Author's response to reviews

Title:  Assessment of the potential impact of a reminder system on the reduction of diagnostic errors: a quasi-experimental study

Authors:
Padmanabhan Ramnarayan (ram@isabelhealthcare.com)
Graham C Roberts (g.c.roberts@imperial.ac.uk)
Michael Coren (michael.coren@st-marys.nhs.uk)
Vasantha Nanduri (v.nanduri@hotmail.com)
Amanda Tomlinson (mandy@isabelhealthcare.com)
Paul M Taylor (p.taylor@chime.ucl.ac.uk)
Jeremy Wyatt (j.wyatt@nice.nhs.uk)
Joseph F Britto (joseph@isabelhealthcare.com)

Version: 4 Date: 7 March 2006

Author's response to reviews: see over
Dear Editor-

Thank you for your comments from the statistical reviewer. We notice that there are no further requests for changes from the other reviewers. We have made changes to the manuscript in response to the statistical reviewer comments. Our responses to the questions raised are covered below:

1. The authors state that "Subjects were used as the unit of analysis [...] in order to account for clustering effects". First, I do not understand what is meant by "clustering effects" in this context. Second, from my point of view aggregating the data for each subjects only can be a valid approach to the data, if each subject had rated the same cases. This is not true as each subject was presented a random subset of cases. Moreover, the number of rated cases differed from subject to subject (within the range from 6 to 12). Consequently, the cumulated number of DEOs strongly depends on this number and, thus, is not comparable between subjects. Similar arguments hold for all other outcome parameters. From my point of view data should be analysed on the basis of each case episode (which is the rating of a case by a subject).

We meant that the 751 case episodes were clustered by both subject and case. We expected that the scores obtained from one particular subject would be highly correlated across all his 12 cases. A similar approach was adopted by Friedman et al (JAMA, 1998) in a study that we based this experiment on. We have now reworded the confusing statement to: “Subjects were used as the unit of analysis for the primary outcome measure”. We decided to aggregate the data for each subject for the following reason: although a set of 12 cases were “randomly” chosen per subject from a total of 24 (and presented in random order), since the underlying rule specified that all 12 specialities had to be represented in the set, there was significant overlap between subjects’ allocated cases. This led to a situation almost similar to subjects assessing the same set of cases. Therefore, we aggregated the data at subject level, despite recognising that greater variance might result from this approach. Data for this is shown in the table below (52 subjects):
We have now analysed all outcome measures on the basis of the subjects who assessed all 12 assigned cases, recognising that the number of rated cases differed from subject to subject in the previous iteration of the manuscript.

2. The authors use a twofold strategy to analyse the data. First they use a repeated measurement ANOVA (or Mixed Models as the authors call it), and second, they apply simple paired t-tests. I suggest to abandon the t-tests but to build a repeated measurement ANOVA model for each outcome parameter extending the original mixed model (by introducing the case as an additional factor). From this model each wanted effect (differences between subjects’ grades, time effects, differences between the complexity of cases, and all interactions) can be estimated and tested.

We have used the two way repeated measures ANOVA (using grade as between-subjects factor, and time effect as within-subjects factor) consistently for all outcome measures
now. The t-tests have been abandoned. Since we aggregated the subjects’ score across all cases, we could not extend the model to include the case as an additional factor.