

Author's response to reviews

Title: Reliability of journal impact factor rankings

Authors:

Darren C Greenwood (d.c.greenwood@leeds.ac.uk)

Version: 3 **Date:** 17 October 2007

Author's response to reviews: see over

Response to reviewers' comments:

Eleanor Pullenayegum

I would like thank the reviewer for her interest in the article, and her clear and insightful comments. I address each one in turn:

1. The impact factor as defined (the number of citations a journal has received in the last complete year for articles published in the two preceding years, divided by the total number of articles the journal published in the two preceding years) is not measured subject to sampling error and hence known rather than estimated. There is thus no need to put a measure of uncertainty around it. The underlying rate at which any particular article in the journal is cited is, however, unknown and hence estimated subject to sampling error. Perhaps the wording throughout the manuscript could be tightened up to reflect this?

Yes, this is the classic argument for not putting confidence intervals around census results, though many still do this to reflect uncertainty, arguing that this is a sample from a set of hypothetical possibilities. I have tightened up the wording to clarify that the underlying rate is unknown, hence estimated and subject to sampling error, and inserted a paragraph in the discussion.

2. The proposed model treats the observed number of citations as Poisson with mean $\lambda_i n_i$, and $\log(\lambda_i) = \lambda + v_i$. Am I correct in thinking that the total number of articles n_i is treated as fixed? Would the results change at all if the number of articles was treated as Poisson (also with random effects to model between-journal variation)? In addition, adopting a random effect for λ_i models between-journal variability, but what about within-journal variability (i.e. some articles are more likely to be cited than others)? Is this accounted for in the model?

The total number of articles n_i has indeed been treated as fixed. When I re-run the model treating this as Poisson and also with random effects, the results are very similar. I would be happy to use this extended model if the reviewer felt strongly about this, but I have left the analysis as is stands, partly because n_i is so large that the additional random effects have little influence, and partly because it is unusual in my experience to treat this total as a random effect – in epidemiological examples I have only ever seen this treated as fixed. In my view, it makes more sense to treat the number of citations as the random variable, leaving the total number of publications open to citation as fixed. The level of the individual article has been ignored in my models, hence within-journal variability is not taken into account. It could potentially be included, however, if the data at the individual article level were available.

3. Under the proposed model, the estimated impact factors (i.e. λ_i) are presumably not the same as the observed impact factors (O_i / n_i) because the introduction of the random effects results in some shrinkage? Could this be clarified in the text and in Figure 1?

Thank you. This has now been clarified.

4. (Last paragraph of Methods) Note that if $\log(\lambda_i)=\lambda+v_i$, then λ is not the mean citation rate, because a log link has been used.

Thank you for noting this. In fact this was a typo, which I have now corrected in the text to show $\ln(\lambda_i) = \ln(\lambda) + v_i$. I have also clarified this throughout the text by labelling λ as the *geometric* mean citation rate.

5. (Second paragraph of Results) To what extent are the observed results a product of the model (i.e. lognormal random effects)? Equivalently, how sensitive are the results to the model assumptions? How well does the model fit the data?

Including a random effect on the total n_i gives very similar. Including an additional level of variation – within journals, would lead to greater shrinkage and wider credible intervals, but to do this analysis would require an enormous data collection effort. I have included some discussion of this in the paper. Excluding any random effects here forces all the journals to be the same, giving very poor fit (DIC). The binomial distribution would also be inappropriate because the number of citations can be far greater than the number of publications. Residual plots demonstrate the shrinkage, but with fitted citation rates very close to observed rates. All priors were uninformative and changing these made negligible difference to the results.

6. Figure 1: Is Figure 1 showing the observed impact factors, the mean values of λ_i or some other summary statistic?

Thank you, this is helpful, I have clarified in the text and figure title that this figure refers to the mean values of λ_i .

7. Figure 2: Similarly, is Figure 2 showing the mean ranks, the median ranks or the observed ranks?

Again, I recognise that I was not particularly clear in the text. These are *mean* ranks. The median ranks are very similar, but of course discrete integers. I have clarified this in the text and figure title.

8. Since you are using a Bayesian approach, perhaps you could identify this in the methods section of the abstract? (Using MCMC methods does not necessarily imply a Bayesian approach).

This suggestion has now been done.

Andy Weale

I would like to thank the reviewer for his thoughtful response. Below I respond to his points and trust these explanations and my additions to the text are satisfactory.

Whilst I would never describe the distribution of data as non-parametric (I prefer to refer to the hypothesis test one might use as non-parametric), I entirely agree with the reviewer that we would never recommend presenting the arithmetic mean of a skewed distribution in medical research because it does not summarise the central tendency of the distribution, giving undue influence to the values in the tail. I have included reference to this (amongst other criticisms of the journal impact factor) in the introduction, and have inserted a new paragraph in the discussion.

The purpose of the project was to indicate the danger (or otherwise) of basing decisions on journal ranks, because of the unacknowledged uncertainty associated with their estimation (hence the title). These ranks are currently based on the journal impact factor, so this was what I modelled, despite its many faults. Ranks based on any other measure, such as the immediacy index, would also have uncertainty associated with them. In the case of the immediacy index there is even greater uncertainty in the ranks than those based on the journal impact factor.

To use an aggregate measure other than those published by ISI would require data from nearly 20,000 individual articles, and this was beyond the scope of the project, and not its main purpose. By modelling the count as a Poisson variable (note that it is the count that is modelled), the credible intervals are valid for this measure, even if the measure used itself is non-sensical. In addition, by working on the log scale, we acknowledge the skewness in the rates, so that λ is the overall *geometric* mean citation rate rather than the *arithmetic* mean.

I have now made my hypothesis more explicit in the introduction.

The implications for use of the impact factor are to avoid ranks and to avoid artificial cut-offs based on ranks. A paragraph has been added to the discussion along these lines.

I have included figure 1 on the log scale as suggested.