

# Defining Adequate Contact For Transmission Of Mycobacterium Tuberculosis In An African Urban Environment

Maria Eugenia Castellanos (✉ [mecastellanos@uga.edu](mailto:mecastellanos@uga.edu))

University of Georgia <https://orcid.org/0000-0002-8181-6321>

**Sarah Zalwango**

Makerere University College of Health Sciences

**Robert Kakaire**

University of Georgia

**Mark H. Ebell**

University of Georgia

**Kevin K. Dobbin**

University of Georgia

**Juliet Sekandi**

University of Georgia

**Noah Kiwanuka**

Makerere University College of Health Sciences

**Christopher C. Whalen**

University of Georgia

---

## Research article

**Keywords:** Tuberculosis, contact, transmission, social network, contact network

**Posted Date:** March 28th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-19461/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on June 9th, 2020. See the published version at <https://doi.org/10.1186/s12889-020-08998-7>.

# Abstract

**Background.** The risk of infection from respiratory pathogens increases according to the contact rate between the infectious case and susceptible contact, but the definition of adequate contact for transmission is not standard. In this study we aimed to identify factors that can explain the level of contact between tuberculosis cases and their social networks in an African urban environment.

**Methods.** This was a cross-sectional study conducted in Kampala, Uganda from 2012-2016. We carried out an exploratory factor analysis (EFA) in social network data from tuberculosis cases and their contacts. We evaluated the factorability of the data to EFA using the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO). We used principal axis factoring with oblique rotation to extract and rotate the factors, then we calculated factor scores for each using the weighted sum scores method. We assessed construct validity of the factors by associating the factors with other variables related to social mixing.

**Results.** Tuberculosis cases (N = 120) listed their encounters with 1,154 members of their social networks. Thirteen variables displayed high inter-correlation (KMO=0.72) and were included for EFA. Two factors were identified, which captured 82% of the variance. The first factor, named 'Setting', involves the type, frequency, duration and ventilation of the usual place of meeting as well the physical proximity among tuberculosis cases and contacts, represented by the sleeping and eating patterns. The second factor, named 'Relationship', was explained by the relationship, its duration, and the level of intimacy among cases and contacts, represented by the strength of knowledge of each other, provision of healthcare, and whether they were travel partners. Setting and Relationship scores varied according to the age, gender and nature of the relationship among tuberculosis cases and their contacts.

**Conclusions.** In this large cross-sectional study from an urban African setting, we identified two factors that can assess adequate contact between tuberculosis cases and their social network members. These findings also confirm the complexity and heterogeneity of social mixing.

## Introduction

For any infectious disease, the transmission of the microorganism is the result of exposure, that is adequate contact between an infectious case and a susceptible host. The nature of adequate contact varies from one infectious disease to another, but regardless of the disease, the definition of adequate contact is not standardized. This lack of definition is especially true for airborne infectious diseases [1–3] for which the sequence, frequency, and duration of contact is challenging to ascertain [4]. Nevertheless, the rate of adequate contact is an important parameter to understand because it is a key determinant of epidemic behavior [5]. Without a clear understanding of the factors that affect the contact rate, it may not be possible to design effective community interventions to minimize transmission.

Tuberculosis is a respiratory disease that is transmitted through the airborne route. Although the global incidence of tuberculosis has declined in the last decade, the disease persists in many low- and middle-

income countries around the world. These countries rely on passive case finding followed by directly-observed therapy as the mainstay of tuberculosis control, yet the disease persists. New approaches are needed. We propose that community-based interventions that reduce the frequency and duration of adequate contact between tuberculosis cases and their contacts would be an effective way to control tuberculosis, but it is challenging to design these interventions without a better understanding of the factors that affect the dynamics of social mixing in a population.

From previous studies of tuberculosis, we know that demographics and the settings of interaction may modify the nature of contact between cases and their contacts [1, 3, 6–8]. At best, these factors give an incomplete description of adequate contact for transmission of *M. tuberculosis*. To develop a more complete understanding of adequate contact, we performed an exploratory factor analysis (EFA) using information collected from index tuberculosis cases and their contacts as identified in their ego-centric social networks. We chose to use EFA because it is a method that can synthesize correlated data [9] into underlying constructs called factors [10], thereby reducing the data into meaningful concepts that provide insight into the nature of adequate contact. We also evaluated the construct validity of the factors we identified by evaluating their association with other variables related to social mixing.

## Methods

We conducted a cross-sectional study of patients with active tuberculosis in Kampala, Uganda. The index cases were patients with microbiologically confirmed, pulmonary tuberculosis recruited from the Uganda National Tuberculosis and Leprosy Control Program. These cases were 15 years or older, residents of Lubaga Division of Kampala, and gave written informed consent to participate in the study.

In a standard interview conducted by trained research assistants, the index cases provided a list of their household and community members (defined as 'extra-household') of their ego-centric networks. A contact was defined as a person with whom the index case had a personal relation and with whom they spent time during the previous three months. The research assistants approached these social network members, and if they agreed, they were enrolled in the study. These participants were then interviewed using a standardized questionnaire to collect epidemiologic information relevant to the nature and duration of contact in the social networks. (Supplementary Material, Table S1).

An item analysis was conducted in the original data set to determine baseline characteristics of the tuberculosis cases, explore the distribution of each variable, and summarized proportions and measures of central tendency. After assessing the distributions, we combined and redefined variables as dichotomous or ordinal to ensure a similar scale across variables and to assign values in which the lowest value represented less contact and the highest value represented the more contact (Supplementary Material, Table S2 & Table S3). Variables were excluded from EFA (Table S3) if they were nominal, had limited distribution, or did not provide additional content. We then conducted an item-analysis of the recoded and nominal variables that allow us to define the relationship between tuberculosis cases and their contacts.

After performing this item analysis, we performed an EFA to determine the underlying factors in the data. By accounting for correlations among variables, this method allowed us to reduce the number of variables needed to define adequate contact. We chose EFA over principal component analysis since we hypothesized that the observed variables could be grouped in underlying constructs, or factors [11], that would provide meaningful insights into the nature of adequate contact. To begin, we estimated the polychoric correlation between all pairs of variables to assess the relations among recoded variables [12]. We evaluated the factorability of the variables using the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) [13], and excluded variables in which this measurement was lower than 0.6 [14].

Once the correlation matrix was estimated, we used the principal axis factoring method to extract the factors [10] and retained factors based on proposed criteria [15]: 1) factors with eigenvalue greater than 1, 2) factors with eigenvalues that appear in the sharp slope of a Scree plot, and 3) factors that account for 80–90% of the variation. Moreover, only factors with at least three variables were retained [16]. Since the factors were likely correlated, we performed an oblique rotation of the factors [17] to arrive at the final set of factors. We then assigned a name to each factor based on their correlated characteristics and literature review [18].

To generate the factor scores for an individual, we used the weighted sum scores method, as it allows that the variables with the highest loadings to have the highest impact in our factor score [19]. To compute factor scores for each study participant, we first multiplied the recorded value for each item by the corresponding estimated factor loading then summed these products across all variables to generate the factor score. We excluded variables with loadings below 0.30 [20]. Factor scores were investigated to determine whether they met the normality assumption and had a unimodal distribution [21]. Based on these analyses, the results of the factor scores are presented as median with interquartile ranges.

To establish the construct validity of the factor scores, we examined the relationship of each factor with relevant variables in a sub-set of the study participants. We stratified these participants according to type of contact (household and extra-household contact) and nature of the relationship between case and control (spouse, child, sibling, friend, co-worker, other relatives, neighbor, other). The median and interquartile ranges of the factor scores were estimated for each stratum. In the process, we examined the variability of the factor scores by gender of contact, age of contact, age of index and sex of index case. To compare the difference in medians among stratified groups we estimated the 95% confidence intervals by bootstrapping, using the package 'boot' for R software [22]. We set the number of bootstraps replicates to 10,000 and calculated the intervals with the adjusted bootstrap percentile (BCa) method [23].

Written informed consent was obtained from all participants prior to study inclusion. Institutional review board clearance was obtained from the Higher Degrees and Ethics Committee at Makerere University School of Public Health, the Uganda Council of Science and Technology, and the University of Georgia.

## Results

The study enrolled 120 index cases with tuberculosis disease and 1,179 household and community contacts to generate social networks. Complete social network information was obtained for 1,154 of 1,179 contacts (98%). The majority of index cases was men (83%), young adults (57%) and had a microbiologically confirmed diagnosis (Table S4). Each case had a median of nine contacts (IQR 8, 12). For male cases, the median number of household contacts was 4 (IQR 2,6) and extra-household contacts was 7 (IQR 4,9), whereas for women, the median number of household contacts was also 4 (IQR 2,6), and extra-household contacts was 5.5 (IQR 4,8).

We initially considered 15 variables from the social network questionnaire for the exploratory factor analysis (Table 1 and Supplementary material, Figure S1). Of these 15 variables, 13 variables had an overall KMO of 0.72, with individual KMO measurements of >0.60. Two variables (“Frequency of meeting since onset of cough” and “Number of other people met in addition to contact”) were excluded from the factor analysis because of a low individual KMO (Table 1).

Of the 13 factors that were identified, only two factors met our *a priori* criteria for selection with eigenvalues greater than one (Supplementary material, Figure S2). The first factor had a loading of 5.7 and the second factor a loading of 1.9. These two factors captured 82% of the variance (61% and 21% respectively). Variable “Contact has cough” produced low factor loadings in both factors, implying that this variable might not contribute particularly to either of them.

The first factor grouped together variables related to the setting and environment of the contact between the index case and his/her contact; we named this factor the “setting” factor. The six variables in this category had factor loadings of 0.60 or more (Table 2). The second factor grouped together variables that corresponded to the intimacy and social relationship of the index case and contact; we named the factor the “relationship” factor. The six variables in this category had factor loadings of 0.50 or more (Table 2).

Among 940 contacts with a full complement of information, we generated setting and relationship scores. In this subset of contacts, the social network findings corresponded well to the findings obtained from the full set of 1,154 contacts (Table S5), thereby reducing the likelihood of selection bias. Both the setting factor and relationship factor followed a multimodal distribution ( $p=0.02$  and  $P<0.0001$ , respectively; unimodality test) (Figure 1). Scores for the setting factor had a median of 10.2 (IQR 7.0, 13.6) with a range of 5.3-18.8; the relationship scores had a median of 7.7 (IQR 6.4, 10.1) with a range of 4.0 to 14.8. Setting and relationship scores varied according to the nature of the relationship among a tuberculosis case and their contact (Figure 2). Spouses had the highest setting score, followed by children and siblings. Altogether, family members had a higher median setting score (13.8, IQR 11.6, 15.7) than non-family members (7.2, IQR 6.2, 9.4). In the case of the relationship factor, spouses, siblings and other relatives had the highest score. The median relationship score in family members (9.9, IQR 7.6, 11.5) was also higher than in non-family members (6.9, IQR 5.6, 8.1).

For both the setting and relationship scores, household contacts (N = 350) had higher scores than extra-household contacts (N = 590, Figure 3,  $p < .0001$ ). For the setting factor, household contacts had a greater median score (14.6, IQR 12.8,16.2) as compared to extra-household contacts (7.4, IQR 6.4, 9.8); the

difference in medians was 7.2 (95% CI: 6.9, 7.6). For the relationship factor, household contacts again had a greater median score (9.8, IQR 7.7, 11.6) as compared to extra-household contacts (7.0, IQR 5.8, 8.6). The difference in medians was 2.8 (95% CI 2.2-3.5). Extra-household family members had a higher median setting score (9.9, IQR 7.3, 11.8) than extra-household non-family members (7.1, IQR 6.2, 9.1). Similarly, extra-household family members had a higher median relationship score (9.4, IQR 7.0, 11.2) than extra-household non-family members (6.8, IQR 5.6, 8.1).

When stratifying the analysis by the sex or age of the index case, the median values of the factors scores among contacts revealed differences according to age and sex of the participants (Table 3). Contacts of male cases had a lower setting score as opposed to contacts of female cases (difference in medians = 2.2 (95% CI 1.4 -3.3)). In contrast, contacts of male and female cases had similar relationship scores.

Contacts  $\leq 4$  years old had the highest setting score, followed by contacts aged 5-14 years and contacts  $\geq 15$  years old. The difference in medians among contacts  $\leq 15$  years of age and contacts  $\geq 15$  years of age was 4.6 (95% CI 3.7, 5.8). Interactions of contacts  $\geq 15$  years old with cases aged 25-44 years were the ones with the lowest setting score. In contrast, interactions of contacts  $\leq 15$  years with cases 15-24 years proved to be the highest. For the relationship score, an inverse association with the age of contact was found.

## Discussion

The goal of this analysis was to develop a working definition of adequate contact between an infectious tuberculosis case and his or her contacts. To do so, we used exploratory factor analysis to identify two underlying constructs related to the social contact pattern between an infectious tuberculosis case and a contact. The first factor characterized the setting and environment of interaction with the index cases, and the second factor described the social relationship between the index case and contact. These two factors explained 82% of the variance in the data, with the setting factor explaining the most variance.

The setting score captured information about ventilation in the meeting place and about the nature of interaction in the venue. As for ventilation, inclusion of this variable provides strong content validity to the score because tuberculosis is transmitted through the airborne route in settings of shared airspace. Transmission is directly related to the concentration of bacteria in the inspired air and to the duration of exposure to contaminated air [24]. Ventilation reduces the concentration of organisms and would therefore be directly related to the likelihood of transmission. This score also captured the complex socio-spatial information about the physical proximity of cases and contacts in their patterns of sleeping and eating together.

The relationship score captured information about the intimacy, social and emotional closeness, of the relationship, especially as it relates to the health of the index case. This intimacy is shown by the trust an index case had for a contact. The score also reveals the reciprocity in the relationship, such that the contact provides care for the index case. This level of trust is not trivial given the stigma that is attached to tuberculosis in Africa because of its association with HIV [25]. The variable about means of

transportation does not seem to fit the underlying construct, and indeed it had the lowest loading in the analysis. There is, however, a cultural interpretation that makes sense. Since transportation in Kampala is crowded and expensive, many residents walk to complete daily chores. When two individuals know one another, they are more likely to walk together. Extra-household contacts have low relationship scores whereas household contacts have higher relationship scores. These scores seem to vary from moderate to high, suggesting different levels of intimacy within the household.

Once the factor scores were estimated, we assessed construct validity by comparing scores according to relationship, age and sex, and household exposure. As for relationship, we postulated that family and relatives would have higher setting relationship scores, and indeed they did. We observed that family members, especially spouses have the higher values as expected when compared to community members. However, it seems that even among these categories of contacts, there are degrees of closeness that should be considered. For instance, some siblings, relatives and children had setting scores as low as six units, but other family members had setting scores as high as eighteen units. Thus, these scores could refine and further characterize the level of contact among contacts of an infectious case. We further postulated that household contacts would have higher setting scores, and as expected, they did. Finally, we evaluated age and sex interactions and found that younger contacts between birth and 14 years had higher setting scores with index cases between 15 and 45 years. This finding suggests age assortment of these child contacts with their parents or older siblings and partially explains the high risk for tuberculosis transmission to children in homes of infectious cases.

Regarding sex of contacts, women had higher setting scores than men. This finding was unexpected but seems to indicate the importance of the household in the transmission of *M. tuberculosis*. In a low- and middle-income settings, women tend to stay at home to fulfill their role as caregiver or because of limited opportunities for them in the formal job market [26]. Moreover, women in Uganda are reported to work 18% more than men in activities at home [27]. The relationship score was less variable across the different types of age and sex categories. There were no differences between female and male index cases. When assessing age assortment, we found that the relationship score was lower for contacts 0–14 years of age for all age groups of the index cases. Considering the nature of some of the questions that comprise this score the findings were as expected. Social intimacy and discussion of health issues seem to be topics that are more relevant to be shared within tuberculosis cases and adults contacts. The relationship score might be helpful to assess the level of social support that an index case receives from their social network. A systematic review of social network analyses in low- and middle- income settings has shown that behavior and health outcomes are associated with the structure and composition of these networks [28].

EFA has been criticized for identifying artificial factors that are not informative [29] to the underlying constructs being evaluated. One of the major strengths of our study is that we minimized this risk by conducting additional analyses that corroborate the robustness of our factors, as it has been recommended [30]. Our scores were consistent with other variables used to describe social network structure, e.g. nature of the relationship among contacts and case, age and sex assortment.

There are several limitations of the study. First, the listed contacts in an index's network may be incomplete. However, the median household size per index case was four, which is similar to the 3.7 reported among Kampala residents in the Uganda National Household Survey 2016/2017 report [31]. Second, there is a risk for recall and response bias because we collected self-reported data. Nevertheless, the nature of the questions and the high dispersion and variability of the factor scores suggest participants did not constrain their answers for social desirability [32]. Finally, there were some nominal variables that we had to exclude or recode as binary variables, thereby limiting the information in the analysis.

In conclusion, our study identified two factors that can be used to assess adequate contact between tuberculosis cases and their contacts, explaining 82% of the variance in the observed variables. As a whole, these findings also confirm the complex and heterogeneous social mixing between cases and contacts [33]. In future studies, we will evaluate the criterion validity of these factors by relating them to the presence, or absence, of tuberculous infection among social networks of tuberculosis cases.

## Abbreviations

EFA=exploratory factor analysis, KMO= Kaiser-Meyer-Olkin, Bca= adjusted bootstrap percentile, IQR= interquartile range

## Declarations

### *Ethics approval and consent to participate*

Written informed consent was obtained from all participants prior to study inclusion. Institutional review board clearance was obtained from the Higher Degrees and Ethics Committee at Makerere University School of Public Health, the Uganda Council of Science and Technology, and the University of Georgia.

### *Consent for publication*

Not applicable

### *Availability of data and materials*

The data that support the findings of this study are available on request from the corresponding author, CCW. The data are not publicly available due to containing information that could compromise the privacy of research participants.

### *Competing interests*

The authors declare that they have no competing interests

### *Funding*

This work is supported by the National Institute of Allergy and Infectious Diseases and the Fogarty International Center of the National Institutes of Health (grants R01 AI093856, NO1-AI95383, D43-TW01004) and by the Schlumberger Foundation Faculty for the Future Fellowship.

### *Author's contributions*

Conception and design: MEC, CCW; data collection: SZ, NK, RK, CCW, data analyses: MEC; interpretation of the study results: MEC, SZ, RK, KKD, MHE, JS, NK, CCW. Review and edit of the manuscript: MEC, SZ, RK, KKD, MHE, JS, NK, CCW. All authors approved the final version of the manuscript.

### *Acknowledgements*

The authors would like to thank the Uganda National Tuberculosis and Leprosy Control Programme, the municipal leaders of Lubaga Division of Kampala, and the Kampala Capital City Authority for their cooperation and support of the field work needed to complete this study. The authors acknowledge the home health visitors for invaluable contributions to the science of this study and the Epidemiology in Action research group of the Global Health Institute, University of Georgia, for their critical insights.

## References

1. Dodd, P.J., et al., *Age- and Sex-Specific Social Contact Patterns and Incidence of Mycobacterium tuberculosis Infection*. Am J Epidemiol, 2016. **183**(2): p. 156-66.
2. Edmunds, W.J., C. O'callaghan, and D. Nokes, *Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections*. Proceedings of the Royal Society of London B: Biological Sciences, 1997. **264**(1384): p. 949-957.
3. Mossong, J., et al., *Social contacts and mixing patterns relevant to the spread of infectious diseases*. PLoS Med, 2008. **5**(3): p. e74.
4. Bansal, S., et al., *The dynamic nature of contact networks in infectious disease epidemiology*. J Biol Dyn, 2010. **4**(5): p. 478-89.
5. Mathema, B., et al., *Drivers of Tuberculosis Transmission*. J Infect Dis, 2017. **216**(suppl\_6): p. S644-S653.
6. Feenstra, S.G., et al., *A qualitative exploration of social contact patterns relevant to airborne infectious diseases in northwest Bangladesh*. Journal of health, population, and nutrition, 2013. **31**(4): p. 424.
7. Melegaro, A., et al., *What types of contacts are important for the spread of infections? Using contact survey data to explore European mixing patterns*. Epidemics, 2011. **3**(3): p. 143-151.

8. Johnstone-Robertson, S.P., et al., *Social mixing patterns within a South African township community: implications for respiratory disease transmission and control*. American journal of epidemiology, 2011: p. kwr251.
9. Pett, M.A., N.R. Lackey, and J.J. Sullivan, *Making sense of factor analysis : the use of factor analysis for instrument development in health care research*. 2003, Thousand Oaks, Calif. ; London: SAGE.
10. Treiblmaier, H. and P. Filzmoser, *Exploratory factor analysis revisited: How robust methods support the detection of hidden multivariate data structures in IS research*. Information & management, 2010. **47**(4): p. 197-207.
11. Suhr, D.D., *Principal component analysis vs. exploratory factor analysis*. SUGI 30 proceedings, 2005. **203**: p. 230.
12. Holgado-Tello, F.P., et al., *Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables*. Quality & Quantity, 2010. **44**(1): p. 153.
13. Kaiser, H.F. and J. Rice, *Little jiffy, mark IV*. Educational and psychological measurement, 1974. **34**(1): p. 111-117.
14. Taherdoost, H., S. Sahibuddin, and N. Jalaliyoon, *Exploratory factor analysis: Concepts and theory*. Advances in Pure and Applied Mathematics, 2014.
15. Berghaus, R.D., et al., *Factor analysis of a Johne's disease risk assessment questionnaire with evaluation of factor scores and a subset of original questions as predictors of observed clinical paratuberculosis*. Prev Vet Med, 2005. **72**(3-4): p. 291-309.
16. Costello, A.B. and J. Osborne, *Best practices in exploratory factor analysis; four recommendations for getting the most from your analysis*. Practical Assessment, 2005. **10**(7): p. 1-9.
17. Gaskin, C.J. and B. Happell, *On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use*. International journal of nursing studies, 2014. **51**(3): p. 511-521.
18. Barbero Garcia, M.I., E. Vila, and F.P. Holgado Tello, *Introducción básica al análisis factorial*. 2013: Editorial UNED.
19. DiStefano, C., M. Zhu, and D. Mindrila, *Understanding and using factor scores: Considerations for the applied researcher*. Practical Assessment, Research & Evaluation, 2009. **14**(20): p. 2.
20. Beavers, A.S., et al., *Practical considerations for using exploratory factor analysis in educational research*. Practical assessment, research & evaluation, 2013. **18**.
21. Ameijeiras-Alonso, J., R.M. Crujeiras, and A. Rodríguez-Casal, *Mode testing, critical bandwidth and excess mass*. arXiv preprint arXiv:1609.05188, 2016.
22. Canty, A. and B. Ripley, *Package 'boot'*. 2017.
23. Puth, M.T., M. Neuhauser, and G.D. Ruxton, *On the variety of methods for calculating confidence intervals by bootstrapping*. J Anim Ecol, 2015. **84**(4): p. 892-7.
24. Riley, E.C., G. Murphy, and R.L. Riley, *Airborne spread of measles in a suburban elementary school*. Am J Epidemiol, 1978. **107**(5): p. 421-32.

25. Katamba, A., et al., *Patients perceived stigma associated with community-based directly observed therapy of tuberculosis in Uganda*. East Afr Med J, 2005. **82**(7): p. 337-42.
26. Onifade, D.A., et al., *Gender-related factors influencing tuberculosis control in shantytowns: a qualitative study*. BMC Public Health, 2010. **10**(1): p. 381.
27. Ortiz-Ospina, E. and S. Tzvetkova. *Working women: Key facts and trends in female labor force participation*. 2017.
28. Perkins, J.M., S.V. Subramanian, and N.A. Christakis, *Social networks and health: a systematic review of sociocentric network studies in low- and middle-income countries*. Soc Sci Med, 2015. **125**: p. 60-78.
29. Shapiro, S.E., M.R. Lasarev, and L. McCauley, *Factor analysis of Gulf War illness: what does it add to our understanding of possible health effects of deployment?* Am J Epidemiol, 2002. **156**(6): p. 578-85.
30. Edefonti, V., et al., *Nutrient-based dietary patterns and laryngeal cancer: evidence from an exploratory factor analysis*. Cancer Epidemiol Biomarkers Prev, 2010. **19**(1): p. 18-27.
31. Uganda Bureau of Statistics (UBOS), *Uganda National Household Survey 2016/2017* 2018.
32. Furnham, A., *Response bias, social desirability and dissimulation*. Personality and individual differences, 1986. **7**(3): p. 385-400.
33. Wallinga, J., P. Teunis, and M. Kretzschmar, *Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents*. American journal of epidemiology, 2006. **164**(10): p. 936-944.

## Tables

Table 1. Individual and Overall Kaiser's Measure of Sampling Adequacy.

Initial results with 15 variables and final selection with 13 variables included in the exploratory factor analysis.

Variable	Kaiser's Measure of Sampling	Kaiser's Measure of Sampling
	Adequacy	Adequacy
	15 items	13 items
Overall	0.55	0.72
Contact happen indoors or outdoors	0.50	0.61
Nature of ventilation at usual place of meeting	0.47	0.63
Case shared TB diagnosis with contact	0.50	0.69
Contact have cough	0.55	0.69
Frequency of shared meals since onset cough	0.56	0.72
Frequency and duration of contact over the past month	0.82	0.73
Care was provided by the contact in the past 3 months	0.54	0.73
Place of usual meeting. Home TB case versus other location.	0.80	0.75
Case trusts contact	0.66	0.75
Length of knowing contact	0.71	0.77
Frequency of sleeping in same room and bed since onset cough	0.55	0.79
How well does the case knows contact	0.55	0.80
Means of transportation used most often with contact. None (walking) versus a type of transportation.	0.80	0.81
Frequency of meeting since onset cough	0.37	NE <sup>1</sup>
Number of other people met in addition to contact	0.12	NE

<sup>1</sup>Not estimated as it was not included in the exploratory factor analysis

Table 2. Factor loadings matrix identified by exploratory factor analysis when two factors were retained. Bold font indicates variables that are grouped in each factor.

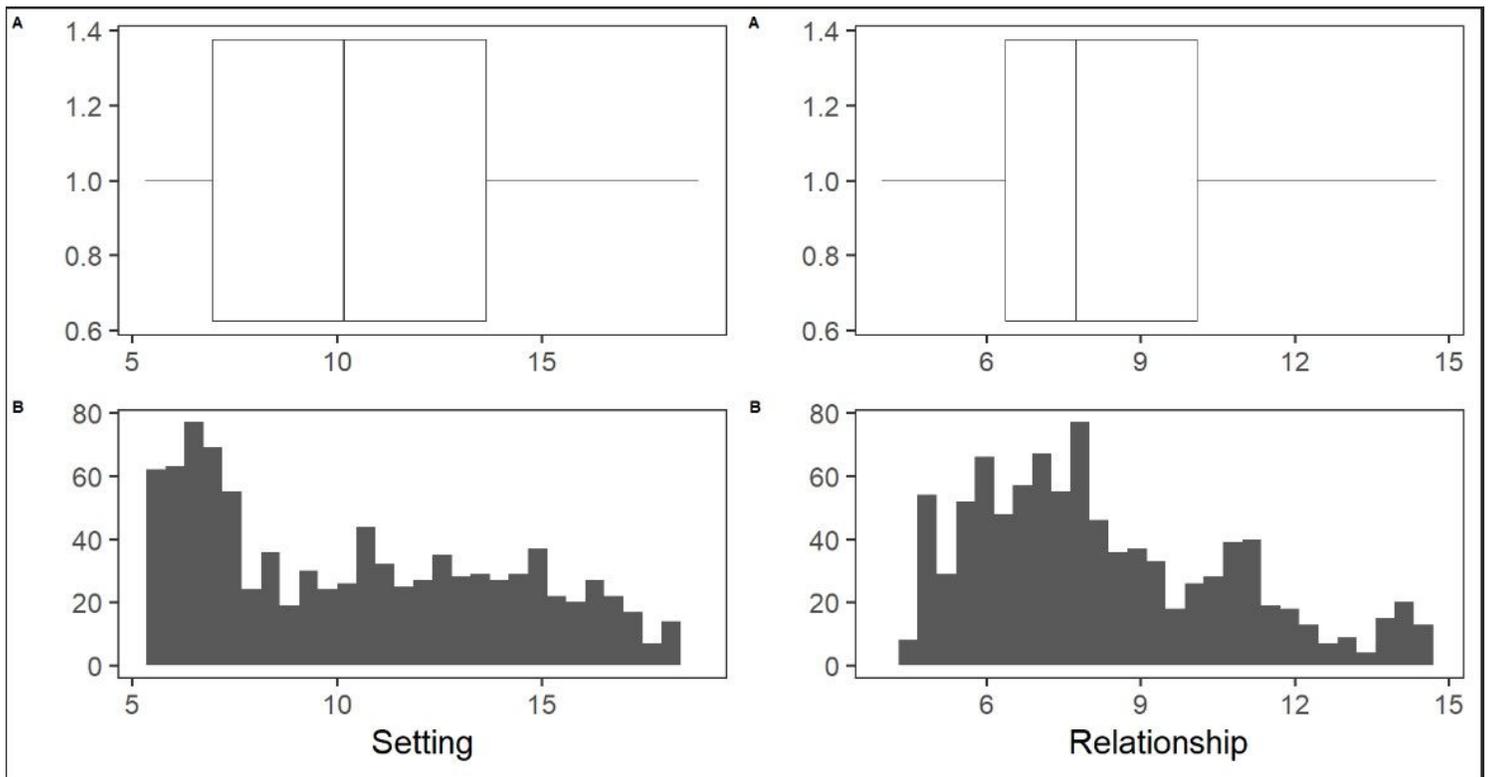
Variable	Factor1 (Setting)	Factor2 (Relationship)
Nature of ventilation at usual place of meeting	<b>0.8247</b>	-0.10068
Frequency of sleeping in same room and bed since onset cough	<b>0.81413</b>	0.07244
Contact happen indoors or outdoors	<b>0.8065</b>	-0.05468
Frequency of shared meals since onset cough	<b>0.76773</b>	0.19686
Place of usual meeting. Home TB case versus Other location.	<b>0.71749</b>	-0.04231
Frequency and duration of contact over the past month	<b>0.62832</b>	0.10417
Case trusts contact	-0.20248	<b>0.94816</b>
Case shared TB diagnosis with contact	-0.13705	<b>0.92354</b>
Care was provided by the contact in the past 3 months	0.24285	<b>0.72162</b>
Length of knowing contact	0.21973	<b>0.55998</b>
How well does the case knows contact	0.38398	<b>0.52836</b>
Means of transportation used most often with contact. None (walking) versus a type of transportation.	0.08259	<b>0.50289</b>
Contact have cough	0.03227	0.15773

Table 3. Median Setting and Relationship scores (with interquartile range-IQR) from the sub-set of 940 enrolled contacts with demographic variables collected.

Overall and stratified by gender and age of cases and contacts.

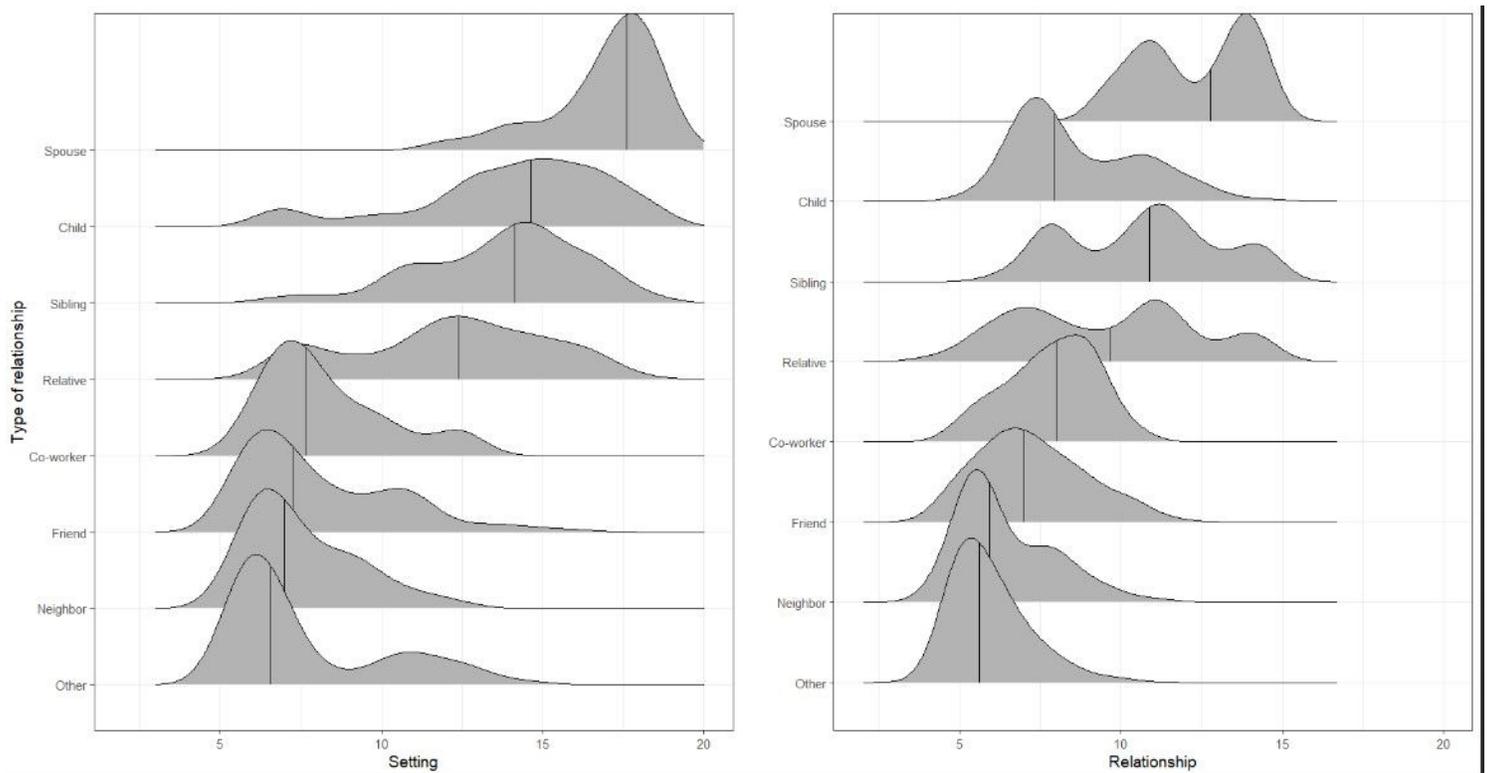
Variable	N (%)	Setting Score Median (IQR)	Relationship Score Median (IQR)
<b>Overall</b>	940 (100)	10.2 (7.0,13.6)	7.7 (6.4,10.1)
<b>Stratified by Gender Index</b>			
<b>Contact with Male Index</b>	661	9.3 (6.8,12.9)	7.8 (6.2,10.1)
00-14 years	137	12.8 (9.1,15.0)	7.2 (6.0,8.4)
15 years and greater	524	8.3 (6.6,12.1)	8.1 (6.2,10.4)
<b>Contact with Female Index</b>	279	11.5 (7.7, 15.2)	7.7 (6.6,9.7)
00-14 years	108	14.0 (9.9,16.3)	7.3 (6.6,8.0)
15 years and greater	171	10.8 (7.0,13.9)	8.2 (6.6,10.9)
<b>Stratified by Age Index</b>			
<b>Contact with index 15-24 years</b>	296	11.3 (7.4,14.8)	7.9 (6.5,10.5)
00-14 years	95	14.2 (9.1,16.0)	7.4 (6.5,8.0)
15 years and greater	201	10.7 (6.9,13.6)	9.0 (6.5,11.2)
<b>Contact with index 25-44 years</b>	521	9.3 (6.7,13.1)	7.7 (6.3,10.0)
00-14 years	119	13.5 (9.4,15.6)	7.3 (6.0,8.6)
15 years and greater	402	8.3 (6.6,12.0)	7.9 (6.4,10.4)
<b>Contact with index 45 years and greater</b>	123	10.2 (7.0,12.8)	7.6 (6.1,9.6)
00-14 years	31	12.2 (6.9,14.0)	7.2 (6.1,7.8)
15 years and greater	92	10.0 (7.0,12.0)	7.7 (6.1,10.3)

## Figures



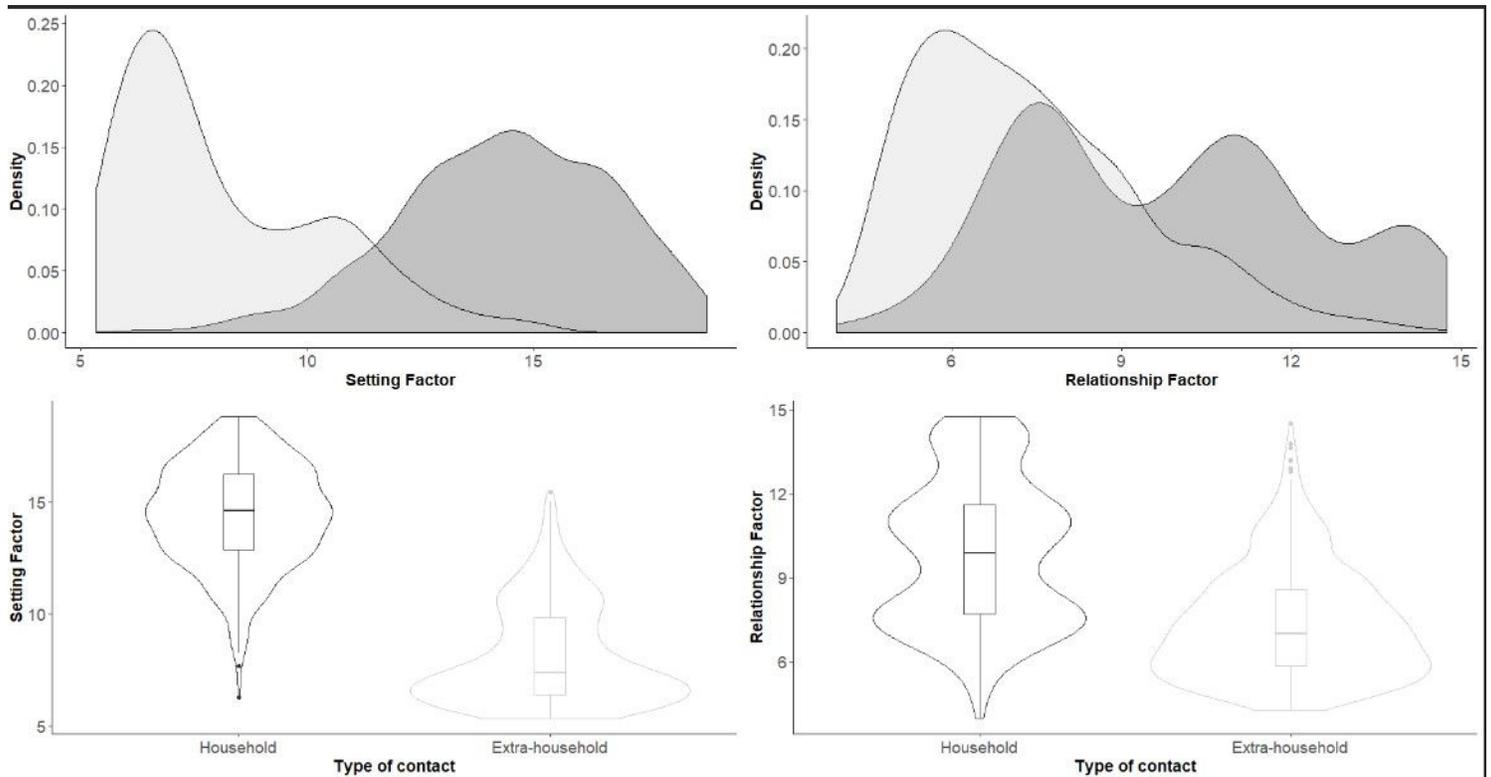
**Figure 1**

Distribution of closeness factors among the study population (n=940). A histogram and a boxplot are shown to study the distribution of the Setting and Relationship factor. Left Panel: Setting Factor. Right Panel: Relationship Factor.



**Figure 2**

Distribution of the Setting and Relationship factors according to the nature of the relationship between a tuberculosis case and their contacts (n=940). A ridgeline plot is shown to study the distribution of the Setting and Relationship factor scores, according to the nature of relationship between tuberculosis case and contact. Vertical lines indicate the median value for each group. Left Panel: Setting Factor. Right Panel: Relationship Factor.



**Figure 3**

Distribution of the Setting and Relationship factors among household and extra-household contacts (n=940). A histogram and a violin plot are presented to describe the distribution of the Setting and Relationship Factor Scores among household and extra-household contacts of the tuberculosis case. The values on the X-axis of the histogram indicates the score of each factor. The Y-axis indicates the relative frequency (density) at each Factor score. Inside each violin plot a box plot is presented. Left panel: Histogram (top) and violin plot (bottom) for the Setting Factor. Right panel: Histogram (top) and violin plot (bottom) for the Relationship Factor.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [CastellanosManuscriptEFAFINALBCMsupmal230320.docx](#)