Author's response to reviews

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

Authors:

Jenny V Freeman (j.v.freeman@shef.ac.uk)
Steve Collier (steve.collier@shef.ac.uk)
David Staniforth (dshgc@btinternet.com)
Kevin J Smith (k.j.smith@shef.ac.uk)

Version: 2 Date: 6 February 2008

Author's response to reviews: see over
Dear Dr Da-Silva,

Re: MS: 7462070701645998 - Innovations in curriculum design: a multi-disciplinary approach to teaching statistics to undergraduate medical students

Detailed in the table below are my responses to the comments of the two reviewers, which I have tried to respond to in full. I would like to thank the reviewers for their comments, which I feel have strengthened the final paper.

I look forward to hearing from you in the near future

Yours sincerely

Dr J Freeman

Reviewer 1 comments in black (Paul Marantz)
Reviewer 2 comments (Peter Petocz)

Where there was overlap in the comments from the two reviewers I have grouped them together

<table>
<thead>
<tr>
<th>Comments</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the &quot;Background&quot; section, it would be reasonable to provide some references for the statements that &quot;students across all disciplines&quot; commonly &quot;dislike and under-perform in courses involving mathematics, numeracy or statistics.&quot; (If medical residents are included as &quot;students,&quot; a recent example of such a reference would be: Medicine Residents' Understanding of the Biostatistics and Results in the Medical Literature; Donna M. Windish, MD, MPH; Stephen J. Huot, MD, PhD; Michael L. Green, MD, MSc; JAMA. 2007;298:1010-1022.)</td>
<td>This has now been added</td>
</tr>
</tbody>
</table>
| Likewise in the "Background" section, I think the italicized comments from the students indicate their dislike of the course, but not that it "does not entirely meet their needs." More appropriate quotes, or relevant data, should be presented. | This has now been amended: Medical students have varied backgrounds and it is increasingly clear that teaching statistics in a traditional didactic way neither engages the students nor meets their needs; comments from previous student evaluations at our institution have included:

‘Medical statistics – YAWN!!! - could be taught in a more interesting way’
‘I did not really find the module valuable, as a lot of the content seemed unimportant and difficult to understand’
‘Medical statistics lectures were poorly presented and not explained well’
‘Statistics sessions were not well presented, too rushed, no handouts’ |
| Whether in the "background" or "methods" section, some description of the educational setting would be of great value to the reader. We know the class size is large, at 260, but that's about it. Anything of interest about the particular school and its curriculum -- location, demographics, number of hours devoted to statistics in the curriculum, where (i.e., what year) this is taught -- would be helpful. | I have added the following text to the background:  

_The current annual intake to the medical undergraduate degree programme in our institution is 260 students. It is a clinically led programme and currently students received only 10 hours of instruction in medical statistics, delivered in their first year. Unless they elect to do a research year, they receive no further training in statistics during their undergraduate training. Thus the challenge is to give students enough instruction to ensure that they understand the basics well, without overloading them with unnecessary detail, within the limited time available._ |

| 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. |  |

| In "methods" -- the bulleted list of "outcomes" of the project don't seem to be outcomes. Would reword as "aims" or "goals." | This has now been changed to goals |

| There are some locutions and spellings that are uniquely British (or at least, not American English), and I think I figured out most of them, but perhaps such phrases as "statistics at 'A' level" and "a general practitioner's surgery" could be replaced by more generic phraseology, or defined for the rest of us. | Have changed 'general practitioner' to 'family practitioner'  
Have added the years that students usually take GCSE and A levels to table 1 for clarification  
Have changed ‘A level’ to ‘High school diploma level’ |

| Please provide some examples of the scenarios used -- it isn't entirely obvious (at least to me) how a GP uses statistics in her/his practice (at least, not involving P values and confidence intervals). Perhaps a text description of one or two scenarios would help -- and maybe even a web link to some of the video clips and animations. | The following has been added to the description of the scenarios (new text in italics):  

_These scenarios featured a trainee family practitioner, chosen to be someone the students might be able to relate to tackling problems that they were likely to encounter in their everyday practice. For example, in one of the scenarios the trainee was initially consulted by an elderly patient with asthma who was on a range of medications to treat it. Following a discussion with one of the other doctors in the practice she decided to look at a recent research paper investigating the impact of stepping down therapy. A second video clip at the end of each lecture showed how the fictitious characters resolved the issues and how statistics played a vital role in diagnosis and_ |
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?)

This is tricky at present as some of the work is copyright for my institution (videos and animations). However, I have placed links to the course manual on my webpage, which I am happy to have added to the paper. The link is http://www.shef.ac.uk/scharr/sections/hsr/statistics/staff/freeman.html

The previous course was indeed a traditional mathematically-based course with little attempt to place any of the examples within a relevant medical context. A brief expansion has now been included on page 5:

Previous teaching had focused on ‘doing’ statistics, including an emphasis on the mathematics underpinning particular methodologies, with little emphasis on its contextualization. However, it was felt that, in line with the recommendations of the GMC, the undergraduate medical students were more likely to become ‘consumers’ or users of statistics.

| MOST IMPORTANT [METHODOLOGIC] COMMENT: the data showing an improvement in students' understanding of the basic concepts is striking, almost too good to be true. For instance, almost NO students provided a textbook answer on CIs the "old style" year, and 20% did the "new style" year. It's great that the assessors were blinded to the student's cohort when grading the answers, and that helps a lot. But my concern is: was the teaching "style" the only difference between the two years? One possibility is that the "textbook" answers were somehow provided to the students more clearly during the "new style teaching" year (e.g., providing new and more precise language regarding the difference in CIs in the class notes). Or perhaps the "set-up" for this student assignment was different in the two years (e.g., "old style" they were asked to fill this out at the last class, and "new style" they did it right after they had studied for the exam). I think it's critical that the authors provide more information related to | There was no difference in the timing of the assessment – neither cohort was closer to the exam. However, I acknowledge that the differences were striking and think that this is in part due to a shift in emphasis between the two styles of teaching. There were many things that differed between the two cohorts in the way that statistics was taught. The old style teaching had small groups of size 32! Teaching ‘style’ was not the only difference between the two groups, I just gave these labels to distinguish between the old course and the new course. I have now changed the labelling to use the word ‘approach’ rather than ‘style’.

There were many differences between the two approaches and without doing a thorough process analysis it is impossible to ascertain what was the most important element in getting students to be better engaged. I think that the main point that I have been trying to make is that in order to engage (non-statistics) students in learning about statistics you need to think very carefully about what you want them to learn, make sure you cover the basics well, embed the learning within a relevant context and use a variety of approaches. We thoroughly revamped teaching, from the rather moribund... |
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?

There was definitely a change in emphasis for the new teaching (much more emphasis on understanding a few basic concepts, rather than rushing through many) and perhaps defining p-value and confidence interval was a loaded metric - in hindsight we should perhaps have got the students to interpret these terms.

I have added the following description to the last paragraph of the introduction:

*The old course used lectures followed by large group seminars (32 students per group). It was a mathematically-orientated course with much theory and formalae.*

In addition I have added a paragraph to the discussion, emphasising the complete overhaul that we did and the change in emphasis, which might account for part of the striking differences:

*The new course did place much greater emphasis on obtaining a basic understanding of key concepts and this might in part explain the big difference in the number of students able to define these, although we would argue that this could not account for all the difference between the groups. We did much more than change the emphasis, we completely remodelled the teaching materials and increased the variety of approaches and it is this multi-dimensional approach that we would recommend to others; in order to engage (non-statistics) students in learning about statistics you need to think very carefully about what you want them to learn, make sure you cover the basics well, embed the learning within a relevant context and use a variety of approaches.*

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

As this reviewer points out, whilst the results with respect to student attitudes are statistically significant, the actual effect is indeed small. However, given that this is a large cohort of students, a change in attitude for 5% of students
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’.

I have changed the phrase ‘significant improvement’ to read ‘statistically significant improvement’.

Since you had 8 students to a tutor, you needed 32+ faculty to teach this. Can you describe anything about their backgrounds and statistical expertise, and how you recruited or incentivized them to do this teaching?

In our department we have an obligation to deliver this teaching and so there was no incentivisation – faculty are required as part of their contracts to deliver it. The following has now been added to the text:

As there were 260 students per year this required approximately 16 faculty giving the same tutorial twice. Not all tutors were statisticians, but all were either statistics PhD students or quantitative researchers with several years experience. An information session was held prior to the commencement of teaching to discuss the main points to be covered in the tutorials and answer any queries that the facilitators might have.

Small point, but since you referenced my

I have now removed this reference to the amount of
paper: you mentioned that in the Marantz study "the amount of teaching time was greater." I'm not sure if this is true, assuming you had 5 lectures followed by 5 tutorials. Even if the tutorials were 1 hour each, that's 10 hours for statistics -- our course has a total of 18 hours for statistics, epidemiology, and clinical research design, so significantly less than 10 hours for statistics. As I said, small point...

| teaching time. |

There are a few minor spelling, punctuation, and grammatical errors that should be corrected

| Have deleted the word criteria |

Page 6/7: last word on p6 is factors and first word on p7 is criteria one should be removed.

| |

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.

| I did seek advice from the chair of the ethics committee within my department and he felt that as this was an evaluation of a revamp of existing teaching, full ethical approval for the evaluation was not necessary. In addition all evaluation questionnaires were anonymous and completion was not compulsory, thus the following text in the statistical methods section has now been amended (new text in italics):

Every student who attended the final statistics session for both cohorts was asked to fill in the evaluation questionnaire and hand it in as they left the lecture theatre. The questionnaires were all anonymous and completion was not compulsory. As completion was anonymous, no attempt was made to link individual student responses to grades. |

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/).