Reviewer’s report

Title: Investigating possible ethnicity and sex bias in clinical examiners: An analysis of data from the MRCP(UK) PACES and nPACES examinations

Version: 2 Date: 10 June 2013

Reviewer: Charles Wiles

Reviewer’s report:

This is an interesting paper which proposes a methodology to make use of the paired examiner system for MRCP – PACES/nPACES to evaluate sex and ethnic bias and hawkish/doveish marking behaviour. My main comments relate to some unstated assumptions about ethnic (and sex) distribution between examination centres and in relation to examiners. The overall well written article might benefit from a slightly clearer exposition of the PACES/nPACES process early in the background.

Major Compulsory Revisions
None

Minor Essential Revisions
1. Methods – Ethnic bias… (p6). I think the Ethnic categorisation (presumably white/non-white?) should be made explicit together with how this was collapsed from the original data collection. Does the number of ethnic divisions make any material difference to the power to determine bias?

2. Although 49.3% of candidates overall were non-white the ability of the study to demonstrate ethnic bias presumably depends on there being a mix of ethnicities within each centre: if for example some international centres had very little white/non-white candidate mix then examiner/coexaminer differences might be minimised unless they circulated significantly between centres. This might not be important if the 9 international centres contributed a small proportion of the overall candidates. So…is heterogeneity between centres (for ethnicity and sex) an issue?

3. Methods (para 3): “Self-declared ethnicity was complete for all examiners” & Results (para 1). The sex and ethnicity distribution of examiners is not stated. Since the study method depends on the evaluation of differences between one examiner and all their coexaminers, one might argue that a more uniform distribution of examiners eg mostly white and male might tend to reduce the likelihood of demonstrating ethnic/sex bias because, on average, they might tend to be biased in the same direction. Whilst this seems unlikely to explain the low level of bias in the current study perhaps the sex and ethnicity of examiners should be stated and commented on if appropriate.

Discretionary Revisions
4. Mentioning the full meaning of the acronyms “MRCP” and “PACES” once in the text would be useful.

5. Background: Those unfamiliar with PACES (though interested in clinical exams) would benefit from a clearer explanation of what PACES and nPACES actually consist of at an early stage. After introducing PACES (current bottom p3) it would be helpful to very briefly describe the carousel of 5 stations, each marked by a pair of examiners blind to each other (i.e. 10 overall); the fact that each examiner has had prior specific training (including on the issue of bias); and that each pair of examiners discusses the key findings and their marking at each station in a process of ‘calibration’ prior to seeing the candidates. This will allow less familiar readers to more easily conceptualise the key points about each examiner being compared with all coexaminers.

6. The Background does not make clear whether examiners always examine the same station in different carousels or whether they are randomly allocated to different stations – so that pairings of examiners are also random.

7. Methods: “For nPACES, the two examiners at each [?insert ‘of the 5’] stations…”

Typos:
Abstract bottom line – comma
Table 1 legend line 4 ‘Bonferroni’
Figure 1 legend line 2 fullstop after ‘(top right)’ and similar Fig 2 legend line 2
Results – 3rd para from bottom – double comma after ‘judgements’

Level of interest: An article of outstanding merit and interest in its field

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:
I declare that I have no competing interests