Author's response to reviews

Title: Predicting Smear Negative Pulmonary Tuberculosis with Classification Trees and Logistic Regression: a cross-sectional study

Authors:

Fernanda C Q Mello (fcqmello@hucff.ufrj.br)
Luiz Gustavo V Bastos (gvbastos@hotmail.com)
Sergio Luiz M Soares (scoperao@bol.com.br)
Valeria M C Rezende (vmrezende@hotmail.com)
Marcus B Conde (marcusconde@hucff.ufrj.br)
Richard E Chaisson (rchaiss@jhmi.edu)
Afranio L Kritski (kritskiagmail.com)
Antonio Ruffino-Netto (aruffino@fmrp.usp.br)
Guilherme L Werneck (gwerneck@nesc.ufrj.br)

Version: 2 Date: 4 December 2005

Author's response to reviews: see over
Dr Peter Newmark  
Editor-in-Chief, BMC-series titles

Dear Dr. Peter Newmark,

On behalf of the authors of the manuscript “Predicting Smear Negative Pulmonary Tuberculosis with Classification Trees and Logistic Regression: a cross-sectional study”, I would like to thanks the reviewers for their thoughtful comments. We have revised the manuscript to take care of their considerations. Following are answers and comments to each of the question they raised. We hope we were able to meet your requirements. Please contact us if you need any further explanations.

Sincerely,

Guilherme L. Werneck, MD, DSc  
Corresponding author

Correspondence should be addressed to Dr Guilherme L Werneck – Instituto de Medicina Social/IMS, Departamento de Epidemiologia, Universidade do Estado do Rio de Janeiro (UERJ), Rua São Francisco Xavier 524, 7\textsuperscript{o} andar, Bloco D - Maracanã - Rio de Janeiro - RJ - Brazil - 20559-900 – Phone: +55-21-25877303 – Fax: +55-21-22641142  
email: gwerneck@nesc.ufrj.br
Reviewer 1 – Dr Brian G Williams

Dr. Williams asked for two additions:

1) Inclusion of 95% confidence intervals for the estimates of sensitivity, specificity etc:

   We have added 95% confidence intervals for all estimates in table 3.

2) Give estimates of the predictive power of the conventional system used to diagnose smear negative TB.

   Unfortunately these estimates are not available, simple because there is not a definite criterion for making such a diagnostic. In general, patients undergo different kinds of approaches, some of them directly being submitted to empirical treatment, some undergoing different kinds of diagnostic tests.

Reviewer 2 – Dr Juan Wisnivesky

Dr. Wisnivesky recommended several revisions:

1) How many of the patients included in the study who had previously two consecutive samples of spontaneous sputum that were acid fast bacilli smear-negative, underwent sputum induction and/or bronchoscopy? How many additional sputum samples were obtained per patient and for how long were these cultures kept to assess MTB growth? If the gold standard for some patients was based on MTB culture from bronchoscopy samples this may not be adequate as the sensitivity of BAL samples is less than for spontaneous or induced sputum. The authors should provide data regarding the number of patients undergoing these procedures.

   All 551 patients underwent sputum induction, but bronchoscopy was performed in 331 patients (60%). Cultures were kept to assess Mycobacterium tuberculosis growth for 60 days. As required we included this description in the methods section (Procedures sub-section).

2) Who reviewed the chest radiographs (radiologists vs investigators) and was the reviewer blinded to the results of TB cultures and clinical predictors?

   One chest physician reviewed the chest radiographs using a standardized form (as described in page 4). Evaluation of chest radiographs was not blinded to clinical predictors, but was for the results of TB cultures. This last phrase was included in the methods section (Procedures sub-section).

3) Was the decision to start treatment standardized? If not, this definition may be weaker. The authors should perform a sensitivity analysis to assess to what extent their
conclusions change by using the more stringent case definition based only on the results of the TB cultures.

Yes, based on the guidelines of the American Thoracic Society - Diagnosis Standards and Classification of Tuberculosis in Adults and Children. Am J Respir Crit Care Med 2000, 161:1376-95. Theses cases were 60 among the 270 confirmed cases of tuberculosis (22.2%). Anyway, results from a logistic regression model considering just the bacteriologically confirmed cases, provided similar results (area under ROC = 68%, sensitivity = 66%, specificity = 61%, PPV = 61%, NPV = 66%).

4) Statistical analysis (Page 6), consider not reporting positive and negative predictive values as these depend on the prevalence of disease, which was considerably high in this cohort (almost 50%).

Although we agree that predictive values are only applicable in settings with such a high prevalence of disease, we think it is important to show the estimates. Readers might consider these values appropriate or under- or overestimated considering the prevalence in their own setting.

5) It’s not clear why the investigators decided to use a significant level of 10% to include predictors into the final logistic model. Additionally, the authors used backward elimination; however automatic model selection may exclude important variables. A manual process may be preferable.

We think that a significant level of 10% is acceptable for building predictive models (not causal models). This criteria has been used elsewhere (e.g., Weber MW et al. Predictors of neonatal sepsis in developing countries. Pediatr Infect Dis J. 2003;22:711-7). Anyway, it should be noted that our evaluation of the performance of the models (sensitivity, specificity etc) uses 95% confidence intervals. We used backward elimination but not in an automatic fashion, we actually did the selection using a manual step-by-step process.

6) Why use two techniques to build the model?

Part of the objectives of the study is to compare such models and verify whether one of them performs much better, and then should be the choice. We could not select one of them, however, because both have similar predictive power. Anyway, as we mentioned in the text, this is just a first tentative of building such models, and we need to validate them in other settings.

7) The authors should give more detailed definition for the clinical predictors to allow reproducibility of their findings. The percentage of HIV positive patients should be also reported. Additional information regarding other clinical predictors (such as fever, self-
reported PPD, TB risk factors, history of shortness of breath) that have been reported as important predictors of TB in other similar studies should be provided.

We understand that most of the clinical predictors used are clearly defined or self-explanatory, just the definitions of alcohol abuse and AIDS were not described in the text, but we give references for the criteria used (references # 24 and # 25). The percentage of HIV patients is 41%, and it can be extracted directly from the data presented in table 1. Regarding other predictors, we described in the text (Results section) that in the logistic regression analysis (before revision):

“Non-specific clinical symptoms such as headache, pain in the face and throat pain were not significantly associated with SNPT diagnosis (data not shown). Neither the length of symptoms nor previous use of antibiotics was significantly predictive of SNPT (data not shown).”

We have actually tested for many other predictors, and we then changed the text to take care of that information. Now the text reads like that (underlined the new data included):

“Non-specific clinical symptoms such as fever, headache, pain in the face and throat pain were not significantly associated with SNPT diagnosis, as was not dyspnea, thoracic pain and sweatiness (data not shown). Neither the length of symptoms nor previous use of antibiotics was significantly predictive of SNPT (data not shown).”

Unfortunately we did not have information on PPD, as we have already mentioned in the Discussion section.

8) The authors used a cutoff for the models that gives a sensitivity of approximately 70% and a specificity of 60%. This cutoff is not very helpful as it would not considerably change the pretest probability of TB. I would suggest that the authors report the sensitivity/specificity and/or LR of different cutoffs or alternatively use a cutoff that either maximizes sensitivity or specificity.

We value Dr. Wisnivesky opinion, but choosing a cutoff that maximizes both sensitivity and specificity is a common procedure and important when considering these models as a first step for building better models that might take them in consideration for comparison. In addition, although the global performance of the models is not quite remarkable, as we mention in the Discussion section, we consider these models useful for separating groups with different levels of probability of being a case, then separating those with high and low chance from those patients in the intermediate range, who should undergo more extensive investigation.

9) The authors should consider discussing in further detail the potential problem of spectrum bias as almost 50% of the patients had TB.
We agree, and decide to include the following statement in the end of the Discussion section:

“Readers should take care when interpreting our results, particularly the predictive values, since prevalence of SNPT in this setting was considerably high.”

Reviewer 3 – Dr Pierre Tattevin

Dr. Tattevin asked for two additions:

1) Smoking seems to be negatively associated with the diagnosis of SNPT in table 1 (p<0.01). However, the authors did not even mention this interesting point in the result section, neither did they comment on this in the discussion. This may be related to the wide range of differential diagnosis in heavy smoker presenting with lung illness, but this has to be developed.

We really found the result interesting, and agree that it might be related to the wide range of differential diagnosis in heavy smoker presenting with lung illness. But it also might be related to confounding effect by age, since prevalence of smoking vary from ~18% (age ≤25 years) to ~50% (age >60 years). Anyway, we did not want to stress the finding because: (1) in multivariate models it did not remain significant (p>0.1), and (2) if we decided to mention that, we think we would have to make comments on all other variables that were significant in the bivariate analysis but were not confirmed in the multivariate analysis, which will increase the length of the paper with too much information not directly related to building a predictive model.

2) Page 3, last line: the term Human Acquired Immunodeficiency Syndrome doesn't need to be detailed (AIDS is now universally accepted as an abbreviation)

We removed the term Human Acquired Immunodeficiency Syndrome leaving just AIDS.

3) Page 7, last paragraph : "...negatively associated with SNPT..." instead of "...negatively associated SNPT..."

Corrected.

4) Table 3, last line, there is a mistake: the authors probably mean CART (validation sample), and not CART (original sample)

Corrected.