Scientometrics (2014) 101:125–158
DOI 10.1007/s11192-014-1423-3

A review of the characteristics of 108 author-level bibliometric indicators

Lorna Wildgaard · Jesper W. Schneider · Birger Larsen

Received: 27 May 2013 / Published online: 6 September 2014

Abstract An increasing demand for bibliometric assessment of individuals has led to a growth of new bibliometric indicators as well as new variants or combinations of established ones. The aim of this review is to contribute with objective facts about the usefulness of bibliometric indicators of the effects of publication activity at the individual level. This paper reviews 108 indicators that can potentially be used to measure performance on individual author-level, and examines the complexity of their calculations in relation to what they are supposed to reflect and ease of end-user application. As such we provide a schematic overview of author-level indicators, where the indicators are broadly categorised into indicators of publication count, indicators that qualify output (on the level of the researcher and journal), indicators of the effect of output (effect as citations, citations normalized to field or the researcher's body of work), indicators that rank the individual's work and indicators of impact over time. Supported by an extensive appendix we present how the indicators are computed, the complexity of the mathematical calculation and demands to data-collection, their advantages and limitations as well as references to surrounding discussion in the bibliometric community. The Appendix supporting this study is available online as supplementary material.

Electronic supplementary material The online version of this article (doi:10.1007/s11192-014-1423-3) contains supplementary material, which is available to authorized users.

L. Wildgaard (✉)
Royal School of Library and Information Science, Birketinget 6, 2300 Copenhagen, Denmark
e-mail: pnm664@iva.ku.dk

J. W. Schneider
Danish Centre for Studies in Research and Research Policy, Department of Political Science and Government, Aarhus University, Bartholins Allé 7, 8000 Aarhus C, Denmark
e-mail: jws@cfa.au.dk

B. Larsen
Aalborg University Copenhagen, A. C. Meyers Vænge 15, 2450 Copenhagen SV, Denmark
e-mail: birger@hum.aau.dk

Keywords Author-level bibliometrics · Research evaluation · Impact factors · Self-assessment · Researcher performance · Indicators · Curriculum vitaes

Introduction

According to Whitley (2000), science operates on an "economy of reputation". Regardless of how scientists and scholars approach their *métier*, they are expected to "cultivate a reputation" and during their career they will successively be assessed individually by committees, e.g. when applying for positions and funding or are nominated for prizes and awards. The pivotal source documenting the accrual of reputation is the curriculum vitae (CV) and perhaps the single most important element in the CV is the section on research publications and thus the researcher's authorship claims. A researcher's reputational status or "symbolic capital" is to a large extent derived from his or her "publication performance". Assessing publication performance is often condensed and summarized by use of a few supposedly "objective" indicators. Especially in the last decade or so, the use of indicators at the individual author-level, for example in CVs, seems to have exploded despite previous warnings from the scientometric community (e.g., Lawrence 2003, 2008; Hirsch 2005). Essentially, there is "individual bibliometrics" before and after the introduction of the Hirsch-index, *h*. After Hirsch (2005), for a time caveats of individual bibliometrics were forgotten and the scientometric community threw themselves into indicator construction especially at the individual level. Recently, the community has returned to a more reflexive discourse where ethical aspects of individual bibliometrics as well as best practices are on the agenda (cf. plenary sessions at the ISSI 2013 and STI 2013 conferences, as well as the topic of one work task in the European ACUMEN research project[1]). In practice, administrators, evaluators and researchers seem to use indicators as never before. Administrators and evaluators for assessment purposes, whereas researchers may add indicators to their CV as a competitive move, in an attempt to show visibility in the academic community as well as the effects of publications (note, for simplicity we use the term end-user in this article to define a non-bibliometrician, who as a consumer of bibliometrics applies indicators to his or her CV).

Today public access to (not always reliable) individual-level indicators such as the *h index* variants is easy through vendors such as Google Scholar or Scopus. Alternatively, such indicators are increasingly being calculated by "amateurs" (i.e., non-bibliometricians, administrators or researchers) using popular tools like *Publish or Perish*.[2] All too often, unfortunately only one indicator is provided and that is usually the most "(in)famous" ones such as the Journal Impact Factor or the *h index*. These are easily accessible and perhaps the only ones many researchers are aware of, but there are many more. Currently, we can count more than one hundred indicators potentially applicable at the individual author-level. The number of indicators seems high given the fact that it is the same few variables that are manipulated though with different algebra and arithmetic. With so many potential indicators and such widespread use, it is important to examine the characteristics of these author-level indicators in order to qualify their use by administrators and evaluators but also researchers themselves. The basic aims of the present article

[1] http://research-acumen.eu/.

[2] http://www.harzing.com/pop.htm.

are to draw attention to the use of multiple indicators which allow users to tell more nuanced stories and at the same time provide a "one stop shop" where end-users can easily learn about the full range of options.

With these aims, it is imperative to examine and compare author-level indicators in relation to what they are supposed to reflect and especially their specific limitations. The usefulness of indicators has been widely discussed through the years. Common themes are disciplinary appropriateness (Batista et al. 2006; Archambault and Larivière 2010; Costas et al. 2010a), the benefits of combining indicators (van Leeuwen et al. 2003; Retzer and Jurasinski 2009; Waltman and van Eck 2009), the construction of novel indicators versus established indicators (Antonakis and Lalive 2008; Wu 2008; Tol 2009; Schreiber et al. 2012), challenges to the validity of indicators as performance is refined through personal and social psychology in recursive behaviour (Dahler-Larsen 2012) and the complexity of socio-epistemological parameters of citations that induces a quality factor (Cronin 1984; Nelhans 2013).

There is to some extent agreement within the scientometric community that performance can only be a proxy of impact and that performance cannot be captured by a single bibliometric indicator. However outside the bibliometric community some indicators are believed to indicate both quality and impact, such as the *h index* (Hirsch 2005) that is commonly added to CVs. The risks of researchers using indicators that condense different aspects of scientific activity in one indicator regardless of disciplinary traits are many, and the debate of the shortcomings of author-level metrics continues (Burnhill and Tubby Hille 1994; Sandström and Sandström 2009; Bach 2011; Wagner et al. 2011; Bornmann and Werner 2012). Also, results of bibliometric assessments have been shown to contribute to both positive and negative culture changes in the publishing activities of individuals (Hicks 2004; 2006; Moed 2008; Haslam and Laham 2009; HEFCE 2009). With this in mind there is a need for indicators to be verified as to whether or not they should be used at the author-level. Depending on the aim of the assessment, a high or low score can affect the individual's chances for receiving funds, equipment, promotion or employment (Bach 2011; HEFCE 2009; Retzer and Jurasinski 2009). As consumers of author-level bibliometrics, researchers can choose the indicators they think best document their scientific performance and will draw the attention of the evaluator to certain achievements. This of course requires knowledge of the advantages and disadvantages of the indicators but also how the many different bibliometric indicators at their disposal are calculated.

Being able to practically calculate the indicator is a major part of communicating the effect of an author's body of work (referred to a as 'portfolio' in the remainder of the article). Complex calculations limit the end-user's choice of bibliometric indicators and hence which effects can be communicated and to what degree of granularity. It is therefore vital when recommending indicators to consider the *usability* of indicators suggested for measuring publications and citations. Bibliometric indicators are based on mathematical foundations that attempt to account for the quantity of publications and the effect they have had on the surrounding community. Effect is traditionally indicated as number of citations or some function hereof. However, the bibliometric indicators proposed or in use are calculated in a large variety of ways. Some of these calculations are simple whereas others are complex and presuppose access to specialised datasets. But the building block of all indicators are paper and citation counts. In addition, some more sophisticated indicators adjust the numbers for variations between fields, number of authors, as well as age or career length. In our analysis we focus, as a novel contribution, on the complexity of the indicators and the consequences for their use by individual researchers. From this point of view we apply a model of complexity to investigate the usefulness of indicators, and to what extent complex calculations limit the usefulness of bibliometric indicators. We argue that the accuracy and completeness of the assessment is limited by the complexity of the applied indicators as a key challenge in recommending bibliometric indicators to end-users. Apart from the actual mathematical foundations, other variables affect the complexity of the calculation of the indicators. For example data access and data collection, including available time and resources, increase the complexity of calculating even simple indicators (Burnhill and Tubby Hille 1994; Ingwersen 2005). Problems with data accessibility, English language bias in citation databases and missing publication and citation data limit the usability of indicators and can directly affect the complexity of the interpretation of the indicator and as such the performance of the researcher (Bach 2011; Rousseau 2006). The goodness of fit of the mathematical model on the bibliometric data relative to end-user profiles within their field, gender and academic position is also important (Alonso et al. 2009; Iglesias and Pecharromán 2007; Wagner et al. 2011). Author-level indicators have been met with a long string of criticisms. The aim of our article is not to passively cultivate this culture of criticism but to actively contribute with objective facts about the usefulness of bibliometric indicators of the effects of publication activity. We are aware of the many caveats but will not discuss them further in this article and focus instead on the issue of complexity. Note also that we limit our study to indicators of the effect of traditional types of publications within the academic community or public sphere, as attempting to review all types of indicators and activities, although needed, is beyond the scope of the present article. Given these aims and caveats, our research questions are:

- *Which author-level bibliometric indicators can be calculated by end-users?*
- *Is it possible to understand what the indicators express?*

The article is structured as follows, the next section provides the background for author-level indicators and the theoretical framework we apply; the subsequent section outlines the methodology of the analysis, including an outline of the analytical framework we use based on Martin and Irvine (1983), and the final two sections contain extensive presentations of and discussions of the analyses and the results.

Methodology

We chose to limit the types of author-level indicators to indicators of the effects of publication activity, resulting in the exclusion of indicators of other important activities such as societal impact, web presence, leadership skills, technical skills, teaching activities, innovation, etc. We included famous indicators that are suggested for use, indicators that are direct adaptations of these known indicators and novel indicators that have been introduced to the bibliometric community but only tested on limited datasets. Novel indicators are included in this review as they are imaginative, attempt to correct for the shortcomings of established indicators and provide alternative ideas to assessment.

Beginning with known works on author-level assessment we identified indicators by exploring the history and development of author-level bibliometrics discussed in Directorate General for Research (2008), Schreiber (2008a), De Bellis (2009), Sandström and Sandström (2009) and Bach (2011). We used citation and reference chasing to find previously unidentified indicators. Supplementary information about the extent the indicators measure what they purport to measure were sourced using the terms (*bibliometri* OR indic*) AND (individual OR micro* OR nano**) in Thomson Reuters Web of

Science® (WoS) and in the Royal School of Library and Information Science's electronic collection of information science journals. Technical papers that analyse the properties of groups of indicators in cluster or factor analyses proved particularly useful. Google Scholar was searched to retrieve for instance national papers, reports, book chapters and other web-based material, such as websites devoted to bibliometric indicators, mediated bibliometric recommendations from ministerial reports, teaching materials and library websites.

Categories of publication indicators

We designed a simple typology of publication and effect indicators that builds on the work of Martin and Irvine (1983). This well-known work recommended thirty years ago a simple model of counting citations and papers to evaluate success and differences in research performance. The simplicity of their model of performance assessment interprets citations as indicators of impact, not quality or importance; presents a range of indicators each focussing on different aspects of research performance and the model clearly illustrates that indicators should be applied to matched research groups, i.e. to compare like with like. We diverge from their model of indicating the performance of research groups, as we extend their model to author-level assessment. We categorize the methods of publication and citation count at the author-level as follows:

1. *Indicators of publication count (output)*: methods of counting scholarly and scientific works published or unpublished depending on the unit of assessment.
2. *Indicators that qualify output as Journal Impact*: impact of a researcher's chosen journals to suggest the potential visibility of the researcher's work in the field in which he/she is active.
3. *Indicators of the effect of output:*
   (a) *Effect as citations:* methods of counting citations, whole or fractional count.
   (b) *Effect of output normalized to publications and field:* Indicators that compare the researcher's citation count to expected performance in their chosen field.
   (c) *Effect of output as citations normalized to publications and portfolio:* Indicators that normalize citations to the researcher's portfolio.
4. *Indicators that rank the publications in an individual portfolio:* indicators of the level and performance of all of the researcher's publications or selected top performing publications. These indicators rank publications by the amount of citations each publication has received and establish a mathematical cut-off point for what is included or excluded in the ranking. They are subdivided into the following:
   (a) *h-dependent indicators*
   (b) *h-independent indicators*
   (c) *h adjusted to field*
   (d) *h adjusted for co-authorship*
5. *Indicators of impact over time:* indicators of the extent a researcher's output continues to be used or the decline in use.
   (a) *Indicators of impact over time normalized to the researcher's portfolio*
   (b) *Indicators of impact over time normalized to field*

The broad categorization of indicators helps us keep the analysis simple and at the same time enables us to identify relationships between the indicators. The indicators identified in the search strategy were grouped according to the aspect of the effect of publication activity that the developers of each specific indicator claim the indicators to measure. As indicators are evolutionary and supplement each other, they cannot in practice be restricted to just one category. Accordingly we agree with Martin and Irvine (1983) that assessment of research performance can be defined in many ways and, particularly in the assessment of publications and citations of individuals, combining indicators from different categories to capture the many different facets of publication activity is recommended.

Judgement of complexity

For each indicator we investigated its intended use, calculation and data requirements. We assume that the end-user has a complete publication list and would only need to find publication data on known documents, citations and calculate the indicator. Each retrieved paper describing the components of indicators was read and the indicators were graded on two aspects of complexity on a 5 point numerical scale namely (1) the availability of citation data and, (2) the intricacy of the mathematical model required to compile the indicator, see Table 1 below. Data requirements were simple to judge, however level of computation proved difficult as mathematical capabilities are individual. Therefore in cases of doubt we calculated the indicator to understand the mathematical foundations and reach consensus about the indicator's level of complexity. All indicators that scored ≤ 3 were calculated to check the complexity score was defendable. As this is a subjective model of scoring complexity, we support our judgements in the extensive appendix that describes the calculations, advantages and disadvantages of each indicator (Online Resource 1). The appendix was our decision tool through-out the judgement process and is published as a supplementary file online.

Our scoring of indicators might result in a set of indicators identified as useful which have lower granularity and sophistication. This represents a balance between, on the one hand, using indicators that are as accurate as possible and measure what they purport to measure, and on the other recommending indicators that not so complex as to deter end-users to use them in practice. The indicators have to measure what they purport to measure of course, however, usability is lost if correct measurement requires data that is not readily available to the end-user, difficult mathematical calculations, and intricate interpretations of complicated data output. We choose to categorise any indicator that scores 4 or above on either of the two complexity criteria as too complex for end-users to apply in practise—and thus not useful.

Results

We identified 108 indicators recommended for use in individual assessment of publication activity and impact. They are presented in tables in the appendix (Online Resource 1) where we briefly describe how each indicator is calculated, provide bibliographic references and discuss what they are designed to indicate, their limitations, advantages, their complexity scores and give comments on their functionality found in related literature. Table 2 below presents an overview of the assessments of complexity, followed by Tables 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and 13 with details about each indicator. Indicators are shown in *italics* in the text.

Table 1 Five point scale used in assessing two aspects of complexity of bibliometric indicators

EASY

Level	Citation Data Collection
1	No citation data needed
2	Individual's citations or ready-to-use journal indicators from structured sources
3	Citing articles, journal, category, field, or world citation data needed, from structured sources
4	Citation data from unstructured sources
5	Citation data not readily available

\+

Level	Calculation of indicator
1	Raw count only
2	Single, simple ratio or linear model
3	Simple, multiple calculation, e.g. repeated simple linear or ratio calculations
4	Advanced multiple calculation, use of weighted parameters gamma or delta that the user must define dependent on the discipline, time interval, velocity or other corrective factors.
5	Advanced multiple calculations and transformation of data

DIFFICULT

Overview of the identified indicators

Out of the 108 indicators we identified as potentially applicable on the level of individual researchers, one third of the indicators are adaptions of the *h index* (35/108). In Table 2 we present the indicator category, the amount of indicators in that category, the number of indicators that scored $\leq$3 in data collection and calculation and in the final column the number of indicators that scored $\geq$4 in either data collection or calculation.

Summary of complexity scores

Overall, our complexity scoring resulted in 79/108 indicators scoring $\leq$3 in both collection of data and calculation, and thus we judged them potentially useful for end-users. The remaining 29 indicators were scored as $\geq$4 in *either* effort to collect citation data or in the calculation itself. Though possibly more accurate and superior measures, these indicators require either special software, e.g. *h index sequences and matrices, hT, co-authorship network analysis*; access to sensitive data, e.g. *knowledge use;* access to restricted data, e.g. *scientific proximity, citations in patents*; no agreement on weighting factors, correcting factors or values of alpha parameters, e.g. *hα, gα, a(t), prediction of article impact;* or advanced multiple calculations, e.g. *hp, hap, DCI, dynamic h, rat h, rat g*. Consequently, these indicators, amongst others, are not considered applicable by an end-user.

The tables in the following analytical summary are limited to the acronym and full name of the indicator; a short description of what it is designed to indicate as defined by the inventor of the indicator, supported with a bibliographic reference and the results of the complexity analysis where **Col** indicates complexity of data collection and **Cal** indicates complexity of data calculation. The indicators that we judged too complex to be useful are highlighted in grey. Primarily indicators that scored $\leq$3 are discussed in the text following each table; however some complex indicators are discussed in categories where no simple indicators were identified. The sections are annotated to help the reader refer back to our categories of publication indicators (see the Methodology section and Table 2).

Table 2 The amount and complexity of indicators in each category

Category		No. of Indicators	Complexity ≤3	Complexity ≥4
1.	Publication Count	15	15	0
2.	Journal Impact	20	16	4
3a.	Effect of output as citations	11	9	2
3b.	Effect of output as citations normalized to publications and field	8	7	1
3c.	Effect of output as citations normalized to publications in the portfolio	6	6	0
4a.	h-dependent indicators	16	10	6
4b.	h-independent indicators	7	4	3
4c.	h-adjusted to field	5	2	3
4d.	h-adjusted for co-authorship	5	2	3
5a.	Impact over time normalized to portfolio	12	8	4
5b.	Impact over time normalized to field	3	0	3
Total		108	79	29

Publication count, category 1

Fifteen indicators of publication count were identified, all with a complexity score ≤2, Table 3. These are simple counting or ratio models that treat contribution to a publication as equally or fractionally distributed across authors.

P is the raw count of publications, while P_{isi}, P_{ts}, count only publications indexed in predetermined sources, which can of course be adapted to any bibliographical database, specific selection of journals or publishers of books. Likewise *weighted publication type* and *patent applications* also account for types of publication judged locally important, showcase specific skills of the researcher or focus on publications deemed as a higher scientific quality relative to the specialty of the researcher. *Dissemination in the public sphere* counts publication and dissemination activities via other channels than academic books or articles. This indicator of publication count is just one of the indicators suggested by Mostert et al. (2010) in their questionnaire tool to measure societal relevance which also includes standardised weighting schemes to accommodate certain activities in the field the researcher is active in. All the aforementioned counting methods assume an equal distribution of contribution across all authors of a publication. The following indicators share the credit for a *publication fractionally* (equal credit allotted to all co-authors), *proportionally* (credit is adjusted to author position on the byline), *geometrically* (twice as much credit is allotted to the ith author as to the $(i + 1)$th author) or *harmonically* (credit is allocated according to authorship rank in the byline of an article and the number of coauthors). *Noblesse oblige* and *FA* prioritize the last and first author in crediting a publication. Correct factional counting should support level of collaboration, not just an integer number symbolizing a share but of course this increases the complexity of the indicator, as data collection would also have to include author declarations. *Co-author* and *co-publication* counts can be extended into analyses of collaboration, networks or even cognitive orientation that identify the frequency a scientist publishes in various fields and if combined with a similar citation study, their visibility and usage. These are, however, outside the scope of this review.

Table 3 Indicators of publication count

Publication count (1)	Designed to indicate	Complexity	
		Col	Cal
P (total publications)	Count of production used in formal communication	1	1
FA (first author counting)	Credit given to first author only	1	1
Weighted publication count	A reliable distinction between different document types	1	1
Patent applications (Okubu 1997)	Innovation	1	1
Dissemination in public sphere (Mostert et al. 2010)	Publications other than scientific and scholarly papers	1	1
Co-publications	Collaboration on departmental, institutional, international or national level and identify networks	1	1
Co-authors	Indicates cooperation and growth of cooperation at inter-and national level	1	1
P (publications in selected databases) e.g. Pisi,	Publications indexed in specific databases, output can be compared to world subfield average	1	2
P_{ts} (publications in selected sources)	Number of publications in selected sources defined important by the researcher's affiliated institution	1	2
Fractional counting on papers	Shared authorship of papers giving less weight to collaborative works than non-collaborative ones	1	2
Proportional or arithmetic counting	Shared authorship of papers, weighting contribution of first author highest and last lowest	1	2
Geometric counting	Assumes that the rank of authors in the by-line accurately reflects their contribution	1	2
Harmonic counting	The 1st author gates twice as much credit as the 2nd, who gets 1.5 more credit than the 3rd, who gets 1.33 more than the 4th etc.	1	2
Noblesse oblige (last author count)	Indicates the importance of the last author for the project behind the paper	1	2
Cognitive orientation	Identifies how frequently a scientist publishes (or is cited) in various fields; indicates visibility/use in the main subfields and peripheral fields	2	1

Table 4 Indicators that ualify output using journal impact factors

Journal impact (2)	Designed to indicate	Complexity	
		Col	Cal
P_{tj} (Rehn et al. 2007)	Performance of articles in journals important to (sub)field or institution	1	2
ISI JIF (SIF) synchronous IF	Average number of citations a publication in a specific journal has received limited to WoS document types and subject fields	2	1
SNIP (Moed 2010; Waltman et al. 2012)	Number of citations given in the present year to publications in the past three years divided by the total number of publications in the past three years normalized to field. Based on Scopus data	2	1
immediacy index	Speed at which an average article in a journal is cited in the year it is published	2	1
AII, aggregate Immediacy Index	How quickly articles in a subject are cited.	2	1
CHL, cited half-life and ACHL, aggregate cited half-life	A benchmark of the age of cited articles in a single journal	2	1
IFmed (Costas et al. 2010a)	Median impact factor of publications	2	2
SJR, Scimago journal rank	Average per article PageRank based on Scopus citation data	2	1
AI, article influence score	Measure of average per-article citation influence of the journal	2	1
NJP, normalised journal position (Bordons and Barrigon 1992; Costas et al. 2010a)	Compares reputation of journals across fields	2	2
DJIF, diachronous IF (Ingwersen et al. 2001)	Reflects actual and development of impact over time of a set of papers	3	2
CPP/FCSm (Costas et al. 2010a)	Impact of individual researchers compared to the world citation average in the subfields in which the researcher is active	3	3
CPP/JCSm	Indicates if the individual's performance is above or below the average citation rate of the journal set	3	3
JCSM/FCSm (Costas et al. 2009; 2010a)	Journal based worldwide average impact mean for an individual researcher compared to average citation score of the subfields	3	3
C/FCSm (van Leeuwen et al. 2003)	Applied impact score of each article/set of articles to the mean field average in which the researcher has published	3	3
Prediction of article impact (Levitt and Thelwall 2011)	Predictor of long term citations	3	4
Co-authorship network analysis (Yan and Ding 2011)	Individual author-impact within related author community	2	5
cf, item oriented field normalized citation score average (Lundberg 2009)	Item orientated field normalised citation score	3	4
%HCP (Costas et al. 2010a)	Percent papers in the 20 % most cited in the field	3	4

Table 5 Indicators of the effect of output as citations

Effect as citations (3a)	Designed to indicate	Complexity	
		Col	Cal
nnc	Number of publications not cited	1	1
Database dependent counting (Scimago Total Cites, WOS, Scopus)	Indication of usage by stakeholders for whole period of analysis in a given citation index	2	1
C + sc (total cites, inc. self-citations)	Indication of all usage for whole period of analysis	2	1
i10 index, Google Scholar metric	The number of publications with at least 10 citations	2	1
C (typically citations in WOS, minus self cites)	Recognised benchmark for analyses. Indication of usage by stakeholders for whole period of analysis	2	2
Sc	Sum of self-citations	2	2
Fractional citation count (Egghe and Rousseau 2008)	Fractional counting on citations removes the dependence of co-authorship	2	2
C-sc (total cites, minus self-cites)	Measure of usage for whole period of analysis	2	2
MaxC	Highest cited paper	2	2
Citations in patents (Okubu 1997)	Citations or use in new innovations	4	1
Knowledge use (Mostert et al. 2010)	Citations in syllabus, schoolbooks, protocols, guidelines, policies and new products	5	1

Qualifying output as journal impact, category 2

Even though journal impact indicators were originally designed as measures of journal or group impact, we have found in the literature that they are applied at an author-level to suggest the visibility of a researcher's work, Table 4. We are aware that many more impact factors are available, and that these are analyzed in detail elsewhere (e.g. Haustein 2012). We therefore only include the journal impact indicators we identified in the literature as used in individual assessment. Thus the recent revisions to the crown indicators, for example the MNCS indicator, are absent from the analysis. Publications in selected journals, $P_{tj,}$.is the only journal impact factor *designed* for use at the author-level; P_{tj} has the advantage that it is entirely independent of subject categories in WoS. It is calculated using journals identified as important for the researcher's field or affiliated institution by the department or university. The journal Impact factors *JIF*, *AII*, *CHL* and *ACHL* are easily available to the end-user through WoS Journal Citation Reports (JCR). *JIF* is the average citation per article, note or review published by the journal over the previous two years calculated using Thompson Reuter's citation data. At the author-level it is commonly used to measure the impact factor of the journals in which a particular person has published articles. *NJP* ranks journals by *JIF* in a JCR subject category. If a journal belongs to more than one category, an average ranking is calculated. The lower the *NJP* for a journal, the higher its impact in the field. Similar to *NJP* is *IFmed*, which is the median value of all journal Impact Factors in the JCR subject category. However, unlike *IFmed*, *NJP* allows for inter-field comparisons as it is a field normalized indicator (Costas et al. 2010a). Misuse in evaluating individuals can occur as there is a wide variation from article to article within a single journal. Hence, it is recommended in JCR to supplement with the *AII*, *CHL* and *ACHL* indicators which indicate how quickly the average article in the journals are cited, i.e. how quickly the researcher's papers are visible in the academic community. An alternative to *JIF* is the *DJIF*, which identifies articles published in

Table 6 Indicators of the effect of output as citations normalized to publications and field

Effect as citations normalized to publications and field (3b)	Designed to indicate	Complexity	
		Col	Cal
Tool to measure societal relevance (Niederkrotenthaler et al. 2011)	Aims at evaluating the level of the effect of the publication, or at the level of its original aim	1	1
Number of significant papers	Gives idea of broad and sustained impact	2	1
Field top % citation reference value	World share of publications above citation threshold for n % most cited for same age, type and field	3	3
E(Ptop) (expected % top publications)	Reference value: expected number of highly cited papers based on the number of papers published by the research unit	3	3
A/E(Ptop) (ratio actual to expected)	Relative contribution to the top 20, 10, 5, 2 or 1 % most frequently cited publications in the world relative to year, field and document type	3	3
IQP, Index of Quality and Productivity (Antonakis and Lalive 2008)	Quality reference value; judges the global number of citations a researcher's work would receive if it were of average quality in its field	3	3
Ptop (percent top publications)	Identify if publications are among the top 20, 10, 5, 1 % most frequently cited papers in subject/subfield/world in a given publication year	3	3
Scientific proximity (Okubu 1997)	Intensity of an industrial or technological activity	5	2

a journal by the researcher in a certain year and the average number of citations received during the 2 or more following years. As a result, *DJIF* reflects the actual development of impact over time of a paper or set of papers. Even though the data collection is more resource demanding, the benefit for the researcher is that it can be calculated for one-off publications, such as books or conference proceedings. *SJR* and *SNIP* (source normalized impact per publication indicator) are journal impact factors based on data from Scopus instead of WoS, and as such include potentially more data on European publications. *SJR* is based on a vector space model of journals co-citation profiles to provide an indication of journal prestige and thematic relation to other journals independent of WoS subject categories. With its longer publication and citation window of three years and the normalization of citations *SNIP* attempts to correct for differences in citation practices between scientific fields.

CPP/FCSm, *JCSm/FCSm* are used together to evaluate individual by Costas et al. (2010a) to indicate the impact profile of individuals. The observed impact of a researcher was indicated by normalizing the *%HCP*, *CPP* and *CPP/FCSm* indicators, while the quality of the journals the individual publishes in was indicated using normalized *IFmed*, *NJP* and *JCSm/FCSm*. As citation rates are increasing and disciplines evolving it is important to normalize the measured impact of researchers to their specialty or discipline. Therefore citations to journals are calculated, as a proxy set for specialty or disciplinary averages using indicators *CPP/JCSm* or *C/FCSm*. Normalization allows for inter-field comparisons (Costas et al. 2010a)

Effect of output, category 3

Effect as citations, 3a

Nine of the 11 identified indicators counting citations were judged useful in assessment, ≤ 3. $C + sc$, and database dependent counting calculate the sum of all citations for the

Table 7 Indicators of the effect of output as citations normalized to publications in the researcher's portfolio

Effect as citations normalized to publications in portfolio (3c)	Designed to indicate	Complexity	
		Col	Cal
%nnc	Percent not cited	1	2
%PNC (percent not cited)	Share of publications never cited after certain time period, excluding self-citations	2	2
CPP (cites per paper)	Trend of how cites evolve over time	2	2
MedianCPP	Trend of how cites evolve over time, accounting for skewed citation pattern	2	2
Age of citations	If a large citation count is due to articles written a long time ago and no longer cited OR articles that continue to be cited	3	2
%SELFCIT	Share of citations to own publications	3	2

period of analysis, while *C*, *C-sc*, adjust the sum for self-citations. Self-citations, *sc*, are relatively simple to collect and calculate but definition can be problematic. *Sc* can be citations by researchers to their own work, but also citations by their co-authors or even affiliated institution. The number of not cited papers, *nnc* is used to illustrate if the citations a researcher has received come from a few highly recognized papers, a stable cited body of work or a group of papers that pull *CPP* in a negative direction. Likewise *MaxC* indicates the most highly cited paper, which can skew indicators based on citation averages but also identify the researcher's most visible paper. Another simple indicator of most visible papers is the *i10 index,* which indicates the amount of papers that have received at least 10 citations each. Just as in *fractional counting of publications*, there are methods to adjust citation count according to the amount of authors to ensure a "fair" distribution of citations, again these assume at the simplest level that authors contribute equally to the paper. Further, they have the benefit of adjusting for the effect of multi-authorship that can in some fields heavily inflate the total amount of citations a researcher receives.

Effect as citations normalized to publications and field, 3b

Identifying the top publications in a field requires the user to design field benchmarks, which is time consuming, or alternatively accept ready-to-use standard field indicators. These standard indicators are based on subject categories in citation indices that may not represent the specialty or nationality of the researcher. Ratio-based indicators account for the amount of citations relative to publications to a fixed field value, *Field Top %*, *E(Ptop)*, *A/E(Ptop), Ptop*.

The 'Index of Quality and Productivity', *IQP*, corrects for academic age, calculates user defined field averages (based on the journals the researcher has published in) and calculates the ratio expected citations to actual citations. This produces indicators of the amount of papers researchers have in their portfolio that perform above the average of the field and how much more they are cited than the average paper. *Number of significant papers* is an indicator on the same theme as *IQP* and uses a field benchmark approach where the number of papers in the top 20 % of the field is considered "significant"; note the caveats for using mechanical significance tests for such decisions (e.g., Schneider 2013, forthcoming). Alternatively a more qualitative approach for identifying *number of significant*

Table 8 Indicators that rank publications in the portfolio, *h*-dependent indicators

h-Dependent indicators (4a)	Designed to indicate	Complexity	
		Col	Cal
h index (Hirsch 2005)	Cumulative achievement	2	2
m index (Bornmann et al. 2008)	Impact of papers in the *h*-core	2	2
e index (Zhang 2009)	Complements the *h index* for the ignored excess citations	2	2
hmx index (Sanderson 2008)	Ranking of the academics using all citation databases together	2	2
Hg index (Alonso et al.2009)	Greater granularity in comparison between researchers with similar *h* and *g* indicators. *The g index* is explained in Table 9	2	3
h^2 index (Kosmulski 2006)	Weights most productive papers but requires a much higher level of citation attraction to be included in index	2	3
A index (Jin 2006; Rousseau 2006)	Describes magnitude of each researcher's hits, where a large a-index implies that some papers have received a large number of citations compared to the rest (Schreiber et al. 2012)	2	3
R index (Jin et al. 2007)	Citation intensity and improves sensitivity and differentiability of *A* index	2	3
$\bar{h}$ index (Miller 2006)	Comprehensive measure of the overall structure of citations to papers	2	3
Q^2 index (Cabrerizoa et al. 2012)	Relates two different dimensions in a researcher's productive core: the number and impact of papers	2	3
Hpd index, h per decade (Kosmulski 2009)	Compare the scientific output of scientists in different ages. Seniority-independent *h* type index	2	4
Hw, citation-weighted h index (Egghe and Rousseau 2008)	Weighted ranking to the citations, accounting for the overall number of *h*-core citations as well as the distribution of the citations in the *h*-core	2	4
hα (Eck and Waltman 2008)	Cumulative achievement, advantageous for selective scientists.	2	4
b-index (Brown 2009)	The effect of self-citations on the *h index* and identify the number of papers in the publication set that belong to the top n % of papers in a field	2	4
hT, tapered h-index (Anderson et al. 2008)	Production and impact index that takes all citations into account, yet the contribution of the *h*-core is not changed	2	5
hrat index, rational h-indicators (Ruane and Tol 2008)	Indicates the distance to a higher *h index* by interpolating between *h* and $h + 1$. $h + 1$ is the maximum amount of cites that could be needed to increment the *h index* one unit (Alonso et al. 2009)	2	5

Table 9 Indicators that rank publications in the portfolio, *h*-independent indicators

h-Independent indicators(4b)	Designed to indicate	Complexity	
		Col	Cal
w index (Wu 2008)	The integrated impact of a researcher's excellent papers	2	2
f index (Tol 2009)	Attempts to give weight/value to citations. *f* is the highest number of articles that received *f* or more citations on average	2	3
g index (Egghe 2006)	The distinction between and order of scientists (Egghe 2006; Harzing 2008)	2	3
t index (Tol 2009)	Attempts to give weight/value to citations. *t* is the highest number of articles that received *t* or more citations on average	2	3
π index (Vinkler 2009)	Production and impact of a researcher is computed by comparing the researcher's citation performance "elite" papers ranked top of his or her field	2	4
Gα (Eck and Waltman 2008)	Based on same ideas as the *g index*, but allows for fractional papers and citations to measure performance at a more precise level	2	4
Rational g-index (g rat) (Schreiber 2008a; Tol 2008)	Indicates the distance to a higher *g index*	2	5

papers is adjusting for seniority, field norm and publication types. However this approach can randomly favour or disfavour researchers. Niederkrotenthaler et al.'s self-assessment *tool to measure societal relevance* attempts to qualify the effect of the publication or its original aim in society by assessment knowledge gain, stakeholders and the researcher's interaction with them. The success of the indicator is dependent on the effort of the researcher to complete the application and assessment forms for the reviewer. It is debateable if this questionnaire is a "bibliometric indicator", but we include it as it attempts to quantify the level of the effect the publication or the original aim has on society by evaluating knowledge gain, awareness, stakeholders, and the researcher's interaction with them.

Effect as citations normalized to publications in portfolio, 3c

The average cites per paper *CPP*, percent self-citations *%SELFCIT* and percent non-cited publications, *%PNC*, are ratio-based indicators which account for the amount of citations relative to the amount of publications in the portfolio. *%PNC* is an indication of articles that have not been cited within a given time frame while *%nnc* is simply the percent papers in the portfolio that have not been cited. The indicator *Age of citations* assesses how up-to-date or "current" a publication is for the academic community by measuring the age of the citations it receives. This indicates if the citation count is due to articles written a long time ago and are no longer cited OR articles that continue to be cited.

The calculation of these indicators is simple, but it is important that the end-user states which citation index the citation count is based on, as a researcher's papers could be uncited in one database but well cited in another dependent on the indexing policy and coverage of the source.

Table 10 Indicators that rank publications in the portfolio, *h*-dependent indicators adjusted to field

h-Adjusted to field (4c)	Designed to indicate	Complexity	
		Col	Cal
n index (Namazi and Fallahzadeh 2010)	Enables comparison of researchers working in different fields by dividing *h* by the highest *h* of journals in the researcher's major field of study to normalize for unequal citations in different fields	2	2
Normalized h-index (Sidiropoulos et al. 2007)	Normalizes *h* to compare researchers' achievements across fields	2	3
h-Index sequences and matrices (Liang 2006)	Singles out significant variations in the citation patterns of individual researchers across different research domains	2	4
hf, generalized h-index (Radicchi et al. 2008)	Allows comparison to peers by correcting individual articles' citation rates for field variation	3	4
x index (Claro and Costa 2011)	Indication of research level. Describes quantity and quality of the productive core and allows for cross-disciplinary comparison with peers	3	4

Table 11 Indicators that rank publications in the portfolio, *h*-dependent indicators adjusted for co-authorship

h-Adjusted for co-authorship (4d)	Designed to indicate	Complexity	
		Col	Cal
Alternative h index (also hi) (Batista et al. 2006)	Indicates the number of papers a researcher would have written along his/her career if they had worked alone	2	2
POP variation individual h index (Harzing 2008)	Accounts for co-authorship effects	2	3
Hp, pure h index (Wan et al. 2007)	Corrects individual *h*-scores for number of co-authors	2	4
h_m-index (Schreiber 2008b)	Softens influence of authors in multi-authored papers	2	4
h_{ap}, adapted pure h index (Chai et al. 2008)	Finer granularity of individual *h*-scores for number of co-authors by using a new *h*-core	2	5

Indicators that rank the publications in the researcher's portfolio, category 4

It is interesting to assess if the publications in the portfolio contain a core of high impact publications. This is done by ranking work within the portfolio by the amount of times cited to create cumulative indicators of a researcher's production and citations. The most commonly used of these is Hirsch's *h index* (Hirsch 2005) which has been corrected and developed since its creation.

h-Dependent indicators, 4a

Ten of the sixteen *h*-dependent indicators scored ≤3 in complexity of calculation and data collection: *h*, *m*, *e*, *hmx*, *Hg*, h^2, *a*, *r*, $\hbar$, Q^2. As these are dependent on the calculation of *h index*, they suffer from the same inadequacies as *h*. The advantages and disadvantages of *h* are explained in detail in i.a. (Costas and Bordons 2007; Alonso et al. 2009; Schreiber

Table 12 Indicators of impact over time normalized to the researcher's portfolio

Impact over time normalized to portfolio (5a)	Designed to indicate	Complexity	
		Col	Cal
AR index (Jin et al. 2007)	Accounts for citation intensity and the age of publications in the core	2	2
m quotient (Hirsch 2005)	*h* type index, accounting for length of scientific career	2	2
mg quotient	*g* type index, accounting for length of scientific career	2	3
AWCR, age-weighted citation rate, AW and per-author AWCR (Harzing 2012)	*AWCR* measures the number of citations to an entire body of work, adjusted for the age, *AW* is the square root of *AWCR* to appropriate *h,* and *per-author AWCR* adjusts *AWCR* for the number of authors of each individual paper	2	3
PI, Price Index (Price 1970)	Percentage references to documents, not older than 5 years, at the time of publication of the citing sources	3	2
c(t), citation age (Egghe et al. 2000)	The age of citations referring to a researcher's work	3	3
h^c, contemporary h-index (Sidiropoulos et al. 2007)	Currency of articles in *h*-core	2	4
h^t, trend h index (Sidiropoulos et al. 2007)	Age of article and age of citation	3	4
Dynamic h-type index (Rousseau and Ye 2008)	Accounts for the size and contents of the *h*-core, the number of citations received and the *h*-velocity	3	4
DCI, discounted cumulated impact (Ahlgren and Järvelin 2010; Järvelin and Person 2008)	Devalues old citations in a smooth and parameterizable way and weighs the citations by the citation weight of the citing publication to indicate currency of a set of publications	3	5

et al. 2012). *A*, $\hbar$, *m* are recommended for comparison across field or seniority. The indicators have subtle differences in their adaptions of the *h index* and which sub-set of publications from a researcher's portfolio is used. *h* ranks publications in descending order by number of citations. *h* is defined where the rank and number of citations are the same or higher. The publications that are ranked equal or higher than *h* are called the *h*-core and regarded as the productive articles. Roughly proportional to *h* is $\hbar$, which is the square root of half of the total number of citations to all publications.

R, *hg*, h^2, *e*, Q^2 and *m* adjust for the effects or discounting of highly cited papers in the calculation of *h*; *e* calculates excess citations of articles in the *h*-core, *A* is the average number of citations to the *h*-core articles whereas *m* is the median number of citations; *R* is the square root of *A*, *hg* is the square root of the sum of *h* multiplied by the *g* index while h^2 is proportional to the cube root of citations; Q^2 is the square root of the sum of the geometric mean of the *h index* multiplied by the median number of citations to papers in the *h*-core. As such Q^2 claims to provide a balanced indication of the number and impact of papers in the *h*-core. Finally, *hmx* simply recommends the researcher refer to their *h index* scores measured across Google Scholar, WOS and Scopus on their CVs.

h-Independent indicators, 4b

Six *h*-independent indicators of cumulative impact were identified, 4 scored a complexity rating of $\leq$3: The Wu index *w*, *f index*, *g index* and the *t index*. *w* is a simple indicator of prestige, tested in physics and recently economics, that states for example a researcher has a *w index* of 1 if 10 of their publications are cited 10 or more times, but they have not achieved a *w index* of 2 because that implies that 20 of their publications have to have been cited 20 or more times. Wu suggests that *w*1 or 2 is someone who has learned the rudiments of a subject; 3 or 4 is someone who mastered the art of scientific activity, while "outstanding individuals" have a *w index* of 10. The *g index* on the other hand is introduced by Egghe (2006) as an improvement of *h*, as it inherits all the good properties of *h* and takes into account the citation scores of the top articles. *g* claims to provide a better distinction between scientists than *h* as it weights highly cited papers to make subsequent citations to these highly cited papers count in the calculation of the index, whereas with *h* once a paper is in the *h*-core, the number of citations it receives is disregarded. Like *h, g* ranks publications after citations in descending order but *g* takes the cumulative sum of the citations and the square root of the sum for each publication. *g* is where the rank and the square root is the same or higher. As such *g* is based on the arithmetic average and ignores the distribution of citations (Costas and Bordons 2007; Alonso et al. 2009), meaning a researcher can have a large number of unremarkable papers and have a large *g index*. Alternative ways to estimate the central tendency of the skewed distribution of citations to core papers are the *f* and *t* indices. These are based on the calculation of the harmonic and the geometric mean and as such suggested as more appropriate average measures for situations where extreme outliers exist, i.e. the few very highly cited papers. Papers are again ranked in descending order of citations, and beginning with the highest cited paper, the harmonic or geometric mean is calculated stepwise until the product is equal or higher than the rank.

h-Adjusted to field, 4c

The indicators in this category claim to adjust different publication and citation habits in different fields and as such present indicators useful for comparing scientists. *Normalized h*

Table 13 Indicators of impact over time normalized to field

Impact over time normalized to field (5b)	Designed to indicate	Complexity	
		Col	Cal
Classification of durability (Costas et al. 2010a, b, 2011)	Durability of scientific literature on the distribution of citations over time between different fields	2	4
a(t), aging rate (Egghe et al. 2000)	Aging rate of a publication	3	4
Age and productivity (Costas et al. 2010a)	Effects of academic age on productivity and impact	3	4

is recommended as an adjustment to *h*. It is calculated as the *h index* divided by the number of articles not included in the *h*-core. Meanwhile the *n index* is the researcher's *h index* divided by the highest *h index* of the journals of his/her major field of study. The *n index* is a theoretical index still awaiting validation, and has only been tested using the Scopus definition of *h* and SCImago Journal and Country Rank website for the journal information.

h-Corrected for co-authorship, 4d

All the six indicators in this category require calculation of the *h*-index in their mathematical foundations. The *alternative h index* or *hi* as it is also known divides *h* by the mean number of authors for each paper in the *h*-core, while *POPh* divides the number of citations by the number of authors for each paper and then calculates *h* using this normalized citation count. Both models give an approximation of the impact authors would have if they had worked alone, however these models treat citations and publications as a single unit that can be evenly distributed.

Impact over time, category 5

Indicators of impact over time indicate the extent a researcher's work continues to be used or the decline in use. Twelve indicators were identified, six potentially useful, complexity $\leq$3. Ten indicators were designed to indicate impact over time relative to the portfolio and two allow comparison to the expected aging rate of the field.

Impact over time normalized to portfolio, 5a

Eight indicators were identified, $\leq$3: The age-weighted citation rates (*AWCR*, *AW* and *per-author AWCR*), *AR index*, *m quotient*, *mg quotient*, *Price Index* and *citation age*, *c(t)*. Of these the age-weighted citation rates (*AWCR*, *AW* and *per-author AWCR*), *c(t)*, *m quotient*, *mg quotient* and *Price Index* are ratio-based models. *AR* is based on the square root of average number of citations per year of articles included in the *h*-core and like the *m quotient* is also *h*-dependent. *m quotient* is the *h index* divided by the length of the researcher's publishing career, which is defined as the number of years since the first publication indexed in the database used to calculate the *h index* to the present year. Similarly, *mg* is the *g index* divided by length of the researcher's publishing career.

Inspired by the *AR index* the *AWCR* calculates the sum of citations to an entire body of work, by counting the amount of citations a paper has received and dividing by the age of that paper. The *AW index* is the square root of the *AWCR* to allow comparison with the *h index*, whereas the *per-author age-weighted citation rate* is similar to *AWCR*, but is normalized to the number of authors for each paper. The *Price Index* is the number of citations less than 5 years old from the time the paper was published divided by the total number of citations, multiplied by 100. *c(t)* also indicates the age of citations referring to a researcher's work. A corrective factor is required if citation rates are to be adjusted for changes in the size of the citing population or discipline.

Impact over time normalized to field, 5b

Indicators of impact over time adjusted to field are sophisticated indicators, all of which we judged as too complex to be useful to the end-user. *The Classification of Durability* is a percentile based indication of the distribution of citations a document receives each year, adjusted for field and document type. It can detect the possible effects durability can have on the performance measurement of an individual. However, at the present time its analysis is limited to WoS. *a(t)* corrects the observed citation distribution for growth, once the growth distribution is known. Costas et al. (2010a) propose combining indicators to produce classificatory performance benchmarks. Their indicator *Age and Productivity* combines the mean number of documents by age and cites per paper (using a three year citation window) in four year age brackets adjusted to field to identify the age at which scientist produce their best research and to some extent the decline in their knowledge production. But the demanding data collection, multiple indicators and dependence on WoS journal categories make it unlikely that an end-user will take the time needed to calculate the indicator.

Discussion

Our initial analysis of the indicators highlighted two problems: (1) The availability and accessibility of publication and citation data does not support the practical application by end-users of the indicators. Many indicators are invented for ideal situations where complete datasets are available and do not cater to real life applications. (2) Some indicators lack appropriate validation and recognition by both the bibliometric and academic community. Judging by the quantity and availability of the indicators we identified, it is obvious that end-user bibliometric assessment has the potential to go beyond the *h index* and *JIF*. This paper has only focused on the effects of publications, which is a small area of scientific activity. Still, even for this one activity, sub-dividing indicators of "effects of publications" into different aspects illustrates how essential it is to recommend groups of indicators to end-users rather than single indicators. Presenting indicators in categories is a way to demonstrate how different aspects of performance can be captured, as each indicator has its own strengths and weaknesses as well as "researcher/field" variables that can be redundant or counter-productive when indicators are used together. Even though our schematic presentation simplifies understanding what the indicators do, recommending useful indicators is still a challenge. The benefit of choosing one over the other is highly dependent on the spread of the end-user's publication and citation data, the academic age of the end-user and the availability of the data. In the following we discuss some of the main issues for each category.

Publication count, category 1

Indicators of publication count provide information of the sum of a researcher's publications produced within a given timeframe, Table 3. We judged all 15 indicators useful for end-user application however they are some limitations that users of these count-based indicators need to be aware of. Count alone provides a distorted picture of the scope of a researcher's output and divulges nothing about the level of contribution to a work unless authorship credit is explicitly stated (Hagen 2010). In the assessment of contribution, validation is required from all authors of actual contribution to a paper, as name order in the by-line can be strategically or politically motivated or simply alphabetical (Bennett and Taylor 2003). If it is normal for the discipline to have many authors per paper rather than single authored papers, correcting for single author contribution is superfluous and perhaps counterproductive. Count can be balanced by weighting different forms of publication, be it patents, books, book chapters, articles, enlightenment literature, conference papers etc., after importance for the field in which the researcher is active. Though which document types and how they are weighted needs to be clear. The value given to a specific type of publication varies from discipline to discipline but on an individual level could be weighted in relation to the mission and resources of the researcher's affiliated institute. Weighting output types should, however, be used with caution as the positive or negative effect this has on publishing behaviour needs further investigation. Also, weighting can make the comparison to normalised national and international standards unreliable as document type has to be compared with the exact same document type, which can result in the preference of some forms of publications to the detriment of others in the computation of the standards.

Qualifying output as journal impact, category 2

Journal and article impact indicators are frequently added by end-users to CVs next to publications as a proxy for the level of quality of a published paper. In assessments for jobs or tenure they are used as a selection parameter to judge applicants' publications and as benchmarks for expected disciplinary performance. They were originally designed to indicate the average impact of articles published in journals in a defined publishing year and with a short citation window or to aggregate the publications of a research group or center. The journal-based citation indicators in Table 4 are dependent on journal performance and have been shown to measure popularity and not prestige. Popularity is not considered a core notion of impact (Bollen et al. 2006; Bollen and van de Sompel 2008; Yan and Ding 2011). As such they are dependent on the disciplinary characterisation of publications and citations, journal aggregation in sub-disciplines in citation databases, the methodology used to estimate citations and the type of papers included (excluded) in the calculation. However, where the individual publishes is considered an important criterion in the assessment of visibility and impact. Yet, as Table 4 illustrates, the construction of the impact factors means they are an indication of researchers publishing success and not the actual use of their articles. In the investigation of the use of journal impact factors at the author-level it is necessary to study if time and impact of journals correlate in the same way in the assessment of individual impact. Our results identified only one indicator of impact designed for assessment at the author-level and simple enough for the researcher to use; P_{tj} (articles published in journals deemed relevant or prestigious by heads of department or institution). *Ptj*, can of course be extended to encompass other types of publications, as to support non-journal based fields, and can also be extended to source deemed authoritative by other that heads of department.

Effect of output, category 3

Indicators of the effect of output can be grouped into three types of aggregation: number of citations, averages or percentiles. Calculations in all approaches are relatively simple but in practice the availability of data makes the feasibility of the end-user using these indicators to produce reliable indications of the true effect of the publications questionable. As field coverage is limited in citation databases, citation indicators are more appropriate in some fields than others. Ideally citation indicators require data collection in multiple sources to provide as complete a picture of "use" as possible, however, the overlap between citation databases requires the end-user to filter out duplicate citations. This immediately adds to the complexity of the indicators.

Effect as citations, 3a

Citations are counted as the amount of times a paper has been used in other published academic papers. For our recommended 9 indicators, the count is limited to citation databases that index citations, e.g. WoS, Scopus, Google Scholar and the resulting count can differ from database to database. Further the count can differ between versions of the same database, dependent on the subscription of the end-user. Count does not reveal if the citations have been positive or negative, the currency of the citations or if the count is due to older articles having more time to accumulate citations. Citations can be interpreted however as the contribution of research to the social, economic and cultural capital of academic society and/or an indication the interaction between stakeholders, how new approaches to science are stimulated, and influence on informing academic debate and policy making (Directorate General for Research 2008; Bornmann 2012). Consequently, a high citation count is desired and the indicator *MaxC* is a proxy for the researcher's most prestigious paper. Likewise the *i10 index* indicates substantial papers. Putting the arbitrariness of a minimum of 10 citations as a cut-off point to one side, the index is based on a Google Scholar, whose database content is not transparent and suggested vulnerable to content spam and citation spam (Jacsó 2011; Delgado Lopéz–Cózar et al. 2014). To understand if the citations are to a few papers out of the end-users entire portfolio, the *nnc* indicator counts the number of papers that have not received any citations. *nnc* is not an indicator of low quality work, but is a useful indicator that helps interpretation of other performance indicators that build on average citations per paper. Publications can be greatly used and of great influence, but never cited. Certain types of publications are important but rarely cited such as technical reports or practice guidelines. Lack of citations can be caused by restricted access to sources, fashionableness of the topic, changes in size of citing or citable population and the citability of different types of publications (Egghe et al. 2000; Archambault and Larivière 2010; Costas et al. 2010b).

Scientists build on previous findings, so self-citation, *sc*, is unavoidable. Excessive self-citations inflate citation count and are considered vanity and self-advertising. In assessment self-citations can affect the reliability and validity of citation count on small amounts of data (Glänzel 2006; Costas and Bordons 2007). However, there is no consensus in the bibliometric community if removing self-citations has any effect on robust indicators or if the removal process can introduce more noise in the citation count than is removed (Harzing 2012). Most citation indexes have the option to remove self-citations but what constitutes a self-citation is undefined, as they can be understood as citations by the researcher to own work, citations from co-authors of the paper or citations from a colleague in the same research group. Alternative indications of the importance of scholarship share the citations between researchers that have contributed to a paper. *Fractional citation count*, i.e. averages - geometric, harmonic and arithmetic - are affected by the skewed distribution of citation data which is why there is a movement in the literature towards the stability and consistency of percentiles (Belter 2012). Consequently, it has been recommended not just to compare results obtained from several databases, but combine citation counts with other methods of performance assessment and only then normalise results of individual performance to academic seniority, active years and field to ascertain excellence (Costas et al. 2010a).

Effect as citations normalized to publications and field, 3b

Percentiles such as *E(Ptop)*, *A/E(Ptop)*, *Ptop* are considered as the most suitable method of judging citation counts normalized in terms of subject, document type and publication year as they attempt to stabilise factors that influence citation rates (Bornmann and Werner 2012). Bornmann argues for their simplicity of calculation, which is debateable, but they are more intuitive to the end-user than average cites-per-paper in that visualization of results in box-charts or bar-charts can provide easy-to-read presentations of performance. Percentages have the further advantage that they are only affected by the skewed distribution of citation data to a limited extent and are adjustable to individual assessments as measures of excellence. *Ptop*, for example, can be adjusted to *Ptop/author* to illustrate the amount of papers a scientist has within the top 5 % papers within a field, and as such indicate excellence (van Leeuwen et al. 2003). Field indicators, *Field Top %*, favour some fields more than others; older articles, senior scientists with extensive publishing careers and are often based in predefined journal-subject categories in citation databases. The degree to which top n % publications are over or under-represented differs across fields and over time (Waltman and Schreiber 2013). Likewise the indicator *Significant Papers* adjusts the number of citations that are considered significant for seniority, field norm and publication types, which results in a subjective indicator that can randomly favour or disfavour researchers. Data-completeness, differences in citation rates between research fields, and the need for a sufficiently large publication output to obtain a useful percentage benchmark at the author-level compromise the simplicity and stability of these comparative measures of excellence. Hence they may not be representative of the response to a researcher's work, but they can prevent a single, highly cited publication receiving an excessively heavy weighting.

To interpret individual researcher impact, end-users compare themselves with peers to understand the level of their performance, however using field normalization to cater for different publication and citation traditions is not without its difficulties. It means that up-to-date reference standards for the field have to be available to the end-user. Reference standards fix the field by calculating normalizing factors using multiplicative correction and other parameters (Iglesias and Pecharromán 2007). Studies have shown that normalized indicators characterise the area but can be disadvantageous for the specific publication patterns of researchers within their sub-field specialty (Ingwersen et al. 2001; van Leeuwen and Moed 2002; Bollen et al. 2006; Yan and Ding 2011). Further, normalization favours highly cited researchers as impact increases in a power law relationship to the number of published papers (Iglesias and Pecharromán 2007) and assume that publication and citations are independent variables. In other words the effect of the publishing size on the citation count has been eliminated. Using journal subject categories is an accessible way to define fields, but it is doubtful if researchers can feasibly indicate their global impact using journal impact-defined field indicators as these are normalized to a field that neither accurately represents the specialty of the researcher or individual researcher demographics, such as seniority or academic age. Antonakis et al. (2008) propose the *IQP* indicator to enable researchers to compare the performance of their papers to other papers within their specialty. *IQP* produces descriptive indicators of the global number of citations a researcher's work would receive if it were of average quality in its specialty, by calculating the ratio actual citations to estimated citations using the journals the researcher publishes in as a proxy for "specialty". As a result the end-user can indicate the number of papers (corrected for subject and academic age) which perform above the expected average for the specialty and how much better than average these papers perform. Acknowledging how time-consuming the indicator can be to calculate for the end-user, they provide a free online calculator and benchmarks for interpretation.[3] The indicator has in tests correlated better with expert ratings of excellence than the *h index*. The indicator is again dependent on subject categories and citation record in WoS which makes the *IQP* more useful only to researchers well represented in WoS.

High scientific quality is not necessarily related to high citation count, but perhaps most important for assessment is to acknowledge that the true impact of a piece of research can take many years to become apparent and the routes through which research can effect behaviour or inform social policy are diffuse. Therefore we include in our analysis Neiderkrotenhaler et al.'s questionnaire tool (2011). The tool indicates the broader impact of publications by combining the interest of societal stakeholders with quantitative indicators of knowledge dissemination and use. It assesses the effect of the publication in non-scientific areas, the motivation behind the publication and efforts by researchers to disseminate their findings.

Effect as citations normalized to publications in portfolio, 3c

The average (mean) cites per paper *CPP,* or *medianCPP* are robust measures for comparisons of researchers to field averages or comparisons between researchers who have been active for different numbers of years. The mean and median are different measures of the central tendency in a set of data, or the tendency of the numbers to cluster around a particular value. In bibliometrics it is desirable to find the value that is most typical. One way of doing this is to find the mean, or average, which is the sum of all the citations divided by the total number of publications. Another way is to find the median, or middle, value, which is the one in the centre of an ordered list of publications ranked after the amount of citations they have received. The average has the disadvantage of being affected by any single citation being too high or too low compared to the rest of the sample. *CPP* seems to reward low productivity. This is why *medianCPP* is taken as a better measure of a mid-point or percentiles are preferred. Percent self-citations *%SELFCIT,* percent non-cited publications after a certain time *%PNC* and percent not cited over all publications *%nnc* are ratio-based indicators which account for the lack of citations or lack of external citations relative to the amount of publications in the portfolio.

The currency of publications can be analyzed by using *Age of citations*. This indicator predicts the useful life of documents over a period of time. Moreover, it helps end-users select the significant (most used) papers and understand how their papers are used—if older papers have first come of age recently and are accumulating citations, if their papers have a short "shelf life" or if they are constantly used.

[3] The IQP calculator can be downloaded from: http://tinyurl.com/nj7s834.

Indicators that rank the publications in the researcher's portfolio, category 4

The indicators in the following categories are an expression of cumulative impact in a single index, as they calculate the quantity and impact of articles into an indication of prestige (Hirsch 2005; Schreiber et al. 2012). To do this comprehensively, the majority are recommended, by their creators, to be combined with other indicators. When used alone the indicators give only a rough measure of quality as the correlation between output, quality and impact remains uncertain (Nederhof and Meijer 1995; Haslam and Laham 2009). To overcome these shortcomings, "quality" is assumed a value of citation count, as a large number of citations are interpreted as "usefulness" to a large number of people or in a large number of experiments. Our results show that attempts to improve *h* can be at the cost of simplicity and usability, Tables 8, 10, 11 and 12. The descendants of *h* are supposedly more precise, yet in many cases their consistency and validity remains problematic. Some have performed well in laboratory studies: *b* (Brown 2009), *IQP* (Antonakis and Lalive 2008), *h index sequences and matrices* (Liang 2006), while others have faltered: *h*, *g*, *r*, h^2 (Waltman and van Eck 2009). Of course the indicators that incorporate *h* in their foundations suffer from the same inconsistencies as *h*: *hg*, Q^2, *normalized h*, *hrat*, *grat*, *a*, *hw*, $\hbar$, *e*, *hpd* and *hmx*. Others give undue weight to highly cited papers, *h*, *f*, *t*, *w*, h^2 (Schreiber 2010) and although some of the sampled the indicators proclaim higher accuracy and granularity, these benefits are lost on the end-user as the complexity of the calcuations mean usability and transparency are reduced. $h\alpha$, *Hpd*, *hw*, *hrat* require multiple and advanced calculations, while *hT* requires special software for computation.

h-Dependent indicators, 4a

The *h index* already plays an important role for end-users (Costas and Bordons 2007) and despite its flaws, is unavoidable as its simplicity and recognisability outweigh debates of its representativeness. Generally, *h*-type indicators are estimated as stable once a scientist has reached a certain level of scientific maturity, >50 papers, otherwise stability issues can lead to misleading results. The exponential growth of the number of papers advocating the advantages and hazards of the *h index* makes it impossible to present a complete reference list. Briefly, the *h index* has been criticised for negatively influencing publication behaviour (Egghe 2006; Harzing 2008), reducing validity in cross-domain comparison and bias towards certain fields (Podlubny 2005; Iglesias and Pecharromán 2007), having granularity issues (Vanclay 2007; Harzing 2008), losing citation information (Waltman and van Eck 2011), under-estimating the achievement of scientists with selective publication strategies, women and researchers who have had taken a break from academia, as well as favouring seniority (Costas and Bordons 2007). Perhaps, most importantly, is the questionable arbitrariness of the *h* parameter (Alonso et al. 2009). Subsequently, the indicators that build on the *h index* suffer the same inadequacies as *h*. All of these criticisms must be known outside of the bibliometric community to produce *informed* end-user assessment.

To compensate for limitations of single indicators, we recommend combining *h*-type indicators, however information redundancy is an issue, as investigated in Panaretos and Malesios (2009) and Bornmann et al. (2011). Their investigations reveal high inter-correlations between the *h*-type indicators and they conclude that the various indicators can be redundant in empirical application. Separating the indicators into categories "fundamental" and "derived" reduces the chance of information redundancy in assessments (Zhang 2009) where, for example, *A* and *R*, are *h*-dependent (derived) and thus have information redundancy with *h*. Both Bornmann et al. (2008) and Schreiber et al. (2012) recommend a more user-friendly approach, which is to categorize and combine pairs of indicators relating to the productive core. Using our identified indicators we recommend combining one of the following indicators of the productive core: h, m, Q^2, h^2 or $\hbar$, with an indicator relating to the impact of papers A, R, AR, m or e to produce insightful results.

h-Independent indicators, 4b

The *g index* is based on the arithmetic average which means it ignores the distribution of citations, Table 9 (Costas and Bordons 2007; Alonso et al. 2009). However the arithmetic mean has the disadvantage that it is disproportionate to the average publication rate meaning that the *g*-index of a scientist with one big hit paper and a mediocre core of papers could grow in a lot comparison with scientists with a higher average of citations (Tol 2009). Attempts to improve the desirable mathematic properties of the *g index* are the *f* and *t* indicators that use the harmonic and geometric mean. These claim to improve discrimination between similar scientists as *f* weights the distribution of citations and *t* is even less effected by highly-cited papers than *f*. Yet in the broad ranking of researchers calculating the *g*, *h*, *f* and *t* indicators, adds more work but no greater insight in a researcher's performance (Tol 2009). h is always $\leq f \leq t \leq g$, similarly, Glänzel and Schubert (2010) suggest using the median of the citations within the core, the *m index*, and show the *m index* and the *f index* to be less affected by outliers than the other measures. *m* is simpler to calculate than *f* and *t* and results in the same or very similar index number. The $G\alpha$ and the *rational g index* allow for fractional papers and citations to measure performance on a more precise level, however they require setting a value of α and interpolating between g and $g + 1$ based on the piecewise linearly interpolated citation curve. Consequently, we scored them too complicated for the end-user to use.

Completely independent of the construction issues of *h* and *g* is the *w index* (Wu 2008). The *w index* is a useful and simple way to assess the integrated impact of a researcher's work, especially the most excellent papers.[4]

h-Adjusted to field, 4c

Field variation creates obstacles to fair assessment of scientific performance. The *n index* and *normalized h index* have been specifically designed for across field comparison and account for the multidisciplinary of researchers, Table 10. Even though these are simple to calculate, they have some severe limitations. The *n index* divides *h* by the highest *h index* of the major journal the researcher publishes in. In many cases, the *h index* will be based on articles in different areas of science and can have no relation to the highest *h index* of the journals of his major field of study, making the calculation impossible. The *normalized h* can only be used in parallel to the *h index* and rewards less productive but highly cited researchers. Other alternatives are the *x*, *hf*, *h index sequences and matrices* indicators but these require advanced multiple calculations, special software and the determination of cut-off values, parameters or stretching the exponential distribution to fit the dataset or field characteristics. These approaches increase confusion over which data is included in the calculation and how it is calculated. If information is lost during the data manipulation the validity is challenged.

A simple option suggested by (Arencibia-Jorge et al. 2008) is to combine *h* type indicators, *h, g* and *A,* to establish quality benchmarks at a lower level of aggregation than

[4] The calculation is explained above Table 9.

the field. They suggest computing successive *h* type indices to account for performance on a "researcher:department:institution" hierarchy. The ranking of researchers at these three levels allows the evaluation at the micro-level, identifies researchers with higher than expected impact as well as aggregated departmental and staff behaviour within the institution and international visibility. Although their solution is interesting, the complexity of data collection increases with the hierarchy and as the indicator becomes a tool for institutional evaluation rather than author-level performance we have not included it in our list of recommended indicators.

h-Corrected for co-authorship, 4d

Assessment of co-authorship is important for the individual researcher in assessment because research collaboration lies at the heart of expressing research activity, knowledge advancement and communication. Simple indicators of *h* adjusted for co-authorship shouldn't be difficult to calculate because the researcher should have all the necessary information - who wrote the articles and their affiliation during publication; homonyms of author and institute names; and the relation between authorship order and contribution. Normalising the *h index* for multi-authorship (*hi*, *POP variation*, *n*, *hm*, *alternative h*, *pure h*, and *adapted pure h*), immediately affects the simplicity of the calculation of *h*. *h* indicators adjusted for co-authorship are calculated in two ways: if the citation count is normalized to the amount of co-authors before or after the h-calculation. For instance, increasing the numbers of papers in the *h*-core affects the precision of the indicator, as in *hm*, while reducing the amount of papers in the *h*-core, *hi*, makes the results sensitive to extreme values and discourages collaborations that can result in multi-authored, highly cited and influential papers.

It is unclear which indicator is best. Egghe et al. (2000) argue that one particular method of adjusting for co-authorship does not contain an absolute truth and that therefore it is unclear which distribution of the credit to co-authors is the correct distribution. In reality authorship can be rewarded as part of departmental publishing deals, or even as a thank you for permission to access data. We will not be discussing "political" authorship agreements in this review, but from the end-users' point of view the desirability of correcting for co-authorship is doubtful as recalculation of the *h*-core can lead to over-correction and thus penalise the researcher under assessment (Rosenberg 2011). The recurring question is, if sharing credit is at all necessary. Realistically, we expect end-users to present the highest number of citations their works have achieved or the highest scoring indicator. If all researchers within a field practice "multiple co-authorship" then sharing the credit is superfluous and in some cases counterproductive. Not only will researchers reduce their performance on their CV, their *h*-indicators will also be reduced. More importantly, future participation in collaborative projects could be discouraged. So even if we agree that harmonic counting gives a more accurate assessment of collaborative scientific productivity and counterbalances the biases of equalization and inflation when issuing author credit (Hagen 2010), it is worth considering if, within the practices of the field, the extra effort is at all necessary.

Impact over time, category 5

Impact over time normalized to portfolio, 5a

It is incorrectly assumed that the chance of a researcher's work being used declines with age because in general its validity and utility decline as well. Usage and validity are not related, and linking usage with validity is unwise (De Bellis 2009). The rate of loss of validity or utility of older documents is not the same in all fields and does not have to same effect on usage. Literature in the natural sciences ages more quickly than literature in the humanities where information in older documents is more readily incorporated elsewhere. Stochastic models allow for the translation of diverse factors influencing aging into parameters that can be estimated from empirical data with a specified margin of error; *Dynamic h*, *AWCR*, *AW*, *DCI*, h^t (De Bellis 2009). However the calculation of ratio or percentile based models are simpler to understand; *c(t)*, *aging rate*, h^c, *m quotient*, *Price Index*, *AR*. Obviously, in these simpler models, the yard stick measure of expected performance is rougher and the illustrated decay of a publication is in some cases steeper, e.g. *AR index*.

Impact over time normalized to field, 5b

The more a field grows the more articles come into existence, acting as competition between "older" articles to get into the reference list of the new ones. Growth has been verified as an influence on impact over time but is not a cause of the obsolescence of publications (Egghe and Rousseau 2000). Therefore, if publications from particular researchers need more time than "normal" to be properly acknowledged by their colleagues, the impact of these researchers may be underestimated with standard citation windows. The rate at which scientific literature ages and the rapidity with which it is cited are important in determining the length of the citation windows used for citation counts making field comparisons complex. Measures of impact over time have to cope with diverse characteristics and fluctuations in usage by local groups. The relative or expected (probabilistic) number of citations an individual article receives over an analyzed time interval, adjusted to the local field and document types, are relevant indicators of sustainability at the author-level. Even though the resulting indicators are more nuanced and allow for a greater granular comparison of research performance over time, we judged the measures too complex for end-user application.

Methodological considerations

This review is limited to a subjective complexity assessment of indicators at the individual level. Our judgements perhaps underestimate the abilities of end-users, especially the end-users that practice using bibliometric indicators and are very knowledgeable about their limitations. Our search for indicators taught us that some researchers are very keen on using bibliometric indicators on their CVs and include a narrative explaining the computations of the indicators they use. Some have gone so far as inventing their own domain specific indices. We have not tested empirically the complexity of each indicator neither have we investigated the applicability, validity, utility, objectivity or the effects on publishing behaviour. Further we have not studied the cause and effect mechanisms inherent to the indicator, or inter-field variations of the indicators when implemented. Neither have we considered the reliability of indicators used by end-users on their CVs. These need to be analysed in future studies involving end-users.

The categorization of indicators covers the basic effects of publications at the author-level. Our simple set of categories, even if they do not converge with other typologies, provides valuable information on the relative merits and weaknesses of the indicators. The qualitative approach was preferred as comprehensive factor analysis was not the purpose of this review.

Conclusions

We did not identify a single indicator that captures the overall impact of a researcher. Our categorization illustrates clearly that author-level indicators only partially capture individual impact as they indicate impact over time, impact normalized to field, impact of a selected number of publications or impact normalized to the researchers' age, seniority and productivity. Only when indicators are used in combination can they approximate the overall impact of a researcher. Hopefully our review will increase awareness of the range of options end-users have to demonstrate the impact of their work and will discourage using a single numerical value to represent the effects of their work. However, choosing the appropriate indicators to combine takes knowledge of which aspect of publication activity the indicator attempts to capture and how the indicators are calculated, including which data needs to be collected. As there is no workable definition of scientific impact, there is no agreement on which combination of indicators best express the impact of a researcher's body of work or which best fit the aim of an assessment of a researcher. But there is at least agreement that using just one indicator is inadequate.

Administrators, evaluators and researchers seem to use indicators as never before and their widespread use has led to the construction of novel indicators as well as variants or combinations of established ones. This paper reviews 108 author-level indicators and exemplifies the complexity of their data collection and calculation in relation to end-user application. Our study attempts to identify which author-level indicators can be calculated by end-users and we succeeded in identifying 79 such potentially useful indicators. The data collection and calculation of these indicators is relatively straightforward, and as such it is clear how they measure or interpret certain aspects of performance. Further, our study shows that superior author-level indicators that claim to produce improved representations of individual performance and more granular distinctions between researchers, were too complicated for end-users to apply them in practise.

As indicators get more refined their complexity increases and as such we assume they are designed for the bibliometric community to use in assessments on the behalf of the individual and not for end-user "self-assessment". The results show that at the current time (1) certain publication activities and effects are more easily evaluated using bibliometrics than others, (2) assessment of publication performance cannot be represented by a single indicator, and (3) it is unwise to use citations as anything other than an *indication* of impact. Our clarification of how the indicators are calculated clearly demonstrates that the majority of indicators are different approximations of the average citations to publications in a dataset. Which indicator is the best approximation of the average is dependent on the data used in the calculation. To choose the best indicator, the end-user has to understand the spread of the data and which indicators present the best model that captures the central tendency. However, unlike statistical models the indicators produce solitary numbers as an estimate of performance, and are presented to the end-user without confidence intervals or minimum/maximum values that would provide contextual information about these point estimates.

Bibliometric indicators are readily available, and will therefore be used in both intended and unintended ways. Using indicators out of their context is a problem in relation to their validity or rather the validity of the use made of the measure. Which indicators are most useful to an end-user in expressing their publication performance requires further study. Taking one indicator alone and interpreting the results out of context of the researcher's field or seniority will result in distorted and useless information. We can conclude that by providing a recommended selection of indicators for end-user assessment, the researcher can reach a better understanding of the impact of their published works and perhaps identify where this can be improved. The success of the indicators are dependent on the completeness of data, which often requires access to comprehensive citation databases and the extraction of unstructured data from the internet or other sources. The knowledge we have about which indicators individuals can employ to reliably measure their performance is limited. They have yet to be properly validated using empirical data from different research fields and their long term effects on scientific behaviour needs to be investigated in prospective studies. However, our extensive tables can contribute to awareness of the possibilities and limitations of bibliometric indicators as well as the data requirements, time and competencies needed to calculate them. Simple indicators are recommended for end-user application as their requirements to bibliographic data are modest and calculations transparent.

Acknowledgments This work was supported by funding from ACUMEN (Academic Careers Understood through Measurement and Norms), FP7 European Commission 7th Framework "Capacities, Science in Society", Grant Agreement: 266632. Opinions and suggestions contained in this article are solely the authors and do not necessarily reflect those of the ACUMEN collaboration.

References

Ahlgren, P., & Järvelin, K. (2010). Measuring impact of 12 information scientists using the DCI-index. *Journal of the American Society for Information Science and Technology, 61*(7), 1424–1439.

Alonso, S., Cabreriazo, F., Herrera-Viedma, E., & Herra, F. (2009). H-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informatrics, 3*(4), 273–289.

Anderson, T. R., Hankin, R. K. S., & Killworth, P. D. (2008). Beyond the durfee square: Enhancing the h-index to score total publication output. *Scientometrics*. doi:10.1007/s11192-007-2071-2.

Antonakis, J., & Lalive, R. (2008). Quantifying scholarly impact: IQp versus the Hirsch h. *Journal of the American Society for Information Science and Technology*. doi:10.1002/asi.20802.

Archambault, È, & Larivière, V. (2010). The limits of bibliometrics for the analysis of the social sciences and humanities literature. In: Caillods F. (Ed.), *World Social Science Report 2010*. UNESCO publishing, pp. 251-254.

Arencibia-Jorge, R., Barrios-Almaguer, I., Ferdandez-Hernandez, S., & Carvajal-Espino, R. (2008). Applying successive h indicators in the institutional evaluation: A case study. *Journal of the American Society for Information Science and Technology, 59*(1), 155–157.

Bach, J. F. (2011). On the proper use of bibliometrics to evaluate individual researchers. Académie des sciences. http://www.academie-sciences.fr/activite/rapport/avis170111gb.pdf. Accessed 5 Apr 2013.

Batista, P., Campiteli, M., Kinouchi, O., & Martinez, A. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics, 68*(1), 179–189.

Belter, C. (2012). A bibliometric analysis of NOAA's office of ocean exploration and research. *Scientometrics,* doi:10.1007/s11192-012-0836-0.

Bennett, D., & Taylor, D. (2003). Unethical practices in authorship of scientific papers. *Emergency Medicine, 15*, 263–270.

Bollen, J., Rodriguez, M., & Van de Sompel, H. (2006). Journal status. *Scientometrics, 69*(3), 669–687.

Bollen, J., & van de Sompel, H. (2008). Usage impact factor: The effects of sample characteristics on usage-based impact metrics. *Journal of the American Society for Information Science and Technology, 59*(1), 136–149.

Bordons, M., & Barrigon, S. (1992). Bibliometric analysis of publication of spanish pharmacologists in the SCI (1984–1989): 2 Contribution to subfields other than pharmacology and pharmacy (ISI). *Scientometrics, 25*(3), 425–446.

Bornmann, L. (2012). Measuring the societal impact of research. *EMBO Reports, 13*(8), 673–676.

Bornmann, L., Mutz, R. & Daniel, H. (2008). Are there better indicators for evaluation purposes than the h-index? A comparison of nine different variants of the h-index using data from biomedicine. *Journal of the American Society for Information Science and Technology.* doi:10.1002/asi.20806.

Bornmann, L., Mutz, R., Hug, S. E., & Daniel, H. (2011). A multilevel meta-analysis of studies reporting correlations between the h-index and 37 different h-index variants. *Journal of Informetrics,* doi:10.1016/j.joi.2011.01.006.

Bornmann, L., & Werner, M. (2012). How good is research really? *EMBO Reports, 14*, 226–230.

Brown, R. (2009). A simple method for excluding self-citations from the h-index: The b-index. *Online Information Review, 33*(6), 1129–1136.

Burnhill, P., & Tubby Hille, M. (1994). On measuring the relation between social science research activity and research publication. *Research Evaluation, 4*(3), 130–152.

Cabrerizoa, F. J., Alonso, S., Herrera-Viedmac, E., & Herrerac, F. (2012). Q2-index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch-core. *Journal of Informetrics, 4*, 23–28.

Chai, J., Hua, P., Rousseau, R., & Wan, J. (2008). The adapted pure h-index. In *Proceedings of WIS 2008: Fourth International conference on webmetrics, informetrics and scientometrics & ninth COLLNET meeting,* Berlin. www.collnet.de/Berlin-2008/ChaiWIS2008aph.pdf.

Claro, J., & Costa, C. (2011). A made-to-measure indicator for cross-disciplinary bibliometric ranking of researchers performance. *Scientometrics,* doi: 10.1007/s11192-010-0241-5.

Costas, R., & Bordons, M. (2007). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics*. doi:10.1016/j.joi.2007.02.001.

Costas, R., Bordons, M., van Leeuwen, T. N., & van Raan, A. (2009). Scaling rules in the science system: Influence of field-specific citation characteristics on the impact of individual researchers. *Journal of the American Society for Information Science and Technology, 60*(4), 740–753.

Costas, R., van Leeuwen, T. N., & Bordons, M. (2010a). A bibliometric classificatory approach for the study and assessment of research performance at the individual level. *Journal of the American Society for Information Science and Technology, 61*(8), 1564–1581.

Costas, R., van Leeuwen, T., & van Raan, A. (2010b). Is scientific literature subject to a sell by date? A general methodology to analyze the durability of scientific documents. *Journal of the American Society for Information Science and Technology, 61*(2), 329–339.

Costas, R., van Leeuwen, T. N., & van Raan, A. (2011). The “Mendel Syndrome” in science: Durability of scientific literature and its effects on bibliometric analysis of individual scientists. *Scientometrics, 89*(1), 177–205.

Cronin, B. (1984). *The citation process: The role and significance of citations in scientific communication*. London: Taylor Graham.

Dahler-Larsen, P. (2012). *The evaluation society*. California: Stanford University Press.

De Bellis, N. (2009). *Bibliometrics and citation analysis: From the science citation index to cybermetrics*. Lanham, MD: Scarecrow Press.

Delgado López-Cózar, E., Robinson-García, N., & Torres-Salinas, D. (2014). The Google scholar experiment: How to index false papers and manipulate bibliometric indicators. *Journal of the Association for Information Science and Technology, 65*, 446–454. doi:10.1002/asi.23056.

Directorate-General for Research. (2008). *Assessing Europe’s university-based research*. Belgium: The European Commission.

Eck, N. V., & Waltman, L. (2008). Generalizing the g- and h-indicators. *ECON Papers.*. http://EconPapers.repec.org/RePEc:dgr:eureri:1765013210.

Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics, 69*(1), 131–152.

Egghe, L., & Rousseau, R. (2008). An h-index weighted by citation impact. *Information Processing and Management*. doi: 10.1016/j.ipm.2007.05.003.

Egghe, L., Rousseau, R., & Van Hooydonk, G. (2000). Methods for accrediting publications to authors or countries: Consequences for evaluation studies. *Journal of the American Society for Information Science, 51*(2), 145–157.

Glänzel, W. (2006). On the h-index—A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics.* doi:10.1007/s11192-006-0102-4.

Glänzel, W., & Schubert, A. (2010). Hirsch-type characteristics of the tail of distributions. The generalized h-index. *Journal of Informetrics, 4*(1), 118–123.

Hagen, N. (2010). Harmonic publication and citation counting: Sharing authorship credit equitably—not equally, geometrically or arithmetically. *Scientometrics, 84*(3), 785–793.

Harzing, A. (2008). Reflections on the H-index. Harzing.com. http://www.harzing.com/pop_hindex.htm. Accessed 5 Apr 2013.

Harzing, A. (2012). Publish or perish user’s manual. Harzing.com http://www.harzing.com/pophelp/metrics.htm. Accessed 11 April 2013.

Haslam, N., & Laham, S. M. (2009). Quality, quantity, and impact in academic publication. *European Journal of Social Psychology,* doi: 10.1002/ejsp.727.

Haustein, S. (2012). *Multidimensional journal evaluation: Analyzing scientific periodicals beyond the impact factor*. Berlin: K. G.Saur Verlag GmbH & Company.

HEFCE. (2009). *Identification and dissemination of lessons learned by institutions participating in the research excellence framework (REF) bibliometrics pilot: Results of the round one consultation.* Higher Education Funding Council for England. https://www.hefce.ac.uk/pubs/rereports/year/2009/bibliometricspilotlessons/ Accessed 11 Apr 2013.

Hicks, D. (2004). The four literatures of social science. In H. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantative sicence and technology research*. New York: Springer/Kluwer.

Hicks, D. (2006). The dangers of partial bibliometric evaluation in the social sciences. *Economica Politica, 33*(2), 145–162.

Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*(46), 16569–16572.

Iglesias, J., & Pecharromán, C. (2007). Scaling the h-index for different scientific ISI fields. *Scientometrics, 73*(3), 303–320.

Ingwersen, P. (2005). Scientometri: Videnskabspublicering og bibliometriske metoder. Ressource document. Biokemisk Forening. http://www.biokemi.org/biozoom/issues/506/articles/2162. Accessed 23 Mar 2014.

Ingwersen, P., Larsen, B., Rousseau, R., & Russell, J. (2001). The publication-citation matrix and its derived quantities. *Chinese Science Bulletin, 46*(6), 524–528.

Jacsó, P. (2011). Google Scholar duped and deduped—the aura of "robometrics". *Online Information Review, 35*(1), 154–160.

Järvelin, K., & Person, O. (2008). The DCI-index: Discounted cumulated impact based on research evaluation. *Journal of the American Society for Information Science and Technology, 59*(9), 1433–1440.

Jin, B. H. (2006). H-index: An evaluation indicator proposed by scientist. *Science Focus, 1*(1), 8–9.

Jin, B. H., Liang, L. L., Rousseau, R., & Egghe, L. (2007). The R and AR indicators: Complementing the h-index. *Chinese Science Bulletin, 52*(6), 855–863.

Kosmulski, M. (2006). A new type Hirsch-index saves time and works equally well as the original h-index. *ISSI Newsletter, 2*(3), 4–6.

Lawrence, P. A. (2003). The politics of publication. *Nature, 422*(6929), 259–261.

Lawrence, P. A. (2008). Lost in publication: How measurement harms science. *Ethics in science and environmental politics, 8*(1), 9–11.

Levitt, J., & Thelwall, M. (2011). A combined bibliometric indicator to predict article impact. *Information Processing and Management.* doi: 10.1016/j.ipm.2010.09.005.

Liang, L. (2006). H-index sequence and h-index matrix: Constructions and applications. *Scientometrics, 69*(1), 153–159.

Lundberg, J. (2009). Lifting the crown—citation z-score. *Journal of Informetrics,* doi: 10.1016/j.joi.2006.09.007.

Martin, B. R., & Irvine, J. (1983). Assessing Basic Research: Some partial indicators of scientific progress in Radio Astronomy. *Research Policy, 12*(2), 61–90.

Miller, C. W. (2006). Superiority of the h-index over the impact factor for physics. arXiv:physics/0608183 [physics.soc-ph].

Moed, H. F. (2008). UK research assessment exercises: Informed judgments on research quality or quantity? *Scientometrics*. doi: 10.1007/s11192-008-0108-1.

Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics, 4*(3), 265–277.

Mostert, S. P., Ellenbroek, S. P. H., Meijer, I., van A. & Klasen, E. C. (2010). Societal output and use of research performed by health research groups. Health Research Policy Systems. http://www.biomedcentral.com/content/pdf/1478-4505-8-30.pdf. Accessed 5 Apr 2013.

Namazi, M. R., & Fallahzadeh, M. K. (2010). N-index: A novel and easily-calculable parameter for comparison of researchers working in different scientific fields. *Indian Journal of Dermatology, Venereology, and Leprology, 76*(3), 229–230.

Nederhof, A. J., & Meijer, R. F. (1995). Development of bibliometric indicators for utility of research to users in society: Measurement of external knowledge transfer via publications in trade journals. *Scientometrics*. doi: 10.1007/BF02020187.

Nelhans, G (2013) *The practices of the citation: Scientific publication as theory, method and research policy.* PhD thesis. https://gupea.ub.gu.se/handle/2077/33516. Accessed 19 Mar 2014.

Niederkrotenthaler, T., Dorner, T. E., & Maier, M. (2011). Development of a practical tool to measure the impact of publications on the society based on focus group discussions with scientists. *BMC Public Health,* doi:10.1186/1471-2458-11-588.

Okubu, Y. (1997). Bibliometric indicators and analysis of research systems: Methods and examples. *OECD Science, Technology and Industry Working Papers 1*. doi:10.1787/208277770603.

Panaretos, J., & Malesios, C. (2009). Assessing scientific research performance and impact with single indicators. *Scientometrics*. doi: 10.1007/s11192-008-2174-9.

Podlubny, I. (2005). Comparsion of scientific impact expressed by the number of citations in different fields of science. *Scientometics, 64*(1), 95–99.

Price, D. S. (1970). Citation measures of hard science, soft science, technology and non-science. In C. E. Nelson & D. K. Pollack (Eds.), *Communication among scientists and engineers* (pp. 3–22). Lexington: Heath Lexington Books.

Radicchi, F., Fortunatoa, S., & Castellanob, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America, 105*(45), 17268–17272.

Rehn, C., Kronman, U., & Wads, D. (2007) Bibliometric indicators—definitions and usage at Karolinska institutet. http://kib.ki.se/sites/kib.ki.se/files/Bibliometric_indicators_definitions_1.0.pdf. Accessed 19 Mar 2014.

Retzer, V., & Jurasinski, G. (2009). Towards objectivity in research evaluation using bibliometric indicators - a protocol for incorporating complexity. *Basic and Applied Ecology, 10*(5), 393–400.

Rosenberg, M. S. (2011). A biologist's guide to impact factors. Rosenberglab.net. www.rosenberglab.net/Pubs/Rosenberg2011_ImpactFactor.pdf. Accessed 5 April 2013.

Rousseau, R. (2006). New developments related to the Hirsch-index. RCLIS.org. http://eprints.rclis.org/7616/1/Hirsch_new_developments.pdf. Accessed 5 April 2013.

Rousseau, R., & Ye, F. Y. (2008). A proposal for a dynamic h-type index. *Journal of the American Society for Information Science and Technology*. doi: 10.1002/asi.20890.

Ruane, F., & Tol, R. (2008). Rational (successive) h-indicators: An application to economics in the Republic of Ireland. *Scientometrics*, doi: 10.1007/s11192-007-1869-7.

Sanderson, M. (2008). Revisiting h measured on UK LIS and IR academics. *Journal of the American Society for Information Science and Technology, 59*(7), 1184–1190.

Sandström, E., & Sandström, U. (2009). Meeting the micro-level challenges: Bibliometrics at the individual level. *Proceedings of ISSI 2009 12th International Conference of the International Society for Scientometrics and Informetrics,* 2, 846-856.

Schneider, J. W. (2013). Caveats for using statistical significance tests in research assessments. *Journal of Informetrics, 7*(1), 50–62.

Schneider, J. W. (Forthcoming). Null hypothesis significance tests. A mix-up of two different theories: the basis for widespread confusion and numerous misinterpretations. *Scientometrics*. doi:10.1007/s11192-014-1251-5.

Schreiber, M. (2008a). An empirical investigation of the g-index for 26 physicists in comparison with the h-index, the A-index, and the R-index. *Journal of the American Society of Information Science and Technology*. doi:10.1002/asi.20856.

Schreiber, M. (2008b). A modification of the h-index: the H(m)-index accounts for multi-authored manuscripts. arXiv:0805.2000[Physics.Soc-Ph].

Schreiber, M. (2010). Twenty Hirsch-index variants and other indicators giving more or less preference to highly cited papers. arXiv:1005.5227v1 [Physics.Soc-Ph].

Schreiber, M., Malesios, C. C., & Psarakis, S. (2012). Exploratory factor analysis for the Hirsch-index, 17 h-type variants, and some traditional bibliometric indicators. *Journal of Informetrics,* doi: 10.1016/j.joi.2012.02.001.

Sidiropoulos, A., Katsaros, D., & Manolopoulos, Y. (2007). Generalized Hirsch h-index for disclosing latent facts in citation networks. *Scientometrics*. arXiv:cs/0607066v1.

Tol, R. S. J. (2008). A rational, successive g-index applied to economics departments in Ireland. *Journal of Informetrics, 2*(2), 149–155.

Tol, R. S. J. (2009). Of the H-index and its alternatives: An application to the 100 most prolific economists. *Scientometrics, 80*(2), 317–324.

van Leeuwen, T. N., Visser, M., Moed, H., Nederhof, T., & Raan, A. V. (2003). The holy grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics*. doi: 10.1023/A:1024141819302.

Vanclay, J. (2007). On the robustness of the H-index. *Journal of the American Society for Information Science and Technology*. doi:10.1002/asi.20616.

Vinkler, P. (2009). The π-index: A new indicator for assessing scientific impact. *Journal of Information Science,* doi: 10.1177/0165551509103601.

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., Rafols, I., Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*. doi: 10.1016/j.joi.2010.06.004.

Waltman, L., & Schreiber, M. (2013). On the calculation of percentile-based bibliometric indicators. *Journal of the Association for Information Science and Technology, 64*(2), 373–379.

Waltman, L., & van Eck, L. (2009). A taxonomy of bibliometric performance indicators based on the property of consistency. ERIM Report. repub.eur.nl/res/pub/15182/ERS-2009-014-LIS.pdf. Accessed 5 Apr 2013.

Waltman, L., & van Eck, N. J. (2011). The inconsistency of the h-index. arXiv:1108.3901v1 [cs.DL].

Waltman, L., van Eck, N. J., van Leeuwen, T., & Visser, M. S. (2012). Some modifications to the SNIP journal impact factor. arXiv:1209.0785.

Wan, J., Hua, P. & Rousseau, R. (2007). The pure h-index: Calculating an author's h- index by taking co-authors into account. ELIS. http://eprints.rclis.org/10376/. Accessed 5 Apr 2013.

Whitley, R. (2000). *The intellectual and social organization of the sciences*. Oxford: Oxford University Press.

Wu, Q. (2008). The w-index: A significant improvement of the h-index. arXiv:0805.4650v1 [Physics.Soc-Ph].

Yan, E., & Ding, Y. (2011). Discovering author impact: A PageRank perspective. *Journal Information Processing and Management, 47*(1), 125–134.

Zhang, C. (2009). The e-index, complementing the h-index for excess citations. *PLoS ONE,* doi:10.1371/journal.pone.0005429.