Ideas in Ecology and Evolution 3: 11-15, 2010
doi:10.4033/iee.2010.3.3.c
© 2010 The Author. © Ideas in Ecology and Evolution 2010
Received 21 April 2010; Accepted 6 June 2010

Commentary

Do 'Faculty of 1000' (F1000) ratings of ecological publications serve as reasonable predictors of their future impact?

David A. Wardle

D.A. Wardle (david.wardle@svek.slu.se) Department of Forest Vegetation Ecology, Swedish University of Agricultural Sciences, SE901-83 Umeå, Sweden

Abstract

There is an increasing demand for an effective means of post-publication evaluation of ecological work that avoids pitfalls associated with using the impact factor of the journal in which the work was published. One approach that has been gaining momentum is the 'Faculty of 1000' (hereafter F1000) evaluation procedure, in which panel members identify what they believe to be the most 'important' recent publications they have read. Here I focused on 1530 publications from 7 major ecological journals that appeared in 2005, and compared the F1000 rating of each publication with the frequency with which it was subsequently cited. The mean and median citation frequencies of the 103 publications highlighted by F1000 was higher than for all 1530 publications, but not substantially so. Further, the F1000 procedure did not highlight any of the 11 publications that were each cited over 130 (and up to 497) times, while it did highlight 14 publications that were each cited between 4 and 9 times. Further, 46% and 31% of all manuscripts highlighted by F1000 were cited less often than the mean and median respectively of all 1530 publications. Possible reasons for the F1000 process failing to identify high impact publications may include uneven coverage by F1000 of different ecological topics, cronyism, and geographical bias favoring North American publications. As long as the F1000 process cannot identify those publications that subsequently have the greatest impact, it cannot be reliably used as a means of post-publication evaluation of the ecological literature.

The past decade has seen a substantial increase in the use of external evaluations to evaluate scientific performance, ranging from the level of the individual to university departments and even whole institutions, and frequently these form the basis for major decisions about allocation of funding and other resources. As the scientific publication is the most tangible output of the scientific process, many evaluations have focused on assessing the relative 'quality' or 'merit' of publications in terms of their likely impact in the field. The true impact of a publication does not usually become apparent until several years after it is published, especially in the case of ecological research. However, evaluations often focus on the most recent publications produced by an individual or organization, because evaluations usually focus on current outputs and thus future potential, rather than on publications from several years previously. As such, there is an obvious need for assessments of 'quality' that can be measured soon after papers are published without waiting for them to garner citations or recognition. Any short term measure of scientific merit has potential drawbacks and may be potentially influenced by various types of bias (Lortie et al. 2007). One such short-term measure used in many evaluations involves rating publications by the impact factor of the journals in which they appear. However, there are numerous criticism of this focus on impact factors (Seglen 1997), the most important arguably being that most publications in high impact factor journals still only attract modest numbers of citations (Seglen 1997, Leimu and Koricheva 2005). The increasing importance being placed on journal impact factors has led some editors to concentrate on enhancing their journal's impact factor (sometimes in rather

questionable ways) at the expense of ensuring that their journal best meets the needs of the scientific community (Monastersky 2005, Aarssen and Lortie 2009). This latter problem has been explicitly identified for some leading ecological journals (Monastersky 2005).

These and other problems with journal impact factors have led to a demand for other means of determining the relative 'quality' of manuscripts shortly after publication. One such approach that has been rapidly gaining momentum is the 'Faculty of 1000' (hereafter F1000) evaluation process (http://f1000biology.com), in which (to use the phrasing on the F1000 website) '*a global panel of experts*' or '*the world's top scientists*' '*select*' '*the most important*' or '*the most interesting*' publications in each of the fields that F1000 covers. The F1000 system also rates selected publications into categories of 'recommended', 'must read' and 'exceptional', corresponding to F1000 scores of 3.0, 6.0 and 9.0 respectively; publications not selected by any members of the panel presumably have a F1000 score of 0.0. This method of rating is rapidly gaining attention, as revealed by the frequency with which scientists list F1000 scores for their publications in personal web sites, CVs, and application packages for positions, tenure and promotions. This measure is also attracting attention from administrators; for example, as noted on the F1000 website, CNRS (the main government-funded research institute in France) '*has given its lab heads access to Faculty of 1000 Biology to better assess the research that CNRS funds and the suitability of applicants*'. If such a system is to be used for these sorts of purposes then it is essential that it can reliably identify those scientific papers which go on to have the greatest scientific impact. Despite the momentum that the F1000 ranking system has been gaining, it has been subjected to little scrutiny by the scientific community; most publications addressing the issue of science evaluation instead offer unquestioning praise for expert evaluations of this type (e.g., Raff et al. 2004, Badjer 2005, Marder 2006, Allen et al. 2009). To my knowledge, no quantitative assessment of the F1000 evaluation process has ever been performed.

If the F1000 rating system is indeed identifying the most important publications, then we would expect those publications that it highlights to be the same ones that subsequently become heavily cited. There are well known limitations with using citation counts of individual publications to assess their impact, and differences in counts between publications of 20% or even two-fold can be ascribed to various reasons other than merit. However, a publication that is cited over 100 times over a five year period would undoubtedly be exerting a bigger impact in its field than another that is cited less than 10 times over the same period. To put it another way, if the F1000 system was performing well, then it would highlight most or all of those publications that go on to be cited hundreds of times over the following few years, and few if any that become cited only a handful of times. In this light, I sought to test whether manuscripts published in 7 major ecological journals throughout 2005 that were highlighted by the F1000 procedure are the same as those that subsequently become the most heavily cited. The journals that I selected were *Ecology Letters*, *Ecology*, *Oikos*, *Oecologia*, *American Naturalist*, *Journal of Ecology* and *Journal of Animal Ecology*, which were chosen to ensure a good cross section of leading American and European based journals. I selected 2005 because that would give sufficient time (5 years) for manuscripts to garner sufficient citations for this analysis, and avoid earlier years during which the F1000 rating system was still becoming established or was less well known. Over 2005 these journals collectively published 1530 articles. For each article I obtained the total number of citations that it had accumulated by March 31 2010 by using the Web of Science (WoS) data base, and using the search tools in the F1000 data base to determine its F1000 score. Publications that were not present in the F1000 data base, and had therefore not been recommended by any members of the F1000 panels, were allocated a F1000 score of 0.0.

Frequency distributions of citation counts from the WoS for all 1530 publications, for the 103 of these publications that were selected by F1000, and for the 26 of these publications with F1000 ratings of 'must read' or 'exceptional' (i.e., scores of >3), are given in Table 1. It is apparent from this table that F1000 evaluations serve as weak predictors of which publications go on to have the greatest impact. While publications selected by F1000 members do on average have a higher mean and median number of citations than those not selected, this difference is not substantial, and the ability of F1000 to predict those papers that go on to have the greatest impact is poor. For example, this analysis reveals that:

(1) None of the 12 most cited publications published over 2005, including all 11 cited between 120 and 497 times, were recommended by any F1000 member.
(2) In contrast, 14 publications that were identified by the F1000 system as recommended or higher (including 3 identified as 'must read') were subsequently cited between 4 and 9 times over this period.
(3) Of those 103 publications rated by F1000 as 'recommended' or higher, 46% were subsequently cited less often than the mean of all 1530 publications, while 31% were cited less than the median. For those 26 identified as 'must read' or 'exceptional' (i.e., the top 1.7 % of publications), the corresponding figures are 35% and 19%.

Table 1. Frequency distribution of citations from publication date until 31 March 2010 (both totals and percentages) for all manuscripts published in seven major ecological journals in 2005 (see text for listing), as well as the manuscripts within that pool either 'selected' by F1000, or allocated a F1000 score >3 (i.e., 'must read' or 'exceptional').

Number of citations	All Publications		Publications selected by F1000		Publications with F1000 score >3	
	Total	Percent	Total	Percent	Total	Percent
200-500	4	0.3	0	0.0	0	0.0
150-199	2	0.1	0	0.0	0	0.0
130-149	5	0.3	0	0.0	0	0.0
120-129	3	0.2	1	1.0	1	3.8
110-119	2	0.1	2	1.9	1	3.8
100-109	3	0.2	0	0.0	0	0.0
90-99	5	0.3	2	1.9	1	3.8
80-89	5	0.3	1	1.0	0	0.0
70-79	15	1.0	2	1.9	0	0.0
60-69	26	1.7	6	5.8	1	3.8
50-59	42	2.7	7	6.8	2	7.6
40-49	72	4.7	12	11.6	4	15.3
30-39	151	9.9	9	8.3	1	3.8
20-29	290	19.0	23	21.3	7	26.9
10-19	508	33.2	25	23.1	5	19.2
0-9	397	25.9	14	13.5	3	11.5
Total	1530	100.0	103	100.0	26	100.0
Median	16.0		23.0		27.0	
Mean	22.3		32.6		37.5	

(4) The three publications of the 1530 identified by F1000 as 'exceptional' (i.e., the top 0.2% of publications) were cited 112 times, 27 times and 13 times, corresponding to a rank (out of the 1530 publications) of 17th, 381st and 905th respectively. As such, while these three publications do include one that has subsequently had a significant impact, they also include one whose citation count was in the lowest 45% of all publications.

Further, we may predict that of those papers actually highlighted by the F1000 panel (i.e., with non-zero scores), those that receive high scores ('must read' or 'exceptional') would go on to have higher citation counts than those with the lowest non-zero scores (i.e., 'recommended'). This could be the case, for example, if the F1000 panel was at least able to reliably rate the relative scientific merit of those manuscripts that they selected for profiling. Although Table 1 does provide evidence for slightly greater mean total citations for manuscripts with high F1000 scores, this difference is not large, and even here it is apparent that many manuscripts that receive low non-zero scores end up being cited much more heavily than many manuscripts that receive high scores.

The issue therefore arises as to why F1000 scores of publications perform so poorly in predicting their subsequent citation count, and crucially, conspicuously fail to predict those highly influential publications that go on to become cited hundreds of times. I identify three reasons.

First, despite suggestions to the contrary on the F1000 website, the F1000 panel does not cover all branches of ecology evenly. There are many examples, but a particularly noticeable one is the 'Community Ecology and Biodiversity' area, which is heavily represented by researchers that work on the biodiversity-ecosystem functioning issue, at the expense of many other topics. As such, publications that represent even relatively minor advances in the diversity-function area (and subsequently garner only modest numbers of citations) often get generous F1000 evaluations, while publications in other branches of community ecology that go on to make a substantial impact in the field are often overlooked.

Second is the issue of cronyism. While the vast majority of panel members undoubtedly take the high road and aim to be scrupulously fair, this is probably not true of all. Cronyism can exist at two levels. First, the F1000 'section heads' appoint the 'faculty members' who perform the evaluations, and there are several examples in ecological fields of these heads appointing their sometime coauthors, collaborators, those who agree strongly with them on controversial topics, and even their recent Ph.D. graduates. Second,

there are several examples of 'faculty members' in ecological fields generously recommending publications written by each other and their sometime coauthors (akin to the type of citation ring highlighted by Keddy 2006), as well as those by the 'section heads' that appointed them. The F1000 system has no obvious checks in place against potential cronyism, and whenever cronyism rather than merit is involved in any evaluation procedure, perverse outcomes are inevitable.

Third is the issue of geographical bias. In this analysis, four of the seven journals have similar impact factors (between 4 and 5) and exercise similar degrees of selectivity; it would therefore be reasonable to expect them to receive comparable coverage by F1000. Two of these, *Ecology* and *American Naturalist*, are published in America, while the other two, *Journal of Ecology* and *Journal of Animal Ecology*, are published in Europe. For all manuscripts published in 2005, 8.4 % and 15.7 % of papers published in *Ecology* and the *American Naturalist* were recommended by F1000, while only 1.6 % and 1.7 % of papers published in the *Journal of Ecology* and the *Journal of Animal Ecology* were recommended. It is notable that in the 'Ecosystem Ecology' area, 26 out of 32 (or 82 %) of 'faculty members' selected by the 'section heads' (themselves American) are from North America. Despite claims on the F1000 website of avoidance of geographical bias, for the 'Ecosystem Ecology' area at least, one gains the impression of American 'section heads' selecting mostly American 'faculty members' who then recommend mostly American publications.

Further, it is relevant that WoS citation counts, despite their obvious limitations, effectively represent a form of scientific evaluation performed by the whole global scientific community, whereas the F1000 ratings are performed by a selected elite subset of this community (or, in the words of the F1000 website, the '*world's top scientists*'). However, given the above points, it is more likely that the very large discrepancies between total citation counts and F1000 scores are due to inherent problems with the F1000 evaluation system rather than because those who are selected to serve as F1000 evaluators are more capable of identifying the best research than are other ecologists. For example, it is not credible that the 12 most cited papers in published 2005 were not selected by F1000 members because they were poorer papers than those that were selected.

The poor concordance between F1000 ratings and subsequent impact of ecological publications may be even more extreme than this analysis suggests. First, it is possible that profiling of publications by F1000 members might in turn cause them to gain more attention and thus greater number of citations than they would otherwise attract; in this light the F1000 process is widely advertised as a means for researchers to filter out the most 'important' and 'interesting' publications from the rest. Second, publications by well known authors are more likely to be read both by F1000 panel members and by the scientific population at large, likely leading to both higher F1000 scores and greater numbers of citations than publications of comparable merit from lesser known authors would attract. This would strengthen the relationship between the two measures.

There is a widely recognized and obvious need for effective post-publication evaluation of manuscripts that avoids the flaws associated with using the impact factor of the journal that the work was published in. However, despite widespread advocacy for the F1000 approach as evaluation tool, this analysis suggests that the current version of this approach has serious limitations for this purpose at least with regard to the ecological literature. Some discrepancy between F1000 scores and subsequent citation counts across publications is inevitable regardless of how well either measure reflects scientific impact. However, it is scarcely credible that all 11 publications cited more than 130 times over the past five years in this analysis (including one cited 497 times), and which all have F1000 scores of 0.0, are less 'important' or 'interesting' than the 14 manuscripts published over the same period that are cited 9 or fewer times (including two each cited 4 times) and have F1000 scores of 3.0 or more. If, as this analysis suggests, the F1000 process is unable to identify those publications that subsequently have the greatest impact while highlighting many that do not, it cannot be reliably used as a means of post-publication quality evaluation at the individual, departmental, or institutional levels.

Acknowledgements

I thank Richard Bardgett for comments on an earlier version of this manuscript and Tom Tregenza and Roosa Leimu for comments on the submitted version; the views expressed are however solely those of the author.

Referees

Tom Tregenza – T.Tregenza@exeter.ac.uk
University of Exeter, United Kingdom

Roosa Leimu – roosa.leimu@plants.ox.ac.uk
University of Oxford, United Kingdom

References

Aarssen, L., and C.J. Lortie. 2009. Ending elitism in peer-review publication. Ideas in Ecology and Evolution 2:18-20. CrossRef

Allen, L. et al. 2009. Looking for landmarks: The role of expert review and bibliometric analysis in evaluating scientific publication outputs. PLoS ONE 4:e5910. CrossRef

Badjer, K. 2005. The faculty of 1000 biology factor will revolutionize scientific evaluation and publishing. The Hypothesis Journal 3:10-12.

Keddy, P. 2006. Putting the plants back into plant ecology: Six pragmatic models for understanding and conserving plant diversity. Annals of Botany 96:177-189. CrossRef

Leimu, R. and J. Koricheva. 2005. What determines the citation frequency of ecological papers? Trends in Ecology and Evolution 20:28-32. CrossRef

Lortie, C.J. et al. 2007. Publication bias and merit in ecology. Oikos 116:1247-1253. CrossRef

Marder, E. 2006. Rejecting arrogance. Current Biology 16:R70. CrossRef

Monasterksy, R. 2005. The number that's devouring science. Chronicle of Higher Education, 52:A12.

Raff, M.C. et al. 2004. Changing scientific publishing. Science 305:945-946. CrossRef

Seglen, P.O. 1997. Why the impact factor of journals should not be used for evaluating research. British Medical Journal 341:498-502.