Cinzia Daraio
Wolfgang Glänzel *Editors*

Evaluative Informetrics: The Art of Metrics-Based Research Assessment

Festschrift in Honour of Henk F. Moed

Springer

Evaluative Informetrics: The Art of Metrics-Based Research Assessment

Cinzia Daraio · Wolfgang Glänzel
Editors

Evaluative Informetrics: The Art of Metrics-Based Research Assessment

Festschrift in Honour of Henk F. Moed

Springer

Editors
Cinzia Daraio
Dipartimento di Ingegneria Informatica, Automatica e Gestionale (DIAG)
Sapienza University
Rome, Italy

Wolfgang Glänzel
Centre for R&D Monitoring (ECOOM) and Dept MSI
KU Leuven
Leuven, Belgium

ISBN 978-3-030-47664-9 ISBN 978-3-030-47665-6 (eBook)
https://doi.org/10.1007/978-3-030-47665-6

© Springer Nature Switzerland AG 2020
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Henk's full first name is Hendrik, his middle initial is F. In official documents, also those related to the Doctorate honoris causa, his name is therefore Hendrik F. Moed. But friends, family and colleagues use the first name Henk, a sort of nickname for Hendrik. The name Henk is also used in Henk's scientific publications.

Contents

Tracing the Art of Metrics-Based Research Assessment Through Henk Moed's Work

Cinzia Daraio and Wolfgang Glänzel

The title of the editorial introduction summarises the main objective of this book. During the ISSI2019 Conference held at the Sapienza University of Rome, on 5 September 2019, we organised a Special Plenary Session in honour of our colleague and friend Prof. Henk F. Moed to celebrate his retirement. Before this special session, a formal ceremony to the conferral of the Doctorate *Honoris Causa* in Industrial and Management Engineering on "Research Assessment Methodologies" was held in the historical Academic Senate House of the Sapienza University of Rome.

We organized this session, since we had the fortune to accompany stages of Henk Moed's career as his colleagues and collaborators, co-authors and friends; younger colleagues enjoyed the opportunity to learn and benefit from the comprehensive knowledge that he has shared with the scholarly community. We embraced the opportunity to commemorate this special occasion and withal *honour* one of the most prominent scholars in the field of scientometrics by editing this book.

The book consists of four parts. The first part presents selected papers by Henk Moed, the second part contains contributed research papers, the third part refers to the ceremony for the conferral of the doctorate *honoris causa* in Research Assessment Methodologies to Henk Moed, and the fourth part includes personal notes.

The first part reports a collection of the most important publications by Henk F. Moed. This selection presents Henk as a scholar with a broad spectrum of activities and a multifaceted research profile. Due to his rich contribution to the advancement of the research assessment methodologies and its application, investigating the development of his career is, to a considerable extent, also a survey of our research field. We

C. Daraio (✉)
Sapienza University of Rome, Rome, Italy
e-mail: daraio@dis.uniroma1.it

W. Glänzel
KU Leuven, Leuven, Belgium
e-mail: Wolfgang.Glanzel@kuleuven.be

© Springer Nature Switzerland AG 2020

C. Daraio and W. Glänzel (eds.), *Evaluative Informetrics: The Art of Metrics-Based Research Assessment*, https://doi.org/10.1007/978-3-030-47665-6_1

Table 1 A selection of the most important publications by Henk F. Moed

Bibliometric databases	*Exploring the use of existing, primarily bibliographic databases for bibliometric purposes has been the most important subject of the first half of Henk Moed's career, although he has made several database-oriented studies also in the second half. It was a topic of great general interest in the field. This topic involves the following sub-topics: the creation of bibliometric databases; combining databases; comparing databases; and the assessment and enhancement of their data quality*
1	Moed, H. F. (1988). The Use of Online Databases for Bibliometric Analysis. In: Informetrics 87/88. L. Egghe and R. Rousseau (eds.), Elsevier Science Publishers, Amsterdam, ISBN 0-444-70425-6, 15-28
2	Moed, H. F., Vriens, M. (1989). Possible Inaccuracies Occurring in Citation Analysis. Journal of Information Science, 15, 2, 95–107. Sage Journals
3	Moed, H. F. (2005). Accuracy of citation counts. In: H.F. Moed, Citation Analysis in Research Evaluation. Springer, Dordrecht (Netherlands). ISBN 1-4020-3713-9, 173–179
4	López-Illescas, C., De Moya-Anegón, F., Moed, H. F. (2008). Coverage and citation impact of oncological journals in the Web of Science and Scopus. Journal of Informetrics, 2, 304–316. Elsevier
5	Moed, H.F., Bar-Ilan, J, Halevi, G. (2016). A new methodology for comparing Google Scholar and Scopus. Journal of Informetrics, 10, 533–551. Elsevier
Journal citation measures	*Journal impact factors and related citation measures are even today probably the most frequently used bibliometric indicators. The articles relate to a critique on existing indicators, proposals for new indicators, and a more reflexive paper addressing criteria for evaluating indicators on the basis of their statistical soundness, theoretical validity, and practical usefulness. Also, one paper examines the effect of the Open Access upon citation impact*

(continued)

Table 1 (continued)

1	Moed, H. F., van Leeuwen, Th. N. (1995). Improving the accuracy of Institute for Scientific Information's journal impact factors. J. of the American Society for Information Science, 46, 461–467 Wiley publisher
2	Moed, H. F., van Leeuwen, Th. N., Reedijk, J. (1999). Towards appropriate indicators of journal impact, Scientometrics, 46, 575-589. Springer
3	Moed, H. F., van Leeuwen, Th. N., Reedijk, J. (1999). Towards appropriate indicators of journal impact, Scientometrics, 46, 575–589. Springer
4	Moed, H. F. (2007). The effect of "Open Access" upon citation impact: An analysis of ArXiv's Condensed Matter Section. Journal of the American Society for Information Science and Technology, 58, 2047–2054. Wiley publisher
5	Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. Journal of Informetrics, 4, 265–277. Elsevier
6	Moed, H. F. (2016). Comprehensive indicator comparisons intelligible to non-experts: the case of two SNIP versions. Scientometrics, 106 (1), 51–65. Springer
Indicators of research performance in science, social science and humanities	*The development of appropriate quantitative research assessment methodologies in the various domains of science and scholarship and various organizational levels has been Henk Moed's core-activity during the first two decades. Bibliometric indicators were applied to research groups, departments, institutions, and countries*
1	Moed, H. F., Burger, W. J. M., Frankfort, J. G., van Raan, A. F. J. (1985). The Use of Bibliometric Data for the Measurement of University Research Performance. Research Policy, 14, 131–149. Elsevier
2	Moed, H. F., de Bruin, R. E., van Leeuwen, Th. N. (1995). New bibliometric tools for the assessment of national research performance: database description, overview of indicators and first applications. Scientometrics, 33, 381–422. Springer

(continued)

Table 1 (continued)

3	Moed, H. F., Hesselink, F. Th. (1996). The publication output and impact of academic chemistry research in the Netherlands during the 1980's: bibliometric analyses and policy implications. Research Policy, 25, 819–836. Elsevier
4	Van den Berghe, H., Houben, J. A., de Bruin, R. E., Moed, H. F., Kint, A., Luwel, M., Spruyt, E. H. J. (1998). Bibliometric indicators of university research performance in Flanders. Journal of the American Society for Information Science, 49, 59–67. Wiley publisher
5	Moed, H. F. (2002). Measuring China's research performance using the Science Citation Index. Scientometrics, 53, 281–296. Springer
6	Moed, H. F., Nederhof, A. J, Luwel, M. (2002). Towards performance in the humanities. Library Trends, 50, 498–520. JHU Press
Theoretical understanding and proper use of bibliometric indicators	*This topic comprises articles of Henk Moed discussing and proposing theories about what citations and other bibliometric indicators measure. Moreover, it includes reflexive articles addressing the issue as to what are appropriate ways to use these indicators in research assessment processes*
1	Moed, H. F. (2000). Bibliometric indicators reflect publication and management strategies. Scientometrics, 47, 323–346. Springer
2	Moed H.F., Garfield E. (2004). In basic science the percentage of 'authoritative' references decreases as bibliographies become shorter. Scientometrics, 60, 295–303. Springer
3	Moed, H. F. (2005). Towards a theory of citations: Some building blocks. In: H. F. Moed, Citation Analysis in Research Evaluation. Springer, Dordrecht (Netherlands). ISBN 1-4020-3713-9, 209–220
4	Moed, H. F. (2008). UK Research Assessment Exercises: Informed Judgments on Research Quality or Quantity? Scientometrics, 74, 141–149. Springer
5	Moed, H. F., Halevi, G. (2015). Multidimensional Assessment of Scholarly Research Impact. Journal of the American Society for Information Science and Technology, 66, 1988–2002. Wiley publisher

(continued)

Table 1 (continued)

Usage-based metrics and altmetrics	*Nowadays publication- and citation based indicators of research performance are not seldom denoted as 'classical', and new, alternative types of indicators are being proposed and explored. Two articles by Henk Moed listed below relate to 'usage' indicators, based on the number of times full text articles are downloaded from publishers' publication archives. A third article discusses the potential of so called altmetrics, especially those that reflect use of social media*
1	Moed, H. F. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. Journal of the American Society for Information Science and Technology, 56, 1088–1097. Wiley publisher
2	Moed, H. F. (2016). Altmetrics as traces of the computerization of the research process. In: C.R. Sugimoto (Ed.), Theories of Informetrics and Scholarly Communication (A Festschrift in honour of Blaise Cronin). Walter de Gruyter, Berlin–Boston. ISBN 978-3-11-029803-1, 360–371
3	Moed, H.F., Halevi, G. (2016). On full text download and citation distributions in scientific-scholarly journals. Journal of the American Society for Information Science and Technology, 67, 412–431. Preprint version available at https://arxiv.org/ftp/arxiv/papers/1510/1510.05129.pdf Wiley publisher
International collaboration and migration	*Scientific collaboration and migration are important phenomena that can be properly studied with bibliometric-informetric methods. Below three contributions by Moed are listed, two on collaboration, and one on migration*
1	Moed, H. F. (2005). Does international scientific collaboration pay? In: H. F. Moed, Citation Analysis in Research Evaluation. Springer, Dordrecht (Netherlands). ISBN 1-4020-3713-9, 285–290
2	Moed, H. F. (2016). Iran's scientific dominance and the emergence of South-East Asian countries as scientific collaborators in the Persian Gulf Region. Scientometrics, 108, 305–314. Preprint version available at http://arxiv.org/ftp/arxiv/papers/1602/1602.04701.pdf. Springer

(continued)

Table 1 (continued)

3	Moed, H. F., Halevi, G. (2014). A bibliometric approach to tracking international scientific migration. Scientometrics, 101, 1987–2001. Springer
The future of bibliometric and informetrics	*The articles of Moed in this section provide a perspective of the future, both in the development of informetric indicators, and in their application in research assessment processes. His monograph Applied Evaluative Informetrics contains several chapters on these topics. Therefore, the executive summary of this book is also listed below*
1	Moed, H. F. (2007). The Future of Research Evaluation Rests with an Intelligent Combination of Advanced Metrics and Transparent Peer Review. Science and Public Policy, 34, 575–584. Oxford University Press
2	Moed, H. F. (2016). Toward new indicators of a journal's manuscript peer review process. Frontiers in Research Metrics and Analytics, 1, art. no 5. Available at: http://journal.frontiersin.org/article/10.3389/frma.2016.00005/full
3	Moed, H. F. (2017). A critical comparative analysis of five world university rankings. Scientometrics, 110, 967–990. Springer
4	Moed, H. F. (2017). Executive Summary. In: H. F. Moed, Applied Evaluative Informetrics. Springer, ISBN 978-3-319-60521-0 (hard cover); 978-3-319-60522-7 (E-Book), https://doi.org/10.1007/978-3-319-60522-7

grouped his publications into seven topics at the intersection of bibliometrics, scientometrics, informetrics and research evaluation. The main topics covered are: 'Bibliometric databases', 'Journal citation measures', 'Indicators of research performance in science, social science and humanities', 'Theoretical understanding and proper use of bibliometric indicators', 'Usage-based metrics and altmetrics', 'International collaboration and migration', and 'The future of bibliometric and informetrics' (see Table 1).

The second part collects 13 original research papers by experts in the field who have worked and collaborated with Henk F. Moed during the last over three decades. We organised these contributions, reported in detail in Table 2, in the three following topics:

- Advancement of bibliometric methodology
- Evaluative informetrics and research assessment
- New horizons in informetric studies.

Table 2 Chapters in Part II

Topic	Authors	Title
Advancement of bibliometric methodology.		
	Braam R.	Citation profiles and research dynamics
	Luwel M., van Eck N. J., and van Leeuwen T.	Characteristics of publication delays over the period 2000–2016
	Pendlebury D. A.	When the data do not mean what they say: Japan's comparative underperformance in citation impact
	Zhao Y., Han J., Du J. and Wu Y.	Origin and Impact: A Study of the Intellectual Transfer of Professor Henk F. Moed's works by Using Reference Publication Year Spectroscopy (RPYS)
Evaluative informetrics and research assessment.		
	Calero-Medina C., Noyons Ed, Visser M. and de Bruin R.	Delineating Organizations at CWTS—A story of many pathways
	Halevi G.	Research Trends—Practical Bibliometrics and a Growing Publication
	Pallari E. and Lewison G.	The evidence base of international clinical practice guidelines on prostate cancer: a global framework for clinical research evaluation
	Robinson-Garcia N. and Ràfols I.	The differing meanings of indicators under different policy contexts. The case of internationalisation.
	Gorraiz J., Martin Wieland M., Ulrych U. and Gumpenberger C.	De profundis: a decade of bibliometric services under scrutiny
New horizons in informetric studies.		
	Costas R. and Ferreira M.R.	A Comparison of the Citing, Publishing, and Tweeting Activity of Scholars on Web of Science
	Torres-Salinas D., Arroyo-Machado W.	Library Catalog Analysis and Library Holdings Counts: origins, methodological issues and application to the field of Informetrics
	De-Moya-Anegón F., Guerrero-Bote V.P. and Herrán-Páez E.	Cross-national comparison of Open Access models: A cost/benefit analysis
	Bar-Ilan J. and Halevi G.	The Altmetrics of Henk Moed's publications

The following gives a content-related summary of the above 13 chapters, most which are very closely related to Henk Moed's ideas, proceeding from, reinforcing or generalising his findings by new examples or contexts, others by using his work as the subject of new bibliometric studies.

Advancement of bibliometric methodology

The chapter by Braam (2020) analyses the citation profiles of individual researchers as reflected by Google Scholar in the light of their dynamics. The author distinguished different types of profiles according to the authors' productivity and prestige. The comparison with expected patterns based on bibliometric theories of publication and citation processes resulted in the identification of three characteristic elements in terms of communication and the reception by the community.

Luwel et al. (2020) study the characteristics of publication delays in the era of electronic scholarly communication in the about last two decades. The study is based on Elsevier publications and conducted at three levels, the subject level, the journal level and the publishing model. Although the publication process has been substantially accelerated, the peer-reviewing still requires considerable amount of time and proved the most time-consuming element in the process.

Pendlebury (2020) examines an interesting phenomenon: Japan's comparative underperformance in citation impact. The analysis is methodically based on several aspects that are usually considered influencing factors of citation impact, including publication language, number of co-authors, international collaboration, mobility, research focus and diversity. The author identified the national orientation of publication venues with an effect of cumulative disadvantage as one possible determinant the resulting in a structural citation-impact deficit. He also argues in favour of a careful interpretation of national citation indicators to avoid misconstruction of their meaning.

The chapter by Zhao, Han, Du, and Wu (2020) focuses on the intellectual transfer of Henk Moed's ideas. In particular, the authors propose the (co-)citation analysis of both the documents cited by Henk Moed and the literature citing his most influential papers. In order to implement this idea, the authors adopt a method previously proposed by Marx et al. in (2014) called Reference Publication Year Spectroscopy (RPYS). By doing so, they characterise Henk Moed as one of the influential contemporary scientists in the field of bibliometrics and informetrics and also provide new methodological insights by connecting bibliometrics with research in history of science.

Evaluative informetrics and research assessment

The chapter by Calero-Medina et al. tackle an extremely important task in evaluative bibliometrics, the identification and harmonisation of entities. They describe the time-consuming and complex process of identifying and harmonising organisation names, which includes the careful cleaning of author affiliations of publications. This work proved an indispensable prerequisite to reliable meso-level research evaluation and university rankings.

Halevi (2020) reflects on Henk Moed work as the editor in chief of "Research Trends", Elsevier's online publication aiming to provide straightforward insights into scientific trends based on bibliometric research. Under Henk Moed's management, Research Trends evolved from kind of 'newsletter' to a full-featured scientific publication organ providing a large spectrum of articles in a variety of topics and disciplines.

Pallari and Lewison (2020) investigate the evidence base of international clinical practice guidelines on prostate cancer. The guidelines are designed to ensure that medical diagnosis and treatment are based on the best available evidence. The authors analyse their cited references in journals processed in the Web of Science as their evidence base. They found, among others, that most guidelines over-cite research from their own country and also differences between countries in the topicality of citations. The authors conclude that citations on the guidelines provide an alternative source of information for the evaluation of clinical research.

In their chapter, Robinson-Garcia and Ràfols (2020) focus on the use of indicators in research evaluation regarding internationalisation policies. In particular, they analyse three examples of indicators in this context. The first example is related to international collaboration and investigates whether a larger extent of internationally co-authored publications exhibits higher citation impact and thereby benefits national science systems. The second one concerns the publication language, particularly the promotion of English language as the dominant language of science. The last example shows the effect of the policy contexts in shaping the use and application of bibliometric indicators, sometimes in a partial way which does not properly reflect the phenomenon under study.

Gorraiz et al. (2020) present and discuss the lessons learned after having provided bibliometric services at the University of Vienna for more than a decade. By comparing their experience and insights with current evaluative practices, with statements of declarations and manifestos, they succeeded in coming up with new recommendations and including the question of to what degree alternative metrics have the potential for being used in research assessment. The authors also plead for going beyond evaluative tasks. Bibliometric services should encourage researchers in improving their publication strategies and enhancing their visibility within and beyond their research communities.

New horizons in informetric studies

Proceeding from Henk Moed's statement that web-based indicators "do not have function merely in the evaluation of research performance of individuals and groups, but also in the research process", Costas and Ferreira (2020) set out to go beyond the evaluative perspective of altmetrics to a more contextualised one in which they conducted a comparative analysis of researchers' citing, publishing, and tweeting activities. They found at the individual researcher level that Twitter-based indicators are empirically different from production-based and citation-based bibliometric

indicators. The authors consider their results a step towards a conceptual shift to a more dynamic perspective that focuses on the social media activities of researchers and propose future research directions based on their findings.

A completely different approach is proposed by Torres-Salinas and Arroyo-Machado in their Sect. "Citer Motivations". Library Catalog Analysis designed as the application of bibliometric techniques to published book titles in online library catalogues can be used to analyse the impact and dissemination of academic books in different ways. The aim of the chapter is to conduct an in-depth analysis of major scientific contributions and to this topic. Beyond the discussion of the original purposes of library holdings and analysis of the principal sources of information, the authors study the correlation between library holdings and altmetrics indicators and the use of WorldCat Identities to identify the principal authors and works in the field of informetrics.

A cost-benefit analysis of Open Access publishing in a cross-national comparison of OA models is presented by de-Moya et al. in their Sect. "Relevancy Versus Impact". The transition from traditional publishing towards OA is internationally dealt with in different ways. In particular, the four OA models, platinum, gold, green and hybrid are compared in terms of scientific impact and costs. The authors found and discuss different country models, with different costs and different results.

Halevi and Bar-Ilan have chosen Henk Moed's work as the subject of their study (2020). His work, embracing collaboration with over 60 authors from 30 countries across all continents and published in more than 30 different journals and a variety of attracted thousands of citations. Hitherto relatively little is known about the altmetric impact in terms of usage, readership, and social media attention of his work. The results obtained from the main altmetric indicators shed light on how his publications are viewed, read, shared and tweeted about within the scholarly community and beyond.

Part III concerns the Conferral of the Doctorate Honoris Causa to Henk Moed and includes the opening address of the Rector of Sapienza University of Rome (Gaudio, 2020), the address by the Coordinator of the Doctoral Program in Industrial and Management Engineering (Tronci, 2020) and Henk Moed's Lectio Magistralis (Moed, 2020). In his Lectio magistralis on "The Application Context of Research Assessment Methodologies", Moed sheds new lights on the complex and controversial role and use of bibliometric or informetric indicators in the assessment of research performance. He highlights the fundamental importance of the application context of these indicators enlightening and further developing the search for best practices in research assessment.

Part IV includes a personal note (Lopez-Illescas, 2020) and concludes the book a nice Interview done by Cinzia Daraio to Bluma Peritz during the ISSI2019 Conference in Rome (Peritz, 2020).

We would like to express our gratitude to all the authors of the chapters of Part II of the book for their valuable contribution. We warmly thank the publishers of the journals and books reported in Part I of the book that kindly allowed us to reproduce the abstracts and executive summary of the selected works of Henk Moed. Finally, our deepest thanks are due to Diletta Abbonato for her precious support in the finalisation

of the book and to Guido Zosimo-Landolfo from Springer Nature for his kind support in the development of this project.

References

Bar-Ilan, J., & Halevi, G. (2020). The Altmetrics of Henk Moed's publications, In this volume.
Braam, R. (2020). Citation profiles and research dynamics. In this volume.
Calero-Medina C., Noyons Ed, Visser, M., & de Bruin, R. (2020). Delineating Organizations at CWTS—A story of many pathways. In this volume.
Costas, R., Ferreira, M. R. (2020). A comparison of the citing, Publishing, and Tweeting Activity of Scholars on Web of Science, In this volume.
De-Moya-Anegón, F., Guerrero-Bote, V. P., Herrán-Páez, E. (2020). Cross-national comparison of Open Access models: A cost/benefit analysis, In this volume.
Gaudio, E. (2020). *Conferral of the Doctorate honoris causa in Industrial and Management Engineering to Hendrik F*. Moed: Opening Address, In this volume.
Gorraiz, J., Wieland, M., Ulrych, U., & Gumpenberger, C. (2020). De profundis: A decade of bibliometric services under scrutiny In this volume.
Halevi, G. (2020). Research Trends—Practical Bibliometrics and a Growing Publication. In this volume.
Lopez-Illescas, C. (2020). Under Bibliometrics. In this volume.
Luwel, L., van Eck, N. J., & van Leeuwen, T. (2020). Characteristics of publication delays over the period 2000–2016. In this volume.
Marx, W., Bornmann, L., Barth, A., & Leydesdorff, L. (2014). Detecting the historical roots of research fields by reference publication year spectroscopy (RPYS). *Journal of the Association for Information Science and Technology, 65*(4), 751–764.
Moed, H. F. (2020). The application context of research assessment methodologies, Lectio Magistralis, Doctorate Honoris Causa Sapienza University of Rome, 5 September 2019 In this volume.
Pallari, E., & Lewison, G. (2020). The evidence base of international clinical practice guidelines in prostate cancer: A global framework for clinical research evaluation. In this volume.
Pendlebury, D. A. (2020). When the data don't mean what they say: Japan's comparative underperformance in citation impact. In this volume.
Peritz, B. (2020). My long time acquaintance with Henk Moed. In this volume.
Robinson-Garcia, N., & Ràfols, I. (2020). The differing meanings of indicators under different policy contexts. The case of internationalisation In this volume.
Torres-Salinas, D., Arroyo-Machado, W. (2020). Library catalog analysis and library holdings counts: Origins, methodological issues and application to the field of Informetrics In this volume.
Tronci, M. (2020). *Conferral of the Doctorate honoris causa in Industrial and Management Engineering to Hendrik F*. Moed: Address by the Coordinator of the Doctoral Program in Industrial and Management Engineering In this volume.
Zhao, Y., Han, J., Du, J., & Wu, Y. (2020). Origin and impact: a study of the intellectual Transfer of Professor Henk F. Moed's works by Using Reference Publication Year Spectroscopy (RPYS) In this volume.

Selected Essays

Selected essays of Henk F. Moed

Cinzia Daraio and Wolfgang Glänzel

Introduction

This part presents a collection of the most important publications by Henk F. Moed. This collection characterises the author as a researcher personality with a broad spectrum of activities and a multifaceted research profile. As Henk Moed has contributed to the advancement of the field in many topics, an overview of the development of his career is, to a considerable extent, also a *survey* of the research field. We grouped his publications into seven topics in the field at the intersection of bibliometrics, informetrics, science studies and research assessment. The main topics are the following.

1. *Bibliometric databases*
2. *Journal citation measures*
3. *Indicators of research performance in science, social science and humanities*
4. *Theoretical understanding and proper use of bibliometric indicators*
5. *Usage-based metrics and altmetrics*
6. *International collaboration and migration*
7. *The future of bibliometric and informetrics.*

The authors would like to thank Sarah Heeffer (ECOOM, KU Leuven, Belgium) for her kind assistance in proofreading the chapter.

C. Daraio (✉)
Sapienza University of Rome, Rome, Italy
e-mail: daraio@dis.uniroma1.it

W. Glänzel
KU Leuven University, Louvain, Belgium

© Springer Nature Switzerland AG 2020

C. Daraio and W. Glänzel (eds.), *Evaluative Informetrics: The Art of Metrics-Based Research Assessment*, https://doi.org/10.1007/978-3-030-47665-6_2

Bibliometric databases

Exploring the use of existing, primarily bibliographic databases for bibliometric purposes has been the most important subject of Henk Moed's work during the first half of his career, although he has made several database-oriented studies also in the second half. It was a topic of great general interest in the field. This topic involves the following sub-topics: the creation of bibliometric databases; combining databases; comparing databases; and the assessment and enhancement of their data quality.

The Use of Online Databases for Bibliometric Analysis. In: Informetrics 87/88. L. Egghe[1] and R. Rousseau[2] (eds.), Elsevier Science Publishers, Amsterdam, (ISBN 0-444-70425-6), 1988, 15–28.

[1]Univ Hasselt, Belgium
[2]Univ Antwerp, Belgium

Abstract

Databases containing bibliometric information on published scientific literature play an important role in the field of quantitative studies of science and in the development and application of Science and Technology indicators. For these purposes, perhaps the most important and probably the most frequently used database is the Science Citation Index, produced by the Institute for Scientific Information. SCISEARCH, the on-line version of the Science Citation Index (SCI), is included in several host computers. However, other databases are used as well, such as Physics Abstracts or Chemical Abstracts. In this contribution, potentialities and limitations of several online databases as sources of bibliometric data in a number of host computers will be discussed. The discussion will focus on the on-line version of the Science Citation Index, and on citation analysis. It will be argued that for several specific bibliometric applications, on-line databases and software implemented in the host computer do not provide appropriate facilities. In fact, for these specific applications, one should first download the primary data from the host into a local computer (PC, Mainframe). Next, dedicated software should be developed on a local level, in order to perform the bibliometric analyses properly. This will be illustrated by presenting a number of applications, related to citation analysis (' impact measurement') and co-citation analysis (' mapping fields of science').

Moed, H.F., Vriens, M.[1](1989). Possible Inaccuracies Occurring in Citation Analysis. *Journal of Information Science* 15, 2, 95–107. Sage Journals

[1]University of Wisconsin, La Crosse, United States

Abstract

Citation analysis of scientific articles constitutes an important tool in quantitative studies of science and technology. Moreover, citation indexes are used frequently in searches for relevant scientific documents. In this article we focus on the issue of reliability of citation analysis. How accurate are citation counts to individual scientific articles? What pitfalls might occur in the process of data collection? To what extent do

'random' or 'systematic' errors affect the results of the citation analysis? We present a detailed analysis of discrepancies between target articles and cited references with respect to author names, publication year, volume number, and starting page number. Our data consist of some 4500 target articles published in five scientific journals, and 25000 citations to these articles. Both target and citation data were obtained from the Science Citation Index, produced by the Institute for Scientific Information. It appears that in many cases a specific error in a citation to a particular target article occurs in more than one citing publication. We present evidence that authors in compiling reference lists, may copy references from reference lists in other articles, and that this may be one of the mechanisms underlying this phenomenon of 'multiple' variations/errors phenomenon of multiple' variations/errors.

Accuracy of citation counts. In: *Citation Analysis in Research Evaluation*. Dordrecht (Netherlands) 2005 From: Moed, H.F. (2005). *Citation Analysis in Research Evaluation*. Dordrecht (Netherlands): Springer. ISBN 1-4020-3713-9, 346, Chapter 13, Pages 173–180

Introduction and Research Questions

Many bibliometric indicators are based on the number of times particular articles are cited in the journals processed for the various ISI Citation Indexes, the Science Citation Index (SCI) being the most prominent. Thus, citation links constitute crucial elements both in scientific literature retrieval and in assessment of research performance or journal impact (Garfield, 1979). The reliability of citation-based indicators strongly depends on the accuracy with which citation links are identified. It is therefore essential to users of citation-based indicators to have detailed insights into the types of problems that emerge and the degree of accuracy that can be achieved in establishing these links. This chapter aims at providing such insights. It builds upon the terminology described in Chap. 6.

The ISI citation indexes, including the SCI and the Web of Science, contain for all documents published in approximately 7,500 journals, full bibliographic data, including their title, all contributing authors and their institutional affiliations, journal title, issue, volume, starting and ending page number. The cited references from source articles are also extracted. These are the publications included in the reference lists at the bottom of a paper. From a cited reference, ISI includes five data fields: the first author, source (e.g., journal, or book) title, publication year, volume number and starting page number.

Generally, the representation of a target document subjected to citation analysis may differ from that regarded as a cited reference. For instance, an author citing a particular target article may indicate an erroneous starting page number or may have misspelled the cited author's name in his or her reference list. The neutral term 'discrepancy' is used to indicate such differences or variations between a target article intentionally cited in a reference and the cited reference itself. A basic problem in

any citation analysis holds: how does one properly match a particular set of target articles to the file of cited references, in order to establish accurate citation links between these targets and the source articles citing them, and how should one deal with discrepancies? This chapter examines the case in which the set of target articles is a set as large as the total collection of source articles processed by ISI during a twenty-year period. In other words, it deals with citation links between ISI source articles, described in Sect. 6.3. The questions addressed in this chapter are: What types of discrepancies between cited references and target articles occur? How frequently do these occur? And what are the consequences of omitting discrepant references in the calculation of citation statistics?

Data and Methods

The Centre for Science and Technology Studies (CWTS) at Leiden University has created a large database of all documents processed during the period 1980–2004 for the CD-ROM version of the SCI and a number of related Citation Indexes on CD-ROM. The database is bibliometric, as it is primarily designed to conduct quantitative, statistical analysis and mapping, and was used in a large series of scientific and commissioned projects conducted during the past 10 years (van Raan, 1996; van Raan, 2004a). The analyses presented below relate to as many as 22 million cited references extracted from all source articles processed in 1999, matched to about 18 million target articles, being the total collection of ISI source articles published during the period 1980–1999.

The methodology applied in this chapter builds upon work described in an earlier paper by Moed and Vriens (1989), and in a paper by Luwel (1999). It focuses on cases showing discrepancies in one data field only. Cited references and target articles were matched in a process involving five match keys, each one based on four out of the five data fields available. In a first round, a match key was applied consisting of the first six characters of the author's family name, his or her first initial, the year of publication, volume number and starting page number. This key can be assumed to be a sufficiently unique characterization of a journal article and will be denoted as 'simple' match key. For reasons of simplicity, cited references matched in this round will be denoted as 'correct'. In a second round, additional match keys were applied, including the journal title, but leaving out the author name, publication year, volume number and starting page number, respectively. Thus, discrepancies in the data field omitted could be analyzed. Cited references matched in this second round will be denoted as 'discrepant'. Discrepancies were reconstructed by finding a 'plausible' explanation for them. Therefore, a classification was designed of 32 types of discrepancies. Discrepancies for which, in the current stage of the work, no plausible explanation could be given, were assigned to a rest category.

Results

Table 13.1 presents the number of matches obtained in applying the various match keys. In the second round, 989,709 discrepant cited references were matched. This number equals 7.7% of the total number of 'correct' references matched in the first round, applying the simple match key. The 32 types of discrepancies were grouped into 11 main types, presented in Table 13.2. Many of the discrepancies showing small variations in a data field can be attributed to inaccurate referencing by the citing authors. However, a substantial part of small variations in author names is not due to inattention or sloppiness, but rather to difficulties in identifying the family name and first names of authors from foreign countries or cultures (Borgman and Siegfried, 1992). A typical example is when Western scientists unfamiliar with Chinese names cites a Chinese author. Moreover, transliteration, i.e. the spelling of author names from one language with characters from the alphabet of another, may easily lead to mismatches. Chapter 3 further discusses problems with author names.

Table 13.1. Matches and discrepancies

Round	*Datafield in which discrepancy occurred*	*No. refs matched*	*Ratio discrepant/ Correct refs (%)*
1	No discrepancy ('correct' reference)	12,887,206	
2	Volume number	207,043	1.6
	Author	272,009	2.1
	Publication year	95,190	0.7
	Starting page number	415,467	3.2
	Total 2nd round	989,709	7.7

Number of ISI source/target articles (1980–1999): approximately 18.4 million. The figure for starting page number includes an estimated 20% of cases in which the cited page number originally contained a character (e.g., p. L115) but was missing in the file used in this analysis.

Table 13.2. Main types of discrepancies

Main type of discrepancy	*N*	*%*
Page number in cited ref missing	165,793	16.7
Small variations in author names	159,503	16.1
Small variations in page numbers	117,683	11.9
Small variations in volume numbers	95,336	9.6
Small variations in publication years	62,837	6.3
Cited page number lies between starting and end page of target	58,853	5.9
Issue number cited rather than volume number	41,369	4.2
Citations to papers by 'consortia'	36,196	3.7
Volume number missing in cited ref (but not in target)	20,323	2.1
Secondary author cited rather than first author	19,281	1.9
Author name in target or cited reference missing	14,754	1.5
Total number of discrepancies explained	791,928	80.0
All other discrepancies in author names	42.275	4.3
All other discrepancies in page numbers	73,138	7.4
All other discrepancies in volume numbers	50,015	5.1
All other discrepancies in publication years	32,353	3.3
Total number of discrepancies not (yet) explained	197,781	20.0
Total number of discrepancies analysed	989,709	100.0

Table 13.2 shows that in the current stage of the work about 80% of the discrepancies could be explained and matched with a very high probability to the intended target. For the remaining 20% of discrepant references no plausible explanation of the discrepancy could yet be given. It is expected that there is a certain percentage of these that was erroneously matched to a target, particularly when they contain discrepancies in two or more datafields. Several types of discrepancies are caused mainly by editorial characteristics of the journals cited, by referencing conventions in particular fields of scholarship, or by data capturing and formatting procedures at ISI, or by a combination of these three factors. This can be illustrated with the following examples.

– When scholars in the field of law cite a paper, they often include in their reference the page number containing the statement(s) they are referring to. Thus, the cited page number is often not the starting page number, but rather a number between starting and end page. There is a striking similarity among reference lists among US law journals in this respect, all showing around 50% of mismatches. Indicating a page number 'in between' also occurs, though less frequently, in references to reviews or data compilations in the natural and life sciences.
– Several journals have dual-volume numbering systems, or publish 'combined' (particularly proceedings) volumes. ISI data capturing procedures do not allow for ranges of numbers in the (source) volume number field, and therefore in a

sense has to choose from several possibilities. Citing authors may make different choices, however, so that volume numbers in cited reference and target article may differ. A similar problem arises with journals of which it is apparently unclear whether the serial numbering system relates to volumes or to issues.

– Journals may publish their articles in a printed and an electronic version, and article identifiers in these versions may differ from one another. Starting and end page numbers may differ, or the electronic version may apply article serial numbers rather than page numbers. Although ISI puts an enormous effort into dealing which such differences, these may hinder proper matching of cited references and target articles, and are expected to become more onerous in the future.
– Particularly in the medical sciences, more and more papers are published presenting outcomes of a joint study conducted by a consortium, task force, survey committee or clinical trial group. Such papers normally do have authors, and ISI includes the first author on the paper in the first author field. However, scientists citing such papers indicate in their reference list mostly the name of the consortium rather than that of the first author. As a result, names in the author fields of target and cited reference do not match. The journal Nature is not the only journal suffering from this type of discrepancy (Anonymous, Nature, 2002).

It is essential to make clear that, due to their systematic nature, the discrepancies between targets and cited references are skewly distributed among target articles. Table 13.3 shows parameters of the distribution of discrepant citations among target articles. Most informative is an analysis by journal, examining the effect of including discrepancies upon its impact factor, and one by country of origin of the target articles receiving discrepant citations (Table 13.4). The journal most affected by ignoring discrepant citations is Clinical Orthopedics and Related Research. The serial numbers attached to this journal are captured by ISI as issue numbers, whereas virtually all cited references to the journal's papers include these numbers in the volume number field. Focusing on the bigger non-Western countries, (former) USSR shows the highest ratio of discrepant/correct citations (21%) followed by China (13%). Among the larger Western countries, Spain and Italy rank top with 7.9 and 7.0%, respectively. USA and Australia show the lowest percentages, 5.7 and 5.3, respectively.

Table 13.3. Distribution of discrepant citations among cited target articles

No. Citations	*Cumm Cited articles (%)*	*Cumm discrepant citations (%)*
1	78.7	51.9
≤ 2	91.3	68.5
≤ 3	95.2	75.9
≤ 10	99.4	91.1
≤ 15	99.7	93.7
≤ 444	100.0	100.0

Table 13.3 demonstrates how the 989,709 references showing a discrepancy are distributed among target articles intentionally cited: 652,419 targets were affected; 78.7% of these received only one discrepant citation, accounting for 51.9% of all cited references showing a discrepancy. About 5% of the targets received at least 4 discrepant citations that account for about 24% of all discrepant citations. About 4,000 targets (0.6%) received more than 10 discrepant citations, accounting for 8.9% of all discrepant citations. The maximum number of discrepant citations to the same target is 444. This is a 'Consortium' paper published by the Diabetes Control Complication Trial (first author Shamoon, H), in New Engl. J. Med, 329 (14) 977–986, (1993).

Table 13.4. Percentile values of the distribution of the ratio discrepant/correct citations among target journals and countries

Percentile	*Ratio discrepant/correct citations (%)*	
	Journals	*Countries*
P10	2.5	5.4
P25	3.4	6.3
P50	4.9	7.8
P75	7.2	9.0
P90	11.6	11.9
P95	18.3	14.2
P99	108.9	41.6

For 2,547 journals (second column) and 99 countries (third column) receiving in 1999 more than 100 'correct' citations to articles published in 1997 and 1998, the ratio was calculated on the number of discrepant and correct citations, expressed as a percentage. The distribution of ratio scores among journals and countries was characterised by their percentile values. The 50th percentile (P50, i.e. the median) is 4.9 for journals and 7.8 for countries. For 127 journals (5%) the ratio discrepant/correct citations exceeds 18.3% (P95), and for 5 countries this ratio exceeds 14.2. For one country it is 41.6%: Vietnam.

Cited References

Borgman, C.L., and Siegfried, S.L. (1992). Getty's Synoname and its cousins: A survey of applications of personal name-matching algorithms. Journal of the American Society for Information Science, 43, 459–476.

Garfield, E. (1979). Citation Indexing. Its theory and application in science, technology and humanities. New York: Wiley.

Lok, C.K.W., Chan, M.T.V., and Martinson, I.M. (2001). Risk factors for citation errors in peer-reviewed nursing journals. Journal of Advanced Nursing, 34, 223–229.

Luwel, M. (1999). Is the Science Citation Index US-biased?. Scientometrics, 46: 549-562.

Moed, H. F., Vriens, M. (1989). Possible inaccuracies occurring in citation analysis. Journal of Information Science, 15: 95-107.

Van Raan, A.F.J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. Scientometrics, 36, 397–420.

Van Raan, A.F.J. (2004a). Measuring Science. In: Moed, H.F., Glänzel, W., and Schmoch, U (2004) (eds.). Handbook of quantitative science and technology research. The use of publication and patent statistics in studies of S&T systems. Dordrecht (the Netherlands): Kluwer Academic Publishers, 19–50.

López-Illescas, C.[1], De Moya-Anegón, F.[2], Moed, H.F. (2008). Coverage and citation impact of oncological journals in the Web of Science and Scopus. *Journal of Informetrics 2*, 304-316. Elsevier

[1]SCImago Research Group, University of Extremadura, Badajoz, Spain

[2]SCImago Research Group, Spanish National Research Council, Madrid, Spain

Abstract

This paper reviews a number of studies comparing Thomson Scientific's Web of Science (WoS) and Elsevier's Scopus. It collates their journal coverage in an important medical subfield: oncology. It is found that all WoS-covered oncological journals (n = 126) are indexed in Scopus, but that Scopus covers many more journals (an additional n = 106). However, the latter group tends to have much lower impact factors than WoS covered journals. Among the top 25% of sources with the highest impact factors in Scopus, 94% is indexed in the WoS, and for the bottom 25% only 6%. In short, in oncology the WoS is a genuine subset of Scopus and tends to cover the best journals from it in terms of citation impact per paper. Although Scopus covers 90% more oncological journals compared to WoS, the average Scopus-based impact factor for journals indexed by both databases is only 2.6% higher than that based on WoS data. Results reflect fundamental differences in coverage policies: the WoS based on Eugene Garfield's concepts of covering a selective set of most frequently used (cited) journals; Scopus with broad coverage, more similar to large disciplinary literature databases. The paper also found that 'classical', WoS-based impact factors strongly correlate with a new, Scopus-based metric, SCImago Journal Rank (SJR), one of a series of new indicators founded on earlier work by Pinski and Narin [Pinski, G., & Narin F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. Information Processing and Management, 12, 297–312] that weight citations according to the prestige of the citing journal (Spearman's rho = 0.93). Four lines of future research are proposed.

Moed, H.F., Bar-Ilan, J[1], Halevi, G[2]. (2016). A new methodology for comparing Google Scholar and Scopus. *Journal of Informetrics*, 10, 533–551. Elsevier

[1]Bar Ilan University, Ramat Gan, Israel

[2]Mount Sinai School of Medicine, NY, USA

Abstract

A new methodology is proposed for comparing Google Scholar (GS) with other citation indexes. It focuses on the coverage and citation impact of sources, indexing

speed, and data quality, including the effect of duplicate citation counts. The method compares GS with Elsevier's Scopus, and is applied to a limited set of articles published in 12 journals from six subject fields, so that its findings cannot be generalized to all journals or fields. The study is exploratory, and hypothesis generating rather than hypothesis-testing. It confirms findings on source coverage and citation impact obtained in earlier studies. The ratio of GS over Scopus citation varies across subject fields between 1.0 and 4.0, while Open Access journals in the sample show higher ratios than their non-OA counterparts. The linear correlation between GS and Scopus citation counts at the article level is high: Pearson's R is in the range of 0.8–0.9. A median Scopus indexing delay of two months compared to GS is largely though not exclusively due to missing cited references in articles in press in Scopus. The effect of double citation counts in GS due to multiple citations with identical or substantially similar meta-data occurs in less than 2% of cases. Pros and cons of article-based and what is termed as concept-based citation indexes are discussed.

Journal citation measures

Journal impact factors and related citation measures are even today probably the most frequently used bibliometric indicators. The articles relate to a critique on existing indicators, proposals for new indicators, and a more reflexive paper addressing criteria for evaluating indicators on the basis of their statistical soundness, theoretical validity, and practical usefulness. Also, one paper examines the effect of the Open Access upon citation impact.

Moed, H.F., van Leeuwen, Th.N[1]. (1995). Improving the accuracy of Institute for Scientific Information's journal impact factors. *J. of the American Society for Information Science (JASIS)* 46, 461–467 Wiley publisher

[1]Leiden University, Leiden, Netherlands

Abstract

The Institute for Scientific Information (ISI) publishes annually listings of impact factors of scientific journals, based upon data extracted from the Science Citation Index (SCI). The impact factor of a journal is defined as the average number of citations given in a specific year to documents published in that journal in the two preceding years, divided by the number of "citable" documents published in that journal in those 2 years. This article presents evidence that for a considerable number of journals the values of the impact factors published in ISI's Journal Citation Reports (JCR) are inaccurate, particularly for several journals having a high impact factor. The inaccuracies are due to an inappropriate definition of citable documents. Document types not defined by ISI as citable (particularly letters and editorials) are actually cited and do contribute to the citation counts of a journal. We present empirical data in order to assess the degree of inaccuracy due to this phenomenon. For several journals

the results are striking. We propose to calculate for a journal impact factors per type of document rather than one single impact factor as given currently in the JCR.

Moed, H.F., van Leeuwen, Th.N.[1], Reedijk, J.[2] (1999). Towards appropriate indicators of journal impact, *Scientometrics* 46, 575–589. Springer

[1, 2] Leiden University, Leiden, Netherlands

Abstract

This paper reviews a range of studies conducted by the authors on indicators reflecting scholarly journal impact. A critical examination of the journal impact data in the Journal Citation Reports (JCR), published by the Institute for Scientific Information (ISI) has shown that the JCR impact factor is inaccurate and biased towards journals revealing a rapid maturing or decline in impact. In addition, it was found that the JCR cited half-life is an inappropriate measure of decline of journal impact. More appropriate impact measures of scholarly journals are proposed. A new classification system is explored, describing both maturing and decline of journal impact as measured through citations. Suggestions for future research are made, analyzing in more detail the distribution of citations among papers in a journal.

Moed, H.F. (2007). The effect of "Open Access" upon citation impact: An analysis of ArXiv's Condensed Matter Section. *Journal of the American Society for Information Science and Technology 58*, 2047–2054. Wiley publisher

Abstract

This article statistically analyses how the citation impact of articles deposited in the Condensed Matter section of the preprint server ArXiv (hosted by Cornell University), and subsequently published in a scientific journal, compares to that of articles in the same journal that were not deposited in that archive. Its principal aim is to further illustrate and roughly estimate the effect of two factors, 'early view' and 'quality bias', upon differences in citation impact between these two sets of papers, using citation data from Thomson Scientific's Web of Science. It presents estimates for a number of journals in the field of condensed matter physics. In order to discriminate between an 'open access' effect and an early view effect, longitudinal citation data was analyzed covering a time period as long as 7 years. Quality bias was measured by calculating ArXiv citation impact differentials at the level of individual authors publishing in a journal, taking into account co-authorship. The analysis provided evidence of a strong quality bias and early view effect. Correcting for these effects, there is in a sample of 6 condensed matter physics journals studied in detail, no sign of a general 'open access advantage' of papers deposited in ArXiv. The study does provide evidence that ArXiv accelerates citation, due to the fact that that ArXiv makes papers earlier available rather than that it makes papers freely available.

Moed, H.F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics 4*, 265–277. Elsevier