Reviewer's report

Title: Developing model-based algorithms to identify screening colonoscopies using administrative health databases

Version: 2 Date: 2 November 2012

Reviewer: Peter Richardson

Reviewer's report:

The approach taken in this work is a very interesting addition to the study of this issue.

There are two major compulsory revisions that I believe need some addressing.

[1] Conceptual issue:

The approach taken is "in the absence of established gold standards". However this may be due to some lack of precision of the notion of "screening" per se, i.e. that there is no clear cut gold standard conceptual definition and hence no gold-standard procedure for discerning screening. Do the authors believe that the conceptual boundaries of the concept are or are not clear cut? The status of the latent classification would hinge on the answer to this; i.e. is the latent classification an attempt to get at a clearcut but unknown classification or is it an underlying "factor" behind existing instruments ("indications") that itself cannot be defined in a clear way but does serve as an explanation of what these indications have in common.

[2] Statistical issue:

The LCA is performed in a Bayesian way which seems like an excellent idea. However, what this should yield is posterior distributions for the latent class probabilities as well as the three indication-related conditionals. As such, the "predicted probabilities" which are then dichotomized to be used as outcomes in the logistic analyses are presumably point estimates of some sort. Are they posterior means, posterior modes or something else.

Also, these posterior distributions have some variation (presumably a good deal less than the priors). Can the authors tell us anything about the variability in the latent-class-predicted-probability posterior? How much could this uncertainty variability affect the stability of the dichotomization-based classifications and thus the results of the logistic analyses themselves?

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

Quality of written English: Acceptable
**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

'I declare that I have no competing interests'.