32% (95% credible interval 7%-63%) per 100 additional young adults (aged 15-29) and 1% (0%-2%) more *Mtb* DNA per 10% increase of relative humidity. Estimated cumulative transmission risks for patients attending the clinic monthly for at least 1 hour range between 9%-29%.

Conclusions: We identified young adults and relative humidity as potentially important factors for transmission risks in healthcare clinics. Our approach should be used to detect transmission and evaluate infection control interventions.

Keywords: transmission, carbon dioxide, modelling, primary care clinic, tuberculosis, biosampling, intervention, humidity, cough, infection control

BACKGROUND

Caused by *Mycobacterium tuberculosis (Mtb)*, tuberculosis (TB) remains a global public health problem and one of the deadliest infectious diseases worldwide. Understanding TB transmission at primary care clinics is of particular public health importance in high TB/HIV burden settings, such as South Africa, and in places with a risk of transmission of multidrug-resistant (MDR) and extensively drug-resistant *Mtb* in clinics [1]. Sub-Saharan Africa is one of the most heavily burdened TB regions. *Mtb* is transmitted by droplet aerosols generated when people infected with TB cough, sneeze, shout, speak, or breathe [2, 3]. For TB transmission to occur, an infected person must expel *Mtb* bacilli from their respiratory tract, and an uninfected person must inhale *Mtb* bacilli-containing aerosols. Although TB control measures have been in place since the beginning of the 20th century, *Mtb* transmission is difficult to measure. Currently, the preferred approach is to measure presumptive transmission by determining secondary cases through molecular and genomic epidemiology [4, 5]. This approach is expensive and not feasible in all settings. Therefore, new approaches to measure TB transmission are needed.

This study piloted a novel approach to estimate transmission risk based on environmental measurements and patient data at a South African primary care clinic. We measured indoor carbon dioxide (CO_2) levels, which indicate the proportion of exhaled, rebreathed air in a room [6-8]. We also captured aerosol droplets containing viable *Mtb* bacilli from contaminated air [6, 9-11] and measured humidity, which is associated with the survival of airborne *Mtb* [12]. We obtained clinical data on patient diagnoses, visit frequency from electronic medical records, and cough counts in waiting areas, and we tracked people's movements through the primary care clinic. Combining the different data allowed us to assess risk factors for airborne *Mtb* transmission in a high TB/HIV burden setting [13].

METHODS

Study design

We previously described the study design in detail [13]. We collected environmental data such as indoor CO_2 levels, relative humidity, frequency of coughs, and presence of *Mtb* DNA in the air, as well as patient data over four weeks from July 25 to August 23, 2019 at a primary care clinic in Cape Town, South Africa.

Study setting

The primary care clinic offers both TB and HIV services and reproductive health and childhood immunization services, Monday to Friday, from 7 am to 4 pm. The clinic is situated within a large settlement of formal and semiformal housing where both TB and HIV are highly prevalent [14, 15]. We delineated three areas within the clinic: the registration area, the waiting room, and the TB treatment room (Figure 1). Further we defined three time periods: morning (7am-10:30am), midday (10:30am-2pm), and afternoon (2-4pm).

Patient data

Tracking data

We used an anonymized movement tracking system (Xovis; Zollikofen, Switzerland) to monitor people's movements (staff members, patients, and other visitors) throughout the clinic (Figure S1). The resulting date- and time-stamped movement data consisted of a person's height, their position recorded as x-y coordinates, and a unique signal for each person while in the clinic (Table S1) [13]. If individuals went out of a sensor's range and subsequently returned, they could contribute multiple signals. Thus, the number of captured signals is higher than the number of unique persons. While in the waiting room, close contacts were defined as other persons passing within a radius of 1 meter.

Clinical data

We extracted clinical data from the electronic patient registry for all patients who visited the clinic during the study period. These data included the date and time of arrival for the clinic visit and when the patient passed by the registration desk and

135   their age, sex, TB diagnostic results, and date of TB treatment start (if applicable).

136

137   ***Environmental data***

138   *CO_2 monitoring*

139   Three CO_2 monitors (Digital CO_2 Monitor Carbon Dioxide Meter XE-2000, XEAST;

140   Guangdong, China) covered the clinic's most crowded spaces. The CO2 monitors

141   were installed in the waiting room, by the registration desk and in the TB treatment

142   room (Figure S1). The monitors recorded indoor CO_2 concentrations (in parts per

143   million [ppm]), temperature, and relative humidity at one-minute intervals (Table S1)

144   [9, 10]. Monitors were regularly auto-calibrated [16].

145

146   *Cough monitoring*

147   We installed a microphone (RØDE NT-USB; Sydney, Australia) near the clinic's

148   waiting room ceiling to continuously record sounds (Figure S1). We used a cough

149   detection algorithm based on MXNET's open-source deep learning software

150   framework to classify audio signals as coughing or other sounds (CoughSense;

151   Seattle, Washington, USA) [17]. In addition, we developed a cough counting

152   algorithm to test for cough in the recorded coughs automatically. We trained, tested,

153   and validated the algorithm model using multiple audio recordings obtained during

154   the study period (Table S1).

155

156   *Bioaerosol sampling and molecular testing*

157   Air was sampled using mobile bioaerosol sampling devices (Dry Filter Unit (DFU) 1000,

158   Lockheed Martin Integrated Systems, Gaithersburg, Maryland, USA). The number of

159   Mtb genomes was ascertained from dried filters using highly sensitive droplet digital

160   polymerase chain reaction (PCR) [11]. We placed one bioaerosol sampling device in

161   the clinic's waiting room and the other in the TB treatment room (Figure S1). During

162   data collection, each bioaerosol sampling device collected air through two filters over

163   two time periods (morning and midday). Each day both devices collected air for about

164   3.5   hours,   totaling   approximately   7   hours   per   day   (Table   S1).

165

166

Linkage of people tracking data with clinical patient data

We applied several criteria to link the movement tracking system data with the clinical data. We included people who (1) passed by registration and (2) had a height of at least 140cm according to the tracking data to exclude children; we included clinical visits of patients aged 15 years and older from the clinical data. We then combined the datasets using the time-stamp of when a person was recorded by the tracking system in the registration area and the time a patient was registered in the electronic patient registry. We identified 2,355 adult patients (≥15 years) whose visits were recorded in the clinic's electronic patient registry from the clinical data. After linking with the movement tracking data, we included 978 unique adult patients, resulting in 1,135 clinical visits.

Statistical analyses and modelling

We used descriptive statistics for the environmental and patient data obtained in the different clinic areas. We calculated the number of individuals in the three clinic areas, the time spent in the waiting room, and the number of contacts an individual had during this time period, thus enabling the identification of highly frequented areas.

As previously described, we calculated the rebreathed air volume and ventilation rates from CO_2 and clinic presence [10, 13]. We summarized the coughs per minute in the waiting room over the three time periods [18, 19]. We described the number of Mtb genome copies present in each filter by time period and clinic area.

We used a negative binomial regression model to assess clinical and environmental factors associated with the number of Mtb genome copies measured in the waiting room air (Table 1). Using the mean for the environmental data (CO_2, and relative humidity), the total number of people present in the clinic, and the total number of coughs, we aggregated the data by the minute to the exact time period of the bioaerosol sampling devices. The model will be estimated with MCMC in a Bayesian framework using Stan, a probabilistic programming language [20]. The results are unadjusted and adjusted risk ratio per unit increase with 95% credible intervals. The model was adjusted for sex, age group (15-29 years, 30-44 years, 45-59 years, and >60), relative humidity, indoor CO_2, and frequency of cough (Table 1).

199   Finally, we calculated the risk of *Mtb* transmission per hour during the day and

200   per each clinical visit as previously described [21]. Briefly, we used the modified

201   Wells-Riley formula considering the work of Rudnick Milton on non-steady state

202   situations to estimate the annual risk of TB transmission, taking into account the re-

203   breathed air volume, time at risk, the infectious quanta of contagion, and the number

204   of people occupying the confined space [6, 8]. The parameters we used to calculate

205   the risk of transmission are given in Table S2.

206   All analyses were performed in R (version 3.6.0) [22].

207

208   **Ethics statement**

209   The University of Cape Town Faculty of Health Sciences Human Research Ethics

210   Committee (HREC/REF: 228/2019); the City of Cape Town (Project ID: 8139), South

211   Africa; and the Ethics Committee of the Canton of Bern (2019-02131), Switzerland

212   approved the study.

213

214

RESULTS

Patient data

Movement of patients

The movement tracking system captured 14,795 unique data points corresponding to people in the clinic between July 25 and August 23, 2019. The median number of unique signals per day was 706 (interquartile range [IQR] 622-803). Most individuals visited the clinic in the morning when the highest density of individuals was found in the waiting room (Figure 2A). The median time spent in the waiting room was 24 minutes (IQR 23-27 minutes).

Patient characteristics

After data linkage, we included 978 unique patients. Their median age was 33 years ([IQR] 26-41), and 757 patients (77.4%) were female. Overall, 171 (17.5%) had a TB diagnosis at some time, among whom 153 (90.6%) had a clinical history of TB, and 16 (9.4%) had active pulmonary TB and were potentially infectious at the time of clinic visit (Table S3). The density of potentially infectious TB patients and all other people was highest in the waiting room (Figure 1). These potentially infectious TB patients were more likely HIV-positive men who had three or more visits during the four weeks (Table S3).

Time in the waiting room

The median time a patient spent in the waiting room was 41 minutes (IQR 17-85 minutes) with a median of 62 (IQR 16-173) close contacts (within a radius of 1 meter). There were no significant differences between potentially infectious TB patients and all other patients in the time spent in the waiting room (41 vs 43) or in the number of contacts (67 vs 66).

Coughing

The median number of coughs per day in the waiting room was 466 (IQR 368-503). The total number of coughs was higher at midday than in the morning (495 vs. 421, Table 2). The median length of coughs was 0.67 seconds (IQR 0.47-0.91).

Environmental data

CO₂ levels

The median CO_2 level in the clinic was 564ppm (IQR 495-646). It was higher in the morning than at midday and in the afternoon (639 vs 568.7 vs 477ppm). We measured the highest CO_2 levels in the waiting room (Table 2, Figure 2B). The share of time people experienced CO_2 levels at/above 1,000ppm of the opening hours was 4.7%.

Rebreathed air volume

The overall median rebreathed air volume was 46.5 L/day (IQR 22.7-74.8), and it decreased over the day (Table 2, Figure 2C). The rebreathed air volume was highest in the waiting room compared to the registration area and TB treatment room (68.1 vs 42.3 vs 9.5 L/day). The ventilation rate in the waiting room was at 12.2 L/h (recommended ventilation rate: 6.0 [23]).

Relative humidity

The overall median relative humidity was 60.6% (53.6-65.8%). It was higher in the morning compared to midday and afternoon (66.2% vs 58.9% vs 54.1%). The relative humidity was highest in the TB treatment room followed by the registration area and the waiting room (63.6% vs 60.9% vs 57.3%) (Table 2, Figure 2D).

Presence of *Mtb DNA copies/μL in the air*

The overall median number of *Mtb* DNA copies/μL per day was 4.2 (1.2-9.5). The median *Mtb* DNA copies/μL throughout the day was slightly higher in the waiting room than in the TB treatment room (Table 2, Figure S2A), and higher in the afternoon than in the morning.

Risk factors for potential transmission

In the univariate analysis, we found an increased presence of *Mtb* DNA copies in the air of 15% (95% credible interval 3-32%) per 100 incremental young adults (aged 15-29 years) visiting the clinic. No other variables were associated with an increase presence of *Mtb* DNA copies in the air (Table 1). In the multivariate analysis, we found an increased presence of *Mtb* DNA copies in the air of 32% (95% credible

282 interval 7%-63%) per 100 incremental young adults (aged 15-29 years) visiting the

283 clinic. For a 5% incremental increase of relative humidity, 1% (95% credible interval

284 0%-2%) more *Mtb* DNA copies were in the air (Table 1). Figure 3 shows the

285 standardized risk ratio per one standard deviation with the 95% credible interval.

286

287 **Risk of infection**

288 We modelled different scenarios using the observed TB prevalence at the clinic and

289 the estimated TB prevalence of 737 per 100 000 people for South Africa with varying

290 infectious quanta [24] (5.5 and 8.2 infectious quanta per hour, Table S2). The

291 observed TB prevalence at the clinic suggested that the risk of *Mtb* transmission

292 during the day was about 3% per hour using 5.5 infectious quanta per hour. It was

293 about 6% per hour using 8.2 infectious quanta per hour (Figure 4A). The risk of

294 infection was lower when using TB prevalence estimates by WHO (Figure 4B).

295 To put this in perspective, a patient coming each month to the clinic for 1 hour (12

296 visits per year) would have a cumulative risk of *Mtb* transmission ranging from 9% to

297 29% depending on the scenario (Figure 5). The cumulative risk was higher for

298 observed TB prevalence at the clinic compared to the TB prevalence estimated by

299 WHO. In an extreme scenario assuming a weekly visit to the clinic of 1 hour (52 visits

300 per year), a patient would have a cumulative risk ranging from 33% to 78%,

301 depending on the scenario.

302

303

DISCUSSION

At this South African primary care clinic, an increased risk of *Mtb* transmission was associated with the presence of young adults and higher room humidity. We estimated the risk of transmission during a clinic visit of one hour to be 3% to 6%, increasing to 9% to 29% for patients making regular monthly visits. Our study suggests that multiple environmental measures and clinical data can be used to assess indoor ventilation quality and evaluate airborne disease transmission control measures in primary care and similar settings.

Our study observed more copies of *Mtb* DNA in the air when young adults visited the clinic than when clinic visitors were older. Several factors might account for this. Behavioral and social contact patterns differ by age, and they might play a role in the risk of *Mtb* transmission [25, 26]. Young index cases (<40 years) have been shown to have more close contacts and contacts with all age groups than older index cases who have fewer contacts (and mainly within their own age group) [26]. And as adolescents and young adults' transition from child to adult health services, they face specific age-related challenges accessing appropriate healthcare [27, 28]. These challenges might result in delayed HIV or TB diagnoses and treatments. A study from Cape Town, South Africa, showed that TB notification was highest among young adults. Among those aged 25-45 years, 63% were HIV-associated TB patients. The study also showed that TB notification rates among HIV-negative people peaked between 20-24 years and a second peak between 45-54 years [29]. We observed that increasing relative humidity was associated with increased copies of *Mtb* DNA in the air, only a few other studies have investigated this relationship. Relative humidity was shown to play an important role in the presence of *Mtb* genome copies in the air [30], and our finding is also in line with results from a more recent study which showed that *Mtb* DNA copies were more likely to be found in health facilities when the relative humidity was above 65% [12].

Studies of different settings have reported that healthcare clinics may be drivers of *Mtb* transmission [31-35]. In low and middle income countries, resource constrained care clinics are often crowded with people sitting close together on benches or standing in passageways. In these clinics, the waiting times are often long and the ventilation is poor. These kinds of conditions favour *Mtb* transmission [9,

13

337 36, 37]. Because of these conditions, exposure to *Mtb* might be prolonged. People

338 with undiagnosed TB or delayed TB diagnoses pose risks of *Mtb* transmission to

339 other individuals at the clinic. In addition, those diagnosed with TB who continue to

340 receive care at a clinic may pose a risk to uninfected people and reinfection in people

341 with a *Mtb* infection [36]. Furthermore, HIV coinfection plays a major role as disease

342 progression is faster in HIV-positive compared to HIV-negative individuals [31, 32].

343 Therefore, it is important to screen people regularly for TB symptoms. Infection

344 control measures are needed, such as improved ventilation and, for presumptive TB

345 cases or anyone who is coughing, wearing masks. Because of the COVID-19

346 pandemic, wearing masks is likely an easy and familiar intervention to implement.

347 Finally, detection of *Mtb* DNA by ppPCR has been shown to be more sensitive than

348 detection by aerosol using traditional culture techniques [11].

349   High indoor CO_2 levels (above 1000 ppm) are indicators of poor ventilation.

350 We found CO_2 levels above 1000 ppm, mainly in the morning in the waiting room

351 area of the clinic. Levels in the TB treatment room were kept lower through measures

352 to minimize occupancy and keeping the doors and windows open to allow ventilation.

353 Since we know that crowed waiting rooms are the most likely infectious place, we

354 focused on this room as well as the TB treatment room where presumptive TB

355 patients are screened, diagnosed and treated there [38]. Previous studies have

356 measured CO_2 levels at different locations and combined these environmental data

357 with social interaction data to model the risk of *Mtb* transmission [9, 10, 21]. The

358 highest annual risk for *Mtb* transmission in another Southern African setting was

359 found in prisons, with descending lower risks for persons in schools, riding public

360 transport, and social halls [21]. These findings complement other studies of high-

361 burden settings, which found that only a small proportion of *Mtb* transmission occurs

362 between household members [39-41]. Using the observed prevalence at the clinic,

363 we found that the risk for *Mtb* transmission during the day was 3 to 5% per hour. A

364 modelling study showed that the annual risk of *Mtb* infection in the waiting room at a

365 clinic with closed windows and doors ranged from 23-34% for chronic patients with

366 monthly visits and from 2.2-3.4% per patient visit [38]. Further, they showed that with

367 good ventilation, the risk of *Mtb* infection was reduced 50-fold.

368   The mathematical models showed that the duration and frequency of clinic

369 visits increased the risk of *Mtb* transmission. However, this could be addressed

370　effectively by relatively simple infection control interventions: improved ventilation

371　through opening windows and decreased room presence, which resulted in very low

372　rebreathed air volume for the room. In settings where airborne transmission is

373　possible, both *Mtb* bacilli and the SARS-CoV-2 virus are transmissible via aerosols

374　[42-45]. In the COVID-19 pandemic, primary care clinics have implemented infection

375　control measures such as increased hand hygiene and physical distancing, and all

376　attendees and clinical staff members are wearing face masks. These infection control

377　measures would likely also decrease the risk of *Mtb* infection and other airborne

378　transmitted diseases at healthcare clinics.

379　　　　The collection of environmental data had several limitations. The video sensor

380　system assigned a new ID whenever a seated person stood up. Therefore we had

381　challenges in tracking people, and we cannot exclude incorrect assignments in these

382　cases. Furthermore, the bioaerosol sampling devices collected data over about 3.5

383　hours, whereas the other data were collected by the minute. By aggregating these

384　data, we lost some information, which may explain why we did not find an association

385　between *Mtb* counts and CO_2 levels. The highly sensitive ddPCR assay we applied

386　detects *Mtb* genome DNA but does not distinguish between viable, *Mtb* bacilli

387　causing infections and dead or noninfectious bacilli and DNA fragments. Moreover,

388　the assay could conceivably be detecting DNA fragments present in the clinic over a

389　long time, and efforts in our laboratory are underway to develop improved analysis

390　and assay approaches that can address this. These caveats notwithstanding, we

391　used a novel and rapid system to study transmission, which goes beyond traditional

392　methods such as molecular genotyping. However, we did not measure actual

393　transmission events, but rather estimated the risk of transmission events using a

394　range of clinical and environmental data, including detection of *Mtb* DNA in the air.

395　　　　Our approach to assessing *Mtb* transmission risks using various

396　environmental and clinical data is novel. It identified young adults and relative

397　humidity as potentially important factors in TB transmission in these settings. A global

398　study using the WHO TB notification database that showed about 17% of all new TB

399　cases were among people aged 10-24 years [46]. Therefore, TB research and public

400　health interventions should have increased focus on young adult health [29, 46, 47].

401　However, we should not only understand the drivers of transmission, but also

402　evaluate interventions [48]. Our multiple measures approach can be used in health

403   care clinics and other congregate settings to evaluate interventions to halt

404   transmission, including the evaluation of infection control measures such as

405   improved room ventilation, increased hand hygiene, or wearing of masks.

406

407

AUTHOR CONTRIBUTIONS

KZ, CM, RW, ME, and LF wrote the concept. KZ and LF wrote the first draft of the paper, which was reviewed by all authors and revised based on the comments received by all coauthors. KZ, CM, KM, and RW coordinated data collection. CM, AK, KM, DW, and RW were involved in laboratory work, and they were involved in extracting the clinical data from the electronic registry. SB did the medical informatics and cough extraction AI. JR and KZ completed the statistical analyses. All authors approved the final version of the manuscript.

COMPETING INTERESTS

All authors declare that they have no conflicts of interest.

REFERENCES

1. Shah NS, Auld SC, Brust JCM, et al. Transmission of Extensively Drug-Resistant Tuberculosis in South Africa. New Engl J Med **2017**; 376:243-53.
2. Rieder HL. Epidemiologic basis of tuberculosis control. In: International Union Against Tuberculosis and Lung Disease, ed. 1 ed. Paris: International Union Against Tuberculosis and Lung Disease, **1999**:1-162.
3. Patterson B, Bryden W, Call C, et al. Cough-independent production of viable Mycobacterium tuberculosis in bioaerosol. Tuberculosis (Edinb) **2021**; 126:102038.
4. Walker TM, Ip CL, Harrell RH, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect Dis **2012**.
5. Cook VJ, Shah L, Gardy J, Bourgeois AC. Recommendations on modern contact investigation methods for enhancing tuberculosis control. Int J Tuberc Lung Dis **2012**; 16:297-305.
6. Andrews JR, Morrow C, Walensky RP, Wood R. Integrating social contact and environmental data in evaluating tuberculosis transmission in a South African township. J Infect Dis **2014**.
7. Myatt TA, Johnston SL, Rudnick S, Milton DK. Airborne rhinovirus detection and effect of ultraviolet irradiation on detection by a semi-nested RT-PCR assay. BMC Public Health **2003**; 3:5.
8. Rudnick SN, Milton DK. Risk of indoor airborne infection transmission estimated from carbon dioxide concentration. Indoor Air **2003**; 13:237-45.
9. Wood R, Morrow C, Ginsberg S, et al. Quantification of shared air: a social and environmental determinant of airborne disease transmission. PLoS One **2014**; 9:e106622.
10. Richardson ET, Morrow CD, Kalil DB, Bekker LG, Wood R. Shared air: a renewed focus on ventilation for the prevention of tuberculosis transmission. PLoS One **2014**; 9:e96334.
11. Patterson B, Morrow C, Singh V, et al. Detection of Mycobacterium tuberculosis bacilli in bio-aerosols from untreated TB patients. Gates Open Res **2017**; 1:11.
12. Sornboot J, Aekplakorn W, Ramasoota P, Bualert S, Tumwasorn S, Jiamjarasrangsi W. Detection of airborne Mycobacterium tuberculosis complex in high-risk areas of health care facilities in Thailand. Int J Tuberc Lung Dis **2019**; 23:465-73.
13. Zürcher K, Morrow C, Riou J, et al. Novel approach to estimate tuberculosis transmission in primary care clinics in sub-Saharan Africa: protocol of a prospective study. BMJ Open **2020**; 10:e036214.
14. Wood R, Middelkoop K, Myer L, et al. Undiagnosed tuberculosis in a community with high HIV prevalence: implications for tuberculosis control. Am J Respir Crit Care Med **2007**; 175:87-93.
15. Middelkoop K, Bekker LG, Myer L, et al. Antiretroviral therapy and TB notification rates in a high HIV prevalence South African community. J Acquir Immune Defic Syndr **2011**; 56:263-9.
16. National Oceanic & Atmospheric Administration. Earth System Research Laboratory; Global Monitoring Division. Available at: https://www.esrl.noaa.gov/gmd/dv/data/index.php?site=CPT&parameter_name=Carbon%2BDioxide,%20accessed%2012%20November%202019. Accessed 12.11.2019 2019.

489 17. Chen T, Li M, Li Y, et al. Mxnet: A flexible and efficient machine learning library
490 for heterogeneous distributed systems. arXiv preprint arXiv:151201274 **2015**.
491 18. Birring SS, Fleming T, Matos S, Raj AA, Evans DH, Pavord ID. The Leicester
492 Cough Monitor: preliminary validation of an automated cough detection system in
493 chronic cough. Eur Respir J **2008**; 31:1013-8.
494 19. Botha GHR, Theron G, Warren RM, et al. Detection of tuberculosis by automatic
495 cough sound analysis. Physiol Meas **2018**; 39:045005.
496 20. Carpenter J, Crutchley P, Zilca RD, et al. Correction: Seeing the "Big" Picture: Big
497 Data Methods for Exploring Relationships Between Usage, Language, and Outcome
498 in Internet Intervention Data. J Med Internet Res **2017**; 19:e347.
499 21. Hella J, Morrow C, Mhimbira F, et al. Tuberculosis transmission in public
500 locations in Tanzania: A novel approach to studying airborne disease transmission. J
501 Infect **2017**; 75:191-7.
502 22. Goodrich B, Gabry J, Ali I, Brilleman S. rstanarm: Bayesian applied regression
503 modeling via Stan. R package version 2.21.1 https://mc-stan.org/rstanarm.
504 23. Centers for Disease Control. Guidelines for Environmental Infection Control in
505 Health-Care Facilities. Available at:
506 https://www.cdc.gov/infectioncontrol/guidelines/environmental/appendix/air.html.
507 Accessed 20 April 2021 2021.
508 24. National Department of Health. The First National TB Prevalence Survey - South
509 Africa 2018. Available at: https://www.knowledgehub.org.za/elibrary/first-national-tb-
510 prevalence-survey-south-africa-2018. Accessed 01/01/2021 2021.
511 25. Borgdorff MW, Nagelkerke NJD, de Haas PEW, van Soolingen D. Transmission
512 of Mycobacterium tuberculosis Depending on the Age and Sex of Source Cases. Am
513 J Epidemiol **2001**; 154:934-43.
514 26. van Geuns HA, Meijer J, Styblo K. Results of contact examination in Rotterdam,
515 1967-1969. Bull Int Union Tuberc **1975**; 50:107-21.
516 27. Dahourou DL, Gautier-Lafaye C, Teasdale CA, et al. Transition from paediatric to
517 adult care of adolescents living with HIV in sub-Saharan Africa: challenges, youth-
518 friendly models, and outcomes. J Int AIDS Soc **2017**; 20:21528.
519 28. Zanoni BC, Sibaya T, Cairns C, Lammert S, Haberer JE. Higher retention and
520 viral suppression with adolescent-focused HIV clinic in South Africa. PLoS One **2017**;
521 12:e0190260.
522 29. Wood R, Lawn SD, Caldwell J, Kaplan R, Middelkoop K, Bekker LG. Burden of
523 new and recurrent tuberculosis in a major South African city stratified by age and
524 HIV-status. PLoS One **2011**; 6:e25098.
525 30. Wells WF. On air-borne infection. Study II. Droplets and droplet nuclei. Am J Hyg
526 **1934**; 20:611-18.
527 31. Pearson ML, Jereb JA, Frieden TR, et al. Nosocomial transmission of multidrug-
528 resistant Mycobacterium tuberculosis. A risk to patients and health care workers. Ann
529 Intern Med **1992**; 117:191-6.
530 32. Centers for Disease Control. Nosocomial transmission of multidrug-resistant
531 tuberculosis among HIV-infected persons--Florida and New York, 1988-1991.
532 MMWR Morb Mortal Wkly Rep **1991**; 40:585-91.
533 33. Andrews JR, Shah NS, Weissman D, Moll AP, Friedland G, Gandhi NR.
534 Predictors of multidrug- and extensively drug-resistant tuberculosis in a high HIV
535 prevalence community. PLoS One **2010**; 5:e15735.
536 34. Dheda K, Limberis JD, Pietersen E, et al. Outcomes, infectiousness, and
537 transmission dynamics of patients with extensively drug-resistant tuberculosis and

home-discharged patients with programmatically incurable tuberculosis: a prospective cohort study. Lancet Respir Med **2017**; 5:269-81.

35. Basu S, Andrews JR, Poolman EM, et al. Prevention of nosocomial transmission of extensively drug-resistant tuberculosis in rural South African district hospitals: an epidemiological modelling study. Lancet **2007**; 370:1500-7.

36. Mzembe T, McLean E, Khan PY, et al. Risk of Mycobacterium tuberculosis transmission in an antiretroviral therapy clinic. AIDS **2018**; 32:2417-21.

37. Wood R, Morrow C, Barry CE, 3rd, et al. Real-Time Investigation of Tuberculosis Transmission: Developing the Respiratory Aerosol Sampling Chamber (RASC). PLoS One **2016**; 11:e0146658.

38. Taylor JG, Yates TA, Mthethwa M, Tanser F, Abubakar I, Altamirano H. Measuring ventilation and modelling M. tuberculosis transmission in indoor congregate settings, rural KwaZulu-Natal. Int J Tuberc Lung Dis **2016**; 20:1155-61.

39. Middelkoop K, Mathema B, Myer L, et al. Transmission of tuberculosis in a South African community with a high prevalence of HIV infection. J Infect Dis **2015**; 211:53-61.

40. Verver S, Warren RM, Munch Z, et al. Proportion of tuberculosis transmission that takes place in households in a high-incidence area. Lancet **2004**; 363:212-4.

41. Glynn JR, Guerra-Assunção JA, Houben RMGJ, et al. Whole Genome Sequencing Shows a Low Proportion of Tuberculosis Disease Is Attributable to Known Close Contacts in Rural Malawi. PLoS One **2015**; 10:e0132840.

42. Binder RA, Alarja NA, Robie ER, et al. Environmental and Aerosolized Severe Acute Respiratory Syndrome Coronavirus 2 Among Hospitalized Coronavirus Disease 2019 Patients. J Infect Dis **2020**; 222:1798-806.

43. Ong SWX, Tan YK, Chia PY, et al. Air, Surface Environmental, and Personal Protective Equipment Contamination by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) From a Symptomatic Patient. JAMA **2020**; 323:1610-2.

44. Liu Y, Ning Z, Chen Y, et al. Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. Nature **2020**; 582:557-60.

45. Lane MA, Brownsword EA, Babiker A, et al. Bioaerosol sampling for SARS-CoV-2 in a referral center with critically ill COVID-19 patients March-May 2020. Clin Infect Dis **2021**.

46. Snow KJ, Sismanidis C, Denholm J, Sawyer Susan M, Graham SM. The incidence of tuberculosis among adolescents and young adults: a global estimate. Europ Respir J **2018**; 51:1702352.

47. Middelkoop K, Bekker LG, Liang H, et al. Force of tuberculosis infection among adolescents in a high HIV and TB prevalence community: a cross-sectional observation study. BMC Infect Dis **2011**; 11:156.

48. Dowdy DW, Grant AD, Dheda K, Nardell E, Fielding K, Moore DAJ. Designing and Evaluating Interventions to Halt the Transmission of Tuberculosis. J Infect Dis **2017**; 216:S654-s61.

49. Escombe AR, Moore DAJ, Gilman RH, et al. The Infectiousness of Tuberculosis Patients Coinfected with HIV. PLoS Medicine **2008**; 5:e188.

50. Andrews JR, Morrow C, Wood R. Modeling the role of public transportation in sustaining tuberculosis transmission in South Africa. Am J Epidemiol **2013**; 177:556-61.

TABLES AND FIGURES

587   **Table 1:** Factors associated with transmission risk (indicated by *Mtb* genome copies

588   in the air) presented as risk ratio per 100 incremental persons with the corresponding

589   95% credible interval.

| Variable	Unit	Unadjusted risk ratio, 95% credible interval	Adjusted risk ratio, 95% credible interval
Sex			
Female	Per 100 incremental persons	1.04 (0.97-1.11)	0.92 (0.75-1.15)
Age groups			
15-29	Per 100 incremental persons	1.15 (1.03-1.32)	1.32 (1.07-1.63)
30-44	Per 100 incremental persons	0.96 (0.87-1.08)	0.80 (0.61-1.03)
45-59	Per 100 incremental persons	1.08 (0.85-1.43)	1.35 (0.89-2.09)
>60	Per 100 incremental persons	0.86 (0.40-2.15)	1.19 (0.22-6.57)
Environmental factors			
Average RH per day	Per 10% incremental increase RH	0.99 (0.99-1.01)	1.01 (1.00-1.02)
Average CO_2 per day	Per 10 incremental increase ppm	0.0 (0.0-0.47)	0.04 (0.0-1.08)
Sum coughs	Per 100 incremental coughs	0.99 (0.83-1.20)	1.11 (0.89-1.34)

590   Abbreviation: RH, relative humidity, ppm, parts per million

591

592

593 **Table 2:** Environmental data collected at a primary care clinic in Cape Town, South

594 Africa, overall and by location.

|	Overall median, (IQR)	Registration area median, (IQR)	Waiting room median, (IQR)	TB treatment room median, (IQR)
CO_2 levels				
Per day	564.3 (495-646)	564 (494-686)	646 (531-765)	471 (447-516)
Time				
Morning	639 (551-753.7)	669.5 (551-823)	747 (623-852)	497 (460-572)
Midday	568.7 (514.5-624)	570 (502-659)	655 (564-742)	468 (445-504)
Afternoon	477 (455.7-517.3)	487 (461-524)	491 (458-565)	453 (437-477)
Rebreathed air volume in litres/day				
Per day	46.5 (22.7-74.8)	42.3 (22.1-74.0)	68.1 (33.0-102.2)	9.5 (0-18.6)
Time				
Morning	46.5 (44.2-98.6)	67.2 (38.1-107.6)	97.1 (61.9-127.2)	13.4 (0-25.0)
Midday	47.7 (30.6-70.3)	39.1 (22.1-63.1)	69.5 (42.1-93.3)	9.5 (0-15.9)
Afternoon	11.6 (0-24.5)	16.7 (0-26.5)	15.9 (0-33.4)	5.8 (0-13.1)
Relative humidity				
Per day	60.6 (53.6-65.8)	60.9 (54.1-66.2)	57.3 (49.9-63.3)	63.6 (57.2-67.7)
Time				
Morning	66.2 (61.6-68.6)	66.7 (62.9-69.4)	63.9 (58.9-66.6)	67.4 (63.5-70.7)
Midday	58.9 (52.1-62.9)	58.8 (52.5-63)	55.2 (48.3-59.9)	62.1 (55.8-66.3)
Afternoon	54.1 (48.2-59.4)	53.7 (48.2-59.4)	50.3 (45.1-56.3)	58.7 (51.2-62.8)
Number of coughs				
Per day	466 (368-503)	-	466 (368-503)	-
Time				
Morning	421 (350.5-487.5)	-	421 (350.5-487.5)	-
Midday	495 (392-514)	-	495 (392-514)	-
Number of *Mtb* DNA copies/μL				
Per day	4.2 (1.2-9.5)	-	4.2 (1.8-9.4)	4.7(0.5-9.5)
Number of observations	*79*	*-*	*38*	*41*
Time				
Morning	3.6 (0.4-7.4)	-	4.2 (1.4-8.0)	2.1 (0.30-6.3)
Number of observations	*39*	*-*	*19*	*20*
Midday	5.6 (2.2-11.8)	-	6.2 (1.8-10.9)	5.5 (2.7-12.2)
Number of observations	*40*	*-*	*19*	*21*

595 Abbreviation: CO_2, carbon dioxide; IQR, interquartile range; *Mtb, Mycobacterium tuberculosis*

596

597

598

599 **Figure 1:** Density of (**A**) potentially infectious TB patients (defined as being

600 diagnosed with TB [clinically or bacteriologically confirmed] one week before the

601 study started or up to three months after the end of the study period [July 18-

602 November 25, 2019]); and (**B**) all other people visiting the primary care clinic over the

603 study period. Data from the movement tracking system were linked with clinical data

604 from the electronic register.
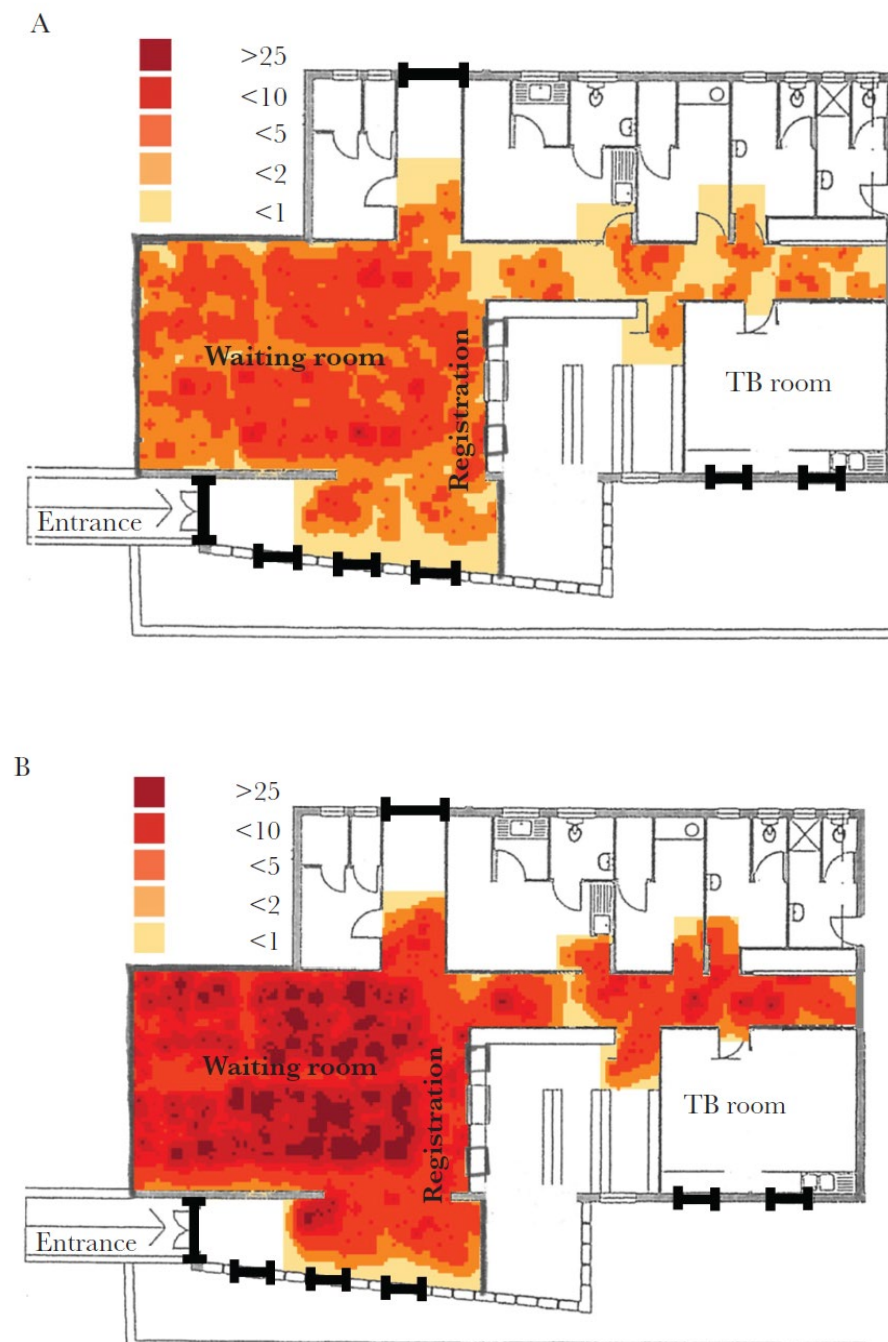
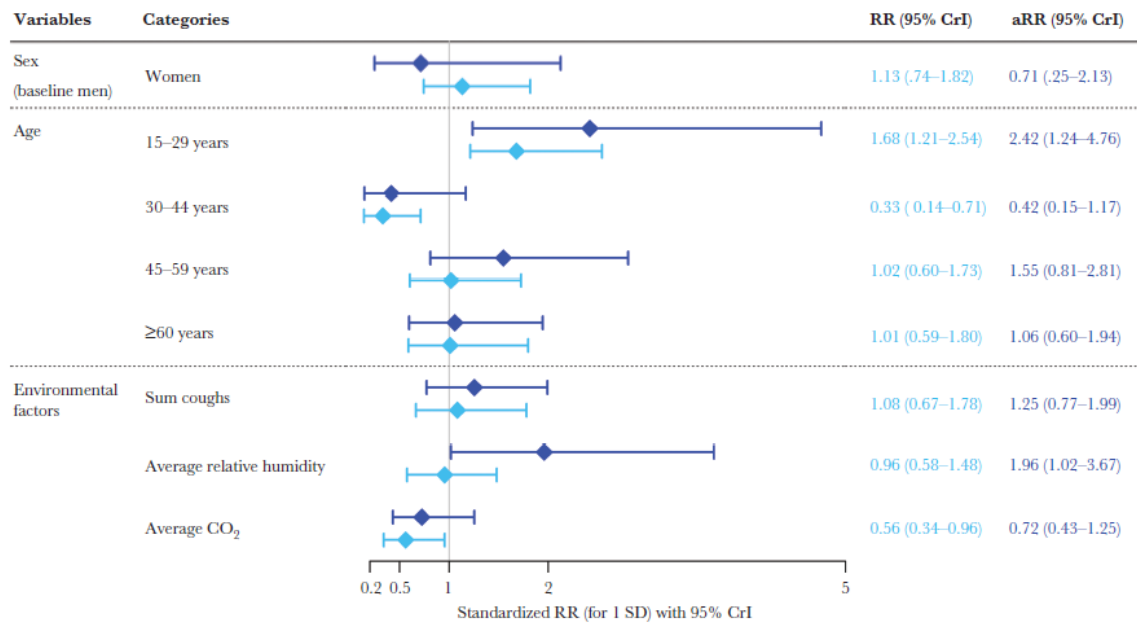607 **Figure 2: Environmental data collected at the primary care clinic.** Average clinic

608 presence, CO_2, rebreathed air volume (RAV), and relative humidity, over time and by

609 location. The solid line is the mean with pale fill the recorded values from the

610 minimum to the maximum.

611 Abbreviation: Co2, carbon dioxide; RAV, rebreathed air volume; ppm, parts per million

612

613
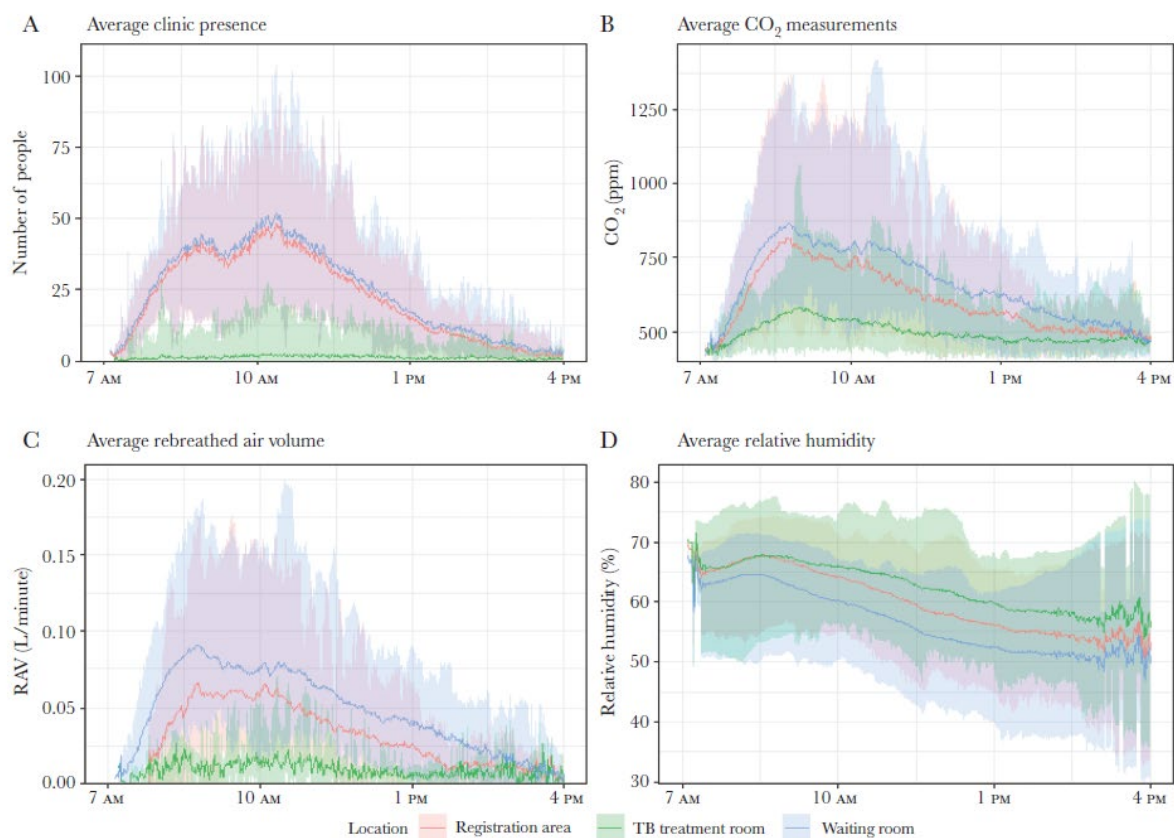
614   **Figure 3:** Patient and environmental factors associated with *Mtb* genome copies in

615   the air, presented as standardized risk ratio from a multivariate analysis.

616   Abbreviation: CrI, credible interval; RAV, rebreathed air volume

Variables	Categories		RR (95% CrI)	aRR (95% CrI)
Sex (baseline men)	Women		1.13 (.74–1.82)	0.71 (.25–2.13)
Age	15–29 years		1.68 (1.21–2.54)	2.42 (1.24–4.76)
	30–44 years		0.33 (0.14–0.71)	0.42 (0.15–1.17)
	45–59 years		1.02 (0.60–1.73)	1.55 (0.81–2.81)
	≥60 years		1.01 (0.59–1.80)	1.06 (0.60–1.94)
Environmental factors	Sum coughs		1.08 (0.67–1.78)	1.25 (0.77–1.99)
	Average relative humidity		0.96 (0.58–1.48)	1.96 (1.02–3.67)
	Average CO_2		0.56 (0.34–0.96)	0.72 (0.43–1.25)

Standardized RR (for 1 SD) with 95% CrI

618

25

619 **Figure 4: The risk of TB infection during a day at the primary care clinic**

620 **estimated based on a mathematical transmission model (21).** Panel A shows the

621 risk of infection based on the observed TB prevalence at the clinic. Panel B shows

622 the risk based on the TB prevalence in the general population as estimated by WHO.

623 The solid line is the mean with pale fill the recorded values from the minimum to the

624 maximum. Estimations for two different definitions of the infectious quanta are shown.

625 The parameters and assumptions for the transmission model are described in Table

626 S2 or described in Hella et al. [21].

628
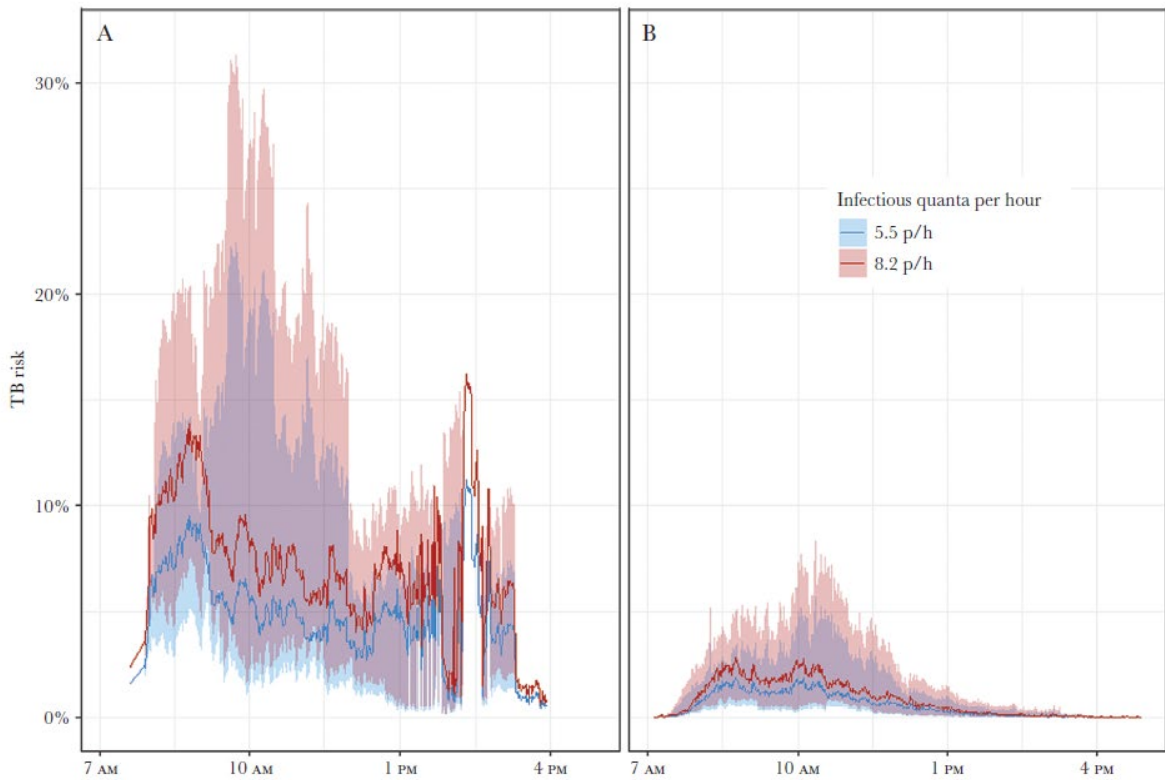
629 **Figure 5: Cumulative risk of TB infection according to the time spent at the**

630 **primary care clinic.** The solid line present the observed prevalence and the hashed

631 line the estimated TB prevalence by WHO.