

„EIN HERAUSRAGENDES KRITERIUM FÜR INNOVATIVE ARBEITEN IST NICHT DER KONSENS, SONDERN DER DISSENS VON GUTACHTERN. BEWERTUNGSKRITERIEN FÜR INNOVATION BENÖTIGEN DRINGEND WEITERENTWICKLUNG UND DISKURS, SCHON, UM DEN SCHADEN DERZEITIGER SCORES ABWENDEN ZU HELFEN!“



Tim Hucho

## Wie sinnvoll sind Wissenschaftsbewertungs-Scores?

■ Es wird so viel Geld für Forschung ausgegeben wie noch nie. Da die Ressourcen trotzdem knapp sind und um das Wissensgebäude möglichst breit aufzuspannen, muss Wissenschaft bewertet und „effizientes Wissen schaffen“ gefördert werden. Der Griff für Wissenschaftsverwalter in die betriebswirtschaftliche Wundertüte ist verlockend. Die Produktionszyklen sind klar: Antrag, Forschungsprojekt, Veröffentlichung, Antrag, ... Die Reduzierung von wissenschaftlichem Wert auf eine Zahl wie den Impact Factor, vielleicht noch die Drittmittelsumme und, einer gegenwärtigen „Mode“ folgend, Lehrstunden. Damit ist dann nun wirklich „vielfacettig“ alles zur Leistungsbeurteilung erfasst, oder etwa nicht?

Die Melodie dieser kleinen Polemik ist sicherlich vertraut. Es ist vermutlich ebenfalls präsent, dass die vom Impact Factor beanspruchte Leistungsbeurteilung rundweg falsch und seit 18 Jahren vielfach widerlegt ist [1]. Bleibt die Frage, was denn eigentlich fehlt in diesen wenigen Zahlen? Das Niveau des Impact-Factor-Wissenschaftsbildes haltend, mag man mit leichter Hand Wissenschaftlerstereotypen in Unterhaltungsmedien heranziehen. Dort findet man weniger Archive und Fließbanddatenproduzenten als überfokussierte, unangepasste, von Neugierde umtriebene, Risiken eingehende, enthusiastische (schlecht frisierte) Wissenschaftsrevolutionäre.

Eine Wahrheit dieser Stereotypen mag sein, dass Wissenschaft eben nicht konformistische Fließbandproduktion ist, sondern auf Innovation zielt. Und so bewerten wir Wissenschaftler andere Wissenschaftler in der Regel nicht nach Masse, sondern nach Validität von Daten, der Kohärenz und Überzeugungskraft der Konzepte sowie des Innovationsgrades. Sind das in Zahlen ausdrückbare Größen? Ist dies schwierige Bewerten von Inhalten (sic!) nicht eben auch der ureigenste Kern des Wissenschaftsprozesses? Wird nicht das wissenschaftliche Weltbild etabliert und nutzbar gemacht durch Wissenschaftler, die

komplexe, häufig hoch abstrakte Daten sinnstiftend in einen Kontext stellen, bewerten und in jahrelanger Arbeit nicht durch Scores, sondern durch die wissenschaftliche Anwendung validieren? Ist der Kern unserer Arbeit nicht genau, der Bewertungsfilter zu sein, der gute von schlechter und wichtige innovative Wissenschaft von unbedeutender Datenproduktion trennt? Werden vielleicht daher Nobelpreise nicht auf der Basis eingereicherter Projektförderungsabschlussberichte, eingeworbener Drittmittel und kumulierter Zitate vergeben?

Während man sich ausgiebig über die hier angerissenen Argumente in die zumeist doch eher gut frisierten Haare bekommen kann, führen die neuen Medien die beiden Aspekte Experten- versus Algorithmen-basierte Bewertung überraschend pragmatisch zusammen. In einem auf der Wissenschaftsplattform ResearchGate (RG) kürzlich eingeführten RG-Score gehen nicht nur die Publikationen, sondern auch Bereiche der Wissenschaftskommunikation (Beteiligung an Foren, Datenupload, ...) sowie die Reputation des Wissenschaftlers (Wichtung der Bedeutung der vernetzten Wissenschaftler, Anzahl von Followers) ein. Aber nicht nur die Aktivität des Einzelnen, sondern insbesondere auch die Bewertung dieser Aktivitäten durch die „Community“ spielt dort eine Rolle. Und nicht zuletzt erlaubt eine Darstellung der Scoreverteilung eine greifbare Einordnung, etwas, was die auf Transparenz und Objektivität ausgerichteten Bewertungsindikatoren bislang durchweg vermissen lassen.

Diesem Score muss sicherlich noch mancher Wirbel glattgekämmt werden. Aber er löst zumindest ein zentrales Problem: Während die expertenbasierte wie auch zumeist Impact-Factor-basierte Bewertung auf in der Regel sehr „kleinen Zahlen“ beruht und somit potenziell hochgradig „stochastisch verauscht“ ist, könnte die Aussagekraft durch die breite Bewertungsbasis enorm gesteigert werden. Derartige Scores werden hoffentlich bald zudem weitere Aspekte der Reputation

zu erfassen suchen – nicht zuletzt z. B. gutachterliche Tätigkeiten und idealerweise deren Relevanz bzw. Trefferquote.

Und wenn schon „Scores“, warum dann nicht ein wenig tiefer in die wirtschaftswissenschaftliche und nobelpreisprämierte Wundertüte greifen? Es sollten dann doch eher, wie kürzlich angeregt, weniger die trivialen statischen als vielmehr abgeleitete dynamische Aspekte in z. B. optionshandelbasierten Modellen eine zentrale Rolle für Wissenschaftsbewertung spielen [2]. Nicht zuletzt ist ein herausragendes Kriterium für innovative Arbeiten eben gerade nicht der Konsens, sondern der Dissens von Gutachtern [3]. Eine Sache ist daher sicher: Bewertungskriterien für Innovation benötigen dringend Weiterentwicklung und Diskurs, schon, um den Schaden derzeitiger Scores abwenden zu helfen! ■

Tim Hucho,  
Universitätsklinikum Köln

### Literatur

- [1] Seglen PO (1994) Causal relationship between article citedness and journal impact. *J Am Soc Inform Sci* 45:1–11
- [2] Hucho C (2012) Der Wert der Wissenschaft. *Gegenworte* 27:54–56
- [3] Benda WGG, Engels TCE (2011) The predictive validity of peer review: a selective review of the judgmental forecasting qualities of peers, and implications for innovation in science. *Int J Forecast* 27:166–182

### Korrespondenzadresse:

Prof. Dr. Tim Hucho  
Experimentelle Anästhesiologie  
und Schmerzforschung  
Klinik für Anästhesiologie und  
Operative Intensivmedizin  
Universitätsklinikum Köln (AöR)  
Robert-Koch-Straße 10  
D-50931 Köln  
Tel.: 0221-478-97760  
Fax: 0221-478-87329  
tim.hucho@uk-koeln.de  
<http://anaesthesie.uk-koeln.de>