Author's response to reviews

Title: Isolation of non-tuberculous mycobacteria (NTM) from pastoral ecosystems of Uganda: Public Health significance

Authors:

Clovice Kankya (clokankya@yahoo.com)
Adrian Muwonge (adrian.muwonge@nvh.no)
Berit Djønne (berit.djonne@vetinst.no)
Musso Munyeme (munyeme@yahoo.co.uk)
John Opuda-Asibo (opas_jn@yahoo.com)
Eystein Skjerve (Eystein.Skjerve@nvh.no)
James Oloya (joloya2001@yahoo.co.uk)
Vigidis Edvardsen (vigidis.edvardsen@vetinst.no)
Tone B Johansen (tone.bjordal-johansen@vetinst.no)

Version: 2 Date: 28 February 2011

Author’s response to reviews: see over
Comments to review and reviewers comments

Title: Isolation of non-tuberculous mycobacteria (NTM) from pastoral ecosystems of Uganda: Public Health significance
Version: 1 Date: 29 December 2010

The paper has been revised according to reviewers’ comments and formatted to conform the journal style. We have corrected typographical and grammatical errors, hopefully to the benefit of the manuscript.

Below are some specific comments related to each reviewer’s comments.

Comments to Reviewer’s report: Reviewer: Rachel Thomson

Introduction.
I found the introduction to be long and not particularly focused to the research question.

The introductions has been shortened down and straightened up.

There could be less generalized comments about NTM disease and more discussion about which NTM have been found to cause disease in humans in Uganda, and how those cases presented and were diagnosed.

There is not much knowledge about NTM in humans in Uganda, but the discussion about NTM disease in general is shortened down.

The aim of the study is a little vague and could be elaborated.

The aim of the study is clarified.

Some discussion about the methods of isolation of mycobacteria is also needed, to support the methods chosen by the authors.

Some general discussion of the methods is included, and a brief comment on the need for combining microbiological methods with epidemiological studies mentioned.

The case definition is poor, and whilst I realize the microbiological confirmation of disease is not always provided in developing countries, some criteria for selection of cases is needed.

This is real problem, as different authors use different approaches. We have tried to clarify this more in the M/M section.
Methods.

The methods are well described, though some of the information provided is superfluous to the discussion of results. E.g. the lat/longitudes and populations of the districts are not particularly relevant unless they are useful in the analysis - so cases per 100 000 population could be provided, and an analysis of yield of mycobacteria against geographical variables etc.

The superfluous information in the methods section is omitted. Knowledge about the cases per 100 000 population does not exist, mainly due to the fact that culture and identification of mycobacteria are very seldom done for these patients. This is now explained in the introduction. Analysis if yield of mycobacteria has not been done, as this is not a quantitative study. Due to the harsh decontamination necessary, quantification from these samples is difficult.

Sample size determination of the questionnaire survey: need to separate NTM and TB incidence/prevalence. As the numbers are likely to be contaminated by cases of M. tuberculosis, you need to be more clear about the case definition.

How did you identify a household with a case – and how was that case diagnosed? What form of disease did they have?

Hopefully this is now more clear. Data from bacteriological analyses are presented linked to sample category. For statistical inference, a case household was defined as a household with NTM presence in water, soil, and or animal fecal matter.

It took me a while to work out that the survey results were being compared to the isolation of NTM from the environment – this could be more clearly articulated in the aims.

The comparison of the survey results being compared to the isolation of NTM is clarified, both in the aim and in the method section.

Questionnaire: How was the questionnaire “standardized”? In these parts of Uganda there are different languages, and the questionnaire was standardized making one questionnaire in English for all participants, and translating it into the different local languages, so that everyone interviewed got the same questions. This is now clarified in the text.

You state “alongside the questionnaire, samples from water...” – were the samples collected the same day as the questionnaire were administered or Separate – what was the time frame?

The samples were collected the same day as the questionnaire was administered, except for valley dams that were visited every fortnight. This is now clarified in the text.

How did the timing of administration of questionnaire and the sample collection relate to the timing of the diagnosis/illness of the index case in the household?

The samples could not be timed linked to any disease situation, and relate to the visit and samples available during the visit(s).
Samples: what was the rational behind the chosen methods of sample size (30ml),
decontamination, concentration and culture and how might these methods have impacted
on your yield?

Ideally the samples size for water samples could have been bigger, but the samples
had to be transported from Uganda to Norway, as there was no possibilities to
culture these kind of samples in Uganda.
The water samples contained a lot of organic material, containing other bacteria and
fungi, and therefore the decontamination had to be the same as for soil and fecal
samples. This could have affected the yield of mycobacteria. This is incorporated in
the discussion.

Surface layer soil was not collected – how deep did you go to collect the soil
samples then, and how close to the water edge were they?

Approximately 5-20cm were moved below the soil surface at each marked spot
before collecting approximately 2g of the soil sample.

It wasn’t clear from reading how the water samples were represented until I read
the supplements at the end. You could include in the text (and refer to the table)
to better describe the number of household vs dam vs stream samples. There
were 231 households, yet only 130 drinking water samples collected ??why
This has now been included and explained in the text.

Given the finding of the seasonal differences in yield, I think you need to show
when the samples were collected – or in the results, show denominators/percent
rather than absolute numbers per month.

This is included in the text, the percentage of samples collected each month in the
methods section, and the number of isolates each month in the result section.

Isolation and identification: why was NCBI Blast chosen over other databases
such as Genbank and RIDOM ? What did you do with isolates that came back as
M. species ? what level of match did you accept?

Obtained sequences were edited and analyzed in Bioedit
(ftp://www.mbio.ncsu.edu/BioEdit/bioedit.html) and sequences were compared
to available sequences in GenBank by the NCBI Blast sequence alignment tool
The isolate was determined to species based on the maximum score and
maximum identity values on NCBI Blast alignment, a maximum score and
maximum identity of ≥ 99% were accepted. We did not use RIDOM, due to the
habit of using NCBI Blast. RIDOM is an alternative of the same quality. Similar
mycobacterial laboratories also use NCBI Blast.
RESULTS.

Microbiological results are not well expressed. What proportion of samples had positive growth, (and subsequent ZN positivity), no growth or contamination/overgrowth by non acid fast organisms?

Forty eight (48) i.e (15.0%) of the 320 environmental samples showed mycobacterial growth while the remaining 272 (85.0%) samples, no growth of mycobacteria were detected. However, of these many other samples had to be discarded without further tests because of massive overgrowth by other non acid fast organisms and fungi. This is included in the text.

As mentioned the yield for each month needs to be expressed rather than absolute numbers. You could examine the yield per month from each type of sample also.

The yield for each month is included in the text. The detailed yield for each month for each type of sample is not included as it was found to be very detailed.

Because of the low yield of mycobacteria and the bias introduced by the seasonal variation in yield, I’m not sure that the findings from the multivariable logistic regression proportion are valid.

There are always uncertainties around these statistical analyses, and hopefully the paper now discusses this properly.

The wording “drinking of untreated water relative to treated” is misleading, as it implies ‘drinking’ the water is a risk factor for disease acquisition, when I think you are meaning that the ‘untreated water was more likely to contain mycobacteria that treated’. The use of the phrase “risk factors for exposure” is also confusing. You can really only say “risk factors for presence of mycobacteria” as exposure requires human activity which you did not assess in detail.

The results concerning the radar and the multivariable regression are clarified. The paper deals with isolation of NTM from the environment around the households and the questionnaire deals with behavior that can expose humans to these NTM. This is now clarified both in the aim, the methods and results.

DISCUSSION.

Need to address the results in more detail – ie factors that may have affected the yield – methods chosen, seasonal sampling, etc and how these may have affected what you can conclude “

A number of factors that could have affected the yield have been precisely discussed in detail in the text.

An average of 15% of the environmental samples contained mycobacteria” – average of what?

It is 15% of the samples, not an average. This is now changed in the text.

Yield from soils around water sources highest in this study – could this be because the decontamination method is best suited for soils, and perhaps be too
strong for drinking water?

This is not really clear, as all these samples represent methodological problems.

The third paragraph is one of the major limitations to this study and needs to be stated as such.

This limitation is stated as major, and discussed as such.

Rather that focus on an exhaustive description of how the species you have found have also been found in humans, you would be better to discuss your own results and the implications they have for the people studied and whether they are generalizable to other populations.

The description of the different species found in humans are shortened down, and the possible generalization of our results are discussed.

The last three paragraphs of discussion need attention. The term “exposure risk factor” is misleading and the first sentence is therefore confusing. All you can conclude is that these are factors that may be associated with the environmental isolation of mycobacteria. You state that “Other studies show that contaminated water sources are known to provide the foci of mycobacterial infections in humans and animals” Be careful here. Contaminated bronchoscope cleanining fluid has been linked to isolation of mycobacteria from human samples – but these are mostly false positive isolates collected through the contaminated scopes. Hot tub lung is a rarer condition that has been linked to exposure to contaminated water. However the vast majority of NTM infections have not necessarily been linked to environmental exposure, is less well defined and the subject of ongoing research. Genotyping studies matching environmental isolates and human isolates are needed and have not been conclusive to date. “stipulates” is the wrong word here.

The comment is correct, and these speculations have been omitted from the paper.

The last two paragraphs are also a problem. Have you or others shown that boiling of drinking water can reduce isolation of NTM?

It is well known that NTM as other mycobacteria are killed by boiling.

What is the evidence that drinking water is the most important route of human and animal exposure to NTM? Soil and water are so closely related that it may be gardening and use of wet soil in agriculture that is more important.

This will always be the case, and it is correctly pointed out that this is a problem.

Hopefully the text has enough discussion also in this point.

There are several minor typographical and grammatical errors that need attention, but given that the paper needs to be revised I have not commented on these at this stage.

Hopefully this has now been sorted out
**Comments to Reviewer’s report, Reviewer: Rebecca R Prevots**

Major Compulsory Revisions

1. **Methods:**
   a. **Study design and sampling:** “…selection of study households from a list of households was done based on a systematic approach…” . Authors should clarify that this is a systematic sampling approach (ie every xth household on the list).

   The selection of study household from a list of households was done based on systematic sampling approach (ie every 5th household on the list was considered).

   This is clarified in the text.

   b. **Collection of samples:** It is not clear how the sources of water ascertained in the questionnaire correlate with the samples taken for mycobacterial isolation. This should be clarified. Is there any way to describe distance from households, frequency of use, etc.

   This is clarified in the text.

   c. **Seasonality:** were the authors interested in ascertaining the seasonality of NTM concentrations? This should be clarified in the introduction; that is, why did they choose to sample across different months, rather than perhaps focusing on just the wet seasons when concentrations would be highest.

   We were not studying the seasonality of NTM presence in the environments, but did the study during the rainy season because others have found that to be a good period to isolate NTM. This is mentioned both in the introduction and in the discussion.

2. **Data analysis:**
   a. The outcome and exposure variables are unclear. Are the authors trying to correlate mycobacterial presence in a given household water supply with the use of that water supply by certain types of animals? The specifics of the data analysis are confusing, and this part needs more detail. In the methods related to logistic regression, the authors state that the outcome of interest is “isolation of mycobacteria in the household environment”, but it appears that the Spiderplot relates to “potential exposures”. Are the authors not relating the actual presence of mycobacteria (yes/no) to the documented exposures ascertained in the questionnaire? This section is critical, and needs clarification. Similarly, the authors state “summary statistics of the explanatory variables with respect to occurrence of exposure routes to NTM in the pastoral community environment were carried out using the tab commend of Stata”- this part needs more clarification. Explanatory variables are being related to exposure routes rather than prevalence of exposure to NTM? Do the authors mean related to water or soil sources with high prevalence of NTM?

   This has now been clarified and the outcome (detection of NTM in any sample in household) more directly stated.
b. The authors state “the validity of the models in explaining usefulness of the explanatory variables”... do the authors mean they are trying to assess the fit of the model? Assessing validity is beyond the application of model fit statistics. Suggest rephrasing.

We only assessed the statistical fit using the Hosmer Lemeshow test. Text has been changed.

3. Results:

a. If the main unit of analysis is households, then at least the main results should be presented in terms of the number and proportion of households that had mycobacteria (either proportion that had any, proportion that had a certain concentration, or proportion that a certain type known to be pathogenic).

The number of households where mycobacteria were detected in their environment is stated in the manuscript.

Currently most of the information at the beginning of the results relates to the environmental samples without relating these to the households.

This is hopefully clarified now.

b. The use of the “radar/spiderweb” method should be clarified; further explanation of this method of analysis would be helpful to understanding the results. What does the numeric scale in the middle refer to? Is that the proportion of water sources or of household with any mycobacteria in the sample?

This has been clarified in the text.

c. Table 4: the title states “multivariable logistic regression analysis showing the household health concern...”. Do the authors mean health concerns or are they relating specific exposures to presence of mycobacteria? This point is not clear.

The heading of table 4 has been changed, and hopefully clarified.