Reviewer’s report

Title: Investigation of Household Contacts of Pulmonary Tuberculosis Patients Increases Case Detection in Mwanza city, Tanzania

Version: 0 Date: 16 Aug 2017
Reviewer: Jason I Bacha

Reviewer's report:

Review Comments on Manuscript Draft for INFD-D-17-00741

Overall Impressions: The authors have written their findings on an important - and often over looked and under-emphasized - part of increase TB case finding, active case finding efforts targeting household contacts of bacteriologically-confirmed TB cases. It is an important topic, and real-world example of successes or failures of active case finding among household contacts in a variety of settings (such as Mwanza, Tanzania) are needed to continue to inform best practices and policies. The authors are to be commended for their efforts in promoting and encouraging effective case finding during this study. While the current manuscript provides a good start, there are major revisions and considerations that need to be done to the writing, data analysis, introduction/results/conclusion sections to improve the quality and readability of the manuscript (as outlined below). Importantly, the standard TB definitions created by WHO (in their 2013/revised 2014 definitions and framework publication) need to be used throughout the manuscript to help the reader better understand their findings. There is also no mention of clinically diagnosed in the manuscript, which remains an important means to diagnose TB in many high burden countries (like Tanzania). Overall, this draft is a solid start, and hopefully with major revisions as described below it can become a high quality manuscript.

Title:

Would recommend making the title more descriptive of what was found in the study. Specifically let your reader know what you discovered about household contacts.

It would be okay to remove the adjective "active" from the title (and throughout the paper), as this paper is not comparing latent TB with TB disease. Therefore, using "pulmonary tuberculosis" would be sufficient enough to convey TB disease of the lungs.

I would also be careful about using the term "cross sectional study" in the title. "Cross sectional" study or analysis generally refers to a study that analyzes data gathered from a specific population at a specific (often single) point in time (e.g. a prevalence study or estimation). In your methods, it is not clear for how long the household contacts were followed or monitored,
but since culture results were reported, it can be assumed that the study cohort was followed for at least 4-6 weeks (depending on the turn around time of cultures in your setting). Thus, one could argue that your study is more of a longitudinal study in that it makes a series of observations on the population of interest (household contacts) over a longer period of time (e.g. at least follows them several weeks to get culture results). Consider a different title that focuses more on the study and its results (so that someone reading the title alone would get a good understanding of your findings).

Abstract:

Background: do you have any information specific to Tanzania to add; e.g. how Tanzania is doing currently in regards to household contact tracing? Would be good to add to abstract, "Currently, Tanzania is doing well/poorly with contact tracing as evident by XYZ…"

Methods: Would be good to mention the setting in the abstract (TB clinic in a hospital, community, tertiary centre, inpatient/outpatient, etc). Also would be good to add how the "contacts were traced" (via what methods, home visit, phone, etc). Contact tracing also involves more than just sputum collection, it involves TB screening questions, physical exam findings, CXR, TST, etc. If more was done to household contacts, it would be good to add in the abstract as currently it seems like only "two samples" were taken from contacts but nothing else. Also, add "sputum samples" so it is clear what kind of samples taken. Also use "analyzed" and not "screened" as the samples were at this point being used for diagnosis (and not screening). Can remove the "Prevalence of active…." sentence as it doesn't add anything to the Methods section. Is it possible to add more details about the predictors used in logistic regression (e.g. which ones, clinical, demographics, how collected, etc)?

Results: per WHO definitions, preferred term is "bacteriologically confirmed TB" (and not "bacteriologically diagnosed"). Would be good to harmonize the entire manuscript with these 2013/14 WHO definitions for TB. Use consistency with decimal point places throughout the paper. Most numbers have the tenths (e.g. 1.1%) but a few only list whole number (3%) - you need to keep your decimal places consistent throughout the paper. Technically culture positive is "Mtbc culture positive" (not TB-culture positive), and AFB-smear positive (not TB-smear positive). You don't need to use the term "bacteriologically diagnosed TB-positive contacts" (it is redundant), but instead say "29 (6.8%) of contacts had bacteriologically confirmed TB". Again, also can remove "active" for active TB throughout since this manuscript is not evaluating latent TB infection (LTBI). Using "pulmonary TB" is sufficient to show it is TB disease of the lungs.

Conclusion: the data suggests that 7 in 100 develop 'bacteriologically confirmed TB" but not "active TB" because active TB would also need to include clinically diagnosed TB (i.e. non-
bacteriologically confirmed, which your study doesn't address or mention these clinically diagnosed TB patients - e.g. patients diagnosed with TB on symptoms alone, or CXR findings, or score chart, etc). Also, for the last sentence, you need to mention "TB" contact tracing and I would not use the term "developing countries" (as some developing countries in the world have little to no TB in them) - a better term to use would be "in high TB burden countries" (like Tanzania is).

Introduction:

To first sentence need to add "primarily transmitted through…" because TB is also highly contagious and transmittable through AFB-filled pus (e.g. draining sinus).

Unclear what the sentence about declining mortality from TB (2000-2015) has to do with the paper or the study? Similarly, it was unclear why you focused on TB/HIV prevalence in the intro (since unrelated to the study findings/message). Both of these are important yes, but these mortality trends and HIV/TB co-infection rates are unrelated to this study (keep the focus and message on active case finding is effective, but not well implemented worldwide, and in Tanzania).

WHO 2016 report also estimates the number of case in high burden countries, which then address the treatment gap. For example, in Tanzania, WHO estimates 164,000 people had TB disease. However, only 62,180 cases were notified of the estimated 164,000 total TB cases (38% of the estimated burden), demonstrating the huge treatment gap (62% of all people with TB in Tanzania were not reported/treated) that exists in Tanzania. This treatment gap strengthens your introduction to show why more aggressive case finding (e.g. household contacts via active case findings) is needed - to close this gap!

Better phrasing for line 21 on page 2 would be: "A systematic review and meta-analysis of 41 studies showed that screening household contacts in middle- and low-income countries increased new case findings by 4.5%[1]."

Page 2 line 23: mentions "the case-finding technique…" What case finding technique? Similar on line 25 it says "prioritize case finding" - what type specifically? Passive? Active? Both/all?

Page 2 line 29: again, for "the case finding technique" what are you referring to here?
Page 2 lines 31-33: need to shorten/rephrase this sentence. For example, "Two studies in Tanzania showed X and Y to be factors associated with delayed case detection".

Page 2 line 39-42: active case finding is important in finding new TB cases or diagnosing new TB cases (but not in "screening active TB:" that wording is not used in the TB literature). Also, by definition "active TB" is when the TB causes signs and symptoms (compared to latent TB where the TB is in the body but controlled and not causing signs and symptoms). Thus, your sentence is difficult to understand as it says "Active TB" but then later "before signs and symptoms develop" (which by definition means "latent TB" for which we treat with IPT).

Your introduction needs more information about the current situation in Tanzania for TB case finding (active, passive or otherwise). What are they doing? Is it working? Why or why not? What is the current data for household contacts? You need to give your reader a better understanding of TB and contact tracing in Tanzania (otherwise your reader won't know if 7 of 100 if high, low, downtrending, uptrending, success or failure?), and it will strengthen your final message of the manuscript.

The introduction also needs to discuss the current literature about which factors are associated with higher rates of household TB cases. The very last sentence (row 45-48) mentions factors will be examined but there is no background data (literature review) introducing your audience to this, why it's important, what is already known, what isn't known (i.e. what knowledge gap you are filling).

Lastly, you hypothesize >5% of household contacts will have bacteriologically confirmed TB. Where did this number come from? No prior information on estimates for TB household contacts was discussed in the introduction; so it seems like a very random number to choose (some may ask, 'why not 1% or 10% or 50%?'); you need to put some background information in your introduction about world wide estimates of TB household contacts, prevalence, etc…so that your can show you are making an educated estimate (and not a randomly selected number).

Methods:

As mentioned, "cross sectional study" may not be the best description of this study (others could argue it longitudinally follows the contacts over time). Instead you could say that "retrospective chart review of XXX was done to identify new bacteriologically confirmed TB cases between…"

Also, why are you only looking at bacteriologically confirmed TB? And why not including also "clinically diagnosed TB?" You need to address this in your methods why you are only looking
at certain TB but not others. Especially in resource limited settings, often the majority of TB is clinically diagnosed!

There is no mention in the methods where "baseline" or "demographic information" of the TB cases was extracted (from charts, registry books, patient interviews, etc)?

How did you define a household contact? Sleep in the same room? Same house? Neighoring house? What if they only visited for 2 days or 1 week (and slept in same house, but only for short time, either before or after antiTB was initiated). We need a clear definition of what you defined as 'household contact.'

Page 2 line 57: "household contacts were screened for active TB" - need to write HOW they were screened for TB (e.g. questions, TST, CXR, etc)

Page 2 line 3-4: "The study team visited all case who accepted." How many refused (did not accept)? And for what reasons did they refuse? Your paper may benefit from a flow chart showing how many TB patients found in the registry books, how many reached by phone (vs how many unable to be reached - thus 'excluded'), how many accepted (vs how many refused - 'excluded') so that we can see the proportions that made it all the way to the home visit and to the sputum collection from the beginning totals.

Page 3 line 8-9: So if they couldn't produce a sputum, they weren't included (and information wasn't collected on those contacts not providing sputum)? That seems like it might be an issue, and you need to say of the XXX total contacts identified, only YYY (% of total contacts) were able to provide sputum and thus included in this study. Again, adding a figure that shows the flow chart of those included/excluded is needed

What screening tools or questions were used? You need to include these (and put the screening tool as "supplemental material" Was it the NTLP questions? NACP TB/HIV questions? WHO TB screening questions?

Page 3 lines 21-22: Please define what "clinical measurements and house parameters" you collected? Again, since no mention of factors associated with TB contacts in the home were mentioned in the intro, it is unclear to your audience which social, economic, family, patient-level, etc factors are important and need to be evaluated.

Study definitions section: you need to use the WHO 2013/14 definitions. It's not "TB positive case" it is "bacteriologically confirmed TB" (which includes AFB, Xpert and/or culture). Your
definition also leaves out "clinically diagnosed TB" patients - need to provide good justification for doing this

Sample size estimation: this section doesn't read well nor make sense. Sample size calculations aren't needed for a study like this that is looking at household contacts. I would recommend removing this as it is not necessary for this type of study, and is difficult to understand what the message is here.

Sample collection: need to make mention of the time from sample collection to the time the lab received and processed the sample. We know that over time the Mt b bacteria will die, thus the longer the sample is waiting, the lower its sensitivity. Can you provide what % were run the same day and what % of samples were stored until processed later. Also will be good to add the name of the lab where all this was done (e.g. was it BMC lab? Or multiple labs?). Why were staff running the samples blinded from the results of the other tests? That doesn't seem necessary, and nothing is mentioned to the reason/justification for doing this. You mentioned that all positive sputum cultures were sent for sensitivity testing - this is not standard practice for NTLP (or WHO for that matter); why was this done? And there is no results or information in the Results section about these sensitivity results? If it was done, you need to report the results. NB: typically, DST is only done when there is suspicion for MDR-TB (e.g. Xpert MTB/RIF showing rif resistance, persistently positive smears, clinical deterioration, or close contact with known MDR-TB case….but it is not universally done for all positive cultures).

Data management and statistical analysis: lines 16-19 mention factors analyze, but it is unclear why these factors were chosen (and others were not); these factors would benefit from a table that would show the various answer choices for these factors (and justification for why those cutoffs used in the answer choices). Also you can remove the "Factors with a 95% CI…" sentence (lines 24-25) as this is well understood knowledge about 95% CI.

Informed consent: NIMR is not included here? Was ethical clearance for this study received from NIMR? If not, why not?

Referral of TB positive contacts: this section is not clear what happened to house hold contacts where screened positive via questionnaire (but either didn't do a sputum, or their sputum was negative) - what happened to these patients (e.g. chronic cough, weight loss, night sweats but sputum negative)? Did they start ATT? Did any of the contacts get IPT? Also "children who presented visible signs" - need more explanation of what this means? What is "visible signs?"

Also need to explain "scoring assessment?" Where did it come from? WHO? NTLP? Other groups? Is it validated score chart? You might want to put the score chart used in your
supplemental materials' so that your readers could review it (since there are many slightly different scoring charts used for childhood TB throughout the world).

Results:

As mentioned before, a figure 1 flow chart of the 93 TB index cases identified would be very helpful (how many agreed to participate vs refused, how many contacts participated vs refused, how many provided sputum vs unable, how many got questionnaires vs unable, etc)

Page 5 lines 7-8: is this data referring to the 93 TB index cases or the 456 contacts?

Page 5 line 25: for children, it says 2.1% were TB positive? But what does TB positive mean? Bacteriologically confirmed? If yes, what test (Xpert, AFB, culture)?

Do we know anything about the HIV status of the TB cases or the household contacts? In the intro, there is a lot mentioned about HIV/TB, but then no HIV testing or mention of HIV again in the results? TB/HIV is a big issues in Tanzania (36% co-infection rate), and needs to be addressed…otherwise it looks like a big gap (and missed opportunity) of the study. Somewhere you need to discuss HIV status of participants, or if there is no data then give very good justification why you did not do HIV testing in a high prevalence HIV country when HIV testing for all TB cases (and presumptive TB) is part of the guidelines.

There is also no mention about the treatment of the 29 bacteriologically confirmed household contacts. Did they get linked to TB care and started on ATT? Were they LTFU? Died? What was their treatment?

Risk Factors associated with TB infection

o It is not "infection" (like LTBI) but "TB disease"

o You need to use better age desegregated data (i.e. follow WHO norms). Using >40yo and <40yo is not helpful at all. Instead should use 0-5, 5-14, 14-19, 19-25, 25-45, 45-65, >65, or something like that. That way you can say "children <5yo had a higher risk of developing TB disease" or "young adults aged 15-24" or elderly adults (>65yo) had higher risks. By using only over/under 40yo, you can make any of those helpful conclusions (and it is no surprise that a dichotomized age cut off like that didn't show any trends/association)
Similarly, you need better categories for distance than the dichotomized <1km vs >1km. That is not a very good distance cut off (unless you can support it with density population data from Mwanza), as I would guess that 99% of patients live >1km from a health care facility. It is normal in Tanzania (with a large rural population) that the majority of patients live >1km (thus it becomes a useless number to compare with). Instead think about 5km increments to use (lives within 0-5km, 5-10km, 10-15km, etc).

I also struggle to understand what you are trying to show (i.e. what knowledge gap are you filling) by looking at the triad of cough/night sweats/weight loss as a factor associated with TB. It is very well known that symptoms of TB include those three (it's in every TB text book), so essentially all you found was "patients with pTB had symptoms of pTB" which is well known and to a degree obvious.

Other issues

- What type of TB did these patients have? Did they all have pTB? Or any EPTB?
- HIV status of participants? Or justification for why not tested/reported?
- Past treatment of TB in these household contacts? Which would show re-activation perhaps? Could be a factor to review in your regression analysis
- What about the sensitivity (DST) results mentioned above (that all positive cultures were sent for sensitivity)? What did it show? Any MDR TB?
- Were any of the index cases children/adolescents? If yes, were they as infectious to their households as adult index cases?

Discussion:

First sentence could say "Using active case finding among household contacts of bacteriologically confirmed adult pTB cases, we found that 6.8% of household contacts who produced a sputum sample also had bacteriologically confirmed pTB"

Page 5, lines 43-46: I am very confused by where this data comes from (the "690 per 100,000 people detection rate achieved by passive diagnosis"). Every year, I read and review the WHO Global TB Report (specifically the Tanzania country profile), and have not seen this data or statistic reported. I re-reviewed the 2015 report (NB: 2016 report provides more comprehensive up to date data), and did not find anything on "passive diagnosis" or this 690 figure for detection
rate in Tanzania. As this is a major point of comparison you make, please provide where in the WHO report this detection rate by passive diagnosis can be found.

Paragraph two (page 5 lines 48-53): this is very well known throughout the literature (that culture performs better than Xpert which performs better than AFB smear). As that is not a focus of your paper, would remove this and just focus on the message of active case finding.

Page 6 line 1-2: you mention HIV in the Discussion; but don't report anything about HIV status in your results. Need to rectify this inconsistency (i.e. you need HIV in results). Otherwise, you cannot mention HIV in Intro and Discussion but ignore it in results/findings of your study.

Page 6 line 8-9: "Creating awareness and boosting health service use can improve this situation" is vague and hard to understand. What is the message you are trying to say here? What kind of 'awareness" needs created? How exactly do you "boost health service use?" Can improve what "situation?"

Page 6, lines 6-10: as mentioned above, your 'distance to health facility' measure is severely limited as you only used 1km cut off. A better analysis would be looking at 5km increments and seeing if a true effect exists. Similarly, far distances cause barriers to care (cannot pass roads, financial burden to travel, don't go too far if local dispensary/pharmacy closer, etc), which have been studied/reported and deserve mention.

Page 6, lines 11-13: "This high proportion may be due to fear and stigma…” It is unclear how this conclusions were made, or how it is supported by the data. If being married was associated with higher risk for contact having TB, how this is related to fear, stigma, and couples seeking care together? One theory that seems possible is that married people share a bedroom with more people (i.e. their spouses), and are in closer contact with more people (i.e. their spouses) than a single person. Thus, perhaps that closer contact to another person (spouse, children, etc) more exposure to an infectious person in the house?

Page 6, lines 16-18: Why did you anticipate this? Thes hypotheses weren't mentioned in the intro? There is good reason to anticipate this, but you need to harmonize this train of thought throughout the manuscript by mentioning other studies or literature that supports these factors as risks in the Introduction, so that you can write something like this in the Discussions (and it follows naturally from what you wrote in the Intro).

Page 6 line 22-23: Unclear how living outside Mwanza city would negate the impact of crowding, relationship of contacts, education level. Whether living in urban or rural, these are
well known risk factors of TB, and so should also be risk whether living in or out of Mwanza city centre.

There is no "limitations of the study" section in Discussions - needs to be addressed

There is no "future directions" mentioned in the study - needs addressed

Again, no mention of clinically diagnosed TB in the discussion, use of questionnaires/screening tools, or score cards and their importance in active contact screening. These are all very important (and WHO-encouraged) tools for screening and diagnosis TB (in household contacts or otherwise), and deserve mention.

Also no specific mention of active contact screening in children vs in adults? Do we need different approaches to active screening in children vs adults? Are there unique challenges between the two? What did you find in children vs in adults? What does your data support, or what conclusions would you make based on your findings about approaches to children vs adults?

Conclusions:

Well written, but would consider leaving out the sentence about AFB microscopy. That wasn't the goal or aim or design of the study to evaluate the performance of the three types of sputum tests. Also, this fact is well known (and well published) that AFB microscopy performs worse than Xpert and culture (so not adding anything new here).

Declarations:

Ethics approval and consent to participate section: this is already covered under Methodology (Informed consent and ethical clearance), so why repeating it again here?

Tables/Figures:

Table 1:
  o Table would benefit from a "p-value" column to show us if significant difference existed for these characteristics.
Need to add a "n=456" at the end of the table title

Age: would desegregate age based on commonly accepted (e.g. WHO) age segregation for children, adolescents, young adults, middle age, elderly

Education: "Education level completed" would be a better title (if that is indeed true)

Marital status: may need to add 'widowed/divorced' or at least explain how any widowed or divorced participant was handled and classified

Family members: need to add a 0-1 category (even if zero had "0-1"). Also it is unclear why these cutoffs were chosen. Typically, all the categories are the same range (e.g. 0-2, 3-5, 6-8). Unclear why 2-5 (4 options) then 6-10 (5 options), then 11-20 (10 options) instead of a standardized multiple of 3 or 4 or 5?

Distance to facility: as mentioned, would be much more valuable using multiple distance ranges (e.g. 0-5km, 6-10km, etc)

Smoking: who does this refer to? The participant him/herself? Or within the house there is a smoker? Need to clarify that.

For symptoms listed: need to explain why these particular symptoms were chosen to be reported. The generally accepted TB screening questions (e.g WHO, NTLP) are cough for >2 weeks, fever > 2 weeks, weight loss, known TB contact, hemoptysis, or excessive nights sweats for > 2 weeks).

Of the 7 symptoms you've examined, only 1 (cough blood) meets these definitions

You need to either explain/justify why the standard symptoms for TB weren't used, or clearly list that it was cough for "greater than 2 weeks", night sweats for "greater than 2 weeks", fever for "greater than 2 weeks"

Otherwise, these symptoms used are not very specific for TB

Table 2:

If categories of characteristics are changed (as suggested above, especially for age, distance to facility) then need to update the analysis and table

Number of people per 5 square meter: why did you dichotomize the answer choices (one person vs >1 person)? Instead of one vs 1-2 vs 2-3 vs >3 (or similar)?

For Triad, need to clarify the following things about these:
Were they all for >2 weeks duration (all three)?

It is written as "AND" - thus you mean that the 4 "yes" had all three (cough, night sweats, weight loss" and not "and/or" combination?

Why is 'fever > 2 weeks" not included? It is also a 'classic' symptom of TB

Figure 1:
- Need to use "Bacteriologically confirmed TB" instead of "TB positive confirmed"

There are no side boxes (pointing to the right) that signify any participants who refused to participate, who were not eligible/excluded, or whose data was unable to be obtained (e.g. moved, died, refused, etc). Currently it looks like you had 100% participation of all those identified, and then 100% participation of all of their household contacts, and then 100% success in gathering all the data for all the patients. That is exceedingly rare in any study, and want to confirm this is indeed true.

This flow chart would also benefit from showing us the 'symptoms' of the 456 contacts. How many had any of the major 4 clinical symptoms? Were any of the bacteriologically confirmed cases asymptomatic.

When you say "contacts tested" what do you mean? I think you mean "contacts identified 456" then the next box would be "contacts screened via TB questionnaire" then two boxes below that "screened positive via questionnaire" and "screened negative via questionnaire" then below the screened positives you'd have "obtained sputum samples" then below that the positive sputum samples (by Xpert, smear, culture)

A final row showing treatment would also look good (e.g. antiTB initiated, or IPT initiated)

Other general comments

As mentioned, it would benefit the authors to use the terminology and definitions as defined by WHO in their 2013 "Definitions and reporting framework for tuberculosis."
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

No

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I recommend additional statistical review

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?
If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal