Divinations of academic success may be flawed

Method for predicting future *h*-index comes under fire.

Elizabeth Gibney

08 November 2013

Scientists reacted with curiosity — as well as disdain — to the launch last year of a nifty tool that, according to its architects, could predict a researcher's future scientific impact. A study published last week questions its power to do so.

The team behind the tool was led by Daniel Acuna, who studies machine learning at the Rehabilitation Institute of Chicago and Northwestern University in Evanston, Illinois. They developed a formula to predict an individual's *h*-index — a common but controversial metric for measuring scientific impact — years into the future. They described this in a Comment in *Nature*[1].

A scientist has an *h*-index of *n* if he or she has published *n* articles receiving at least *n* citations each[2]. Acuna and his team proposed that an individual's future *h*-index value could be predicted using just five parameters: a scientist's current *h*-index; the number of years since first publication; the total number of articles published; the number of articles published in top journals; and the number of distinct journals in which the scientist has published.

Questionable predictions

But a new study led by physicist Santo Fortunato of Aalto University in Helsinki says that Acuna's team significantly overestimated the predictive power of models such as theirs. Writing in *Scientific Reports*[3], Fortunato and his colleagues describe a methodological flaw in the study — that the input values the model uses are cumulative. Because these figures never go down, they are always, to some extent, correlated with the previous value.

In measurements such as the R^2 test, which quantifies how accurately one variable predicts another, using cumulative values will make the model seem highly predictive, says Raj Kumar Pan, a colleague of Fortunato's at Aalto University and a co-author of the study.

"It's just like if I measure someone's height. In a year, it might increase from 150 centimetres to 152, then to 155 and more. So if I compare the new value with his previous height, of course you will see some correlation," he says — regardless of how good the model is at predicting the increment.

Although Acuna's model used a range of parameters, and not just a scientist's current *h*-index, its apparent success at predicting the paths of neuroscientists' careers — yielding an R^2 value of 0.66, where 1 means a perfect prediction — was largely attributable to this statistical artefact, says Pan.

Fortunato and his co-authors also criticise Acuna's model for lumping all scientists in together, when the model's ability to predict the future varies greatly according to the career age of the scientists, being least accurate in the early years.

Miguel García-Pérez, a psychologist at the Complutense University of Madrid who was not involved in either study, says that his research also revealed the same basic flaw in Acuna's model. In particular, a study[4] he authored re-ran Acuna's model using career data from his field and found it to be a poor predictor of future *h*-index.

His calculations show that one can predict a researcher's future *h*-index just as accurately as Acuna's team does simply by assuming that the researcher will continue to publish, and be cited, as much as they have in the past few years.

Future developments

Acuna acknowledges the criticism in the Fortunato paper, but stresses that his five-parameter method is nonetheless better at predicting future *h*-index than using current *h*-index alone. "Our paper is saying how well can we predict the future values and comparing that to what people are already using, to the status quo," he says.

And the model is not intended to be universal, he says, instead being no more than a framework for developing similar models for groups other than mid-career neuroscientists, which his data covered.

Although the formula can doubtless be improved, Acuna maintains that a model that plugs objective data into a simple equation could be an important part of avoiding the biases that plague subjective opinion in decisions about jobs and funding.

References

1. Acuna, D. E., Allesina, S. & Kording, K. P. *Nature* **489**, 201–202 (2012).

2. Hirsch, J. E. *Proc. Natl Acad. Sci. USA* **102**, 16569–16572 (2005).

3. Penner, O., Pan, R. K., Petersen, A. M., Kaski, K. & Fortunato, S. *Sci. Rep.* http://dx.doi.org/10.1038/srep03052 (2013).

4. García-Pérez, M. *Scientometrics* **96**, 901–909 (2013).