Some thoughts on Journal Impact Factor

A colleague just e-mailed me to ask me how I felt about journal impact factor being such a big part of the Academic Ranking of World Universities - they say that 20% of the ranking weight comes from # of papers published in Nature and Science. So what do I think?

On evaluations

I'm really not a big fan of rankings and evaluations in the first place. This is largely because I feel that evaluations are rarely objective. For one very specific example, last year at MSU I got formal evaluations from both of my departments. Starting with the same underlying data (papers published, students graduated, grants submitted/awarded, money spent, classes taught, body mass/height ratio, gender weighting, eye color, miles flown, talks given, liquid volume of students’ tears generated, committees served on, etc.) one department gave me a "satisfactory/satisfactory/satisfactory" while the other department gave me an "excellent/excellent/excellent." What fun! (I don't think the difference was the caliber in departments.)

Did I mention that these rankings helped determine my raise for the year?

Anyhoo, I find the rating and ranking scheme within departments at MSU to be largely silly. It's done in an ad hoc manner by untrained people, and as far as I can tell, is biased against people who are not willing to sing their own praises. (I brought up the Dunning-Kruger effect in my last evaluation meeting. Heh.) That's not to say there's not serious intent -- they are factored into raises, and at least one other purpose of evaluating assistant professors is so that once you fire their ass (aka "don't give them tenure") there's a paper trail of critical evaluations where you explained that they were in trouble.

Metrics are part of the job, though; departments evaluate their faculty so they can see who, if anyone, needs help or support or mentoring, and to do this, they rely at least in part on metrics. Basically, if someone has lots of funding and lots of papers, they're probably not failing miserably at being a research professor; if they're grant-poor and paper-poor, they're targets for further investigation. There are lots of ways to evaluate, but metrics seem like an inextricable part of it.

Back to the impact factor

Like faculty evaluations, ranking by the impact factor of the journals that university faculty publish in is an attempt to predict future performance using current data.

But impact factor is extremely problematic for many reasons. It's based on citations, which (over the long run) may be an OK measure of impact, but are subject to many confounding factors, including field-specific citation patterns. It's an attempt to predict the success of individual papers on a whole-journal basis, which falls apart in the face of variable editorial decisions. High-impact journals are also often read more widely by people than low-impact journals, which yields a troubling circularity in terms of citation numbers (you're more likely to cite a paper you've read!) Worse, the whole system is prone to being gamed in various ways, which is leading to high rates of retractions for high-impact journals, as well as outright fraud.

Impact factor is probably a piss-poor proxy for paper impact, in other words.

If impact factor was just a thing that didn't matter, I wouldn't worry. The real trouble is that impact factors have real-world effect - many countries use impact factor of publications as a very strong weight in funding and promotion decisions. Interestingly, the US is not terribly heavy handed here - most universities seem pretty enlightened about considering the whole portfolio of a scientist, at least anecdotally. But I can name a dozen countries that care deeply about impact factor for promotions and raises.
And apparently impact factor affects university rankings, too!

Taking a step back, it's not clear that any good ranking scheme can exist, and if it does, we're certainly not using it. All of this is a big problem if you care about fostering good science.

The conundrum is that many people like rankings, and it seems futile to argue against measuring and ranking people and institutions. However, any formalized ranking system can be gamed and perverted, which ends up sometimes rewarding the wrong kind of people, and shutting out some of the right kind of people. (The Reed College position on the US News & World Report ranking system is worth reading here.) More generally, in any ecosystem, the competitive landscape is evolving, and a sensible measure today may become a lousy measure tomorrow as the players evolve their strategies; the stricter the rules of evaluation, and the more entrenched the evaluation system, the less likely it is to adapt, and the more misranking will result. So ranking systems need to evolve continuously.

At its heart, this is a scientific management challenge. Rankings and metrics do pretty explicitly set the landscape of incentives and competition. If our goal in science is to increase knowledge for the betterment of mankind, then the challenge for scientific management is to figure out how to incentive behaviors that trend in that direction in the long term. If you use bad or outdated metrics, then you incentivize the wrong kind of behavior, and you waste precious time, energy, and resources. Complicating this is the management structure of academic science, which is driven by many things that include rankings and reputation - concepts that range from precise to fuzzy.

My position on all of this is always changing, but it's pretty clear that the journal system is kinda dumb and rewards the wrong behavior. (For the record, I'm actually a big fan of publications, and I think citations are probably not a terribly bad measure of impact when measured on papers and individuals, although I'm always happy to engage in discussions on why I'm wrong.) But the impact factor is especially horrible. The disproportionate effect that high-IF glamour mags like Cell, Nature and Science have on our scientific culture is clearly a bad thing - for example, I'm hearing more and more stories about editors at these journals warping scientific stories directly or indirectly to be more press-worthy - and when combined with the reproducibility crisis I'm really worried about the short-term future of science. Journal Impact Factor and other simple metrics are fundamentally problematic and are contributing to the problem, along with the current peer review culture and a whole host of other things. (Mike Eisen has written about this a lot; see e.g. this post.)

In the long term I think a much more experimental culture of peer review and alternative metrics will emerge. But what do we do for now?

More importantly:

**How can we change?**

I think my main advice to faculty is "lead, follow, or get out of the way."

Unless you're a recognized big shot, or willing to take somewhat insane gambles with your career, "leading" may not be productive - following or getting out of the way might be best. But there are a lot of things you can do here that don't put you at much risk, including:

- be open to a broader career picture when hiring and evaluating junior faculty;
- argue on behalf of alternative metrics in meetings on promotion and tenure;
- use sites like Google Scholar to pick out some recent papers to read in depth when hiring faculty and evaluating grants;
- avoid making (or push back at) cheap shots at people who don't have a lot of high-impact-factor papers;
- invest in career mentoring that is more nuanced than "try for lots of C-N-S papers or else" - you'd be surprised how often this is the main advice assistant professors take away...
- believe in and help junior faculty that seem to have a plan, even if you don't agree with the plan (or at least leave them alone ;)

What if you are a recognized big shot? Well, there are lots of things you can do. You are the people who set the tone in the community and in your department, and it behooves you to think scientifically about the culture and reward system of science. The most important thing you can do is think and investigate. What evidence is there behind the value of peer review? Are you happy with C-N-S editorial policies, and have you talked to colleagues who get rejected at the editorial review stage more than you do? Have you thought about per-article metrics? Do you have any better thoughts on how to improve the system than 'fund more people', and how would you effect changes in this direction by recognizing alternate metrics during tenure and grant review?
The bottom line is that the current evaluation systems are the creation of scientists, for scientists. It's our responsibility to critically evaluate them, and perhaps evolve them when they're inadequate; we shouldn't just complain about how the current system is broken and wait for someone else to fix it.

**Addendum: what would I like to see?**

Precisely predicting the future importance of papers is obviously kind of silly - see this great 1994 paper by Gans and Shepherd on rejected classics papers, for example -- and is subject to all sorts of confounding effects. But this is nonetheless what journals are accustomed to doing: editors at most journals, especially the high impact factor ones, select papers based on projected impact before sending them out for review, and/or ask the reviewers to review impact as well.

So I think we should do away with impact review and review for correctness instead. This is why I'm such a big fan of PLOS One and PeerJ, who purport to do exactly that.

But then, I get asked, what do we do about selecting out papers to read? Some (many?) scientists claim that they need the filtering effect of these selective journals to figure out what they should be reading.

There are a few responses to this.

First, it's fundamentally problematic to outsource your attention to editors at journals, for reasons mentioned above. There's some evidence that you're being drawn into a manipulated and high-retraction environment by doing that, and that should worry you.

But let's say you feel you need something to tell you what to read.

Well, second, this is technologically solvable - that's what search engines already do. There's a whole industry of search engines that give great results based on integrating free text search, automatic content classification, and citation patterns. Google Scholar does a great job here, for example.

Third, social media (aka "people you know") provides some great recommendation systems! People who haven't paid much attention to Twitter or blogging may not have noticed, but in addition to person-to-person recommendations, there are increasingly good recommendation systems coming on line. I personally get most of my paper recs from online outlets (mostly people I follow, but I've found some really smart people to follow on Twitter!). It's a great solution!

Fourth, if one of the problems is that many journals review for correctness AND impact together, why not separate them? For example, couldn't journals like Science or Nature evolve into literature overlays that highlight papers published in impact-blind journals like PLOS One or PeerJ? I can imagine a number of ways that this could work, but if we're so invested in having editors pick papers for us, why not have them pick papers that have been reviewed for scientific correctness first, and then elevate them to our attention with their magic editorial pen?

I don't see too many drawbacks to this vs the current approach, and many improvements. (Frankly this is where I see most of scientific literature going, once preprint archives become omnipresent.)

So that's where I want and expect to see things going. I don't see ranking based on predicted impact going away, but I'd like to see it more reflective of actual impact (and be measured in more diverse ways).

--titus

p.s. People looking for citations of high retraction rate, problematic peer review, and the rest could look at one of my earlier blog posts on problems with peer review. I'd be interested in more citations, though!

**Comments!**
There is incredible inertia within the publishing house system which is dependent entirely upon researchers to submit, review and pay (both to publish and to read). We hold all of the cards but continue with the status quo. So unless researchers (or their funders) are willing to force change themselves, we will be stuck with the current mechanisms. There are nascent efforts such as pre-print servers, PubPeer etc. but these need more mainstream support. Moreover, we should recognize that we do see value in the top tier journals or they really wouldn't exist (and open access isn't the panacea - eLife is rapidly becoming a glamour journal). Journal hierarchy may well distort the numbers but we seem to like and accept these flawed metrics associated with impact - perhaps because citations lag so far behind publication and we don't have time to read everything. If we were able to better identify/grade significant of individual pieces of work based on their individual merit, problem solved. In an age of virtually free information distribution, we are still driving around in horse-pulled carriages.

Personally, I don't see significant, disruptive change unless the funding flows are altered or a new mechanism is invented and becomes accepted. There is plenty of opportunity given the corrupting influences associated with current methodologies and their economic drivers.
Some thoughts on Journal Impact Factor