Reviewer's report

**Title:** Innovations in curriculum design: a multi-disciplinary approach to teaching statistics to undergraduate medical students

**Version:** 1  **Date:** 28 January 2008

**Reviewer:** Peter Petocz

**Reviewer's report:**

Review of "Innovations in curriculum design: a multi-disciplinary approach to teaching statistics to undergraduate medical students" by Jenny V Freeman, Steve Collier, David Staniforth and Kevin J Smith, submitted to BMC Medical Education.

Recommendation: accept after minor revisions.

This paper describes a project that revised and re-conceptualised a syllabus for teaching statistics to undergraduate medical students. A novel feature of the approach was a joint involvement by a statistician, a clinician and an educational expert: the statistician presented the material, the clinician contextualised it, while the educational expert ensured that the pedagogical approach was sound. New teaching materials, emphasising applying statistics and interpreting data, were developed and produced for this course. The approach was evaluated using student questionnaires with the last cohort of the old-style course and the first cohort of the newly-developed course and some encouraging differences were found.

There is much to be positive about in this paper and the project on which it is based. The actual materials developed sound interesting. The explicit acknowledgement that students have different learning styles, and the development of a corresponding range of flexible learning materials. The contextualisation of statistical education and the explicit link to students' future professional worlds. The analysis of what the medical students would need to know about statistics in their future profession. The combination of different experts: statistician, pedagogue and clinician. The evaluation of the effects of the curriculum change and the careful analysis of the results of this evaluation.

On page 7, we learn that the course comprised 5 one-hour lectures (and follow-up tutorials, in groups of 8, a real luxury!). Maybe it should be made more clear earlier on that this is a relatively short module rather than a complete one-semester first course in statistics.

It would be useful to have the learning materials in particular, the video clips and animations, but also the written materials illustrated in some way (or to provide links some of them online, especially as an online resource is currently being developed, p14). Obviously, a key part of the success of the new approach was the quality of these materials. And related to this, the details of the previous
course seem a little sketchy (was it a traditional mathematically-based course, as implied on p13?).

On page 8, the authors describe the evaluation questionnaire that was completed by the students who attended the final session for each cohort. It would be usual to state that the questionnaire and the survey process had been approved by an ethics committee and that students had the opportunity of not participating without this influencing their results in any way assuming that this did in fact occur in this way.

Assessment of students’ statistical understanding by asking them to define a p-value and a confidence interval, with the highest grade given for a “text-book perfect” answer, may have been convenient but could also be very dependent on the material addressed in the classes and the expected study approach. The notion of “text-book perfect” makes it sound as though the students were encouraged to memorise such definitions, and a comment about this may be in order. In particular, it seems that there was only one student of the old cohort who could give a text-book definition of confidence interval and none who could similarly define a p-value, whereas with the new cohort there were about 25 who could carry out the first task and 20 the second task. In what way is this change related to the specifics of the teaching approach?

Assessment of students’ attitudes may also be a little problematic since there seem to be modest (though in some cases statistically significant) improvements. For instance, the percentage of students who agreed or strongly agreed with each statement seems to have risen from 80 to 90 for the first two questions, 80 to 85 for the third and 50 to 55 for the fourth. But this is a common problem in this type of evaluation: small improvements (less than one-quarter of a step on the five-point scale, on average) are significant due to the relatively large numbers.

Of course, the whole basis of the comparison in the evaluation is tricky. While adjustments for age and sex can be made statistically, the differences in entry criteria and background perceptions of the medical profession are not so easy to deal with. Further, the “Hawthorne effect” could also play a part in the more positive results obtained with the new cohort. Of course, practically no pedagogical (pseudo-) experiment is free of such problems, but it would be prudent to acknowledge these limitations rather than writing comments such as we demonstrated that we have made a significant difference to both students’ understanding of some key concepts and their ability to view statistics in relation to their medical practice. This statement seems to run into the common problem of the two interpretations of the term “significant”, for although the authors seem to intend the statistical use of the term, their sentence seems to imply the common usage of “important” or “large”.

However, the paper has much to recommend it, and I believe that it should be published after the authors have had a chance to respond to these (and any other) comments on their submitted version.

A few small comments/typos:
Page 6/7: last word on p6 is factors and first word on p7 is criteria â## one should be removed.

References: several of these are available online â## in particular, numbers 2, 3, 6, 14, 15 and 21 can be found on the JSE website (http://www.amstat.org/publications/jse/) or the IASE website (http://www.stat.auckland.ac.nz/~iase/).