Author's response to reviews

Title: Can American College of Radiology In-training Examination Scores be used to Predict Canadian Radiology Licensing Examination Results? A retrospective study.

Authors:

Trent H Orton (trentorton@gmail.com)
Matthew D McInnes (mcinnes.matt@gmail.com)

Version: 2 Date: 2 January 2013

Author's response to reviews: see over
Dear Mr Ulep and Dr Archer,

Re: Requested Revisions for Manuscript Number 5948557748 294140.

Thank you for considering our paper: “Can American College of Radiology In-training Examination Scores be used to Predict Canadian Licensing Examination Results? A retrospective study,” for publication.

Please find below the following responses to your concerns:

**Referee 1, Chris McManus:**

**Major compulsory revisions**

1. *For readers who are not Canadian radiologists, or even radiologists or from North America, it might be useful to provide a clear timeline of where these various assessments fit within the careers of the Canadian radiologists.*

This is now addressed in a new introductory paragraph (1st paragraph) in the background (board exam timeline) and in the second paragraph of the background on page 3 (ACR administration timeline).

2. *Much of the problem with these data is not the fault of the authors but the fact that their only outcome measure is a binary pass-fail, with most doctors passing. That makes any calculation of correlations difficult. The authors do well to use instead ROC curves, which do address the problem. However it might be that they need to explain why they are not using continuous scores for outcome (and I presume the College of Radiology will not provide them?). Either way, something should be said.*

This is now addressed in the first paragraph of the methods section.

3. *The correlations provided with binary outcome will be heavily constrained. Strictly what is needed here is a proper biserial correlation (i.e. equivalent in some ways to a tetrachoric correlation, and specifically not the point-biserial correlation which is the Pearson correlation), where the assumption is that a bivariate distribution has been dichotomised along one dimension. If the authors could provide that then it would be useful. Calculation is straightforward (and can be found in Howell’s Statistical Methods for Psychology). I have done it for PGY2 to PGY5 and Average and get values of .640, .633, .562, .664 and .692. These are noticeably higher than the Pearson R values and can be interpreted as the correlations if the College of Anaesthetists had provided full marks, rather than merely pass and fail.*

These have been calculated based on the formula provided in Howell’s Statistical Methods for Psychology (p. 287) and values of rb for pgy 2 through 5 and the
average are provided in Table 1. This now replaces the discussion of pearson/spearman correlation coefficients in the methods, results and background and the strength of the correlation has been upgraded from moderate (0.3 to 0.5) to strong (0.5 to 1).

4. A key question underlying these data is the extent to which behaviour in these residents remains constant across time (in other words, does performance ‘track’, so that poor performance at one time is related to poor performance at another). For that it would be helpful to have the correlations of the (continuous) scores available for PGY2, PGY3, PGY4, and PGY5. A clear prediction would be that they will be similar to the biserial correlations described above. If so then there is strong stability during the course.

Pearson correlation coefficients were calculated between consecutive ACR scores of all 4 years of radiology training (ie 3 Pearson Rs produced). The description was included in the Methods (2nd paragraph, last 2 sentences, page 4), Results (4th paragraph, last sentence, page 6) and Discussion (new 4th paragraph page 8).

Minor corrections

5. P.8. “decreasing the skew from one inconsistent ACR score”. I’m not really sure that it is skew here. Essentially the error terms are being averaged out whereas the signal stays constant, and hence, via the Central Limit theorem the average is both more accurate and more likely to be normally distributed.

This phrase has been removed from the discussion.

6. Table 1. I had presumed that the values after the plus/minus are standard errors (or are they 95% CIs?). However they are remarkably similar when N is 50 or 8, suggesting either that the SDs in the groups are very different or that these are SDs. Either way, a) it should be stated, and b) Standard Deviations/SEMs would be useful as well.

These are 95% CIs. To make it clearer, we have labeled them in Table 1 and used traditional absolute values of the lower and upper limits in square brackets.

Referee 2, Arnold C Paulino:

Major Compulsory Revisions

The authors should mention what the RSPSC exam in Diagnostic Radiology consist of. Does it have a written and oral examination component. If there is an oral examination component, it will be good to reanalyze the data according to the written and oral exams. The ACR in-training exam is a written test so one would predict that residents who do well on this will do well on the written but not
necessarily the oral exam component.

This is now addressed in the first paragraph of the background (page 3) and in the discussion of limitations (10th paragraph of the discussion page 10).

Mr Aldrin Ulep:

Also, kindly state the name of the ethics committee who approved the study and include it in the Methods section.

Added in the first sentence of the Methods section (page 4).

Additionally, the references now conform to the journal formatting.

We look forward to your feedback.

Kindest regards,

Matt McInnes and Trent Orton.