

## POINT/COUNTERPOINT

*Suggestions for topics suitable for these Point/Counterpoint debates should be addressed to Colin G. Orton, Professor Emeritus, Wayne State University, Detroit: ortonc@comcast.net. Persons participating in Point/Counterpoint discussions are selected for their knowledge and communicative skill. Their positions for or against a proposition may or may not reflect their personal opinions or the positions of their employers.*

### The future *h*-index is an excellent way to predict scientists' future impact

Daniel E. Acuna, Ph.D.

*Department of Biomedical Engineering, Rehabilitation Institute of Chicago and Northwestern University, Chicago, Illinois 60611  
(Tel: 312-238-6515; E-mail: daniel.acuna@northwestern.edu)*

Orion Penner, Ph.D.

*Laboratory of Innovation Management and Economics, IMT Institute for Advanced Studies, 55100 Lucca, Italy  
(Tel: 39 338 520 8665; E-mail: orion.penner@gmail.com)*

Colin G. Orton, Ph.D., Moderator

(Received 11 July 2013; accepted for publication 11 July 2013; published 2 October 2013)

[<http://dx.doi.org/10.1118/1.4816659>]

#### OVERVIEW

Typically, when scientists are being considered for funding, appointment, or promotion, a committee reviews their publications and often estimates the importance of these using a metric such as the *h*-index.<sup>1</sup> This just represents *past* accomplishments, however, and not potential *future* impact, which is probably more important. A new metric, the future *h*-index,<sup>2</sup> has been introduced and is claimed to be an excellent way to predict scientists' future impact. This is the premise debated in this month's Point/Counterpoint.



Arguing for the Proposition is Daniel E. Acuna, Ph.D. Dr. Acuna received his Bachelor's and Master's Degrees from the University of Santiago, Chile and his Ph.D. from the University of Minnesota, Twin Cities, Minnesota. He is currently a Research Associate in the Sensory Motor Performance Program at the Rehabilitation Institute of Chicago and Northwestern University, specifically

working in Dr. Konrad Kording's Bayesian Behavior Lab. He currently studies large data sets to make sense of the science itself—so called “science of science.” Dr. Acuna and his colleagues try to find statistical regularities in large, unstructured data from heterogeneous sources to better understand publication, funding, and teaching activities. By using machine learning techniques, they hope to distill the rules that tell apart successful from less successful ways of doing science and predict quantities such as the *h*-index, yearly funding costs, and students' teaching evaluations.



Arguing against the Proposition is Orion Penner, Ph.D. Dr. Penner obtained his B.Sc. (Hons.) from the University of Manitoba, an M.A. from Boston University, and his Ph.D. from the University of Calgary, all in Physics. His thesis research focused on Complex Networks and information theoretic approaches to biological sequence alignment. He is currently an Assistant Profes-

sor at the IMT Institute for Advanced Studies, Lucca, Italy, and a member of the Laboratory of Innovation Management and Economics. His research now focuses on novel applications of data centric approaches to problems from Innovation Economics and the Science of Science.

#### FOR THE PROPOSITION: Daniel E. Acuna, Ph.D.

##### Opening Statement

Hiring, promotion, and funding decisions are to a large extent driven by predictions about scientists' future impact. A scientist's future impact can be defined with respect to many different objectives, such as research publications, funding, teaching ability, and outreach.<sup>3</sup> Publications and their citations relate to a scientist's ability to do research and are of great importance to decision makers in research focused institutions. This discussion focuses on a specific kind of impact estimation, the estimation of future publication impact.

The excellence of a prediction is defined by the quality of the decision that it supports. Indices like the future *h*-index are noisy: some aspects of excellence are not visible in the

publication record. They may also be biased: some aspects of excellence, e.g., interdisciplinary training, may not be observable. However, estimates of committees are also noisy. For example, members may have highly individual favoritisms. They may be discriminatory: gender, race, and other types of prejudices may cloud judgments.<sup>4</sup> How useful each source of information is depends on the sizes of biases and noises, and these problems can be solved by predicting the future *h*-index.<sup>2</sup>

The *h*-index is one of the simplest and easiest metrics of publication impact,<sup>1</sup> making it attractive to use. It is robust and hard to manipulate.<sup>5</sup> However, the *h*-index does not account for the age of scientists and individual contributions of each co-author, and self-citation introduces biases in favor of authors with many co-authors.<sup>6</sup> Also, it varies substantially from field-to-field, making it hard to compare researchers across disciplines.<sup>4</sup> There are alternatives that fix these problems (e.g., Aziz and Rozing<sup>6</sup>) but they usually require access to more detailed information, making them impractical. However, both noise sources and biases can be addressed by statistics, and the future *h*-index is one such development.

Committees of peers and experts are the current standard for predicting future scientific success. These committees increasingly use metrics,<sup>1</sup> necessary because of the growing number of applicants per tenure-track job or funding opportunity.<sup>7</sup> Why should we not try to find the statistically optimal combination of multiple features to achieve better predictions? For example, we could have an index that combines many indices such as funding, teaching, and other metrics from social media (e.g., “Altmetrics”<sup>8</sup>) and maybe try to predict them in an optimal manner, similar to Acuna *et al.*<sup>2</sup> True excellence in future predictions comes from describing weaknesses of approaches, and then statistically finding ways of correcting them. The future *h*-index is just one small step in this direction.

## AGAINST THE PROPOSITION: Orion Penner, Ph.D.

### Opening Statement

This proposition can be parsed into three questions. First, assuming 100% prediction accuracy, is the future *h*-index a good measure of a scientist’s future impact? Second, how accurate is the future *h*-index model proposed by Acuna *et al.*?<sup>2</sup> Third, how should its likely use as an evaluation tool shape the criteria we use to judge future *h*-index?

The first question appears to challenge the *h*-index as a measure of research impact, but it does not necessarily do so. What it does challenge is the logic of associating an increase in the *h*-index between years  $t$  and  $t + \Delta t$  with research impact during those years. That increase is driven not only by citations to papers published between  $t$  and  $t + \Delta t$  but also citations to papers published before  $t$ . In fact there is a good evidence that the increase is largely driven by citations to previous work.<sup>9</sup> Indeed, even though the mathematician Paul Erdős passed away in 1996, his *h*-index increased by 9 between 2001 and 2010, one of the largest increases of any mathematician over that period.

Turning to the technical aspects of predicting the future *h*-index, the model of Acuna *et al.* is a good starting point but suffers a number of shortcomings. Collaborators and I discovered its predictive power depends heavily upon “career age.”<sup>10</sup> In forthcoming work we further demonstrate much of the model’s “predictive power” arises from the cumulative and increasing character of the *h*-index itself.<sup>11</sup> These challenges will likely be overcome by improving the model, but it is hard to overlook the fact that a sufficiently powerful model does not yet exist.

The third question is the one most often overlooked. To most practicing scientists these models and measures will be curiosities, but to people making decisions on tenure track hires, fellowships, institutional and national tenure, etc. they will be tools.<sup>12,13</sup> The true measure of the future *h*-index comes down to its suitability as a decision making tool. In that context a model must do more than simply fit the data, i.e., produce a high  $R^2$ , it must also produce accurate rankings. Further, for late career awards (e.g., election to the National Academy of Sciences) it is desirable to measure a scientist’s future impact based on papers published before and after year  $t$ . But in the case of a tenure track hire it is critical to identify the scientist whose work after  $t$  will have the greatest impact. Again, the future *h*-index cannot discriminate between the future impact of previously published papers and the future impact of work yet to be published.

The future *h*-index is an excellent contribution to the scientific community and pushes the discourse in an important direction, but a great deal of technical and conceptual refinement is required before it is an excellent way to predict scientists’ future impact.

### Rebuttal: Daniel E. Acuna, Ph.D.

It seems that there is a deep level of agreement about the central issues that Dr. Penner so eloquently put into the three central questions.

The *h*-index is not perfect at measuring research productivity and it would be better to predict something more meaningful for decision makers—once such a measure has been properly defined. In fact, it would be useful to better understand the aims of the various players in the academic market. Methods related to utility elicitation<sup>14</sup> could help by allowing a scientific approach for measuring what we should predict.

The features that we used in the formulation of the future *h*-index were based only on the publication record, which is not overly indicative early on in the career.<sup>10</sup> Early career predictions could be significantly improved by adding other features such as courses taken, grades obtained, and text analysis of letters of recommendation. Since we did not have access to such information, we focused on the publication record only—which seems to work relatively well a few years into the scientific career, even in other scientific domains.<sup>15</sup> Our approach and feature set is just a first step on the path of providing useful predictions of impact.

Predictions are ultimately used for decision-making, rendering rank predictions or other model assessments more

appropriate than simply demonstrating a high  $R^2$ . For example, granting tenure may be about future publication success, funding, and teaching abilities, whereas election to the National Academy of Sciences may be about career-length impact. Data-driven predictive tools offer an opportunity to impartially help during this process. The way we see our own approach is a starting point that can be improved by both better formulating what matters and by better predicting the variables that matter.

Finally, quantitative approaches need to be compared to the alternatives. Human experts are known to be biased in many ways and, just like any algorithm, might not optimize the variables that truly matter for the decision. A deeper analysis about how to improve these approaches would be an exciting topic for a future discussion.

### Rebuttal: Orion Penner, Ph.D.

Indeed I agree with Dr. Acuna that the future  $h$ -index stands as a good first step towards the use of quantitative approaches in hiring and advancement decisions in academia. However, there is still a great deal more careful and rigorous work to be done before a mature suite of tools is devised. In my opening I highlighted several key flaws that must be addressed, but there are other factors that will play a role in determining whether or not quantitative approaches will be widely accepted. As Dr. Acuna points out, a good framework must be able to integrate information on many other facets of the academic career, including funding, teaching, communication and leadership skills. Any successful approach must also be flexible in how it integrates these data, such as being able to handle input that plays a nonlinear role in determining future impact. Perhaps most importantly it is critical that any framework explicitly produces easily interpreted confidence bounds for its predictions, clearly indicating to the user when it is being stretched beyond its “comfort zone.” Satisfying these criteria in one modeling framework is a challenge to be certain, but it does not seem that any one piece is impossible.

If one huge challenge does lie on the path to an appropriate modeling framework, I speculate it is not a matter of the model at all but rather the availability of highly accurate career data for a huge number of scientists. Indeed, few quan-

titative studies of full and complete career trajectories have surpassed 5000 careers. Development of a suitably flexible and accurate prediction framework will require data sets that stretch, at the very least, into the hundreds of thousands and be spread across all academia. With such diverse data it may be possible to develop a model of academic careers capable of being fully validated, but at the moment the lack of such data represents a massive road block.

Dr. Acuna thanks Dr. Konrad Kording for helpful discussions.

- <sup>1</sup>J. E. Hirsch, “An index to quantify an individual’s scientific research output,” *Proc. Natl. Acad. Sci. U.S.A.* **102**(46), 16569–16572 (2005).
- <sup>2</sup>D. E. Acuna, S. Allesina, and K. P. Kording, “Future impact: Predicting scientific success,” *Nature (London)* **489**, 201–202 (2012).
- <sup>3</sup>A. Abbott, D. Cyranoski, N. Jones, B. Maher, Q. Schiermeier, and R. Van Noorden, “Metrics: Do metrics matter?,” *Nature (London)* **465**, 860–862 (2010).
- <sup>4</sup>K. A. A. Dwan *et al.*, “Systematic review of the empirical evidence of study publication bias and outcome reporting bias,” *PLoS ONE*. **3**(8), e3081 (2008).
- <sup>5</sup>J. K. Vanclay, “On the robustness of the  $h$ -index,” *J. Am. Soc. Inf. Sci. Technol.* **58**(10), 1547–1550 (2007).
- <sup>6</sup>N. A. Aziz and M. P. Rozing, “Profit (p)-index: The degree to which authors profit from co-authors,” *PLoS ONE* **8**(4), e59814 (2013).
- <sup>7</sup>A. Jones, “The explosive growth of postdocs in computer science,” *Comm. Assoc. Comput. Mach.* **56**(2), 37–39 (2013).
- <sup>8</sup>J. Priem, P. Groth, and D. Taraborelli, “The altmetrics collection,” *PLoS ONE* **7**(11), e48753 (2012).
- <sup>9</sup>A. Mazloumian, “Predicting scholars’ scientific impact,” *PLoS ONE* **7**, e49246 (2012).
- <sup>10</sup>O. Penner, A. M. Petersen, R. K. Pan, and S. Fortunato, “The case for caution in predicting scientists future impact,” *Phys. Today* **66**(4), 8–9 (2013).
- <sup>11</sup>O. Penner, R. K. Pan, A. M. Petersen, K. Kaski, and S. Fortunato, “On the predictability of career achievement in science,” *Proc. Natl. Acad. Sci. U.S.A.* (submitted).
- <sup>12</sup>R. Mannella and P. Rossi, “On the time dependence of the  $h$ -index,” *J. Informetrics* **7**, 176–182 (2013).
- <sup>13</sup>Agenzia Nazionale di Valutazione del sistema Universitario e della Ricerca, “Abilitazione scientifica nazionale - la normalizzazione degli indicatori per l’età accademica,” (2013) [[http://www.anvur.org/attachments/article/253/normalizzazione\\_indicatori\\_0.pdf](http://www.anvur.org/attachments/article/253/normalizzazione_indicatori_0.pdf)].
- <sup>14</sup>K. J. Bennett and G. W. Torrance, “Measuring health state preferences and utilities: Rating scale, time trade-off, and standard gamble techniques,” in *Quality of Life and Pharmacoeconomics in Clinical Trials*, 2nd ed., edited by B. Spilker (Lippincott-Raven, Philadelphia, PA, 1996), pp. 253–265.
- <sup>15</sup>M. A. García-Pérez, “Limited validity of equations to predict the future  $h$  index,” *Scientometrics* **February**, 1–9 (2013).