INT J TUBERC LUNG DIS 23(12):1286–1292

http://dx.doi.org/10.5588/ijtld.18.0819

Evidence for supercoughers in an analysis of six tuberculosis cohorts from China, Peru, The Gambia and Uganda

A. Handel,[1] L. Martinez,[1] J. N. Sekandi,[1] S. E. Bellan,[1] L. Zhu,[2] C. Chen,[2] Q. Liu,[2] S. Donkor,[3] J. Sutherland,[3] P. C. Hill,[4] R. H. Gilman,[5,6] L. Grandjean,[5,7,8] C. C. Whalen[1]

[1]Department of Epidemiology and Biostatistics, and Health Informatics Institute and Center for the Ecology of Infectious Diseases, University of Georgia, Athens, GA, USA; [2]Department of Chronic Communicable Disease, Center for Disease Control and Prevention of Jiangsu Province, Nanjing, People's Republic of China; [3]Vaccines and Immunity, Medical Research Council Unit The Gambia at the London School of Hygiene & Tropical Medicine, Banjul, The Gambia; [4]Centre for International Health, School of Medicine, University of Otago, Dunedin, New Zealand; [5]Universidad Peruana Cayetano Heredia, Lima, Peru; [6]Johns Hopkins Bloomberg School of Public Health, Department of International Health, Baltimore, MD, USA; [7]Wellcome Centre for Clinical Tropical Medicine, Imperial College London, London, [8]Institute of Child Health, University College London, London, UK;

SUMMARY

BACKGROUND: It is very difficult to observe tuberculosis (TB) transmission chains and thus, identify superspreaders. We investigate cough duration as a proxy measure of transmission to assess the presence of potential TB superspreaders.

DESIGN: We analyzed six studies from China, Peru, The Gambia and Uganda, and determined the distribution of cough duration and compared it with several theoretical distributions. To determine factors associated with cough duration, we used linear regression and boosted regression trees to examine the predictive power of patient, clinical and environmental characteristics.

RESULTS: We found within-study heterogeneity in cough duration and strong similarities across studies. Approximately 20% of patients contributed 50% of total cough days, and around 50% of patients contributed 80% of total cough days. The cough duration distribution suggested an initially increasing, and subsequently, decreasing hazard of diagnosis. While some of the exposure variables showed statistically significant associations with cough duration, none of them had a strong effect. Multivariate analyses of different model types did not produce a model that had good predictive power.

CONCLUSION: We found consistent evidence for the presence of supercoughers, but no characteristics predictive of such individuals.

KEY WORDS: cough duration; superspreading; TB

FOR MANY INFECTIOUS DISEASES, the distribution of secondary infections caused by an infectious host is heavily skewed and not well characterized by the average. Most infectious individuals infect none or few susceptible contacts, whereas a small proportion of individuals infect the majority of secondary cases. The latter are commonly referred to as superspreaders.[1–3]

In the case of tuberculosis (TB), there is some evidence for superspreaders.[4–8] However, it is difficult to directly measure the number of secondary infections for TB due to the long and variable latent period among contacts and the inability to determine the strain causing their latent infection. Instead, it might be possible to reconstruct this quantity through indirect means. The transmission potential of an individual can be quantified by an individual's reproductive number, *R*. *R* is the product of the rate at which an infectious person has contact with susceptible individuals, the per contact probability that transmission occurs, and the total duration of the infectious period. Heterogeneity in all of these components is likely among TB patients.[9,10] In the present study, we focus on the duration of the infectious period. We used the duration of cough before treatment (hereafter, referred to as cough duration) as a proxy for the duration of the infectious period.[11]

Previous studies have found that cough duration can be considerable for TB patients (see e.g., the reviews of[12–17]). We aimed to analyze the cough duration distribution in greater detail. Using data from six different TB patient populations, we examined the distribution of cough duration to determine the extent of heterogeneity among patients within and between studies and the factors associated with cough duration.

Correspondence to: Andreas Handel, Department of Epidemiology and Biostatistics, University of Georgia, 101 Buck Rd, Athens, GA 30606, USA. e-mail: ahandel@uga.edu

Article submitted 12 December 2018. Final version accepted 21 June 2019.

Table 1 Summary description of studies included in analysis

Study name	China*	Peru	The Gambia	U-Cohsonet	U-Kawempe	U-Steps
Country	China*	Peru	The Gambia	Uganda	Uganda	Uganda
Year	2014	2010–2013	2002–2004	2012–2015	1995–2006	2014
Setting	Hospital	Hospital	Clinics	Hospital	Hospital	Clinics
Design	Retrospective cohort	Household contact	Household contact	Household contact	Household contact	Retrospective cohort
Diagnostic method	Smear or culture	Smear or culture	Smear or culture	Smear or culture	Smear or culture	Smear or culture
Participants, *n*	178	603	316	68	539	264
Predictor variables, *n*	13	14	5	23	27	20
Age, years, median (range)	41.0 (18–85)	NA	28.5 (14–86)	28.0 (17–59)	29.0 (0–67)	30.0 (18–70)
Smokers, %	34.3	15.6	34.2	13.2	21.5	12.9
HIV-positive, %	NA	5.3	NA	4.4	46.6	31.4
Smear-positive, %	NA	90.7	100.0	94.1	97.0	NA

* Cases were confirmed using either culture or smear; however, data on culture or smear outcomes were not recorded.
NA = not available; HIV = human immunodeficiency virus.

METHODS

Study setting, population and data collection

Data from six studies conducted in different populations, calendar years, and settings were used for this analysis. The studies spanned a 20-year period from 1995 to 2016. Three of the studies were conducted in Kampala, Uganda, one in the Banjul region of The Gambia, one in Lima, Peru, and one in Nanjing, China.

While some of the studies were different in design, the data used for this analysis were obtained in essentially the same way in each study. In household contact studies, we only included index cases diagnosed in a hospital or clinic and then queried about their cough duration status. The patients in the retrospective study designs were also recruited from hospitals or clinics. The method of patient recruitment is therefore the same. Patients that had been diagnosed with TB and started on treatment were enrolled and asked a set of questions, including the duration of their cough prior to survey administration. Questions were asked at diagnosis or within 2 weeks of treatment initiation for each study.

All studies used microbiologic evidence for diagnosis of TB, either sputum microscopy or mycobacterial culture, or both. For more details, see [18,19] for the 'U-Kawempe' study,[20] for the 'U-Steps' study,[21] for the 'The Gambia' study,[22] for the 'Peru' study and [23] for the 'China' study. The 'U-Cohsonet' study has not been published yet; the design and data collection on index cases are identical to the 'U-Kawempe' study.

All study participants provided informed consent, and all studies were approved by their respective ethics and Institutional Review Board committees.

Data analysis

Data were cleaned and pre-processed as described in the Supplementary Data. The size of the final data set for each study used for analysis is shown in Table 1, a list of all predictors included for the analysis of each study are shown in Supplementary Table S1. Codebooks and data cleaning R scripts for each data set are provided as Supplementary Data and contain explanations as comments to the code describing how data was cleaned and which predictors were removed and the reasons for doing so.

We used a descriptive analysis to investigate the distribution of cough duration for each study population. To illustrate the heterogeneity of cough duration, we sorted patients based on cough duration and plotted the cumulative cough duration for all patients as a function of the cumulative number of patients.

We investigated the potential underlying generating process of the observed cough duration by fitting exponential, gamma, Weibull, and log-logistic distributions to the data. An exponential distribution corresponds to a constant hazard of receiving treatment, independent of the time since onset of cough. A Weibull distribution has an increasing hazard as cough duration increases. A Gamma distribution is the sum of exponential distributions and can represent multiple care-seeking steps before treatment,[20] each with constant hazard. A log-logistic distribution allows an initially increasing and subsequently decreasing hazard of receiving treatment. Model fit was compared using Akaike's Information Criterion.

For the analysis of correlations between outcome and predictors, we fitted the continuous variable of cough duration (measured in days) to all predictors available for a given study. As the studies were conducted by different teams in different locations, it is likely that even predictor variables which claim to measure the same quantity might not be fully comparable. We therefore did not feel confident that an analysis of data across studies was justified and decided not to pursue it. Instead, we analyzed each study independently. We first performed bivariate analyses to check for correlations. We then tested the predictive power of the different variables with two types of statistical models. First, we built linear models for each study. A genetic algorithm was used to perform efficient subset selection of models with

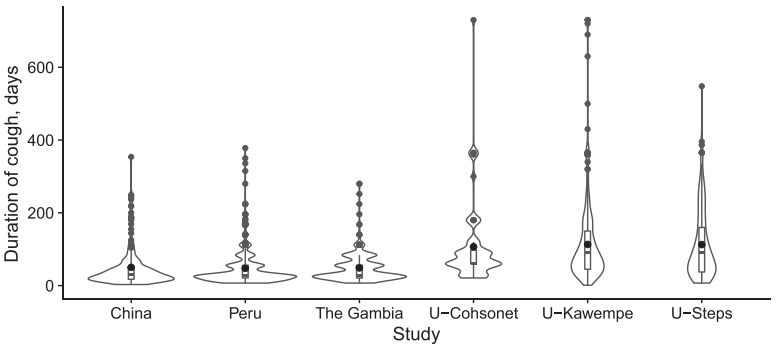

Figure 1 Cough time distribution for each study. Distribution of data are indicated by box and violin plots. Black dots indicate mean of cough duration for each study. The shape for the data from Peru and the Gambia studies stem from the fact that the data were reported in weeks. Conversion to days kept this discretisation, with values being multiples of 7.

different combinations of predictors included or excluded. Performance of each sub-model was assessed using 10-fold cross validation, 10 times re-sampled.[24] Models were chosen based on minimization of the mean squared error (MSE) on the hold-out set in the cross-validation approach. A model with a certain number of predictors was considered superior to another model with different predictors if the mean across all 100 samples of the MSE on the hold-out set was smaller. Model performance was reported using the coefficient of determination (R^2). We also fitted each data set to a boosted regression tree model, a machine-learning approach that typically provides strong predictive performance.[24] This model was fitted and tuned using the same approach as for the linear model. All analyses were performed in *R* software (R Computing, Vienna, Austria). The statistical and machine learning analysis was done using the mlr package.[25] All analysis code is available in the Supplementary Data.

RESULTS

Summary of studies

The six study populations are summarized in Table 1. All patients were recruited in the same manner. Once they presented to either a clinic or hospital and were confirmed to have TB, they were approached and recruited to enter the study. The average age was similar among the African study samples, and approximately 10 years older for the study from China; however, the reported age range was similar. The prevalence of smoking varied by study group, with the highest proportion among patients in The Gambia and China. Co-infection with human immunodeficiency virus (HIV) differed across groups, with the highest prevalence among two study populations in Uganda.

Distribution of cough duration

The distribution of cough duration was skewed for each of the study populations (Figure 1, Supplementary Figure S1), with the mean above the median (Supplementary Table S2). Most patients reported a cough duration less than the group mean, while a few patients had a much longer cough duration. This heterogeneity in cough duration is also seen when the cumulative amount of cough time is plotted as a function of the cumulative number of patients, after patients are sorted by descending order of cough duration (Figure 2).

For each of the studies, the heterogeneity in cough duration was such that approximately 20% of patients contributed 50% of cough time, and 50% of patients contributed 80% of cough time (indicated by the dashed lines). Thus, a small number of patients contributed a large amount to the total cough time. The median cough duration in the three Ugandan studies was between 2 and 3 months, while the mean was close to 4 months. The median in the remaining three studies was around 1 month and the mean around 1.5 months. The maximum duration for the three Uganda studies was 1.5–2 years, while it was around $\leq$1 year for the other three studies.

While there is strong heterogeneity in the duration of cough within each study, and absolute values differ between studies, there is a remarkable consistency in the shape of the cumulative cough time distribution across studies (Figure 3).

Characterization of cough duration distribution

To gain insights into potential processes leading to the observed cough duration distribution, we fitted several theoretical distributions to each data set (Figure 4, Supplementary Figure S2). For five of the six studies, the distribution of cough duration is best described by a log-logistic function with initially increasing hazard, with a maximum value at around 10–12 weeks for the Uganda studies and around 4–6 weeks for the other studies, followed by a declining hazard.

Variables associated with cough duration

We first performed a bivariate analysis of each

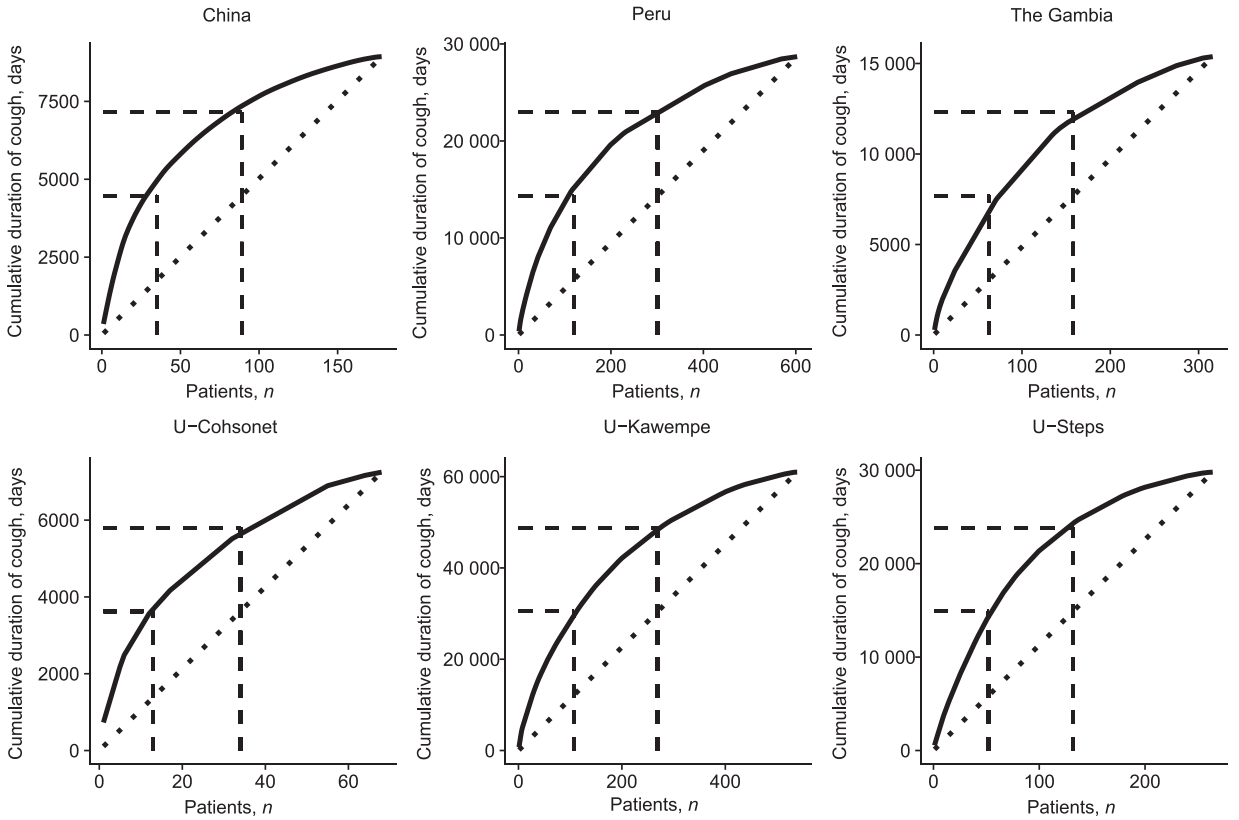

Figure 2 Cumulative cough time distribution as function of a cumulative number of patients for each of the study. Dotted line indicates expected distribution if every patient had the same cough duration. Solid line shows data. Dashed vertical and horizontal lines show the level of 20% patients contributing 50% of cough time and 50% patients contributing 80% cough time.

predictor with cough duration as the outcome for each data set. While several predictors showed statistical significance ($P < 0.05$), the strength of the association (as measured by R^2) was weak for all predictors (Supplementary Table S1). Only three predictors had a value above 0.1. These were the Karnovsky score in the U-Cohsonet study and the variables measuring fraction of visits with either non-TB healthcare providers or social contacts before final diagnosis in the U-Steps study. Full results for all predictors are shown in Supplementary Table S1.

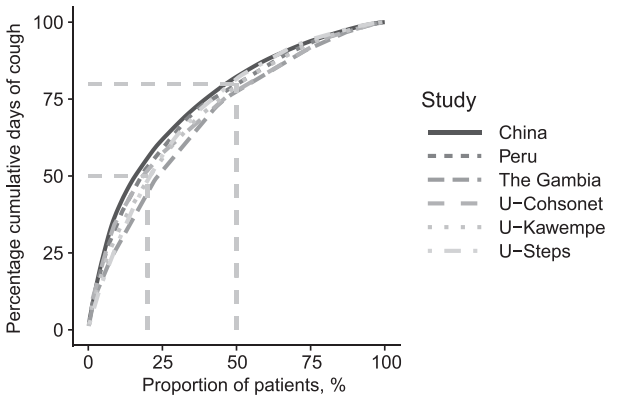

Figure 3 Percentage of cumulative cough time distribution as function of the percentage of patients for all studies. Dashed vertical and horizontal lines show the level of 20% patients contributing 50% of cough time and 50% patients contributing 80% cough time.

In a next step, we wanted to determine if a combination of variables might be predictive of cough duration. To assess this, we chose a cross-validation approach in which the ability of a model to predict variation in cough duration is assessed using independent data not used for model building. The evaluation of a model on a part of the data not used for model building provides a more conservative estimate of the strength of the association and is used if one wants to make predictive inferences—which would be valuable from a practical point of view to plan interventions.

Due to differences between studies, we analyzed each study independently and used a study-internal validation process through repeated cross-validation. For each data set, the performance of the best linear model (LM) was poor (Table 2). Only the U-Steps study has a value that is meaningfully larger than zero, with the R^2 value, indicating that the predictor variables explain below 30% of the variability in cough duration. The gradient boosted regression tree model (GBM) showed a similar pattern (Table 2). This suggests that it is the lack of predictive power of the variables available in each study, and not an

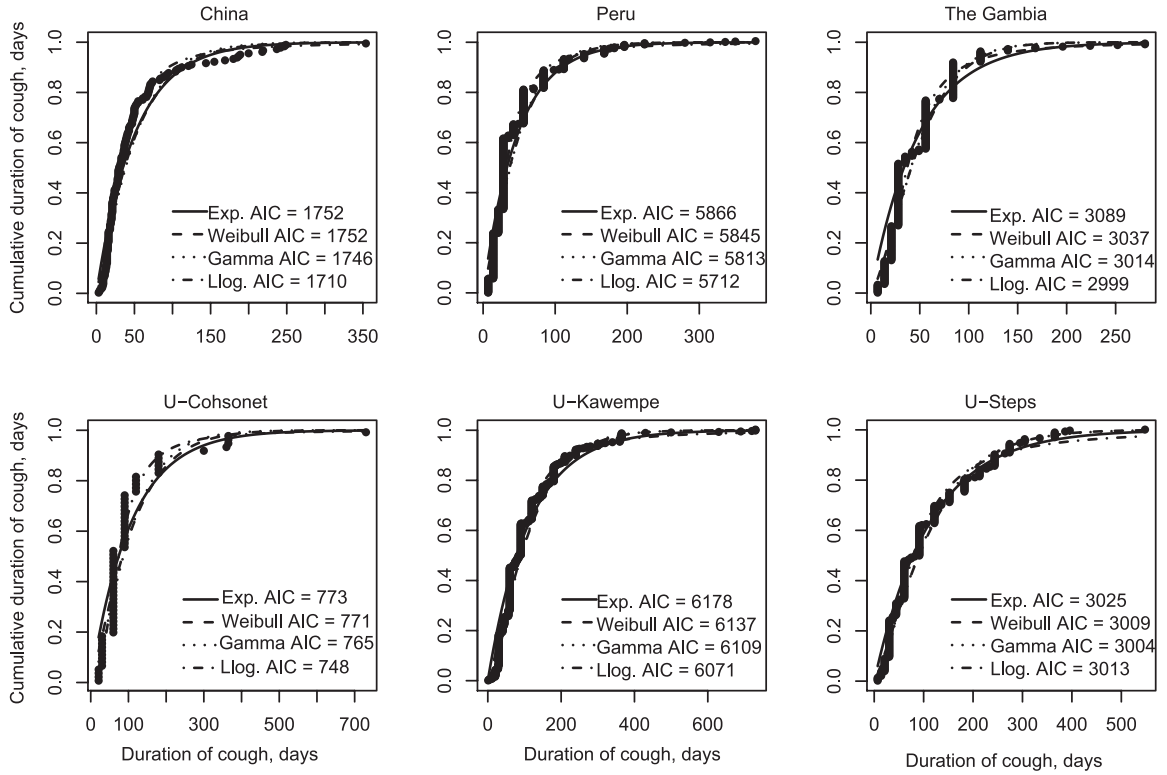

Figure 4 Cumulative cough time distribution and fit of an exponential, Weibull, Gamma and log-logistic distribution. Legend show AIC for each fit and each data set. Smaller AIC indicates statistically better model. Exp = exponential; AIC = Akaike's Information Criterion; Llog = log-logistic.

inherent limitation of the simple linear model, that failed to produce good predictions. Both for the linear model and GBM, the U-Steps variable with the most predictive power used to measure the fraction of visits a person made to health providers who were not trained to provide TB care (e.g., herbal healers, pharmacies). As this information can only be obtained after a patient has been diagnosed, it could not be used in a truly predictive manner.

DISCUSSION

Identifying and targeting infectious individuals who contribute a high proportion to transmission is a potentially promising and efficient control strategy. It is difficult to identify chains of TB transmission, and thus patients who might be responsible for the bulk of transmission. For this reason, other methods are necessary to identify whether some individuals account for more transmission than others. The duration of infectiousness is likely an important factor of overall transmission potential.

We analyzed cough duration as a surrogate measure of infectiousness duration. For all analyzed studies, we found consistent within-population heterogeneity in cough duration, with a pattern that could be summarized as the "20/50/80" rule, with 20% of patients responsible for 50% of the cumulative cough time and 50% of individuals contributing 80%. The shape of the cumulative cough duration distribution was similar across studies. The theoretical distribution which best described the observed cough duration was a log-logistic function with initially increasing hazard, followed by a declining hazard. This might be because the majority of the people become increasingly concerned as their coughing prolongs and seek care after several weeks of coughing, while a small fraction of individuals does not get diagnosed and treated for an extended period of time. However, as most of the theoretical distributions provided visually similar fits, the process leading to the observed distribution needs further investigation.

Cough duration could be affected by a variety of factors, e.g., comorbidites, smoking status, type of TB strain, sex. In line with previous studies (see the reviews of[12,13,15–17]), we found statistically signifi-

Table 2 Model performance as measured by cross-validated R^2 for best performing LM and GBM for each data set*

Study	R^2 (LM)	R^2 (GBM)
China	−0.012	−0.014
Peru	0.000	−0.004
The Gambia	0.004	0.000
U-Cohsonet	0.059	−0.015
U-Kawempe	0.027	0.004
U-Steps	0.288	0.265

* Values close to zero indiciate no predictive power of the models. Negative values indicate over-fitting on the training data.
LM = linear model; GBM = gradient boosted regression tree model.

cant correlations between cough duration and several patient and environmental characteristics. However, none of the individual predictors had a strong effect on the outcome. Our analysis of different multivariate models did not find any combination of variables that performed well at predicting cough duration. The variable measuring the fraction of care-seeking steps with providers not trained in TB care from the U-Steps study had some predictive power. This quantity can only be determined once a person is diagnosed and as such is of no use in early identification.

In all analyzed studies, cough duration is self-reported, which means bias might be present. Patients that cough for longer will likely remember the onset of cough with less precision. However, over- and underestimation of cough duration seem equally likely. Survival bias in the data might have been present if the risk of death increases with the duration of cough, which would lead to less individuals contributing to the right tail of the observed distribution.

To equate supercoughers with superspreaders, one needs to assume that cough is a good indicator of infectiousness. While transmission without cough is likely possible,[26] increased coughing[27] and delay in receiving treatment for TB has been found to lead to increased transmission.[28] Cough of patients with pulmonary TB contain *Mycobacterium tuberculosis*,[29] and smear-positive cases are more likely to infect their contacts than smear-negative cases.[30,31] Estimates for the duration of infectiousness are often many months or even years,[32,33] which matches the long cough duration we found in some individuals. If the supercoughers we identified are indeed superspreaders, the ability to identify them would be very valuable. However, none of the many measured host or environmental variables were able to predict individuals with longer cough duration.

Conflicts of interest: none declared.

References

1 Galvani A P, May R M. Epidemiology: dimensions of superspreading. Nature 2005; 438(7066): 293–295.
2 Lloyd-Smith J O, Schreiber S J, Kopp P E, Getz W M. Superspreading and the effect of individual variation on disease emergence. Nature 2005; 438(7066): 355–359.
3 Stein R A. Super-spreaders in infectious diseases. Int J Infect Dis 2011; 15(8): e510–e513.
4 Ypma R J F, Altes H K, Soolingen D van, Wallinga J, Ballegooijen W M van. A sign of superspreading in tuberculosis: highly skewed distribution of genotypic cluster sizes. Epidemiology (Cambridge, MA, USA) 2013; 24(3): 395–400.
5 Walker T M, Ip C L C, Harrell R H, et al. Whole-genome sequencing to delineate mycobacterium tuberculosis outbreaks: A retrospective observational study. Lancet Infect Dis 2013; 13(2): 137–146.
6 Gardy J L, Johnston J C, Ho Sui S J, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med 2011; 364(8): 730–739.
7 Escombe A R, Moore D A J, Gilman R H, et al. The infectiousness of tuberculosis patients coinfected with HIV. PLoS Med 2008; 5(9): e188.
8 Melsew Y A, Gambhir M, Cheng A C, et al. The role of superspreading events in mycobacterium tuberculosis transmission: evidence from contact tracing. BMC Infect Dis 2019; 19(1): 244.
9 Trauer J M, Dodd P J, Gomes M G M, et al. The importance of heterogeneity to the epidemiology of tuberculosis. Clin Infect Dis 2019; 69(1): 159–166.
10 Melsew Y A, Doan T N, Gambhir M, Cheng A C, McBryde E, Trauer J M. Risk factors for infectiousness of patients with tuberculosis: a systematic review and meta-analysis. Epidemiol Infect 2018; 146(3): 345–353.
11 Jones-López E C, Acuña-Villaorduña C, Ssebidandi M, et al. Cough aerosols of mycobacterium tuberculosis in the prediction of incident tuberculosis disease in household contacts. Clin Infect Dis 2016; 63(1): 10–20.
12 Storla D G, Yimer S, Bjune G A. A systematic review of delay in the diagnosis and treatment of tuberculosis. BMC Public Health 2008; 8(1): 15.
13 Li Y, Ehiri J, Tang S, et al. Factors associated with patient, and diagnostic delays in chinese TB patients: a systematic review and meta-analysis. BMC Med 2013; 11(1): 156.
14 Sreeramareddy C T, Panduru K V, Menten J, Van den Ende J. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature. BMC Clin Infect 2009; 9(1): 91.
15 Cai J, Wang X, Ma A, Wang Q, Han X, Li Y. Factors associated with patient and provider delays for tuberculosis diagnosis and treatment in Asia: a systematic review and meta-analysis. PloS One 2015; 10(3): e0120088.
16 Sreeramareddy CT, Qin Z Z, Satyanarayana S, Subbaraman R, Pai M. Delays in diagnosis and treatment of pulmonary tuberculosis in India: a systematic review. Int J Tuberc Lung Dis 2014; 18(3): 255–266.
17 Getnet F, Demissie M, Assefa N, Mengistie B, Worku A. Delay in diagnosis of pulmonary tuberculosis in low-and middle-income settings: systematic review and meta-analysis. BMC Pulm Med 2017; 17(1): 202.
18 Martinez L, Sekandi J N, Castellanos M E, Zalwango S, Whalen C C. Infectiousness of HIV-seropositive patients with tuberculosis in a high-burden african setting. Am J Respir Crit Care Med 2016; 194(9): 1152–1163.
19 Whalen C C, Zalwango S, Chiunda A, et al. Secondary attack rate of tuberculosis in urban households in Kampala, Uganda. PloS One 2011; 6(2): e16137.
20 Sekandi J N, Zalwango S, Martinez L, et al. Four degrees of separation: social contacts and health providers influence the steps to final diagnosis of active tuberculosis patients in urban Uganda. BMC Infect Dis 2015; 15: 361.
21 Hill P C, Jackson-Sillah D J, Fox A, et al. Incidence of tuberculosis and the predictive value of EliSPOT and Mantoux tests in Gambian case contacts. PloS One 2008; 3(1): e1379.
22 Grandjean L, Gilman R H, Martin L, et al. Transmission of multidrug-resistant and drug-susceptible tuberculosis within households: a prospective cohort study. PLoS Med 2015; 12(6): e1001843; discussion: e1001843.
23 Martinez L, Xu L, Chen C, et al. Delays and pathways to final tuberculosis diagnosis in patients from a referral hospital in urban china. Am J Trop Med Hyg 2017;
24 James G, Witten D, Hastie T, et al. An introduction to statistical learning. Berlin, Germany: Springer-Verlag, 2014.
25 Bischl B, Lang M, Kotthoff L, et al. mlr: Machine learning in R. J Machine Learning Res 2016; 17(170): 1–5. http: //jmlr.org/papers/v17/15-066.html
26 Drain P K, Bajema K L, Dowdy D, et al. Incipient and subclinical tuberculosis: a clinical review of early stages and progression of infection. Clin Microbiol Rev 2018; 31(4): e00021-18.

27 Turner R D, Birring S S, Darmalingam M, et al. Daily cough frequency in tuberculosis and association with household infection. Int J Tuberc Lung Dis 2018; 22(8): 863–870.
28 Golub J E, Bur S, Cronin W A, et al. Delayed tuberculosis diagnosis and tuberculosis transmission. Int J Tuberc Lung Dis 2006; 10(1): 24–30.
29 Fennelly K P, Jones-López E C, Ayakaka I, et al. Variability of infectious aerosols produced during coughing by patients with pulmonary tuberculosis. Am J Respir Crit Care Med 2012; 186(5): 450–457.
30 Behr M A, Warren S A, Salamon H, et al. Transmission of mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. Lancet 1999; 353(9151): 444–449.
31 Martinez L, Shen Y, Mupere E, Kizza A, Hill P C, Whalen C C. Transmission of mycobacterium tuberculosis in households and the community: a systematic review and meta-analysis. Am J Epidemiol 2017; 185(12): 1327–1339.
32 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(1): 87–93.
33 Corbett E L, Bandason T, Cheung Y-B, et al. Prevalent infectious tuberculosis in harare, zimbabwe: Burden, risk factors and implications for control. Int J Tuberc Lung Dis 2009; 13(10): 1231–1237.

RÉSUMÉ

CADRE : En matière de tuberculose (TB), il est très difficile d'observer les chaînes de transmission et donc d'identifier les super contaminateurs. Nous examinons la durée de la toux comme mesure indirecte de transmission afin d'évaluer la présence de super contaminateurs potentiels.

SCHÉMA : Nous avons analysé six études de Chine, du Pérou, de Gambie et d'Ouganda et déterminé la distribution de la durée de la toux et l'avons comparée à plusieurs distributions théoriques. Pour déterminer les facteurs associés à la durée de la toux, nous avons utilisé une régression linéaire et des régressions arborescentes renforcées afin d'étudier le pouvoir prédictif des caractéristiques du patient, de la clinique et de l'environnement.

RÉSULTATS : Nous avons trouvé une hétérogénéité intra étude, en termes de durée de la toux, et de fortes similitudes entre études. Environ 20% des patients ont contribué à 50% du total des jours de toux et autour de 50% des patients ont contribué à 80% du nombre total de jours de toux. La distribution de la durée de la toux a suggéré une chance de diagnostic initialement croissante puis croissante. Si certaines variables d'exposition ont mis en évidence des associations statistiquement significatives avec la durée de la toux, aucune d'elles n'a eu un effet puissant. Les analyses multivariates de différents types de modèle n'ont pas produit un modèle qui ait un bon pouvoir de prédiction.

CONCLUSION : Nous avons trouvé des preuves cohérentes de la présence de super tousseurs, mais aucune caractéristique prédictive de ces individus.

RESUMEN

MARCO DE REFERENCIA: En el caso de la tuberculosis (TB) es muy difícil la observación de las cadenas de transmisión y con ello se complica el reconocimiento de los casos superdiseminadores. Se investigó la duración de la tos como una medida indirecta de transmisión, con el fin de evaluar la eventual presencia de casos de TB superdiseminadores.

MÉTODO: En seis estudios de China, Perú, Gambia y Uganda se determinó la distribución de la duración de la tos y se comparó con varias distribuciones teóricas. Con el propósito de determinar los factores asociados con la duración de la tos, se utilizó una regresión lineal y árboles de regresión potenciados que examinaban la capacidad pronóstica de las características del paciente, del cuadro clínico y del medio ambiente.

RESULTADOS: Se observó heterogeneidad intraestudios y similitudes considerables entre los diferentes estudios con respecto a la duración de la tos. Cerca de 20% de los pacientes aportaban 50% del total de días de tos y alrededor de 50% de los pacientes representaban 80% del total de días de tos. La distribución de la duración de la tos indicó una posibilidad de diagnóstico que aumentaba inicialmente con el tiempo y luego disminuía. Aunque algunas de las variables de exposición se asociaron con la duración de la tos de manera estadísticamente significativa, ninguna ejerció un efecto fuerte. Los análisis multivariantes de diferentes tipos de modelos no generaron ningún modelo con una potencia pronóstica adecuada.

CONCLUSIÓN: Se encontró evidencia constante en favor de la presencia de casos de TB superdiseminadores, pero no se reconoció ninguna característica que permita pronosticar cuáles son estas personas.