Grumpy Scientists

the Ecological Conscience of a Nation

Edited by Daniel Lunney, Pat Hutchings and Harry F. Recher
Royal Zoological Society of New South Wales
Grumpy Scientists
The Ecological Conscience of a Nation

Edited by
Daniel Lunney, Pat Hutchings and Harry F. Recher

Royal Zoological Society of New South Wales 2013
This book is copyright. Apart from any fair dealing for purposes of private study, research, criticism or review, as permitted under the copyright Act, no part may be reproduced by any process without written permission.

The National Library of Australia Cataloguing-in-Publication entry:
Grumpy Scientists: the ecological conscience of a nation
ISBN 978-0-9874309-0-8 hardcopy
978-0-9874309-1-5 ebook
http://dx.doi.org/10.7882/FS.2013.001

Lunney, Daniel, Hutchings, Pat and Recher, Harry F.
Royal Zoological Society of New South Wales.
Wildlife research; climate change; animal ethics; science policy; journal impact factors; Australian fauna; ecological conscience; biodiversity crisis; extinction, taxonomists, museums

Published July 2013

Published by

Royal Zoological Society of New South Wales
P.O. Box 20, Mosman NSW 2088 Australia
Phone: (02) 9969 7336
office@rzsnsw.org.au
www.rzsnsw.org.au
for publications rzsnsw.metapress.com

Design and layout by Scribblygum Design, printed in Australia.

Cover montage: Large coal mines in the Hunter region of New South Wales and Australian pelicans on Lake Eyre Basin. R.T. Kingsford.
Introduction

It is fun to imagine song writers being assessed in the way that scientists are today. Bureaucrats employed by DAFTA (Ditty, Aria, Fugue and Toccata Assessment) would count the number of songs produced and rank them by which radio stations they were played on during the first two weeks after release. The song writers would soon find that producing junky Christmas tunes and cosying up to DJs from top radio stations advanced their careers more than composing proper music. It is not so funny that, in the real world of science, dodgy evaluation criteria such as impact factors and citations are dominating minds, distorting behaviour and determining careers. Lawrence (2007)

Imagine a time, perhaps not so far off, when scientists are ranked like tennis players, measured by their number of papers, impact factors of the journals concerned, their position in the author list and the number of citations their papers receive — put these numbers into a computer and watch it generate your publicly available ranking as number 2,340 in the world! Indeed, a tendency to rank like this already exists, causing biomedical scientists to focus more on their careers and less on understanding nature and disease. Lawrence (2002)

Taxpayers fund research in universities and government departments. Therefore there is legitimate concern that these funds are spent wisely, efficiencies are encouraged, research effort is matched to research need and that professional standards in the public research community are reviewed and improved regularly (Oswald 2010; d’Angelo et al. 2011). Such review is expensive using traditional peer review, so bibliometrics - methods to quantitatively analyse the quality and impact of scientific or technical literature (e.g., Corsi et al. 2010; OECD 2010) - offer an economical alternative accompanied by the apparent objectivity of numbers as opposed to the relative subjectivity of peer review (Adler et al. 2008; d’Angelo et al. 2011). In Italy, Abramo et al. (2013) noted that the rankings produced by the Valutazione Triennale Della Ricerca (VTR) after expensive peer review correlated strongly with those calculated using bibliometric data available freely online. They concluded that, for the sciences at least, similar conclusions would be reached at far lower cost by using bibliometrics rather than peer review.

Others express concerns that ranking schemes may lead people to play to good scores rather than sound science (Burler 2007) by focusing on short-term, high-profile projects (Box 2010; Corsi et al. 2010; Northcott and...
Linacre 2010) to the detriment of research diversity and clinical or applied research that influences professional practice, not citations (Roa et al. 2009; Shewan and Coats 2006; Sampson and Comer 2010; Calver et al. 2013c). This can lead to a clash of values when ranking schemes are implemented, especially if the ranking statistics are deployed inappropriately or insensitively.

As Lawrence (2002, 2007) noted in the opening quotes, one may satirise the mismatch of evaluation tools with the task they supposedly perform, but because of the consequences for science it is actually not funny. Nevertheless, satire can highlight the problems, just as dandruff will show on a dark collar. To do this, I’m going to call for help from four employees at the Department of Environment and Natural Resources (DENR, pronounced denner) in the fictitious Australian mainland state of Transcendentia. One, Research Scientist Frank Lamming, is being interviewed for promotion to Senior Research Scientist. Frank is thirtyish, talented, successful – and he knows it, which sometimes makes him too brash for his own good. Sue Wallace from Media and Public Relations is chairing his interview panel (which tells us something about how DENR views its science). She’s long on patience and has an open mind, good qualities to have when speaking to Frank. Then there’s Frank’s immediate boss Peter Ericson, Head of Weed Control, Vertebrate Pests and Wildlife Conservation. He has a tolerant streak too. Patience, tolerance and an open mind are somewhat lacking in the last panel member, Joe Tarazza from Human Resources. Joe considers himself well-informed on bibliometrics, which goes to prove Alexander Pope’s classic dictum that a little learning is dangerous. Through the panel discussions, I will explore some of the critical issues underpinning assessments of the quality and impact of research, before offering suggestions for using bibliometric statistics validly when building a research profile.

Common misuses of bibliometrics or publication data

The value of baseline research and natural history

*There is still a great need for conservationists to understand the basic ecology of individual species … . In the end our ability to arrest the decline of woodland birds depends on how well we know the individual species, and understand the ecological processes leading to decline, such as dispersal, interspecific competition and predation. Surely there are enough ornithologists, ecologists and conservationists in Australia for us to be able to achieve this!* Ford (2011)

Sound natural history data and careful descriptions of behaviour, distributions and communities are essential for forming new hypotheses in animal ecology, or for providing baselines against which changes can be measured (Erskine 1994). Yet such studies are not published by major international journals in animal ecology or conservation, which place a premium on novelty and interest to a broad readership as well as scientific quality in evaluating contributions (e.g., Meffe 2006). While national journals were once suitable outlets for basic Australian research, over the past decade Australian journals including *Emu, Wildlife Research* and *Austral Ecology* have shifted their editorial policies to favour experimental studies, theoretical explorations and overseas reviewing or content as part of repositioning themselves internationally. This is at the expense of descriptive work, which is increasingly difficult to publish (Bryant and Calver 2012). As Frank finds in his interview, the view that baseline or descriptive publications are beneath a good scientist is influencing promotion panels as well.

Joe:

Frank, on a quick look, your publication record is rather good for a new Research Officer – 12 papers over the last five years. But research finds new information and shapes a discipline. Isn’t what you’ve published while at DENR basically just baseline information and reviews?

Frank:

You never know when baseline information will become valuable. Schuch’s recent paper in *Biological Conservation* caught my eye.¹ He went back to some work from East Germany in the 1960s – published in some obscure museum journal. But it documented in detail the insect fauna at nearly 50 reserves in East Germany, with exhaustive methods. Schuch’s team could go back and do an incredibly important retrospective study – but only because the original study was done and documented.

Joe:

That was probably a lucky one – most of them must vanish. Let’s get to the point. You’ve written a lot documenting established facts about gliders and local environments. But I found over 2,000 references to gliders and other tree-dwelling mammals in one quick search online and I’m no expert – why don’t you read that for your background information and get into the field to find original information about gliders?

Frank:

Does that mean I get an iPhone?

Joe:

What?

Frank:

An iPhone – so I can check all those online references while I’m in the field doing original research.

Joe:

Don’t be impertinent!

Frank:

Uh – sorry. But I just wanted to make the point that field research projects don’t deliver publishable field data from day one. Setting up experiments and running them takes months or years, especially if seasonal replication is needed. To keep up a publication record I write reviews and descriptive papers while waiting for the experimental data to come through. And those reviews pull out what’s relevant for the local situation – makes it easier for locals to find what’s

---

¹ Schuch et al. (2012).
important here. I also get a chance to collate our departmental grey literature and summarise it in a peer-reviewed paper with wider circulation. That lifts awareness of what we do.

The paper Frank mentions, Schuch et al. (2012), makes use of original data from Schiemenz (1969). He documented meticulously the auchenorrhynchan (a sub-order of the true bugs, Hemiptera) fauna of 48 protected dry, grassland sites in Eastern Germany, as well as his detailed methodology. Given the detail, Schuch et al. (2012) resurveyed 26 of Schiemenz’s sites in a long-term study of changes in the auchenorrhynchan fauna. In a recent editorial, Schuch (2012) pointed out that opportunities for further long-term studies were almost lost when a departmental library was closed and important archival data were almost discarded. Closer to home, Short (2004) re-evaluated data from Shortridge (1909) on the distributions of southwestern Australian marsupials. Such examples of significant contemporary studies based on original surveys up to a century old indicate the enduring value of such work, even though this might not be apparent immediately.

Ranking journals

Our results indicate that the notion of scientific impact is a multi-dimensional construct that cannot be adequately measured by any single indicator; although some measures are more suitable than others. The commonly used citation Impact Factor is not positioned at the core of this construct, but at its periphery, and should thus be used with caution. Bollen et al. (2009)

Journal ranking is most commonly done using the Journal Impact Factor (JIF) (Garfield 2005) which, for a journal in year x, is the average number of citations to papers in the journal in the two years prior to x. Although numerous other measures are possible and some may well be superior (see Calver and Bryant 2008; Bollen et al. 2009; Colledge et al. 2010; Calver et al. 2012; Chang and McAleeer 2012; Vanclay 2012), journals often display JIF on their web pages. Editors may even adjust editorial policies to manipulate the JIF of their journals (Ashkanasy 2007; Falagas and Alexiou 2008). Many departmental administrators also use JIF as a convenient measure of the quality of the journals in which their faculty publish (Adler et al. 2008).

For many, the logic of applying the JIF to the question of the quality of a journal or the papers it publishes is seductively simple. The higher the JIF, the higher the quality of the journal and the better the quality of the papers it contains. Others argue that the JIF itself is limited and represents only part of a journal’s full quality or impact profile (Adler et al. 2008; Chang and McAleeer 2012), or that one cannot judge a paper by the journal in which it appears (Seglen 1997; Adler et al. 2008). These views clash at Frank’s interview.

Joe: Well, if we can’t establish quality that way, let’s try something different. Can you name three prominent journals in your field?

Frank: Not sure everyone would agree with these, but I’ll suggest Conservation Biology, Biological Conservation and Diversity and Distributions.

Joe: And papers in these journals have a real impact – they shape the thinking in your discipline?

Frank: Yes.

Joe: And the standard of science in papers in those journals is uniformly high?

Frank: On the whole, yes – the odd strange paper gets through the reviewers, but most are very good.

Joe: Yes, and you don’t publish in them. In fact, you publish in mid-tier local journals – Pacific Conservation Biology, Wildlife Research, Austral Ecology, Australian Zoologist, even book chapters – doesn’t this mean that your research lacks impact and the science is low standard?

Frank: Um, I’m not sure that Wildlife Research and Austral Ecology would see themselves as local – their current editorial policies aim them at international audiences.

Joe: Deal with the question, please.

Frank: I submitted some of my first papers to Conservation Biology and Biological Conservation. They said that the work was good but regional, lacking the broad general appeal they prefer to publish. They advised me to submit to prominent regional journals. It was good advice. I had good reviewing sensitive to local issues and published in journals read by local practitioners. As to impact, my papers are available online through the departmental research repository.

Last time I looked the biggest download was 170, the smallest 20 for the most recent paper. People read my papers.

Joe: You’re avoiding the point. If you don’t publish in the best journals, you don’t shape the discipline and your work isn’t of a standard I’d expect of a senior research scientist.

Frank: I doubt that the editors of some of the best journals would agree. Take Conservation Biology – they reject over 80% of the papers they receive. When was it? Early noughties I think – then editor Gary Meffe wrote an editorial explaining the selection policy.

Yes, papers had to be scientifically excellent, but they also had to be novel and with broad appeal. Can’t remember all the categories – but he said that some papers he’d call ‘critical studies’, ‘important’ and ‘breakthrough’ failed the novelty or broad appeal tests and he wouldn’t publish them. When the last of our exclusion experiments is done we should have something right for a big journal. But for the moment we’re solving local problems and publishing good work in local journals. And don’t rubbish local journals – Meffe himself coined the term ‘internationally recognised regional journal’ for those focusing on a region but still relevant internationally.

--

2. Institutional research repositories are seen increasingly as valuable tools to promote research online (Jain 2011).

Joe: The fact remains that you publish in low impact factor journals.
Frank: Well, where you work in the library . . .
Joe: No! I'm from Human Resources!
Frank: Sorry, I forgot. It's just that the Journal Impact Factor was developed to help librarians, so I thought you might be one. It was used to prioritise selection of journals for cash-strapped libraries – I bet the DENR librarians use it. All the fluff about quality –
Joe: It is not fluff! It's about standards!
Sue: Joe, let Frank finish his answer.
Frank: As I was saying, the Journal Impact Factor isn't a good surrogate for quality. It's based on a two-year window – disciplines like mine often don't really start accumulating citations until well beyond that. It's based on the Web of Science database, which excludes a lot of regional journals. Pacific Conservation Biology and Australian Zoologist don't cut it, but stuff we need is published there. Even the raw data are unreliable. There are plenty of papers documenting the errors, in lots of disciplines. Impact factor may have been 'state of the art' 15 years ago, but it's outdated now. Adler and his colleagues argue that judging a journal by its impact factor is like assessing a man's health by his weight – I reckon they're right!
Joe: Errors and history lessons aside, papers in the top journals by impact factor must have a greater impact than those in low impact factor journals.
Frank: Not necessarily. I agree with those who say that a paper can't be judged by the journal in which it appears. The distribution of citations in any journal is skewed.
Peter: Excuse me, skewed?
Frank: Not evenly distributed about the mean. In any journal, a few papers attract most of the citations. Most papers aren't that well cited. That means that a well-cited paper in a low impact journal may have more citations than many papers in a high impact journal. People shouldn't regard highly cited papers in 'low impact' journals as less important than poorly cited papers in 'good' journals. And that assumes that citations are a good measure of research impact or quality anyway – maybe they just show discussion. Maybe the ease of access of the work is more important than quality in determining whether it gets cited.

Frank and Joe are not the only ones to get worked up over the JIF. Vanclay's (2012) criticism of the JIF drew a blistering response: This paper reflects on the most current and some of the recent contributions of JK Vanclay, focusing on his methods, findings, and criticism about the journal impact factor and the h-index. It is argued and demonstrated that some of the recent papers of the author about scientometric issues, measures and sources show so much demagoguery, ignorance and arrogance, have so much prejudice and bias, so profound errors in using the databases, calculating metrics, and interpreting search results that the papers are very unlikely to be meant as a genuine contribution from an academic who is a graduate of many others-Oxford University, professor and dean in a respected university, a well-published and well-cited author and a recipient of the Queen's Award (all the above in forest science). The papers are much more likely to serve as props for a staged, mock-up scenario based on slipshod research in an experiment, to illustrate the deficiencies in the processes and in the assessment of scholarly publishing productivity and impact in order to present a more idealized solution of Vanclay: using the h-index, portrayed as the Prince, mounted on the shoulder of the White Horse. Jacsó (2012).

It is therefore unsurprising that Buela-Casal and Zych (2012) called the JIF 'a highly polemic topic'. Rather than contribute to the polemical debate, they surveyed 1 704 scientists from 86 countries regarding their opinions of the JIF. The responses were deemed 'neither positive nor negative', with respondents taking a position slightly above neutral on most questions. Nevertheless, there was a significant positive correlation between respondents' perception of the importance of the JIF and the use of the JIF in ranking research performance in their home countries. There was also a significant negative correlation between the number of papers respondents had published and their opinion of the JIF, suggesting that more experienced researchers had less respect for the JIF.

Frank and Joe represent the extremes of opinion regarding the JIF as a measure of journal quality. Bensman (2012) has the most measured position: 'abandoning it would be counterproductive because of its demonstrated ability—even with its defects—to identify small important journals like review journals, giving it an important role in science evaluation and library collection management.' There is also general agreement that JIF should not be used as a

---

4. Frank is not quite right here, although on the right track. Eugene Garfield, who developed the Journal Impact Factor together with Irving Sher, put it this way:

In the early 1960s, Irving H. Sher and I created the journal impact factor to help select journals for the new Science Citation Index (SCI) . . . From this simple exercise, we learned that initially a core group of large and highly cited journals needed to be covered in the new SCI (Garfield 2005).

5. For criticisms of the journal impact factor; see Vanclay (2012). For defenders, see Bensman (2012) and Jacó (2012).


10. Amancio et al. (2012).
surrogate for the quality or impact of the papers within a journal (e.g., Seglen 1997; Bloch and Walter 2001; Adler et al. 2008), nor to evaluate individual scientists (Krell et al. 2012), although these concepts persist amongst at least some departmental administrators (Adler et al. 2008; Hodge and Lacasse 2011).

**Ranking authors**

The drive towards more transparency and accountability in the academic world has created a "culture of numbers" in which institutions and individuals believe that fair decisions can be reached by algorithmic evaluation of some statistical data; unable to measure quality (the ultimate goal), decision-makers replace quality by numbers that they can measure. Adler et al. (2008)

Over time, the mean number of authors on scientific papers has been rising (Kennedy 2003; Harrison 2006). While this suggests more collaboration, it can cause conflict over right to authorship and also over the ordering of authors’ names on co-authored publications. Codes of practice such as NHMRC (2007) give guidance on these important matters. In addition, various unwritten conventions have arisen about the significance of the placement of an author’s name in a list of authors. As Frank discovers during his interview, his view of this varies from that of members of the interviewing panel.

Joe: You’re first author on only two of your 12 papers over the last five years and last author on only one of them. Isn’t it the case that the only substantive contributions are recognised by first or last authorship? Aren’t you commonly an ‘add on’ to an authorship list?

Frank: Now, look - (controls himself). DENR’s clear in its authorship policy – we follow the NHMRC guidelines.\(^{11}\) Authors have to take a substantive role in planning, execution, analysis or writing to the point where they can take responsibility for part of the work. I can for every paper where I’m a named author.

Joe: Of course I know the DENR policy. But it’s also well known that first or last are the spots taken by the most influential authors on any paper.

Frank: I’m with Linda Butler on this one – that paper in the OECD proceedings on research assessment.\(^{12}\) Some people share your view. Lots of others – me included – don’t, or at least don’t accept all of it. Last authorship means nothing to me other than the person who made the smallest (but still significant) contribution. First authorship usually indicates the driver in the paper, unless there’s a student involved. If there is, then I’m with those who say let the student lead.

Unlike many of the other contentious issues in this interview, this one does have a clear solution:

Any attempt to fractionate publications across collaborating authors/institutions is fraught with difficulties, particularly if an attempt is made to weight shares according to level of contribution. Practices vary too widely to allow this to be done automatically. Until such time as authors routinely ascribe percentage contributions, it is unlikely that fractionation can be used without seriously disadvantaging some authors and/or disciplines. Butler (2010).

It seems that Frank is right - it is inappropriate to place great weight on author order when assessing a researcher’s contributions because there is no single convention that applies across all disciplines.

Citations, though, are a different matter and the frequency with which a paper is cited is often taken as an indicator of its quality or impact. In sum, the pattern of citations of an author’s work is also taken as an indication of the author’s scientific standing, and numerous indices have been developed in an attempt to reduce citation patterns across many papers to a single index (Adler et al. 2008; Chang and McAleer 2012). The most well known of these is the Hirsch index. Papers are ranked from the most highly cited to the least highly cited, with the Hirsch index \( h \) being the highest ranked paper where the number of citations equals or exceeds that paper’s rank (Hirsch 2005).

Such assessments of research performance are far removed from the original intention of citations, which was to assist in literature searches by enabling researchers to trace papers citing key work (Garfield 2005). From there, they came to be seen as a useful tool for librarians in deciding on subscriptions and, most recently, in evaluating research performance (Buela-Casal and Zych 2012). It is this latter use that causes sparks to fly in Frank’s interview.

Peter: Joe, Sue – this may be a good time for me to ask some questions?

Joe: (curtly) Yeah, I’m done.

Sue: Fine, go ahead, Pete.

Peter: I’d like to return to your comments about citations. Leaving aside impact factors, we can use the Hirsch index to assess the acceptance of your papers by your scientific peers. How does your Hirsch index compare to those of other research scientists and senior research scientists in DENR?

Frank: Could you clarify, please – why choose the Hirsch index for comparison?

Peter: It’s the yardstick – the single established metric.

Frank: Huh. That reminds me about the matron at the censorship conference saying ‘but surely there’s just the one position for sexual intercourse’.

Joe: Frank!

Frank: Err – sorry. But I’ve lost track of all these citation indices. There are so many that there’s now a collective acronym for them – RAMs – research assessment measures – let’s see, (counts on his...
Frank: If I must, but I’d rather not. I’m an ecologist. Ecologists work with multivariate data sets all the way from single species to communities. We know that a single index is often incredibly misleading. Yet when it comes to assessing the ‘impact’ of my papers, you want me to use one index out of many. Surely it’s a multivariate problem needing multivariate statistics. That’s what I liked so much about that paper by Bollen’s team – the one in PLoS One – that looked at a raft of citation measures for assessing research altogether. They even looked at non-citation measures too – download statistics were one of them. And funny thing – the first division in their cluster analysis was between the non-citation measures and the citation measures. They measure fundamentally different things! Maybe we should do something like that, with the guys from biostats to help us.

Peter: OK. But I still want you to pick a measure to assess you.

Frank: Alright, one. I’m 30 – early in my career. I’m in the first division in their cluster analysis with the non-citation measures and the citation measures. For the number of co-authors. I’ll look best with maybe the contemporary Hirsch index. The Hirsch index, the contemporary Hirsch index, individual Hirsch indices – even Google Scholar is in on the act with its h5-index and h5-median. They all have to reference Hirsch, of course, so he’s headed for a citation classic.

Peter: That’s still two indices.

Frank: (laughs) Got me. I just can’t shed the idea that one index alone won’t cut it. And I’m not alone. There have been studies of what researchers actually think about citation indices,
which suggest that they think that they aren't particularly useful as indicators of scientific contribution. Researchers ought to know – they're the ones doing the citing!20

Peter: OK, no more questions from me. I'll follow up on your scores with those indices. Sue!

The first critical issue emerging from this exchange is that assessing research impact or research quality is a multivariate problem, irrespective of whether it applies to the evaluation of journals, papers or individuals (Bollen et al. 2009; Calver et al. 2010). Many assessment indices are possible - Chang and McAleer (2012) describe 13, and that list is incomplete (Bollen et al. 2009; Li et al. 2010). Some measures correlate highly with each other, while others do not and therefore provide complementary information (Calver and Bryant 2008; Chang and McAleer 2012). Furthermore, not all measures are citation based, allowing for the real possibility in applied disciplines that people may read and act on the recommendations in papers but not necessarily publish themselves and cite the papers (Shewan and Coats 2006). In an online environment, social network analyses and download data provide alternative measures of which papers are being read, which may lead to very different conclusions as to uptake than citation data Bollen et al. (2009). Different citation measures may also have varying relevance to different disciplines and even to different research approaches (Ouimet et al. 2011; Rafols et al. 2012). Multivariate analyses are an obvious solution to this complexity (e.g., Bollen et al. 2009; Perakakis et al. 2006; Calver et al. 2010), but the prevalence of simple univariate statistics or additive combinations of them in the major online databases such as Web of Science, Scopus and SCImago and in much of the research literature (e.g., Buelsa-Casal et al. 2007) indicates that the uptake is low. This places the univariate studies at risk of specific biases in citing (MacRoberts and MacRoberts 1996), or the problem that research such as taxonomic papers or descriptive floras can be influential without attracting high citations at all (MacRoberts and MacRoberts 2010).

The second issue is that there is often uncritical acceptance of the accuracy of the information in databases, despite cautions from experienced bibliometricians (Jacsó 2009a,b; Walters 2011; d’Angelo et al. 2011; Gagolewski 2011; de Sutter and van den Oord 2012) and the documented experience of zoologists (Calver et al. 2013a). If one is using databases as a literature search tool to trace papers on a particular topic, mistakes in attributing a paper to the correct author are at most a minor irritation. But if the aim is to evaluate authors, journals, institutions or even countries, then such errors can lead to flawed conclusions (Jacsó 2012).

Abbott (1998) affords an unusual example. The address line for this paper indicates the author’s affiliation as ‘Science and Information Division, Dept. Consrv. Land Mgmt., Locked B., Bentley, WA 6983, United States’, although the author was (and still is) based in Western Australia where he completed the work. Presumably, the original manuscript used the abbreviation ‘WA’ for Western Australia, which a copyeditor misinterpreted as meaning Washington State and ‘corrected’ the final version. This is certainly not an isolated error, another instance being Shearer et al. (1981) where the authors’ address is given as ‘Forests Department, Research Branch, Dwellingup, 6213, W.A., United States’. Before dismissing this as trivial, it is worth noting that SCImago’s SIR database includes the statistic ‘percentage international collaboration,’ which is the percentage of papers from an institution that have authors from more than one country. In the 2009 SIR (http://www.scimagoir.com/pdf/sir_2009_world_report.pdf), the four Western Australian public universities ranked 4th, 5th, 8th and 11th out of 36 Australian public universities for international collaboration. Could the high scores for Western Australian universities on this measure relative to those from other states be in part an artefact of such copyeditors’ errors?

Finally, empirical evidence suggests that researchers in science and sociology have been cautious about using citations as indicators of research contribution. Based on a survey of 204 American scientists and sociologists, Hargens and Schuman (1990) concluded that the scientists did not regard citations as useful indicators of research contribution while the sociologists were only marginally positive. More recently, Aksnes and Rip (2009) found that Norwegian scientists were ambivalent about the utility of citations for assessing research performance. However, Hargens and Schuman (1990) were well in advance of the success of the Hirsch index (Hirsch 2005) and Aksnes and Rip (2009) published just after it, so these studies may have missed a sea change in attitudes.

Fighting back
People do not realize that when it comes to arguing their case for more funding, scientists who do basic research are the least articulate, least organized, and least temperamentally equipped to justify what they are doing. In a society where selling is so important, where the medium is the message, these handicaps can spell extinction. Arthur Komberg, quoted in Lawrence (2009)

Why fighting back is important
Ranking systems are intended to encourage and reward good research and good researchers, facilitating responsible and accountable allocation of public money. Ranking statistics, despite potential problems, offer objectivity and the opportunity to change researchers’ behaviour through a desire to perform well in rankings (Holden et al. 2005). However, should they be chosen poorly, the resulting changes in what researchers do, how they do it and their attitude to their work can damage individual careers and also science’s ability to provide the knowledge most needed for the community (Lawrence 2007; Adler and Harzing 2009, Calver et al. 2013c). Marsh et al. (2012) called undesirable consequences arising when individuals or organizations seek to optimize scores on ranking systems ‘perverse incentives’.  

20. Hargens and Schuman (1990); Aksnes and Rip (2009)
Kohn (1993) outlines the research evidence for the general dangers arising from ranking and reward systems in any area of human endeavour. Those of particular concern to science include: damaging of collegial relationships (colleagues become tools providing key skills or data, or even competitors); a focus on the score and not the task, discouraging risk-taking and diversity (surprisingly, when a reward is at stake people take the tried and true approach to reaching it rather than innovating); higher levels of anxiety (ironically decreasing performance); a greater chance of cheating; and a loss of intrinsic motivation (researching and publishing for one’s personal satisfaction, not just to score on a ranking scheme).

While Butler (2010) cautions against accepting anecdotal or perception-based criticism of ranking systems, there are already signs in science that an obsession with them can promote negative consequences that fit the model described by Kohn (1993). For example, research on characteristics of highly cited papers indicate that they are more likely to be reviews or techniques papers, to be longer, to have many authors, to include at least one highly cited author amongst the authors, to address one or more ‘hot topics’ in the current literature and to be positive studies (i.e., finding support for the hypothesis tested). They are also unlikely to be negative studies (finding no support for the hypothesis tested) or confirmatory studies (endorsing results published elsewhere) (see the short review in Calver and Bradley 2010). Therefore, authors writing to maximise their citations will concentrate on a narrow range of types of paper. Ironically, this is the antithesis of the conclusions of many authors who have studied the characteristics of highly cited papers, some of whom ‘...advocate that to be prominent in science, the most important goal is to be innovative, make good research and write high-quality articles’ (Padial et al, 2010), or ‘Ultimately, researchers should focus on improving the state of knowledge in their fields rather than becoming highly cited’ (Parker et al. 2013).

Nevertheless, Falagas and Alexiou (2008) argue that journal editors are aware of the characteristics of highly cited papers and consider them when deciding what to publish to boost the JIF for their journals, while Brown (2007) gives a detailed case study. Furthermore, selection priorities in highly regarded journals viewed favourably by ranking schemes require accepted papers to be excellent scientifically, novel, of broad appeal and making a significant advance (Meffe, 2006; Watson et al., 2007, Primack, 2009). Meffe (2006) acknowledges that many ‘critical’, ‘important’ and ‘breakthrough’ papers will lack novelty or broad appeal and hence be rejected by top journals.

Consequently, Lawrence (2007, p. R383) complains that: ‘...over the last twenty years a scientist’s primary aim has been downgraded from doing science to producing papers and contriving to get them into the “best” journals’. Research topics unsuitable for top journals may not be attempted or the data left unpublished (Roa et al. 2009; Sampson and Comer 2010, Marsh et al. 2012). This in turn may cause the demise of some local journals because of lack of copy or institutional support (Gowrishankar and Divakar 1999; Steele et al. 2006; Andersen et al. 2008), or changes in editorial policies to aspire to international status that reduce opportunities to publish local work (Bryant and Calver 2012).

A narrowing of research diversity, demise of some local journals, reluctance to publish negative results or confirmatory studies and a perception that some ‘critical’, ‘important’ and ‘breakthrough’ papers are less worthy because they are not in top journals is not good for science, scientists or society (Lawrence 2007; Adler and Harzing 2009). Some have taken the stance that they ‘...want no part of this dangerous and misguided exercise’ (Andersen et al. 2008). For those without that option, or who see that reform rather than rejection is the route to follow, the following suggestions may help.

**Fighting back in presenting individual profiles**

Here, the question is really how to present oneself fairly and honestly, while also speaking up for important forms of research and important publications that do not rate well under the current approaches. An analogy that may appeal to experimental zoologists is justifying conclusions from an unconventional study, where either logistical problems have dictated a less than ideal approach or unforeseen crises during an experiment such as storm damage to experimental plots have compromised the analysis. In this context, Underwood (1997) advised:

> Above all, be self-critical. If some experimental test of an hypothesis is problematic because of difficulties with controls, independence, replication and other aspects of design, the person most responsible for explaining this is the person who did the experiment. Then an argument can be advanced to explain why the results and interpretation should be accepted, despite the problem. This argument will rest on ancillary evidence, inductive notions based on experience, analogy, etc. Some of it may be compelling. All of it needs to be aired. Otherwise, readers, referees, editors, etc., elsewhere in the world all working on different problems are entitled (and have a duty) to reject the findings.

This advice can be followed profitably when presenting one’s research achievements using bibliometric measures. Even if a record is not outstanding in terms of many publications in top journals or a high Hirsch index, there may be convincing arguments that it has been influential. Less well-known points that could be considered are:

- **Download statistics.** If your institution has a research repository, you may find that download statistics for your work are high, indicating that people are reading the work although not necessarily publishing and citing it – an outcome claimed for many articles aimed at practitioners (Shewan and Coats 2006). Downloads may also give a completely different assessment of ‘important’ papers than citations. For example, the last time I checked downloads of my work in my institutional repository the top 10 papers by downloads included only two of the top 10 papers by citations in the Scopus database. Furthermore, the most highly downloaded paper – a ‘how to do it’ teaching publication – had never been cited. Presumably, teachers were finding it useful in the classroom although it wasn’t making a contribution to educational research.

- **WorldCat holdings.** As an alternative to citations for assessing the impact of a book, WorldCat (http://www.worldcat.org/advancedsearch) indicates libraries around
the world that hold copies of a title. Torres-Salinas and Moed (2009) and White et al. (2009) argue that librarians choose books carefully to meet the needs of patrons while containing costs, so library holdings are analogous to authors citing papers. Therefore WorldCat listings or those in online national library catalogues are publicly available indicators of the uptake of books. White et al. (2009) proposed ‘licitation’ for counts of library holdings. They also point out that books can be classified into Library of Congress class (LC classes), so books can be ranked according to their relative licitations within a class. Calver et al. (2013b) give a basic approach to using library holdings to assess uptake of books in natural history, biological conservation and zoology.

- ‘Corrected’ indices. A metric that corrects for the number of co-authors may improve the profile of an author who does not publish in large groups, or a metric that allows for career breaks can assist a researcher with a break in track record for family reasons (see Harzing http://www.harzing.com/pop.htm for convenient freeware approaches to such assessments using Google Scholar data).

- Profiling international citations. The Cited Reference Search option in the Web of Science database allows one to profile citations to a particular author by the countries of the citing authors. Even if citations are limited, many of them may be international, indicating a wide readership (see Appendix 1 for directions).

- Presenting citations to books and book chapters. Books and book chapters were only added to the widely used Web of Science database in 2012, with coverage for the previous five years (Thomson Reuters 2011). However, citations to books and book chapters outside this range can be retrieved with a Cited Reference Search within Web of Science. This option retrieves citations from sources listed in Web of Science to any source, irrespective of whether or not the cited source is also in Web of Science. Thus it does retrieve citations to books and book chapters prior to 2005. Similarly, while the Scopus database does not include books and book chapters (except for books in a named series) (Elsevier 2011), citations from entries in Scopus to unlisted items can be retrieved using the ‘secondary documents’ option. Instructions for completing a Cited Reference Search or using the secondary documents option in Scopus are given in Calver et al. (2013b). Google Scholar can also be used to retrieve citations to books and book chapters, but see the comments on ‘Google Scholar problems’ below.

- Alternative metrics. Dissatisfaction with conventional citation analysis and the obsession with the Journal Impact Factor is driving rapid developments in the field of ‘alternative’ research metrics, which purport to access a range of web-based indices of research impact other than citations or to place citation data in a comparative context. Here are the beta sites for three:

- ImpactStory - http://impactstory.org/create. This site gives alternative metrics for journal papers. These include citations in the context of citation norms (e.g., this paper has 8 citations, which places it in the top 3% of cited papers published in 2012), the incidence with which papers are saved, and the frequency of online discussion in scholarly and public sources.


- Altmetric - http://www.altmetric.com. One feature of this site is a ‘bookmarklet’ users install in their web browser. Clicking this when visiting a web page linking to a journal paper opens a menu displaying a range of metrics. They include tweets, Facebook entries, posts, blogs, and Mendeley (free reference manager) readers.

Whatever approach is chosen, it is important to beware logical flaws in the analysis. These can be subtle, but devastating to conclusions (Butler 2011). Pitfalls include, but are not limited to:

- Inappropriate benchmarking. Comparing your own record with another’s is risky, because without knowledge of the other person’s full CV your search will be biased by errors and omissions that only the author could detect (Jacsó 2008, 2009a, Calver et al. 2013a).

- Mixing and matching databases. Different databases may give very different citation counts in response to a literature search because of the breadth of material they include or the range of years scanned. If results from different databases are compared or combined without taking this into account, flawed interpretations result. Jacsó (2005) contrasts the important differences between the major databases Web of Science, Scopus and Google Scholar. Although all databases have undergone major developments since then, his article is still a useful starting point in appreciating the differences.

- Web of Science comes in different flavours. The widely used Web of Science database comes in differing subscription versions (Jacsó 2008, 2009, Calver et al. 2013b). Different databases may give very different citation counts in response to a literature search because of the breadth of material they include or the range of years scanned. If results from different databases are compared or combined without taking this into account, flawed interpretations result. Jacsó (2005) contrasts the important differences between the major databases Web of Science, Scopus and Google Scholar. Although all databases have undergone major developments since then, his article is still a useful starting point in appreciating the differences.

- Google Scholar problems. Despite its value for finding a wide range of literature online, especially open access versions (Walters 2011), Google Scholar is criticised heavily for its failings as a bibliometric tool. Jacsó (2008b) (not one to pull his punches, as noted in his comments on the lack of its value for finding literature) has said that ‘…the deficiencies in Google Scholar are so voluminous, unscholarly, and often so hidden that its hit counts and citation counts should not be accepted even as a starting point for evaluating the research output of real scholars’. Although Google Scholar has its defenders as a bibliometric tool (e.g., Harzing and van den Wul 2008), a case based on it alone may be attacked vigorously.

- Not defining time windows. Citation statistics are a moving target – an author’s cumulative citation profile at the end of one month will not match that at the end of the next month. It is important to define the time window used in any analysis or comparison.

Frank, despite his manner, does manage to convince some of his interview panel of the merit in alternative approaches.
Sue: Well, that’s all my questions. Pete, Joe – any more from you?

Joe: (Shakes his head)

Peter: No, thanks for your time, Frank.

Sue: Frank, do you have any questions for us?

Frank: Not really. Where do we go from here?

Sue: After you leave, we’ll have a chat and prepare a recommendation for the full promotions committee. You’ll be advised of the result in writing.

Frank: That’s clear - thanks.

Sue: I’ll show you out.

Everyone rises except Joe. Frank shakes hands with Peter and Sue, looks uncertainly at Joe, and exits stage left with Sue. Sue returns. She and Peter resume their seats.

Joe: Impudent young pup! Speaking to us like that.

Peter: He’s a rough diamond, that's for sure. But I think there’s some substance there – he has done a lot at DENR and he just taught me a few things about research evaluation. I’m wondering if we need to review our procedures.

Joe: Poppycock. I can’t imagine how he deals with the plantation managers if he can’t respect a senior DENR sub-committee.

Sue: Joe, I can see why you’re annoyed – he rubs me the wrong way too. But we’re here to judge a record, not a personality.

Joe: I am discussing a record! We need to be concerned about reputation. That means publishing in high impact journals. He doesn’t! Until I see a string of papers from him in the top journals, I can’t recommend promotion.

Sue: Pete?

Peter: (looking at his papers) Well, I can see a publication record that is above the other research scientists in DENR and overlaps with two of the three seniors. His citations aren't outstanding, but he has me wondering now about how relevant they are. The policy people have an interest in his work and folk outside DENR are downloading his papers. Ultimately, local policy is what we’re about. We want to keep and encourage him. Promote him.

Joe: We’re hung, then. I’m not budging.

Sue: OK, what about a compromise. In our report we play up the contribution to policy rather than journal impact factors and those other measures. And we make the promotion conditional on giving some seminars to management too on those – what did he call them – RAMs? Pump up his service to DENR profile.

Peter: Sounds good to me. And we don’t need to give up on citations. Remember that bit in his application about profiling his international citations? He’s right - he does attract international citations. That’s evidence we can use.

Joe: (reluctantly) But what about some top journal papers?

Sue: We’re shifting the emphasis to policy and service. We have discretion to do that. And we’re weighting those international citations as more important than the journals he publishes in.

Peter: I agree.

Joe: Alright, but our reputation will suffer.

Sue: That remains to be seen. Thank you, gentlemen. Done.

Fighting back in committees and by example

Beyond the measures suggested for strengthening the presentation of one’s own record that Frank used successfully in his promotion application, there is a larger issue of how the experienced researchers who find themselves evaluating research for granting bodies, tenure committees and so on respond to inappropriate obsessions with simple numbers to assess research and set examples for younger researchers. Adler et al. (2008) cite three examples from many they claim to have received:

Example 1: My university has recently introduced a new classification of journals using the Science Citation Index Core journals. The journals are divided into three groups based only on the impact factor. There are 30 journals in the top list, containing no mathematics journal. The second list contains 667, which includes 21 mathematics journals. Publication in the first list causes university support of research to triple; publication in the second list, to double. Publication in the core list awards 15 points; publication in any Thomson Scientific covered journal awards 10. Promotion requires a fixed minimum number of points.

Example 2: In my country, university faculty with permanent positions are evaluated every six years. Sequential successful evaluations are the key to all academic success. In addition to a curriculum vitae, the largest factor in evaluation concerns ranking five published papers. In recent years, these are given 3 points if they appear in journals in the top third of the Thomson Scientific list, 2 points if in the second third, and 1 point in the bottom third. (The three lists are created using the impact factor.)

Example 3: In our department, each faculty member is evaluated by a formula involving the number of single-author-equivalent papers, multiplied by the impact factor of the journals in which they appear. Promotions and hiring are based partly on this formula.

Common to all of them is an obsession with journal ranking as a basis for simple numerical assessments, supporting Lawrence’s (2003) claim that one important reform would be to ‘break up the cult of the journal’. By this he means that established authors should publish in a wide range of appropriate journals, not just the most highly regarded ones. This ensures that a range of relevant work is presented, with major breakthroughs being published in top journals and sound work of regional relevance or lesser scope in appropriate
disciplinary or regional journals. An example is also set for younger researchers that publication is valuable at all levels. This position is reiterated by Colledge et al. (2010), who point out that prolific authors often publish in journals across a wide range of impact factors, directing papers to the journals addressing their desired target audience.

Lawrence (2003, 2007) also asks granting bodies and departmental administrators to use hindsight. Unlike journal editors, who must judge a manuscript on the basis of their own experience and input from a small number of colleagues, these evaluating track records can consider a wide range of information about the significance and impact of a researcher’s work that go beyond the standard bibliometric measures of citations and journal rankings (Roa et al. 2009; Lane 2010; Lane and Bertuzzi 2011). Possibilities include publication in languages other than English to reach local practitioners (Adler and Harzing 2009), or demonstrating that the results of research have been adopted by professional societies or local government departments (Witten and Hammond 2010). These approaches address the pertinent criticism that research can be influential, yet poorly cited or even uncited (MacRoberts and MacRoberts 2010). Many of these ideas coalesce in ‘The San Francisco Declaration on Research Assessment’ (DORA) (http://am.ascb.org/dora/). This document, signed by many prominent editors and publishers, lays down recommendations for research evaluation. The overarching general recommendation is: ‘Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist’s contributions, or in hiring, promotion, or funding decisions.’ A specific recommendation for institutions is: ‘For the purposes of research assessment, consider the value and impact of all research outputs (including datasets and software) in addition to research publications, and consider a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice.’ The DORA site documents the growing number of signatories to the declaration, as well as supporting editorials from leading editors. (Then run on immediately to ‘Some of Frank’s interviewing panel …’) Some of Frank’s interviewing panel were reaching in that direction. It’s an effort that should be emulated.

Acknowledgements

I thank Dan Lunney and Harry Recher for many detailed discussions that were the genesis of this paper, and two reviewers for comments on the first submission. However, these people do not necessarily agree with all the opinions expressed. Mark Preston, a graphics designer at Murdoch University, adjusted the screen captures in Appendix 1 to a suitable resolution for printing. Kayley Morgan, of Surrey Beatty & Sons, gave permission to reproduce the last figure of Appendix 1 (copyright Surrey Beatty & Sons).

References


Calver, M. C. and Bradley, J. S., 2009. Should we use the mean citations per paper to summarise a journal’s impact or to rank journals in the same field? Scientometrics 81: 611-615.


Calver, M.C. and Bradley, J.S., 2009. Should we use the mean citations per paper to summarise a journal’s impact or to rank journals in the same field? Scientometrics 81: 611-615.


Lawrence, P. A., 2009. Real lives and white lies in the funding of scientific research: The granting system turns young scientists into bureaucrats and then betrays them. PLoS Biology 7:


Profiling an author's international citations in the Web of Science database.

1. In Web of Science, select the 'Cited Reference Search' tab. Enter the author details, then click 'Search'.

2. In the output, ensure that the checkbox is selected against all papers to be included in the search - only include publications from the target author.
3. The output shows all the papers that cite works by the chosen author. Click ‘Analyze Results’ at the upper right.

4. In the results analysis pane, select ‘Countries/Territories’ in the left hand menu. Choose how many records to display (set this to be larger than the number of records displayed at the top left) and the minimum count needed (i.e. the minimum number of references) before a record is included (here set to 2, 1 is probably better). Click ‘Analyze’.

5. The output, which can be exported, shows the percentage of citations by authors from different countries. These percentages do not sum to 100%, because a single citation may involve authors from multiple countries. In the example below, 56% of all citations included an author from Australia, 24.7% an author from the USA and so on. (This figure is reproduced from Calver and Bryant (2008), with permission from the publisher, Surrey Beatty & Sons).
## Appendix I

<table>
<thead>
<tr>
<th>Region</th>
<th>No. of Records</th>
<th>Percentage of Total Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>886</td>
<td>56.0</td>
</tr>
<tr>
<td>USA</td>
<td>391</td>
<td>24.7</td>
</tr>
<tr>
<td>Europe</td>
<td>275</td>
<td>17.4</td>
</tr>
<tr>
<td>New Zealand</td>
<td>221</td>
<td>14.0</td>
</tr>
<tr>
<td>Asia and Pacific</td>
<td>74</td>
<td>4.7</td>
</tr>
<tr>
<td>Canada</td>
<td>58</td>
<td>3.7</td>
</tr>
<tr>
<td>South America</td>
<td>37</td>
<td>2.3</td>
</tr>
<tr>
<td>Africa</td>
<td>33</td>
<td>2.1</td>
</tr>
<tr>
<td>Middle East and India</td>
<td>28</td>
<td>1.8</td>
</tr>
<tr>
<td>Central America and Caribbean</td>
<td>22</td>
<td>1.4</td>
</tr>
</tbody>
</table>