Reviewer's report

Title: Incorporating natural variation into IVF clinic league tables: The Expected Rank

Version: 1 Date: 22 February 2009

Reviewer: Mohammed Amin Mohammed

Reviewer's report:

Review by Dr M A Mohammed

of “Incorporating natural variation into IVF clinic league tables” by Lingam et al submitted to the BMC (Feb 2009).

Major Comments

The authors start from the position that comparative performance of healthcare providers is often presented using league tables. They note that such league tables do not incorporate the natural variation in the rankings and propose a method for “improving” the ranking process to allow for chance variation.

The working premise in this paper is that ranking can be “improved”, when the arguments against ranking (empirical evidence that ranks are not reliable and are meaningless under chance causes and even more meaningless under assignable causes of variation) are overwhelming. Ranking is flawed! (Lancet 2001; 357: 463–67; Statist. Med. 2005; 24:1185–1202)

Fortunately, there are appropriate ways of presenting comparative data which avoid the pitfalls associated with ranking (Human Fertility 2006; 9(3): 145 – 151). The reader would benefit from the being made aware of the above, and also about ways of presenting data which are statistically sound and avoid ranking altogether.

The rationale for the expected ranking procedure suggested by the authors is not well established in the introduction – in other words, why do we need another method of adjusting ranks for natural variation when at least two other methods already exists? Some justification is needed.

The authors make use of random effects modelling to arrive at expected ranks and also make a probability statement of one clinics performance against another. Whilst a statistical review is necessary, the fundamental limitation appears to be that this approach does not distinguish between clinics that are varying purely by chance - ie even if we set up a test data set with variation in outcomes between clinics being purely random, then the proposed method will still give a rank and a probability – but these statistics would not be meaningful for a random variable. Clinic ranked 1 and clinic ranked 10 are all the same because they arise from the same random process and it is not clear how the proposed method overcome this issue?
The outputs of the proposed methods are not materially different from the ones arrived at by simple ranking. This could be due to the selected data set or may be a feature of the method. It seems to me that a more a rigorous and comprehensive evaluation (eg after adjusting for case-mix factors) is required to determine the “operating” characteristics of this method versus other approaches before this method can be routinely used.

**Level of interest:** An article of insufficient interest to warrant publication in a scientific/medical journal

**Quality of written English:** Needs some language corrections before being published

**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'