Reviewer's report

Title: Novel study design to assess the utility of the COPD Assessment Test in a primary care setting

Version: 2 Date: 28 January 2013

Reviewer: Sarah Chapman

Reviewer's report:

Major compulsory revisions

1) I would find more detail on the methods used helpful to support the claim that they were 'robust'. In particular, how the physicians were recruited including rates of consent, how many took part, number of assessors, country of origin, training. For example, the manuscript states that 2 or 4 locations were used, depending on the number of physicians required, but it isn't clear how this number was reached. Also, I would like to find out more about how physicians in the CAT+ arm were trained. Davidson et al (2003) in Implementation Science, gave guidelines for reporting the content of behaviour change interventions which could help to frame this description.

2) The guidance/training given to assessors on how to rate consultations should be summarized within the manuscript. For example, what consisted 'acceptable' understanding. For patients with only 4 issues, the maximum A score might be lower- how this was dealt with should be discussed. The authors might wish to give examples of physician consultations where issues were addressed or not to assist with understanding.

Minor essential revisions

1) The authors might wish to consider the impact of different countries on physician communication skills. Under physician recruitment they state that 'location is unlikely to affect physician communication skills'- I would think that the different healthcare systems and social norms regarding healthcare systems in different countries might have an impact on consultation content. Further discussion of relevant literature around this would help to put these results in context.

2) 'physicians were consented' should be 'physicians consented'.

3) The authors might wish to include more discussion of the hierarchical nature of their data. My understanding of the method is that the assessments of physician performance could be clustered within actors, participant physicians, physician group, country and assessor. Any implications that this might have for analysis strategies (e.g. using ICCs) or power should be discussed.

4) The researchers use the term 'patient issues' in several places. A definition of this term, and description of how these issues differ from the medical histories section of Table 1 is needed.
Discretionary revisions

1) The authors state that physicians were unaware of the goals of the study but from the recruitment strategy it seems possible to me that some of the participants might have been able to guess what the study goals were. If there is any information available to support the assumption e.g. physician checks, this would help support this claim.

2) My understanding is that some actors and assessors were used to evaluate consultations in countries which they were not native to. This approach might have some strengths (e.g. allowing for the same actors/assessors in different countries) and some limitations (e.g. if assessors were more or less well-disposed to physicians from different countries). A description of where the overlap took place, and discussion of these strengths and limitations, and any steps to overcome these effects could assist the reader in understanding the validity of these techniques.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Not suitable for publication unless extensively edited

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

No competing interests.