Reviewer's report

Title: A decision aid to rule-out pneumonia and reduce unnecessary prescriptions of antibiotics in primary care patients with cough and fever

Version: 5 Date: 27 January 2011

Reviewer: carmela cappelli

Reviewer's report:

Major compulsory revision:

1) The authors state that they snipped off branches of the classification tree to obtain a simple decision rule. Actually it is well known that tree based methods tend to overfit the data; thus, snipping off branches (the correct word is "pruning the tree") is meant to remove the overfitted branches of the so called maximal tree that can be seen as an "exploratory tree" (that describes the data) in order to get a "decision tree" (that can be used to predict the response class of new observations). Of course pruning also simplifies the tree structure but, as just said, it is mainly employed to get a reliable decision rule (to get an insight into this issue the author can give a look to the paper: Cappelli et al 2002, CSDA).

The author should clarify this relevant point: which pruning procedure they employed (classical CART??) and, since usually pruning produces a sequence of increasingly pruned subtrees, how they selected the final decision rule that is displayed and tested in the paper.

2) It's well known that a major advantage of tree based methods is the automatic selection of the most relevant covariates (explanatory or independent variables) thus, there's no point in pre-selecting the variables to employ in growing the classification tree. I strongly recommend that the authors run the procedure with as much variables as possible among those available (of course I expect that they exclude those corresponding to diagnostic tools whose employment they aim to avoid) ; they might find a better decision rule. On the contrary for the logistic model, as for any regression model, there's a problem of selecting the so called "parsimonious model" i.e. a model with a limited number of relevant variables but a satisfactory explanatory power; the authors should clarify if they have employed some variable selection procedure and which one; in case they did not they should. Indeed, at least from a statistical point of view, the model they derived seems too "large" and thus impractical.

3) how they tested the difference in antibiotics prescription with and without the classification rule? The author should spend some word on this point.

Minor Essential Revisions:

page 3 (two lines from bottom): the sentence "...safely rule of pneumonia t in patients..." what is letter t??

page 7 lines from bottom: the sentence "...I.e. The diagnostic..." why the capital
letter for "the"?

Discretionary Revisions
From page 5 on: the author employ the term "multivariable model" it's not a statistical terminology; statisticians rather use "multivariate" or "multiple but maybe it's medical science terminology.

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

'I declare that I have no competing interests