Opening scientific dialog: preprints and published peer review

Jessica Polka

@jessicapolka | @ASAPbio_

ASAPbio

Postdocs' #1 concern about preprints: I'm going to get scooped

ie: preprints are public but not obviously well-respected

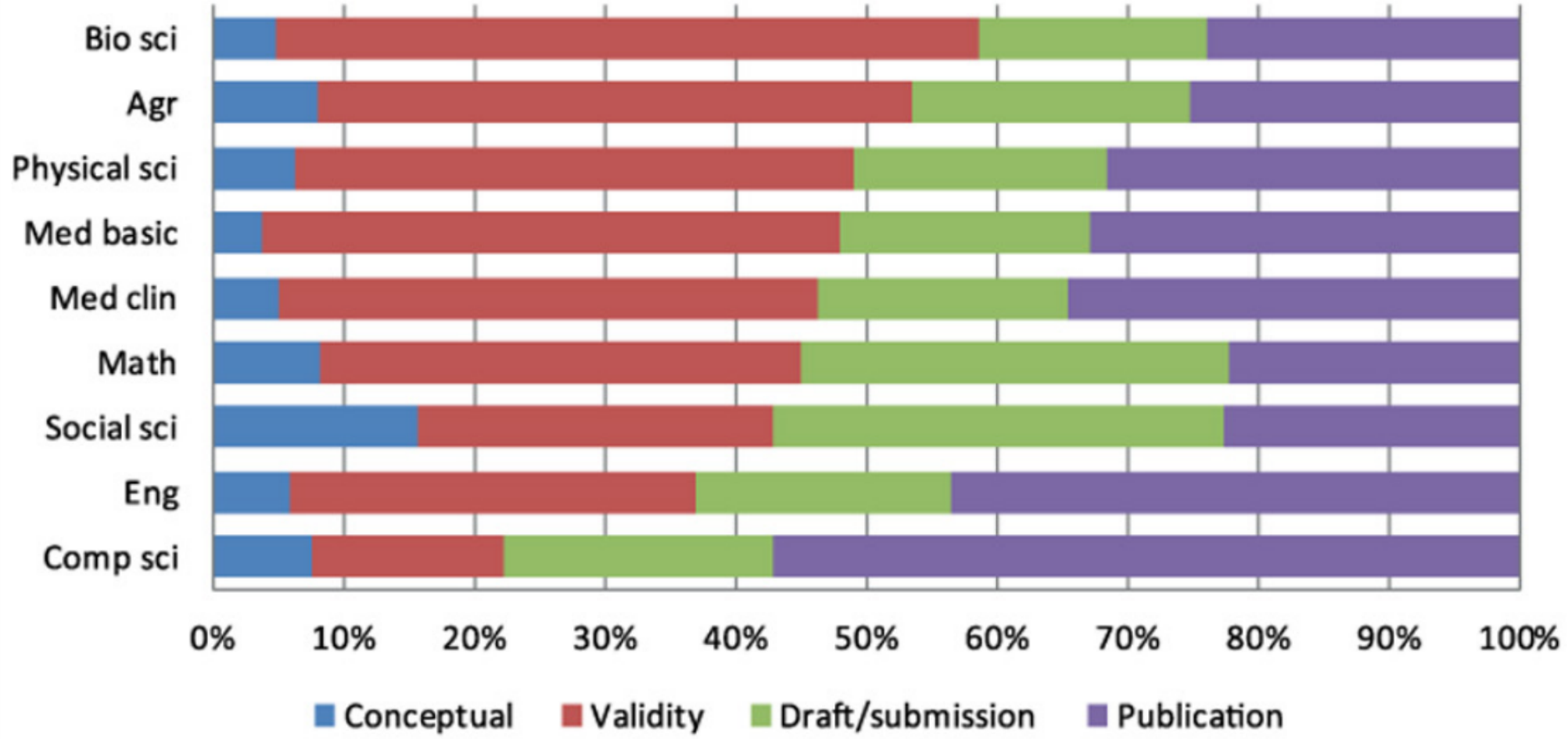
39 responses (EMBO Postdoc Fellows meeting, 2016)

Paul Ginsparg, founder of arXiv on scooping:

"It can't happen, since arXiv postings are accepted as date-stamped priority claims."

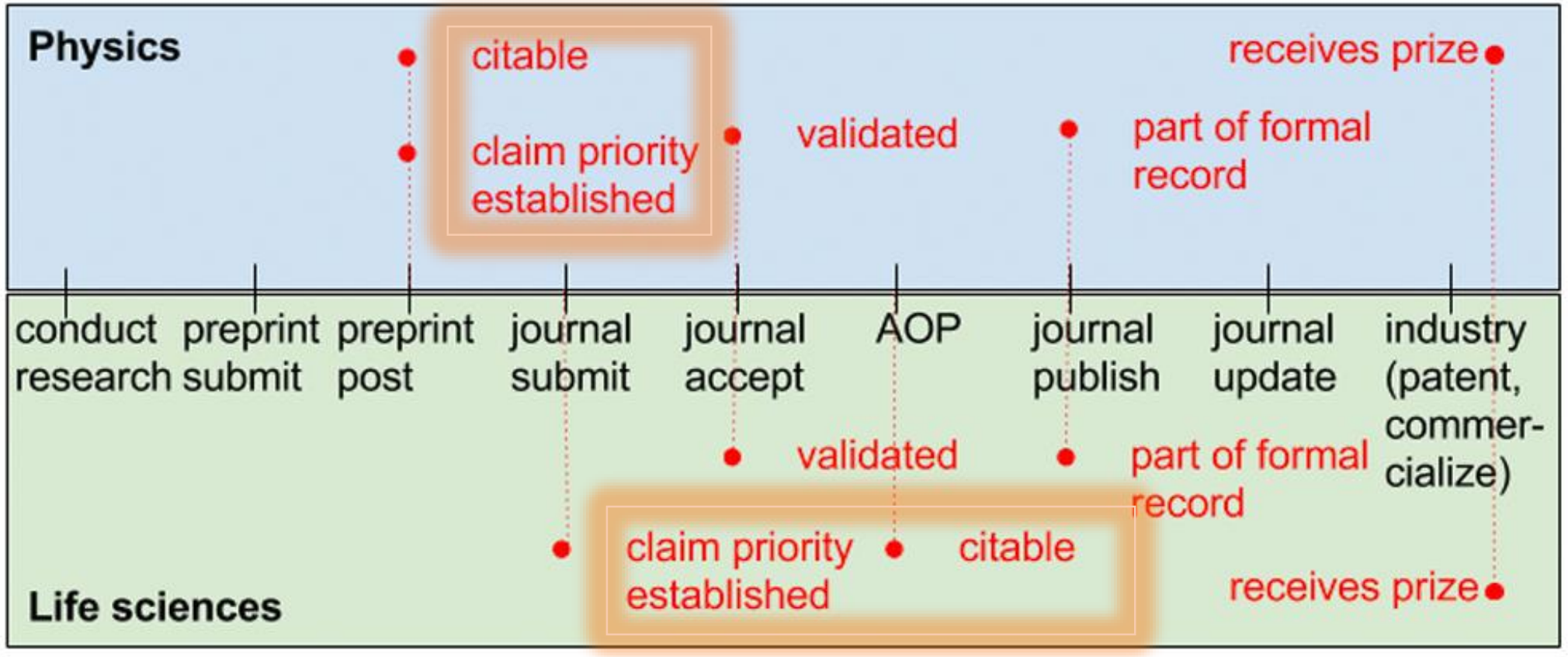
asapbio.org/preprint-info/preprint-faq

Ironically, biologists share early



Thursby, Haeussler, Thursby, Jiang: http://advances.sciencemag.org/content/4/5/eaar2133.full

But much of this sharing is informal



Neylon, Pattinson, Bilder, Lin: https://f1000research.com/articles/6-608/v1

Making preprints “count”

- ❑ Funders asapbio.org/funder-policies
- ❑ Jobs
- ❑ Subsequent journal publication
- ❑ Are appropriately cited

What does “appropriately” mean?

One argument: preprints should not be cited

But –

- Citations mean different things
- Policies that disallow citations invite plagiarism
- Underlying assumption: peer reviewed = true

Responses to these arguments by Tennant: http://fossilsandshit.com/should-we-cite-preprints/

Second class citation

Nucleic Acids Research

"We do not allow formal citation of preprints in the reference list, but they can be cited in the main text, for example: (BioRxiv: https://doi.org/10.1101/xxxxxx)."

https://academic.oup.com/nar/pages/Ms_Prep_Submission

Another argument: preprints must be labeled

NIH-recommended citation format:

- *Example*: Bar DZ, Atkatsh K, Tavarez U, Erdos MR, Gruenbaum Y, Collins FS. Biotinylation by antibody recognition- A novel method for proximity labeling. BioRxiv 069187 [**Preprint**]. August 11, 2016 [cited 2017 Jan 12]. Available from: https://**doi**.org/10.1101/069187.

https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-050.html

Let's make peer review visible as well

(Anonymous) peer review is worth publishing

- Encourage reviewers to be civil & constructive
- Expose “predatory” journals
  - Support/select journals based on how constructive or rigorous the peer review is
- Help readers understand…
  - Debates in the field
  - How much (and what parts of) the paper has been scrutinized
  - What good (and bad) peer review looks like
- Enable the systematic study of peer review

Concerns

- Weaponization – discredit science
- Amplify bias
- Deter reviewers
- Change reviewer quality

Does it work?

Effect on peer review of telling reviewers that their signed reviews might be posted on the web: randomised controlled trial

BMJ 2010; 341 doi: https://doi.org/10.1136/bmj.c5729 (Published 16 November 2010)

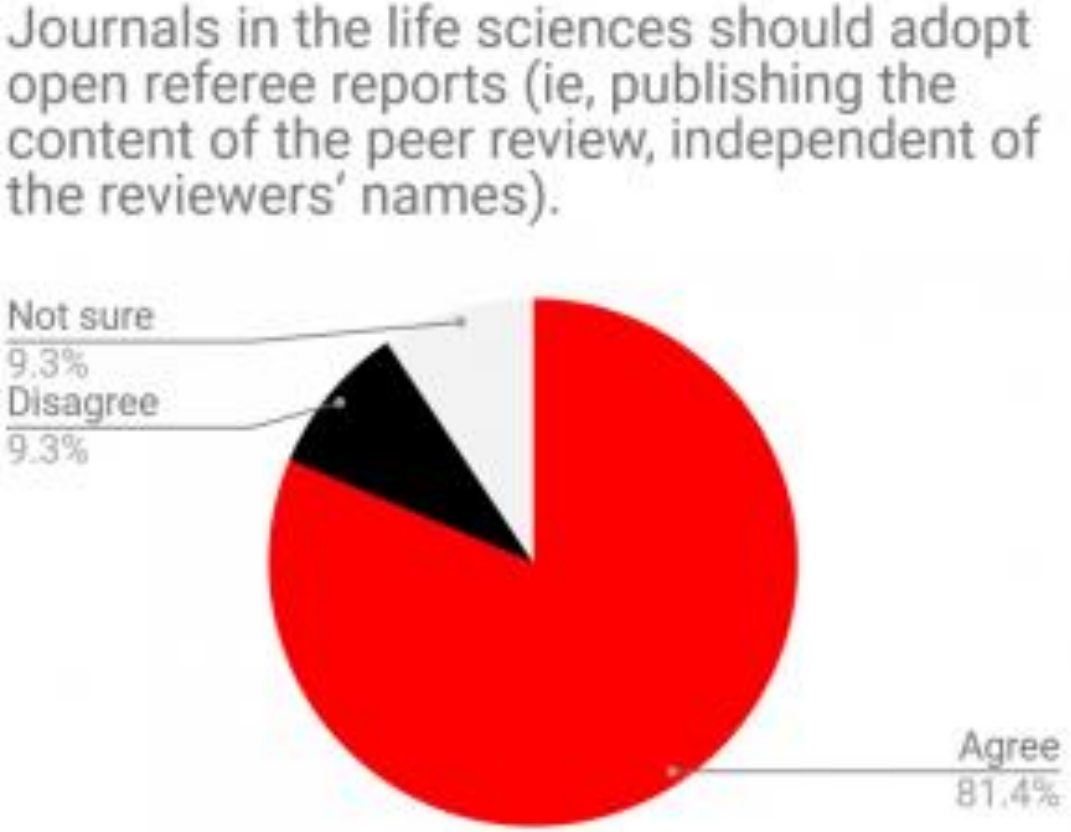
471 papers

Conclusion Telling peer reviewers that their signed reviews might be available in the public domain on the *BMJ*'s website had no important effect on review quality. …high refusal rate among potential peer reviewers and an increase in the amount of time taken to write a review

Intervention	Control
Time in minutes (SD)	*Time in minutes (SD)*
182 (135.2)	157 (101.9)

#bioPeerReview – HHMI/Wellcome/ASAPbio

Feb 7-9, 2018



http://asapbio.org/peer-review/summary

Another benefit of open peer review reports?

Normalizing public, constructive scientific dialog

~10% of bioRxiv papers have comments

asapbio.org/biorxiv

Preprint feedback benefits authors

Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP

Daniel S. Quintana[1*] and Donald R. Williams[2]

https://osf.io/sgpe9/

https://www.facebook.com/groups/853552931365745/permalink/1349684805085886/

Daniel Quintana shared a link.
April 10

Just posted a preprint on Bayesian alternatives for common null-hypothesis significance tests that may be of interest to the group. Our goal was to put together a non-technical walkthrough using JASP for those unfamiliar with Bayesian alternatives. Would appreciate any feedback

osf.io
OSF.IO

Like Share

13

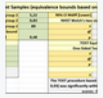
Uli Schimmack I thought this would be a tutorial about picking alternative hypothesis to carry out a Bayesian statistical analysis because this is an important additional and new step that researchers are not familiar with. Unlike NHST where you only need to specify H0, default effect size = 0, Bayesian hypothesis testing requires also to specify H1 because BF provide information about the relative support for H0 and H1 given the data.

Alah, this is just another "tutorial" with all the wrong claims about p-values, a focus on hypothesis testing, when we really want to know how effective drugs are (effect sizes) and a total neglect of Bayesian and frequentist ways to assess the probabilty that a drug is not effective.
Daniel Lakens

http://daniellakens.blogspot.ca/.../tost-equivalence...

Excuse me, if this is a bit harsh, but we have been discussing these issues for over a year now and I think it is fair to request a balanced and informative review of options to draw inferences from data.

Stop bashing p-values and provide some guidelines for researchers how they can pick a sensible alternative and how they BF have to be interpreted in the light of prior odds of H0 and H1.



TOST equivalence testing R package (TOSTER) and spreadsheet
I'm happy to announce my first R package 'TOSTER'.
DANIELLAKENS.BLOGSPOT.COM

Like · 1 · April 10 at 6:42pm

Daniel Quintana Thanks for the feedback, glad to hear this while it's still a preprint. We actually cited Daniel Lakens' excellent TOSTER paper but I guess we can make this clearer.
Like · 1 · April 10 at 11:46pm

Daniel Lakens Hi Daniel Quintana, I read the first few pages, and I have good news and bad news. The good news is if the reviewers are all Bayesians, it will be accepted. The bad news is there are quite important misunderstandings of p-values and Bayes factors in the paper.

The hypothesis you describe in the intro (is the null true, or is there an effect larger than 0) can only be tested with p-values. It is underspecified for Bayesian stats. In Bayes, the alternative is 'is there a true effect between x and y with the distribution like z'. So the intro is an argument against Bayes factors. They don't allow you to test the hypothesis you seem interested in.

Then I stopped reading where you said Bayes factors could quantify the size of an effect. It is not true. You need to provide an effect size estimate with a Bayes factor. You can't only report a Bayes factor - it tells you nothing about the size of an effect. This is such a basic misunderstanding, I stopped reading, but you might want to reconsider getting an expert on board?

Finally, you misunderstand p-values. You are re-hashing arguments by p-value bashers. But not by experts on p-values (e.g., Benjamini, Nickerson, Frick). P-values are ONLY used for error control. Not mentioning that in the intro is the last reason this paper should not be read by novices.

Now it will be read, like crazy, because everyone thinks they need to report Bayes Factors. As I have blogged, equivalence tests outperform Bayes factors for testing the absence of any effect you care about. But to quote your excellent podcast: there are acadamic hipsters. They want to twist their mustaches, drink machiato's, and report Bayes factors.

There are thousands of 'intro to Bayes' factors resources. And there are 2 Intermediate Bayes factorsresources. Everybody wants to know what it is, but no one really goes on to use it. Think about that.

Daniel Lakens Here is the critical misunderstanding error (you'll need to remove the criticisms on cohen's d from the paper, or admit you need effect sizes in addition to bayes factors) - also, the Bayes factor can not provide evidence for the presence of an effect.... See More
Like · April 11 at 1:54am · Edited

Daniel Quintana This is very good feedback, great to have extra pairs of eyes looking over this before submission. Looking forward to discussing this topic on our podcast!
Like · 1 · April 11 at 2:05am

Kyle Morrissey There are thousands of intro to Bayes factors resources? That was not my experience :S

Though I finally did have someone run me through the conceptual basics in person the other day, and it made sense.
Like · April 11 at 8:28am

Daniel Lakens Kyle , -1 for not saying that the intro in my MOOC was all you needed. You can lead a horse to the water, but you can't make them drink.
Like · April 11 at 8:58am

Stephen Martin P-values really aren't used for error control. That's conflating NP and Fisherian approaches, no?

Piggy backing off this comment thread.... See More
Like · April 12 at 12:33am · Edited

Stephen Martin After reading Donald Williams' response, I thought I should just clarify: I'm all for papers giving 'new' (or at least, newly applied) perspectives on old topics, along with critiques of old perspectives on old topics. I intended my reply to be a critique moreso of BFs and some of the specific arguments, not as a critique of you or your intentions. I realized I never actually made that explicit in my reply above.
Like · 2 · April 12 at 12:36am

Matt Williams >The hypothesis you describe in the intro (is the null true, or is there an effect larger than 0) can only be tested with p-values. It is underspecified for Bayesian stats. In Bayes, the alternative is 'is there a true effect between x and y with the distribution like z'. [Daniel]

>More importantly though, the p(Model | D) can only be interpreted in the family of models that you're testing, but I think people interpret it as "probability I'm correct", [Stephen]

I agree given the standard interpretation of Bayes factors (where the prior on effect size is treated as part of the H1 model itself). But if you separate out the H1 "hypothesis" from the statistical model/prior the problem becomes sort-of resolvable. This is what I was banging on about in my recent blog: ://thepathologicalscience.blogspot.com/.../separating...

PS. Like Stephen Martin I'm also a Bayesian who doesn't really like Bayes factors, but I'm working on a manuscript at the moment where I've been asked to write an introduction to them for a special issue on methods in a particular sub-area of psych. It's been bloody difficult trying to produce a "balanced" view of Bayes factors (i.e., balancing reasonable views of frequentists, pro-BF people, and Bayesians who prefer estimation). Thanks Daniel Quintana for provoking a discussion that has been helpful to me in making final revisions.

Separating model from hypothesis in the Bayes factor test
Premise When using statistical analyses, we will often test a statistical model that has one or more parts that we regard as forming an hy...
THEPATHOLOGICALSCIENCE.BLOGSPOT.COM

Like · April 12 at 4:09pm

Daniel Quintana That blog post is really handy, thanks for sharing! We're working on an update now based on everyone's great feedback
Like · April 13 at 4:48am

Donald Williams Hi Daniel Quintana. To all providing comments, I think it is important to remember the likely readership of this article. I imagine this paper is targeted to those in more clinical fields who have not been exposed to much Bayesian stuff. That said, I am not sure I see this as an introduction to Bayes factors, and especially not Bayesian statistics. Instead, I think this is more of an introduction into the doctrine of Rouder, Wagenmakers..etc (i.e., the BF crew) in psychology. Now that there approach has become more common, this has also resulted in finding several limitations in their approach and downright rebukes of their use of statistics (e.g., our paper: Uli Schimmack and Rickard Carlsson). That said, I think the BF crew does a lot great research, but has also oversold BF and feel as though they have sought extreme examples to show how BF and p differ, but always in favor of their method being superior. That said, rather than introduce this approach circa a few years ago, I see this as a unique opportunity to introduce what might be a "new" method to a field, but also include the recent critiques and other ways of using Bayesian statistics. In this way, we have a fair and balanced paper, and not one slanted towards the BF crew's philosophy that has dominanted psychology. Not that Dominant means the approach is necessarily good (or bad), just that they were shouting the loudest and often publishing things that were not novel other than computing a Bayes factor. This resulted in a flurry of opportunistic Bayes factor publications. Those days are hopefully winding down, although now the challenge is that more people are using JASP without really understanding what is going on. I cannot blame them, as the ease with which BF can be manipulated is not really described in any amount of detail--e.g., the infamous prior odds on Bem's ESP. As for the paper, I would steer away from critiquing p-values and instead think of ways we can think about using them. For example, p can be considered as a kind of model fit indices, not for the observed data, but to the null sampling distribution. That is, if we set up a null model (or envision a hypothetical null model), p gives us a measure of departure from that model. The question then becomes contexts in which this is useful, or what needs to accompany p to ensure it is valid and allows for rich inferences--there are lots and lots of assumptions that may or may not make sense depending on the situation, but no less sensible than any statistical quantities assumptions. While much attention has been paid to the Bayesian prior, what is less considered is the chosen likelihood, which is a modeling based decision both frequentist's and Bayesian's make, but Bayesian more explicitly so. That said, Bayesian's do not often examine the influence of distributional departures from the chosen likelihood on the resulting posterior (to my knowledge). These are important issues, as they directly affect the density with which Bayes factors are computed. How does non-normality, unequal variances, treating a count variable as continuous influence the resulting Bayes factor, for example? This says nothing about the importance of fully understanding that BF is a model comparison metric. It provides relative evidence. This generally comes with even odds on the null and alternative. This does not makes much sense, but I have also made this assumption in some of my work. I am not sure this is more unreasonable than testing the value of zero in a frequentist framework, so proceeded but with effect size estimates and intervals on those effects (quantities not provided by Bayes factors). These are important issues, and I see that you have a unique opportunity to introduce the current state of Bayesian methods to your field (prior odds, the importance of the prior, and inferences obtained from the posterior...etc.). This also comes with great responsibility, and I think it would be a shame to align yourself so heavily with the BF crew in their use of not only Bayesian statistics, but also their arguments against p-values.
Like · 5 · April 12 at 12:57am · Edited

Donald Williams Let me also say that I too made many of the similar arguments against p-values in the past. Since then, I learned that p is not evil, and that Bayes factors are not great. They simply are what they are, and the problem really arises from misuse or misunderstandings.
Like · 4 · April 12 at 12:36am

Daniel Quintana Thanks for these comments. In earlier versions of the manuscript we went into a lot more depth (including the importance of the chosen likelihood) but were squeezed for space. The tricky thing here is to make this paper approachable to those who are more clinically oriented, while also appropriately covering all the important issues (and keeping within word limits).
Like · 1 · April 12 at 2:26am

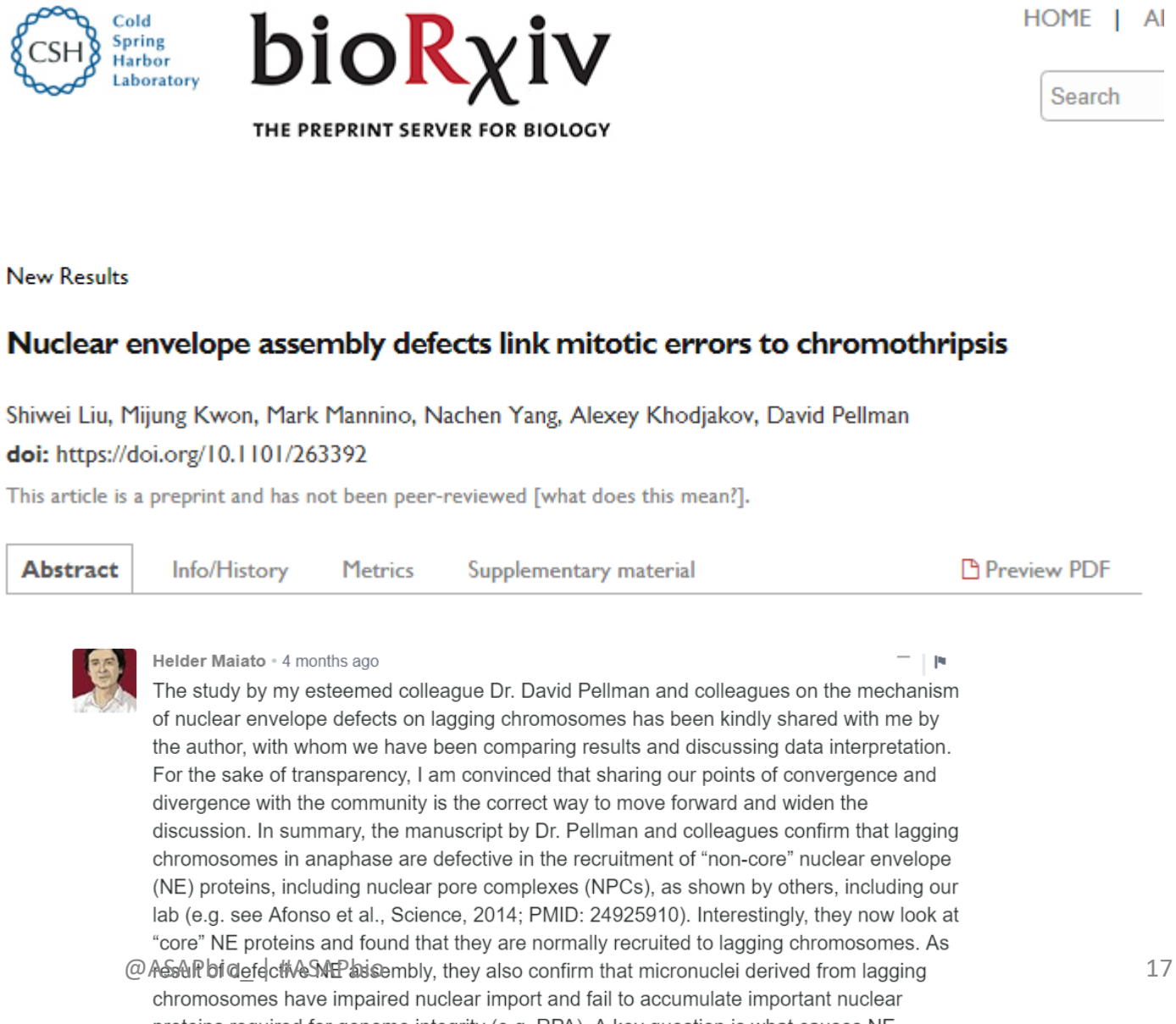
Donald Williams [illegible] to mention is whether in clinical oriented work we even care about model selection via bayesian null hypothesis testing? For example, for making treatment decisions, what is more informative: d = 0.30, 95-% CI

Dan Quintana @dsquint… 15h
Replying to @dsquintana @jessi…
....I reached out to one of the people who wrote some of the critical feedback and asked if he wanted to join as a co-author.

Dan Quintana @dsquint… 15h
Replying to @dsquintana @jessi…
He agreed 🎉 So with his input and re-writes, along with input from others, the paper was updated to its current version.

Dan Quintana @dsquint… 15h
Replying to @dsquintana @jessi…
Now the paper is under review at a top journal. I also mentioned in the cover letter that the preprint had been downloaded 700+ times

Preprint feedback benefits readers



[Link to paper]

Preprint feedback benefits students

- Meaningful exercise: send feedback to authors to improve their paper
- Teach students how to write a review

PREreview

Prereview.org

Prachee Avasthi at the University of Kansas Medical Center draws material for her "Analysis of Scientific Papers" course exclusively from preprint servers. She's generously shared her syllabus and introductory slide deck, and the students' reviews can be found on the Winnower.

See more examples: http://asapbio.org/preprint-journal-clubs

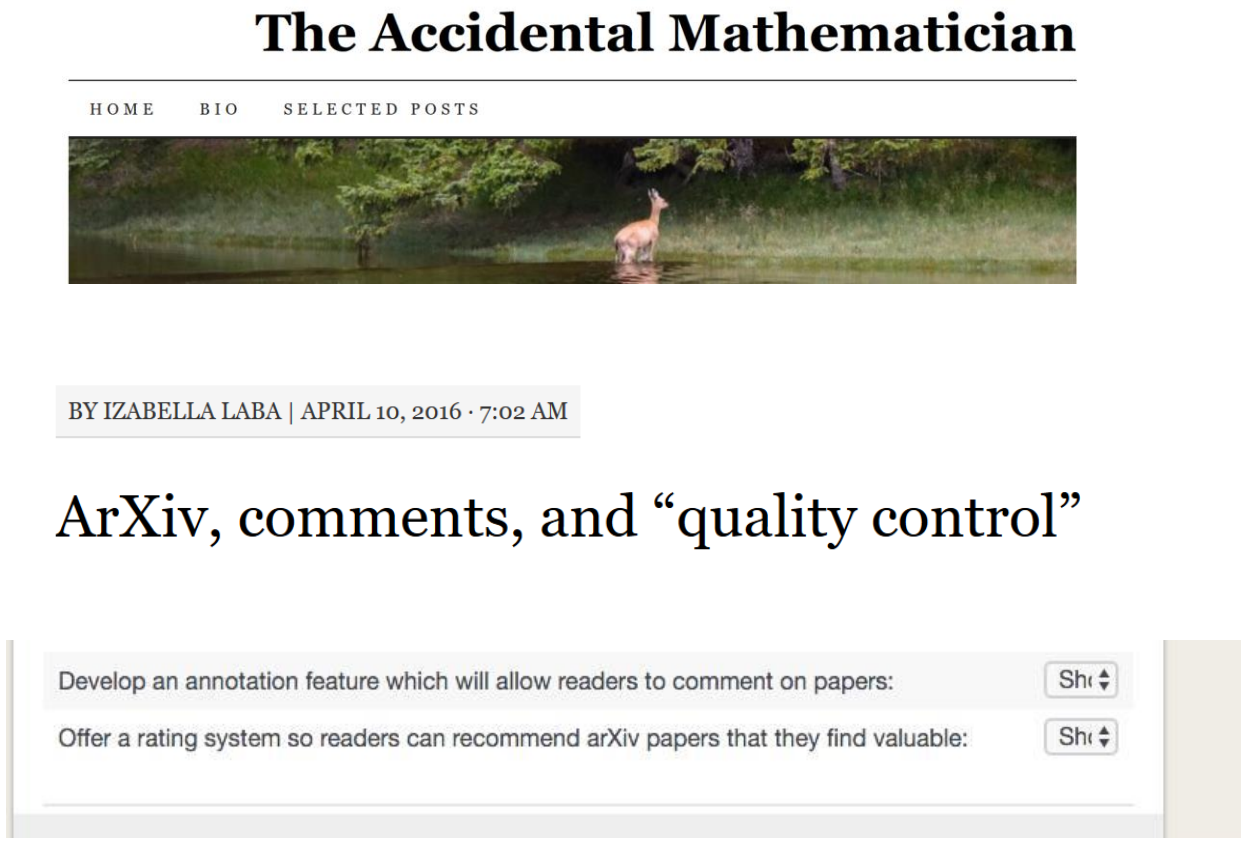
asapbio.org/10-ways

Preprint feedback can inform journal decisions

"In addition, the journal reserves the right--but is not obligated--to consider the comments made to manuscripts posted to preprint servers and factor these comments into final decisions at any stage of the peer review process."

http://www.fasebj.org/site/misc/edpolicies.xhtml#Preprint_Submissions

The dark side of comments



"Internet comment sections are in decline everywhere you look. They are mocked, ridiculed, despised. Many websites have closed them already; others have seen their comments become a racist, sexist bog of eternal stench from which any reasonable person is best advised to stay away."

"Women, in particular, get far too many comments questioning our competence[…] We're also subject to gendered standards of "professionalism" that do not allow us to respond in kind and give as good as we get. But if you tell me that men, too, can get inane, confused, or malicious comments–why, yes, I agree. More reason to refrain from making the arXiv more like YouTube."

https://ilaba.wordpress.com/2016/04/10/arxiv-comments-and-quality-control/

Approaches to increasing quality

- Banning anonymous commentary
- Moderation

Preprint commenting venues

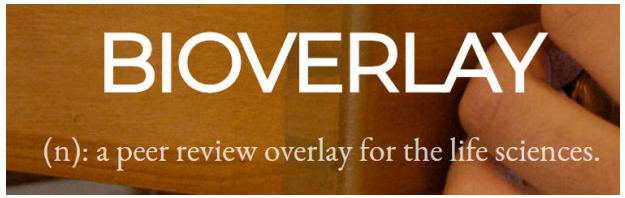
ACADEMIC KARMA

PREreview

épisciences.org

PUBPEER
The online Journal club

Peer Community In
Evolutionary Biology
Free and transparent preprint and postprint recommendations in evolutionary biology

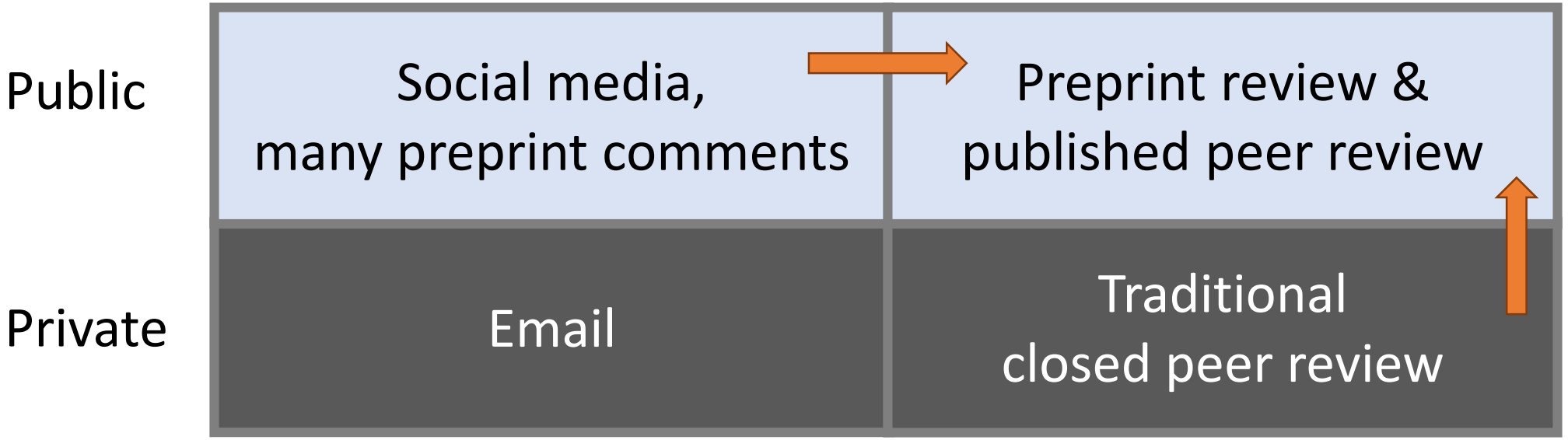


publons

PEERIODICALS
Select the best science beta

scienceOPEN.com
research+publishing network

Getting the best of both worlds



Thank you!

- ASAPbio board
  - Ron Vale (Founder/President)
  - Cynthia Wolberger
  - Jaime Fraser
  - Prachee Avasthi
  - Heather Joseph
  - Harold Varmus
  - Daniel Colon-Ramos
  - Tony Hyman
  - Harlan Krumholz
  - Dick Wilder (non-voting)

- Funding
  - Alfred P. Sloan Foundation
  - Simons Foundation
  - Laura and John Arnold Foundation
  - Gordon and Betty Moore Foundation
  - Leona M. and Harry B. Helmsley Charitable Trust