ACUMEN

academic careers understood through measurement and norms

FP7 Grant Agreement	266632

Deliverable No and Title	D5.8 - Novel bibliometric indicators
Dissemination Level	PU (public)
Work Package	WP5
Version	1.0
Release Date	29-04-2014
Author(s)	Birger Larsen Lorna Wildgaard Jesper W. Schneider

Project Website	http://research-acumen.eu/

European Commission
7th Framework Programme
SP4 - Capacities
Science in Society 2010
Grant Agreement: 266632

Preface

Deliverable 5.8 reports on WP5 of the ACUMEN project. The work package has investigated to what extent bibliometric indicators can be used in the evaluation of individual researchers. WP5 has analysed a wide range of bibliometric indicators such as indicators of production, citations, production & citations, production adjusted for time, production adjusted for field and several measures that describe different aspects of a researcher's publishing portfolio as a whole. WP5 has also assessed the need for the creation of new bibliometric indicators for the assessment of individuals and discussed ethical aspects. In addition the work package has also carried out a study of the feasibility of predicting later star researchers given early citation data. A main result of WP5 is the recommendation of a set of bibliometric indicators the researchers can use for self-assessment and which can be included in the ACUMEN portfolio along with indicators from other work packages. The indicators have been tested empirically on samples drawn from the joint ACUMEN dataset.

Deliverable 5.8 consists of a number of reports and publications reporting on the different tasks of WP5.

Part 1 reports on Task 5.1 and is a state-of-the-art literature review of bibliometric indicators that potentially can be used on the level of individual researchers, as well as on Task 5.2 that examines the need for development of new bibliometric indicators for this level. A main conclusion of the review is that there is no pressing need for the development of new bibliometric indicators for the individual level as there is a very large number in existence. Part 1 consists of an article submitted to the journal *Scientometrics*, where a revised version is currently under review.

Part 2 reports on the study of the feasibility of predicting later star given early citation data. This thus covers one part of Task 5.3 (selection of a sample of successful researchers) and analyses if bibliometric indicators can predict these later stars when compared to normal researchers (part of Task 5.4).

Part 3 reports on Task 5.3 - the selection of samples for the main empirical study of applying bibliometric indicators on a large sample of the ACUMEN shared data set covering four scholarly fields. It also discusses how non-experts can best collect publication and citation data.

Part 4 reports on Task 5.4 from the perspective of the researcher and discusses how to develop guidelines for a codex of behaviour when carrying out self-evaluation using bibliometric indicators and how to best report the results. Part 4 analyses current evaluation practices and provided input for the ACUMEN Portfolio and Good Evaluation Practices.

Part 5 reports on Task 5.4 and is an analysis of the consequences of applying bibliometric indicators derived from Google Scholar on the sample of researchers selected in Part 3.

Part 6 reports on Task 5.4 and is an analysis of the consequences of applying bibliometric indicators derived from Web of Science on the sample of researchers selected in Part 3. The indicators tested in Part 6 are draw from Part 1.

Part 7 reports on Task 5.4 and summarises and compares the conclusion from Part 5 and Part 6.

Table of Contents

D5.8 Part 1 – Literature Review 1

D5.8 Part 2 – Bibliometric Indicators of Young Authors in Astrophysics: Can Later Stars be Predicted?..... 65

D5.8 Part 3 – Selection of Samples 91

D5.8 Part 4 – Consequences of Indicators. Effects on the users 188

D5.8 Part 5 – Consequences of Indicators: using indicators on data from Google Scholar 204

D5.8 Part 6 – Cluster analysis of bibliometric indicators of individual scientific performance 213

D5.8 Part 7 – Comparison of indicators in Google Scholar and Web of Science 253

ACUMEN
academic careers understood through measurement and norms

FP7 Grant Agreement	266632

Deliverable No and Title	D5.8 Part 1 – Literature Review
Dissemination Level	PU (public)
Work Package	WP5 Bibliometric Indicators
Version	1.0
Release Date	25th September 2013
Author(s)	Lorna Wildgaard Birger Larsen Jesper Schneider

Bibliometric Self-Evaluation: A review of the characteristics of 114 indicators of individual performance

L. Wildgaard[a]*, J. W. Schneider[b], B. Larsen[c]

[a] *Royal School of Library and Information Science, Birketinget 6, 2300 Copenhagen, Denmark*
[b] *Institut for Statskundskab - Dansk Center for Forskningsanalyse, Bartholins Allé 7, 8000 Aarhus C, Denmark*
[c] *Royal School of Library and Information Science, Birketinget 6, 2300 Copenhagen, Denmark*

*Corresponding author.
Email Addresses: Lew@iva.dk (L. Wildgaard), jws@cfa.au.dk (J.W.Schneider), Bir@iva.dk (B. Larsen)

The political use of bibliometrics as a form of 'psuedo peer review' has raised concerns in the bibliometric community regarding the misuse of indicators and the inaccurate interpretation of bibliometric results. In this paper we consider the potentials for researchers to use bibliometrics themselves to counterbalance quick and dirty background checks in the competition for tenure or funds. We compare the advantages and limitations of 114 bibliometric indicators that purport to measure academic performance at the individual level. This comparison results in the identification of 64 indicators researchers can use themselves to contextualize the scientific activities listed on their curriculum vitae, categorized as: scientific impact, quality, output, outcome, sustainability, innovation and societal benefits or research infrastructure. Rather than conclusions, this study has led to further questions. The indicators require empirical analysis to establish their stability and usefulness, but specifically the ethical and behavioural issues in using bibliometrics in self-evaluation, both from the perspective of the researcher and the evaluator, demand further investigation.

Keywords
Individual bibliometrics; Research evaluation; Impact factors; Self-evaluation; Researcher performance; Indicators; Curriculum Vitae;

Introduction

The field of bibliometrics has reached such maturity that policymakers are considering using indicators in concrete evaluations of the individual. This has created discontent with researchers who regard external bibliometric review as monitoring. They are mistrustful of how the results of evaluations will be used eg. the effects of quantitative evaluation on scientific behavior and methodological favouritism for domains that are easier to assess (Hicks, 2012); if some scientific activities will be prioritized by policy makers, and how the results of evaluation rounds will affect the distribution of investment in research projects. For the same reasons, bibliometricians are cautious of evaluation at the micro-level, as the context and variables affecting the results of analyses are many, and often unsatisfactorily explored. Hence, the debate on the shortcomings of performance indicators generated by bibliometric methods at the micro-level continues (Bach, 2011; Bornmann & Werner, 2012; Burnhill & Tubby Hille, 1994; Sandström & Sandstrøm, 2009; Wagner et al., 2011). Researchers thus need to thoroughly investigate the effects on and changes in research behaviour in an extended retrospective study before micro-level evaluation is officially implemented.

In practice, evaluation and benchmarking requires the individual to document research activities with bibliographical data. Correct assessments by critical and rational evaluationalists (not politically motivated agents), provide reliable quantitative data, but only when interpreted in context and combined with qualitative evaluation such as interviews or peer-review, i.a. (Directorate-General for Research, 2008; EFC, 2010). To compensate for the limitations of indicators and to capture the nuances of scientific activities the combination of assessment methods is vital (Mostert, Ellenbroek, Meijer, van A., & Klasen, 2010). Despite of the concerns from the bibliometric community, evaluation of the individual through bibliometric indicators is already being performed as a form of ‘pseudo peer review’ in selection of candidates for tenure, in background checks of potential employees’ publication- and citation impact, and in appraisal of funding applications. We the authors do not support this use of individual bibliometrics, but recognise that the researcher can use them to strengthen presentation of his or her CV in the competition for tenure or funds, and to counterbalance quick and dirty background checks.

The chosen bibliometric method of individual self-evaluation has important implications as indicators alone are not informative and variables that affect the performance of indicators, such as field variation, academic seniority, gender or length of scientific career, are not always adequately accounted for. In addition to ethical issues, we have found four reoccurring themes in the literature concerning individual bibliometrics. First, how can researchers ensure an objective analysis of all of their dissemination activities for a complete assessment of their entire body of work (Hicks, 2004)? Second, how can researchers be discouraged from ‘pimping their CVs’ thus embellishing results of their activities? Third, how stable are indicators when computed on a small amount of publication or citation data (De Bellis, 2009)? And fourth, how to account for differences in publishing and citing traditions across scientific fields. Failure to fit indicators to these variables can lead to a distorted indication of scientific activities, counter productively effecting the researcher’s CV or falsely boosting achievements (Archambault & Larivière, 2010; Batista, Campiteli, Kinouchi, & Martinez, 2006; Iglesias & Pecharromán, 2007).

The lack of agreement on how to measure bibliometrically the research activities of an individual is made worse by the lack of qualified and validated indicators that are actually designed for this purpose (Bollen, Van, Hagberg, & Chute, 2009). The validity of bibliometric indicators at the micro-level demands attention in order to establish what the indicators and resulting data represent and do not represent (Bornmann & Werner, 2012). Further, guidelines for both using indicators and the results of an evaluation need establishing. To address this, the ACUMEN collaboration[1] is developing a portfolio of indicators that account for age, gender, discipline and different scientific activities. The recommended indicators are based on empirical studies using bibliographic data from 2000 researchers in the fields and sub-fields of astronomy, environmental science, philosophy and public health. The aim is to present the researcher with indicators purposefully designed for self-evaluation. They are thus not reliant on large datasets for stability or complex calculations. As they have been tested empirically, the indicators can be explained in the context of scientific behaviour within the research field and academic seniority they are implemented. Ideally, this approach will strengthen the researcher's CV and improve understanding of the limits and strengths of indicators used in individual evaluation, how they supplement external review and contribute to Good Evaluation Practice. Ultimately, the portfolio will provide useful and qualified indicators to overcome policy blindness in extended evaluations and unwise comparisons with peers. Giving more control and insight to the researcher will hopefully reduce the fear of monitoring or the "publish or perish" mentality. Consequently the contextual interpretation and understanding of individual performance will improve.

This review is a preliminary study in the development of the aforementioned ACUMEN portfolio. The purpose is to collocate and compare bibliometric indicators that are feasible in an assessment of the individual's performance and can be undertaken by the researcher themselves. The methodological considerations to accomplish this haven't changed since Moravcsik identified them in 1986, in that science and technology have many different goals, aims and justifications and in the case of the individual, it must be specified which ones of these are taken into account and with what weight (Moravcsik, 1986). It follows that the researcher's activities will, in this review, be assessed as multidimensional. Because of the contributing variables and links between activities, no one indicator is expected to fully express an activity.
The validity of indicators will be discussed, because the results can be affected by: errors (Bollen, Van, Hagberg, & Chute, 2009; Franceschet, 2009), subjectivity (Bach, 2011), scope of citations indicators where data is sourced (Archambault & Larivière, 2010; Hicks & Wang, 2009), motivations to cite (Costas, Bordons, van Leeuwen, & van Raan, 2009; Leimu, 2005), the aim of the assessment and the extent of author contribution in multi-authored papers (Franceschet, 2010; Schreiber, 2008b). To attempt assessment of the quality of scientific output, it is necessary to obtain an unambiguous evaluation that accounts for the critical nuances at an individual level (Bach, 2011; Retzer & Jurasinski, 2009). This is not achievable using a single indicator, hence the desirability of combining indicators to obtain a global view of scientific output, (Costas, van Leeuwen, & Bordons, 2010a; Glänzel, Debackere, Thijs, & Schubert, 2006; van Leeuwen, Visser, Moed, Nederhof, & Raan, 2003).

In summary, this review will 1) identify which indicators are useful in individual self-evaluation to document activities listed on the CV and contextualize publication performance, 2)identify which scientific activities it is possible to measure and with which indicators, and 3) analyse the applicability of these indicators by discussing the strengths and weakness of each one.

Method

Bibliometric indicators were identified in a three-tiered search approach. The approach was designed

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

to establish which indicators can be and are currently included in research assessment, accordingly indicators implemented in practice and novel ones yet to be applied empirically were identified.

In level one, current guidelines for research evaluation by European Research Agencies were searched for performance reports on units of assessment from 2006-to present. Guidelines are often built on three or four year trial periods to enable assessment of the successes, failures and effects of the implemented strategy across institutions, disciplines and levels of aggregation. Therefore a broad time interval was chosen to capture these nuances. The aim was to 1) assemble a typology of research activities and 2) map the activity under evaluation to the indicators and identify if supplementary evaluation methods were used. The following agencies were included: Austria (ERA), Belgium (ULB), Denmark (Action Plan for research Evaluation), Finland (AALTO/UH RAE), France (AERES), Germany (CHE Ranking, Initiative for Excellence), Hungary (Maintainer Agreements), Italy (CIVR), Netherlands (SEP), Sweden (A New Model for Allocation of Resources) and the UK (REF 2014, HEFCE). An overview is presented in Appendix 1.

Level two explored the history, the development and the relationships between indicators through reference and citation chasing, beginning with known works by (Bach, 2011; De Bellis, 2009; Directorate-General for Research, 2008; Sandström & Sandstrøm, 2009; Schreiber, 2008a).

Finally in level three, previously unidentified indicators and supplementary information about the extent indicators measure what they purport to measure, were sourced using the terms (bibliometri* OR indic*) AND (Individual OR micro*) in Thomson Reuters Web of Science and in The Royal School of Library and Information Science's electronic collection of information science journals. Google Scholar was searched to retrieve i.a. national papers, reports, book chapters and other web-based material. Searches were supplemented with terms *impact, quality, co-author, co-authorship, collaboration, durability, obsolescence, ethics, societal, social, humanities* and *humanist* to focus the search and improve specificity where needed.

Definition of categories of scientific activity

The indicators identified in the search strategy were categorised according to the aspect of scientific activity they claim to measure. As indicators are evolutionary and supplement each other, they cannot in practice be restricted to just one category. The un-granular categorisation is for schematic purposes. The authors acknowledge that evaluation of an individual researcher requires combining indicators from different categories to capture the many different facets of scientific activity.

Scientific activity can be defined in many ways. Our groupings are based on categorical definitions already applied by research evaluation agencies in qualitative and quantitative assessments. These are: *output, outcome, quality, research infrastructure, impact, innovation and social benefits,* and *sustainability*.

Output or production is countable works, published or unpublished dependent on the unit of evaluation.

Outcome is the extent a researcher's work is used in the scientific community and thus contributes to the advancement of scientific knowledge. Usage is measured as citation count.

Quality is understood as an *indication* of the level and performance of research conducted by the researcher within normalized standards for the field (Alonso, Cabreriazo, Herrera-Viedma, & Herra, 2009).

Research infrastructure is a reflection of the scientist's collaboration; people, organizations and countries, and to which extent, these are citing the scientist's work.

Impact uses a combination of output and outcome indicators to formally suggest the visibility of the researcher's work in the field in which he/she is active.

Innovation and social benefits is the contribution of research to the social, economic and cultural capital of society. An indication of the innovation and social benefits of a researcher's work is gained in an evaluation of interaction between stakeholders, how it stimulates new approaches to social issues, and its influence on informing public debate and policy making (Bornmann, 2012; Directorate-General for Research, 2008).

Sustainability is the extent a researcher's output continues to be used or the decline in use.

We do not assume that our categorisation is the only correct aggregation of aspects of scientific activity. The categories were selected a priori, and, in restricting the placement of an indicator to one category only, it was clear that we could only judge the main function of the indicator. It is an interesting challenge to investigate if categories previously defined for qualitative evaluation, e.g. innovation and societal benefits, can be assessed using bibliometrics by the individual researcher. Placement of the indicator within a category was suggested independently and together we argued for this placement until consensus was reached. This qualitative approach was preferred as comprehensive factor analysis is not the purpose of this review. Further, we induct that as these categories are implemented in evaluation they are recognisable to the individual under evaluation. It would be futile for the researcher in self-assessment to use a typology that does not correlate with the evaluator's typology. We could, for example, have based the categories on a domain analysis of scientific communication within different disciplines, drawn a map of scientific activities and subsequently chosen the set of indicators for measuring the identified activities such as input, output, productivity, progress, function, importance, quality and impact and so on pertaining to each discipline. It is not possible to say which approach is better, nevertheless as the indicators of these various aspects of scientific activities are clearly described in the literature, our simple set of categories, even if they do not converge with other typologies, provide valuable information on the relative merits and weaknesses of the indicators.

Judgement of complexity

The usability of indicators is a major consideration in this review therefore the complexity of each indicator was assessed. The indicators were graded on a 5 point numerical scale to assess 1) the availability of citation data and, 2) the intricacy of the mathematical model required to compile the indicator, Table 1. This assessment might result in a reduction of the granularity and sophistication of the indicators we identify as useful for the researcher, and might even encourage the use of rougher measures over more accurate ones. The indicators have to measure what they purport to measure, however, usability is lost if correct measurement requires data that is not readily available to the researcher, difficult mathematical calculations, and intricate interpretations of complicated data output.

We assume the individual has a complete list of their publications and would only need to source citations and calculate the indicator.

Table 1. Scoring matrix for levels of complexity

DATA COLLECTION / CALCULATION		1 No citation data needed	2 Single citation index, structured data	3 Multiple citation indexes, structured data	4 unstructured citation data	5 No readily available citation data
1	Raw count					
2	Simple ratio or linear model					
3	Multiple calculations, simple*					
4	Multiple calculations, advanced†					
5	Advanced multiple calculations and transformation of data					

*Multiple simple calculations include repeat simple linear or ratio calculations in the mathematical foundation.
†Advanced calculations incorporate weighted parameters such as gamma or delta that the researcher has to define according to the discipline or time interval under analysis, defining velocity or other corrective factors in their mathematical foundations.

Results

The search found 114 indicators recommended for use in individual assessment. Sixty-nine of the indicators are implemented in practice while forty-five are theoretical constructs, the majority of these are corrections to the h-index (82%) and are placed in the quality (26/28 indicators), research infrastructure (6/12 indicators) and sustainability (5/14 indicators) categories. Due to the amount of collected indicators and the deliberations surrounding them, a detailed overview of these indicators, their definitions, purpose, advantages and limitations, complexity scores and additional comments is available electronically in Appendix 2.

Sixty-four of the 114 indicators scored score ≤3 in complexity in both collection of data and calculation, and where hence judged potentially useful for researchers to use themselves to support or strengthen their CV in an evaluation. An analytical summary of these indicators follows.

Output

11 indicators of output were identified and all can be easily used by the individual in self-assessment, complexity score ≤ 2. All are simple counting or ratio models. P is a raw count of output, while P_{isi}, P_{ts}, adjust for publishing source and 'weighted publication type' accounts for types of publication judged locally important or of a higher scientific quality relative to the specialty of the researcher. The remaining 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 *i*th 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. Only co-publication counting encourages identification of the level of collaboration rather than an integer number symbolizing a share.

Table 2 Bibliometric indicators used to assess the quantity of a researcher's output

Output	Designed to indicate	Complexity	
		Col*	Cal*
P (total publications)	Count of production used in formal communication	1	1
P_{isi} (publications processed in ISI)	Calculation of impact compared to world subfield citation average based on ISI citation data.	1	2
P_{ts} (publications in selected sources)	Number of publications in selected sources defined important by the researcher's affiliated institution.	1	2
Co-publications	Collaboration on departmental, institutional, inter- or national level & identify networks.	1	1
Fractional counting on papers	Shared authorship of papers gives 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	Indicates the importance of the last author for the project behind the paper.	1	2
FA (First author counting)	Credit given to first author only	1	1
Weighted publication count	A reliable distinction between different document types.	1	1

* Col. = data collection, Cal. = calculation

Outcome

Fourteen citation-based indicators of output were identified and all were judged useful for the researcher in self-evaluation, ≤3. The majority are ratio-based indicators which account for the amount of citations relative to publications, %SELFCIT, CPP, %PNC, Ptop, A/E(Ptop), and 'Number of significant papers'. Just C+sc and STC calculate the sum of all citations for the period of analysis, while C, C-sc, adjust the sum for self-citations. A measure of excellence is attempted with Ptop, A/E(Ptop), and 'Number of significant papers' all of which require a field reference standard. The effect of age on the publications and corresponding citations is adjusted for in 'Age of citations' and 'Age and productivity'. All these indicators require one or more citation index to source the data to enable comprehensive results.

Table 3 Bibliometric indicators used to assess the outcome (citation count) of a researcher's output

Outcome	Designed to indicate	Complexity	
		Col*	Cal*
C + sc (total cites, inc. self-citations)	Indication of all usage for whole period of analysis	3	1
C (citations in WOS, minus self cites)	Recognised benchmark for analyses. Indication of usage by stakeholders for whole period of analysis	2	2
Scimago Total Cites (STC)	Indication of usage by stakeholders for whole period of analysis	2	1
C-sc (total cites, minus self-cites)	Measure of usage for whole period of analysis	3	2
% SELFCIT	Share of citations to own publications	3	2
CPP (cites per paper)	Trend of how cites evolve over time	3	2
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
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
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	1
Number of significant papers	Gives idea of broad and sustained impact	3	1
Age and productivity (Costas, van Leeuwen, & Bordons, 2010a)	Effects of academic age on productivity and impact.	2	3
%Pnc (percent not cited)	Share of publications never cited after certain time period, excluding self-citations	3	1

* Col. = data collection, Cal. = calculation

Quality

Twenty-eight indicators of quality were identified, fourteen potentially useful to the individual researcher, score ≤3 in collection and calculation. Twelve of these are dependent on the calculation of h index which means they suffer from the same inadequacies as h: e, r, $\bar{h}$, m, hg, normalized-h, h^2, a, w, Q^2, h, and hmx. The remaining two are h-independent: g, and the 'index of quality and productivity'. The indicators measure quality as cumulative impact, and use is dependent on the variable they aggregate. Q^2 and the 'index of quality and productivity' account for field and amount of publications, a general indication of cumulative impact is achieved with h or hmx (which ranks academics by their maximum h measured across GS, WOS and Scopus), while r, g, hg, h^2, e, w account for the effects of highly cited papers. Meanwhile, for across field or seniority comparison normalized h, a, $\bar{h}$, m can be employed.

Table 4 Bibliometric indicators used to assess the quality of a researcher's output

Quality	Designed to indicate	Complexity	
		Col*	Cal*
h-index (Hirsch, 2005)	Cumulative achievement	3	2
g-index (Egghe, 2006)	The distinction between and order of scientists (Egghe, 2006; Harzing, 2008)	3	3
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	3	4
Generalized h-index hf (Radicchi, Fortunatoa, & Castellanob, 2008)	Allows comparison to peers by correcting individual articles' citation rates for field variation	3	4
h-index sequences and matrices (Liang, 2006)	Singles out significant variations in individual scientists citation patterns across different research domains	3	4
Hg-index (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009b)	Greater granularity in comparison between researchers with similar h- and g-indicators.	3	3
hα (Eck & Waltman, 2008)	Cumulative achievement, advantageous for selective scientists.	3	4
Gα (Eck & Waltman, 2008)	Based on same ideas as g-index, but allows for fractional papers and citations to measure performance at a more precise level.	3	4
Normalized h-index (Sidiropoulos, Katsaros, & Manolopoulos, 2007)	Normalizes h to compare scientists achievement based across fields	3	3
H(2) index (Kosmulski, 2006)	Weights most productive papers but requires a much higher level of citation attraction to be included in index.	3	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, Malesios, & Psarakis, 2012)	3	3
R-index (Jin, Liang, Rousseau, & Egghe, 2007)	Citation intensity and improves sensitivity and differentiability of A index	3	3
Citation-weighted h-index (hw) (Egghe & 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.	3	4
ħ-index (Miller, 2006)	Comprehensive measure of the overall structure of citations to papers	3	3
m-index (Bornmann, Mutz, & Daniel, 2008)	Impact of papers in the h-core	3	2
π-index (Vinkler, 2009)	Production and impact of scientist	3	4
Tapered h-index (hT) (Anderson, Hankin, & Killworth, 2008)	Production and impact index that takes all citations into account, yet the contribution of the h-core is not changed.	3	5
Rational h-indicators hrat Index (Ruane & 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).	3	5
Rational g-index grat, (Schreiber, 2008a; Tol, 2008)	Indicates the distance to a higher g-index	3	5
e-index (Zhang, 2009)	Complements the h-index for the ignored excess citations	3	2
f-index (Tol, 2009)	Attempts to give weight/value to citations. Highest number of articles that received *f* or more citations on average.	3	4
t-index (Tol, 2009)	Attempts to give weight/value to citations. Highest number of articles that received t or more citations on average	3	4
Hmx-index (Sanderson, 2008)	Ranking of the academics using all citation databases together.	3	2
w-index (Wu, 2008)	The integrated impact of a researcher's excellent papers.	3	2
Index of Quality and Productivity (Antonakis & Lalive, 2008)	Quality reference value; judges the global number of citations a scholar's work would receive if it were of average quality in its field.	3	3
x-index (Claro & Costa, 2011)	Indication of research level. Describes quantity and quality of the productive core and allows for comparison with peers.	3	4
H per decade (Hpd-index) (Kosmulski, 2009)	Compare the scientific output of scientists in different ages. Seniority-independent Hirsch-type index.	3	4
Q^2 –index (Cabrerizoa, Alonso, Herrera-Viedmac, & Herrerac, 2012)	Relates two different dimensions in a researcher's productive core: the number and impact of papers	3	3

* Col. = data collection, Cal. = calculation

Research Infrastructure

Twelve indicators of research infrastructure were identified, ten deemed useful for the researcher, complexity ≤ 3. Five indicators require calculation of the h-index in their mathematical foundations; hi, POPh, n, alternative h, Pure h, two indicators are purely citation-based, 'count of co-citations' and 'fractional counting'. Three are publication based indicators: number of co-authors, co-publications, and cognitive orientation.

A comprehensive and structured citation index is required to calculate co-citations, n-index, and cognitive orientation, however authors per paper, co-publications, fractional counting, hi and POPh, h, alternative h and Pure h can, with varying degrees of difficulty be calculated using information in Google Scholar. Likewise visual representation techniques illustratively map collaboration and activity networks and their complexity also varies according to the software available to the researcher. The researcher can choose to present areas of collaboration with 'number of co-authors', 'cognitive orientation' and 'visual representation' relative to his or her position within the field, or represent level of co-authorship using either ratio-based models; 'fractional counting on citations', POPh or n, or mean-based models; hi, alternative h. Moreover, these models treat citations and publications as a single unit that can be evenly distributed. An alternative is normalizing using the square root of h as in pure h or pure r.

Table 5 Bibliometric indicators used to assess the infrastructure linked to a researcher's output

Research Infrastructure	Designed to indicate	Complexity	
		Col*	Cal*
Number of co-authors	Indicates cooperation and growth of cooperation at inter- and national level;	1	1
Co-citations	Thematic networks and influence and impact of researcher.	3	1
Fractional counting on citations	Designed to remove the dependence of co-authorship (Egghe, 2008)	3	2
hi-index (Batista, Campiteli, Kinouchi, & Martinez, 2006)	Indicates number of papers with at least h citations scientist would have written if worked alone.	3	3
POP variation individual H-index (Harzing, 2008)	Accounts for co-authorship effects	3	3
n-index (Namazi & Fallahzadeh, 2010)	Enables comparison of researchers working in different fields:	2	2
H_m-index (Schreiber, 2008b)	Softens influence of authors in multi-authored papers	3	4
Alternative H index (Batista et al., 2006)	Indicates the number of papers a researcher would have written along his/her career if worked alone.	3	2
Pure h-index (Hp) (Wan, Hua, & Rousseau, 2007)	Corrects individual h-scores for number of co-authors	3	3
Adapted pure H-index (h_{ap}) (Chai, Hua, Rousseau, & Wan, 2008)	Finer granularity of individual h-scores for number of co-authors by using a new h-core.	3	5
Cognitive orientation	Identify how frequently a scientist publishes or is cited in various fields; indicates visibility/usage in the main subfields and peripheral subfields.	3	1
Visual representation techniques	Based on bibliographic data graphical representations are generated of publishing, collaboration, citations, growth and activity in research field.	3	1

* Col. = data collection, Cal. = calculation

Impact

In judging the complexity of impact indicators, we assumed the researcher's needs were met in Science Citation Index (SCI), Journal Citation Reports (JCR), EigenFactor, Scimago, Web of Science or Scopus databases. Twenty-seven impact indicators were identified, twenty-one judged simple enough for the researcher to employ in self-evaluation, ≤ 3: ISI JIF, Diachronous IF, Y factor, SJR, Eigenfactor, P_{tj}, CPP/JCSm, JCSm/FCSm, C/FCSm, AI, Normalised journal impact, JFIS, DIF, IFmed, NJP, FCS, FCSm, JSCS or JRV, JSCm, JCSm/FCSm, CPP/FCSm. However, although used as benchmarks in evaluation, twenty of these twenty-one indicators were designed as indications of journal impact or impact at a higher level of aggregation than a single researcher, such as research groups or institutions. Only one indicator is actually designed for use at the micro-level; $P_{tj.}$ 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.

Table 6 Bibliometric indicators used to assess the impact of a researcher's output

Impact	Designed to indicate	Complexity	
		Col*	Cal*
ISI JIF (SIF) **Synchronous IF**	Average number of citations a publication in a specific journal has received limited to ISI document types and subject fields.	2	1
Diachronous IF (Ingwersen, Larsen, Rousseau, & Russell, 2001)	Reflects actual and development of impact over time of a set of papers.	3	2
Weighted PageRank rating of journal status (Bollen, Rodriguez, & Van, 2006)	Indicates relative importance of journal within a journal citation network	2	5
Y Factor (Bollen, Rodriguez, & Van, 2006)	Scientific impact defined as a combination of popularity and prestige	2	2
Scimago Journal Rank (SJR)	Average per article PageRank based on Scopus citation data	2	1
EigenFactor	Journal's total importance to the scientific community	2	1
Article influence score (AI)	Measure of average per-article citation influence of the journal	2	1
Co-authorship network analysis (Yan & Ding, 2011)	Individual author impact within related author community	2	5
Normalised journal impact	Mean impact value of all the normalized citation counts for publications in a specific journal	2	2
Journal to field impact score (JFIS) (van Leeuwen & Moed, 2002)	Journal to fields citation score that indicates relative impact of a journal	3	2
Discipline Impact Factor (DIF) (Hirst, 1978)	Number of times a journal is cited by the core literature of a single subfield rather than a complete set of ISI journals.	2	3
Median impact factor (IF med)	The aggregate Impact Factor for a subject category	2	2
Normalised journal position (NJP) (Bordons & Barrigon, 1992)	Compare reputation of journals across fields	2	2
Item oriented field normalized citation score average ($\bar{c}f$) (Lundberg, 2009)	Item orientated field normalised citation score.	3	4
Field citation score (FCS)	Represents the number of citations expected for a paper of the same type, published in all journals within a specific field in the same year.	2	3
Field Citation Score Mean (FCSm)	Weighted average for comparison of impact in different subfields	2	3
JSCS or JRV Journal citation score (journal reference value)	Worlds average of citations to publications according to type and age.	2	3
Normalised Journal Citation Score (JSCm)	Reference value accounting for type of paper and years in which papers were published.	2	3
JCSM/FCSm (Costas, Bordons, van Leeuwen, & van Raan, 2009)	Journal based worldwide average impact mean for an individual researcher compared to average citation score of the subfields	3	2
Crown Indicator CPP/FCSm	Individual performance compared to world citation average to publications of same document types, ages, and subfields.	3	3
Prediction of article impact (Levitt & Thelwall, 2011)	Predictor of long term citation	2	4
P_{tj} (publications in selected journals)	Performance of articles in journals important to (sub)field or institution.	1	2
CPP/JCSm	Indicates if the individual's performance is above or below the average citation rate of the journal set.	3	2
JCSm/FCSm (Gaemers, 2007)	Relative impact level of the journals compared to their subfields	2	3
C/FCSm (van Leeuwen, Visser, Moed, Nederhof, & Raan, 2003)	Applied impact score of each article/set of articles to the mean field average in which the researcher has published	3	2
Logarithm based citation z-score (Lundberg, 2009)	Accounts for citation rate variability of different fields and skewed distribution of citations over publications on an item level.	3	5
Usage Impact Factor (UIF) (Bollen & Sompel van de, 2008)	Average local usage rates for the articles published in a journal	4	5

* Col. = data collection, Cal. = calculation

Innovation and social benefits

Eight measures of innovation and social benefits were sourced in the literature, four judged potentially useful for the individual, score 1 in citation collection and calculation. 'Knowledge exchange' and 'Dissemination in the public sphere' are counts of publication and dissemination activities that can include standardised weighting schemes to accommodate certain activities in the field the researcher is active in.

It is debateable if the questionnaire 'A tool to measure societal relevance' is a bibliometric indicator, but its results can be used bibliometrically if enough data is collected. 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. This approach is also used in 'Knowledge use' and 'Usage log data', but these are judged too complicated for the researcher to calculate alone as the required citation data is not readily available. 'Patent application' is a measure the researcher can easily utilise if he or she uses patents, however measuring the extent of being cited in patents and 'scientific proximity' requires access to patent and sector specific databases.

Table 7 Bibliometric indicators used to assess the level of innovation and societal benefits of a researcher's output

Innovation and Social Benefits	Designed to indicate	Complexity	
		Col*	Cal*
Knowledge exchange (Mostert, Ellenbroek, Meijer, van A., & Klasen, 2010)	Knowledge production, knowledge exchange, knowledge use and earning capacity	1	1
Dissemination in public sphere (Mostert, Ellenbroek, Meijer, van A., & Klasen, 2010)	Impact and use in public sphere (knowledge transfer)	1	1
Knowledge use (Mostert, Ellenbroek, Meijer, van A., & Klasen, 2010)	Impact on learning in stakeholders' environment.	5	1
Patent applications (Okubu, 1997)	Innovation	1	1
Citations in patents (Okubu, 1997)	Impact on or use in new innovations	5	1
Scientific proximity (Okubu, 1997)	Intensity of an industrial or technological activity	5	2
Usage log data (Bollen, Biet-Arie, & Van de Sompel, 2006)	User activity that expresses interest or preference	5	3
Tool to measure societal relevance (Niederkrotenthaler, Dorner, & Maier, 2011)	Aims at evaluating the the level of the effect of the publication, or at the level of its original aim	1	1

Sustainability

Fourteen indicators were identified, nine potentially useful for the researcher, complexity ≤3. Four indicators were designed to indicate sustainability at a journal level; Price index, immediacy, aggregate immediacy and cited or aggregated half-life. The remaining five are designed for use at the micro-level; c(t), m-quotient, AR-index, classification of durability and age-weighted citation rate (AWCR, AW and per-author AWCR). Of these five the age-weighted citation rate (AWCR, AW and per-author AWCR), c(t) and m quotient, which is h-dependent, are ratio-based models, AR is based on the square root of average citations per year and is also h-dependent. Classification of durability is a percentile based indication of the distribution of citations a document receives each year, adjusted for field and document type.

Table 8 Bibliometric indicators used to assess the sustainability of a researcher's output

Sustainability	Designed to indicate	Complexity	
		Col*	Cal*
Citation age c(t) (Egghe & Rousseau, 2000)	The age of citations referring to a researcher's work.	3	3
Aging rate a(t) (Egghe & Rousseau, 2000)	Aging rate of a publication.	3	4
Contemporary h-index h^c (Sidiropoulos, Katsaros, & Manolopoulos, 2007)	Currency of articles in h-core.	3	4
Trend H index h^t (Sidiropoulos, Katsaros, & Manolopoulos, 2007)	Age of article and age of citation.	3	4
Dynamic H-type index (Rousseau & Ye, 2008)	Accounts for the size and contents of the h-core, the number of citations received and the h-velocity.	3	4
M-quotient (Hirsch, 2005)	H type index, accounting for length of scientific career	3	2
AR-index (Jin, Liang, Rousseau, & Egghe, 2007)	Accounts for citation intensity and the age of publications in the core.	3	2
Discounted Cumulated Impact (DCI) (Ahlgrena & Järvelin, 2010; Järvelin & 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
Price index – PI (Price, 1970)	Percentage references to documents, not older than 5 years, at the time of publication of the citing sources	3	2
Immediacy index	Speed at which an average article in a journal is cited in the year it is published	2	2
Aggregate Immediacy Index (AII)	How quickly articles in a subject are cited	2	2
Cited half-life (CHL) & Aggregate Cited Half-Life (ACHL)	A benchmark of the age of cited articles in a single journal	2	2
Classification of durability (Costas, van Leeuwen, & Bordons, 2010; 2010b; 2011)	Durability of scientific literature on distribution of citations over time among different fields	2	3
Age-weighted citation rate (AWCR, AW & per-author AWCR) (Harzing, 2012)	AWCR measures the number of citations to an entire body of work, adjusted for the age of each individual paper	2	3

* Col. = data collection, Cal. = calculation

In summary, of the 114 indicators presented in this study, thirty, though possibly superior measures, require either special software, access to restricted data or demanding calculation (complexity score ≥4 in *either* effort to collect citation data or calculation). Consequently, these indicators are not considered useful for the individual researcher in self-assessment.

The remaining eighty-four indicators are judged potentially useful as they are rated ≤ 3 in *both* effort required in data collection and complexity of calculation. However, twenty of the twenty-one impact indicators were originally designed as measures of journal or group impact. Further studies are required to investigate their utility as performance benchmarks in evaluation at the micro-level. Seventeen indicators, from the quality and research infrastructure categories, are h-dependent and consequently suffer from the same inadequacies as h. Forty-five indicators are purely theoretical and not used in practice in evaluations hence their effects on the individual's performance remain unclear. Further, due to the added complexity of their foundational models and demands on data collection only 22 of these were judged useable by the researcher in self-assessment. These are: Costas age and productivity index (outcome), hg, normalized h, h_2, A, R, $\overline{h}$, m, e, hmx, w, Antonakis' index of quality and productivity, Q^2 (quality), hi, n, alternative h, pure h (research infrastructure), tool to measure societal relevance (innovation and social benefits), m-quotient, AR index, contemporary h and the variants of AWCR (sustainability).

Discussion
The significance of evaluation at the individual level has led to a flux of new indicators as well as new variants or combinations of established ones. However, it can be deduced from the literature used in this review that the development of new indicators appears to outweigh their practical implementation even though they proclaim to be (theoretically) superior. As indicators get more refined their complexity appears to increase. The benefits for the user of these more refined indicators are uncertain.

Within each of our categories of scientific activity there are many choices of indicator. Some are ready to be used, some need adaption to the context of evaluation, some measure the same thing and are information redundant if used together, while some can be improved by combining them to fit a particular situation using a locally defined benchmark or presented in context of academic age or gender. In presentation of their CVs, researchers must consider the affect the indicators have for their performance. These points, plus the advantages and limitations of the indicators, are discussed in the next section.

Output
Indicators of output provide information of the sum of a researcher's publications produced within a given timeframe. Data collection is simple and the indicators easily calculable by the researcher, however publications included in the count have to be verified with bibliographic data to support the credit. Clearly, 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 an equitable sharing of authorship credit is applied (Hagen, 2010). Yet if the field norm is multi-authorship, correcting for single contribution at an individual level is superfluous and perhaps counterproductive. The approach of harmonic counting fits ethical criteria of assessment at an individual level, this is when publication credit is shared proportionally among all authors, or the first author gets most credit, or the greater the number of authors the less credit per author. Meanwhile, arithmetic counting allots twice as much credit to the 1st author when there are only two coauthors but has no fixed ratio of allotment when the number of authors increases. First author credit decreases rapidly and continuously, whereas last author credit initially increases and thereafter decreases slowly as the number of authors increases. In the evaluation 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 even just alphabetical (Bennett & Taylor, 2003).

Count must be balanced by valorisation of different forms of publication, be it patents, books, book chapters, articles, enlightenment literature, conference papers etc., within the field in which the researcher is active. 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 scientific behaviour needs further investigation. Also, weighting makes the comparison to normalised national and international standards unreliable as type has to be compared with type, and this in some emerging fields, is impossible to do.

Outcome
Outcome indicators can be grouped into two methodologies: citation averages or percentiles. Calculations in both approaches appear relatively simple but the availability of data makes it questionable if the individual researcher can use them to produce reliable results. As field coverage is limited in citation databases, outcome indicators are more successful in some fields than others. Consequently worthwhile calculations of indicators based on citations require data collection in

multiple sources to provide as complete a picture as possible. This immediately adds to the complexity of the indicator. Clearly indicators that build solely on citation data are not comprehensive, stable or reliable and this questions their validity. Furthermore 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).

Percentiles such as Ptop, E(Ptop), A/E(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 & Werner, 2012). Bornmann argues for their simplicity of calculation, which is questionable, but they are more intuitive to the reader than average cites per paper in that visualization of results in box-charts or bar-charts can provide easy-to-read presentations of a researcher's performance. Percentages have the further advantage that they are scarcely affected by the skewed distribution of citation data and are adjustable to individual assessments as measures of excellence. Ptop, for example, can be adjusted to Ptop/researcher to illustrate the amount of papers a scientist has within the top 5% papers within a field, as presented in a comparative analysis of indicators of scientific excellence by (T. N. van Leeuwen, Visser, Moed, Nederhof, & Raan, 2003).

Returning to the issue of their simplicity, from the perspective of the individual calculating reliable percentages of performance to field or specialty is difficult and time-consuming. Also, field indicators favour some fields more than others; older articles, senior scientists with extensive publishing careers and often based in predefined subject categories in citation indicators. Hence may not be representative of the response to a researcher's work. The inclusion or exclusion of self-citations has a direct effect in citation counts in individual evaluation, and a policy decision supported by a statement of what exactly constitutes a self-citation needs to be established. In addition, 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 an individual level compromise the simplicity and stability of these comparative measures of excellence. Subsequently, is has been recommended not just to compare results obtained from several databases, but combine citation counts with other methods of performance evaluation and first thereafter normalise results of individual performance to academic seniority, active years and field to ascertain excellence (Costas, van Leeuwen, & Bordons, 2010a).

Consequently, citation counting requires extreme prudence and the ethical issues with constructing measures that account for the effect of age or gender of the researcher on productivity and outcome requires careful consideration.

Quality

Indicators of quality are an expression of cumulative impact in a single index, as they take the quantity and impact of articles into account (Hirsch, 2005; Schreiber, Malesios, & Psarakis, 2012). To do this comprehensively, the majority are recommended, by their creators, 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 (Haslam & Laham, 2009; Nederhof & Meijer, 1995). To overcome these shortcomings, quality is assumed reflected in citation counts as a large number of citations are interpreted as "usefulness" to a large number of people or in a large number of experiments.

The h-index already plays an important role in evaluation of scientific output at an individual level (Costas & Bordons, 2007) and despite its flaws, is unavoidable in self-evaluation as its simplicity and

recognisability outweigh debates of its representativeness. 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 (Iglesias & Pecharromán, 2007; Podlubny, 2005), having granularity issues, (Harzing, 2008; Vanclay, 2007), losing citation information (Waltman & 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 & Bordons, 2007). Perhaps, most importantly, is the questionable arbitrariness of the h parameter (Alonso, Cabreriazo, Herrera-Viedma, & Herra, 2009). Subsequently, the indicators that build on the h index suffer the same inadequacies as h, which could be problematic for twelve of the fourteen indicators of quality we deemed useful for the researcher. All of these criticisms must be accommodated for to produce a *valid* evaluation of the individual. Hence, the development of supplementary indicators to h aim to give a complete picture of 'quality' and novel indicators that are h-independent or correct for the flaws of h.

In this review, attempts to improve h can be seen to be at the cost of simplicity and usability. The descendants of h are supposedly more precise, yet their consistency and validity remains problematic. Some have performed well in laboratory studies: b (Brown, 2009), index of quality & productivity (Antonakis & Lalive, 2008), h-index sequences and matrices, (Liang, 2006), while others have faltered: h, g, r, h2 (Waltman & van Eck, 2009). Of course the indicators that incorporate h in their foundations suffer from the same inconsistencies as h: hg, q2, normalized h, Hrat, grat, a, hw, $\bar{h}$, e, hpd and hmx. Some indicators that are not related to the h-index also have inconsistency problems: π, f, t, hα, gα, hT. Others give undue weight to highly cited papers, h,f,t,w,h^2 (Schreiber, 2010). Generally, 'quality' 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: hw, w, hf, and x.

We judged fourteen out of twenty-eight indicators easily calculable, score ≤3, assuming the necessary information in citation databases was available. Twelve of these are h-dependent. The other indicators require multiple and advanced calculations: x, gα, hα, Hpd, π, hw, hf, t, f, hrat, grat, while two even require special software for calculation: hT, h-index sequences and matrices. Although the indicators proclaim higher accuracy and granularity, these benefits are lost on the end-user as usability and transparency are reduced. Also, the determination of cut-off values, parameters, stretching the exponential distribution to fit the dataset or field characteristics increases consistency problems as well as confusion over which data is included in the calculation. Not least, if information is lost during data manipulation, validity is challenged and comparability of index values unwise (Iglesias & Pecharromán, 2007).

Information redundancy must be addressed as it is recommended to combine h variants to compensate for limitations of single indicators (Panaretos & Malesios, 2009). Even without statistical analyses, we can observe a large overlap between indicators presented in the table. Our observation is supported in (Bornmann, Mutz, Hug, & Daniel, 2011) who investigate correlations and convergent validity of h and 37 variants. The findings of this meta-analysis reveal high inter-correlations between h and its variants, and conclude that the various indicators can be redundant in empirical application. Separating the indicators in categories "fundamental" and "derived" reduces the chance of information redundancy in evaluations (Zhang, 2009) where, for example, a and R, are h-dependent (derived) and thus have information redundancy with h. Both Bornmann and Schreiber recommend a

more user-friendly approach, that is to categorise and combine pairs of indicators relating to the productive core: h,m,q,h2 w, h(2), h, $\tilde{h}$, f, t, $\bar{h}$, ht, x, with indicators relating to the impact of papers a, r,ar, hw, a, g, $\tilde{g}$, m, hw, r, π and e to produce insightful results (Bornmann, Mutz, & Daniel, 2008; Schreiber, Malesios, & Psarakis, 2012).

The indicators discussed in this section all stand for slightly different dimensions of quality of output. The integer number that is h disguises a lot of variations in ratios in the h-core, such as 'quantity of publications to citations', 'age of citations to publications' and 'highly cited to mediocre publications'. Consequently, the choice of h-type index is confusing, and the benefit of choosing one over the other is, for the researcher in self-evaluation, negligible. There is an acute need to validate these indicators empirically using researchers from different academic seniorities and disciplines and to assess the stability of the indicators differing amounts of publications. The use of h type indicators that establish quality benchmarks at a lower level of aggregation than field standards has been suggested by (Arencibia-Jorge, Barrios-Almaguer, Ferdandez-Hernandez, & Carvajal-Espino, 2008). They aggregate successive h indexes to account for performance on a 'researcher:department:institution' hierarchy. We suggest extending this in grouping experts in a specialty or adapting to on even lower level of aggregation, 'gender:academic seniority:department'.

There is clearly no need to introduce more h-index variations until it has been proven that the existing ones are redundant in real examples (Alonso, Cabreriazo, Herrera-Viedma, & Herra, 2009). If a single index has to be used, the simpler ones appear to work just as well as the complex one (Schreiber, Malesios, & Psarakis, 2012), especially, as suggested in this review, if simple ones from each category are combined to give a comprehensive picture of scientific activity**.**

Research Infrastructure

Assessment of research infrastructure is important for the individual researcher in evaluation because it lies at the heart of expressing research activity, illustrating knowledge advancement and identifying communication and possible collaborations. The indicators concern 1) collaboration and attributing fair credit for contribution to papers, and 2) illustration of the visibility and usage of a researcher's work. Both of these approaches require detailed collection of citation and publication data from multiple sources, while the latter also requires specialist network analysis software and user-instruction. This in turn increases the complexity. Application and reliability in areas with poor coverage in citation databases requires consideration prior to implementation.

Simple indicators of research infrastructure shouldn't be difficult to calculate because the author 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 its calculation and understanding of what the indicator actually represents. Hence usability is reduced. 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 evaluating the infrastructure of a scientist's papers 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 practice authorship is often rewarded on the background of 'political' or publishing agreements, or simply as thanks for

access to data. From the evaluees point of view, the desirability of correcting for co-authorship is doubtful. Such a researcher is Rosenberg who pleads for indicators that avoid recalculation of the h-core as they can lead to over-correction and thus penalise the author under evaluation (Rosenberg, 2011). One guide to choosing an indicator is referencing uniform requirements to manuscripts in the field the researcher is active. The International Committee of Medical Journals requires, for example, author's rank in the by-line is decided by level of contribution which is verified in an authorised statement of intellectual contribution. Hopefully, this approach will reduce academic doping, that is, collegial 'under the table publishing agreements' which can drastically inflate publication and citation counts (Hessel, 2013). Bibliometrically, an authorised level of contribution could be used to weight publication and citation count. Thus a fair choice of indicator is one that fits these requirements, and adheres to ethical criteria presented in (Hagen, 2010), as previously discussed in *output*. The question is, if sharing credit is at all necessary. Realistically, researchers in self-assessment will write the highest number of citations their works have achieved. If all authors 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 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

Indicators of impact 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. Impact indicators need to be designed specifically for the individual level and account for the context of application and correlate with peer review assessments. However, there is a major conceptual flaw that needs to be corrected before indicators of individual impact can be established and that is: What is impact?

Impact and quality are not identical concepts, just as the impact and utility of research to users in society are different variables (Nederhof & Meijer, 1995; Satyanarayana, 2010). Yet impact continues to be used as a proxy for quality and the impact factor is mistakenly regarded as a useful yardstick measure of performance of individual publications within the discipline. Without normalisation for field, subject category, document type, and publication year impact figures have very little meaning (Bornmann & Werner, 2012). Normalised impact factors such as, Y, JCSM/FCSm, CPP/FCSm, CPP/JCSm, JCSm/FCSm, C/FCSm (plus the recently named Leiden Mean Citation Score MCS and Mean Normalised Citation Score MNCS) were not designed for evaluating the individual researcher's activities, but for comparing research groups or journals to the mean citation rate of a subfield to suggest the expected performance of a paper published in a discipline (Glänzel, 2003). They say nothing about the impact of a single article independent of journal performance. Likewise synchronous and diachronous impact factors, normalised JIF, JFIS, disciplinary IF, NJP, median impact factor, field impact and FCS are not impact measures of "citedness" but in fact represent a related measure, that is the chance for citedness resulting from the relative contribution of the *journal* to the overall impact of an entire *set of journals*. Clearly, the "traditional" impact factors are not suitable measures or benchmarks of an individual's impact and their correlation with peer review is questionable (Waltman, Eck, Leeuwen, Visser, & Raan, 2011).

Using impact factors out of their context is a problem when discussing their validity or rather the validity of the use made of the measure (Lundberg, 2009). Studies illustrate that in an assessment of the validity or applicability of journal impact indicators it is crucial to take into account the context of

the application, particularly the research questions and the policy issues addressed (van Leeuwen & Moed, 2002). The same considerations must apply in impact judgements of the individual. As there is no workable definition of scientific impact, there is no agreement on which combination of indicators best express the impact of an individual's body of work or which best fit the aim of an evaluation of the impact of an individual. But there is at least agreement that using just one indicator is inadequate. This inadequacy is discussed in both Bollen et al in a cluster analysis of 37 impact factors and in van Leeuwen et al in a comparative analysis of indicators of scientific excellence (Bollen, Van, Hagberg, & Chute, 2009; van Leeuwen, Visser, Moed, Nederhof, & Raan, 2003). Interestingly, Bollen et al showed that scientific impact can be roughly categorised as rapid or delayed when based on usage data or citations (Bollen, Van, Hagberg, & Chute, 2009). But, as the investigation was based on *journal* impact it is necessary to study if time and impact of journals correlate in the same way in individual impact. Consequently, it needs to be investigated if the reliability of "prediction of article impact" could give a fairer evaluation of a young set of an individual's papers. This approach is however limited to well-established article-based fields.

Alternatively, usage-based measures are considered beneficial in calculating an individual's impact, here "use" is equated with downloads or views, thus activity outside of the "journal network" such as online (non)scientific websites or blogs can contribute to impact judgements and provide insight into social impact. According to Yan & Ding (2011) social impact is illustrated in the intensity, patterns and origin of online usages. The main advantage with usage measures, Weighted PageRank, Closeness and Betweeness Centrality, are that they perform as indicators of prestige, in contrast to journal-based citation indicators ISI JIF, Scimage Journal Rank, cites per doc, which are dependent on journal performance and have been shown to measure popularity. Popularity is not considered a core notion of impact (Bollen, Rodriguez, & Van, 2006; Bollen & Sompel van de, 2008; Yan & Ding, 2011).

In this review only one indicator of impact was identified as designed for evaluation at the individual 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, to support non-journal based fields. Although interesting $\bar{c}f$ and the logarithm based citation z-score, (the indication of local impact accounting for field variability at an item level) were excluded due to the complexity of calculation hence the utility of both these indicators to the researcher in self-evaluation is compromised.

In conclusion, Impact indicators must be used with care if used as benchmarks of individual performance, (Moed, 2005) especially if they are normalized to a field and attempt to account for research questions and other methodological variables. It is doubtful if researchers can feasibly indicate their global impact though indicating local impact using P_{tj} is one answer. But this gives a one dimensional measure of impact, and in an evaluation it is important to define which part of impact is best expressed with which combination of which indicators. This review attempts to answer that by encouraging the expression of impact as combined measures from the categories output, outcome, quality, research infrastructure, sustainability, and innovation and social benefits as a collective representation of a researcher's overall impact.

Innovation and social benefits

Despite the societal character of research investment, scientific quality is evaluated bibliometrically through publication count and citation impact. This is under active revision as both publication count and citation impact are limited to communication within the scientific community and underplay the communication, and use, in relevant industrial, private and public sectors (Mostert, Ellenbroek,

Meijer, van A., & Klasen, 2010). Societal impact is an attempt to judge the social, cultural, environmental and economic returns from publically funded research (Meulen van der & Rip, 2000; Okubu, 1997). Current evaluation procedures attempt this by combining contextually relevant qualitative and quantitative indicators constructed in dialogue with the individual under assessment (Rymer, 2011). In the working methods for the Research Excellent Framework 2014, appendix 1, HEFCE recommend case studies and peer review to provide evidence-based evaluation of social benefits of the research (REF2014, 2012). Another approach is the combination of qualitative measurements of knowledge production, knowledge exchange, use and earning capacity with quantitative analysis of citations, reference lists and footnotes of laws, patents, protocols, regulations and guidelines (Mostert, Ellenbroek, Meijer, van A., & Klasen, 2010; Spaapen, Dijstelbloem, & Wamelink, 2007). Yet the credibility of correlation between papers and patents is uncertain, as patents serve a legal purpose and authors can attempt to conceal content from their competition. Therefore, opinion is divided on the importance and significance of citations in patents.

More recently, Neiderkrotenhaler et al (2011) suggested a simple questionnaire-based tool to indicate the societal impact of publications in the biomedical sciences by combining the interest of societal stakeholders with quantitative indicators of knowledge dissemination and use. They attempt to assess the effect of the publication in non-scientific areas, the motivation behind the publication and efforts by the authors to translate their findings. This tool has the potential to translate well in to other fields as it is flexible enough to allow for the differences of societal aspects between disciplines in connecting the aims of research to the perceived value of their outcomes. The different types of societal impact are suggested to be impact on beneficiaries (individuals, organisations, communities, regions, processes, behaviour or practices), society, culture and creativity, economy, commerce, public policy and services, production, practitioners and services, and the environment whether regionally, nationally or internationally. Claims must be supported by evidence and indicators take different forms depending on the type of impact they support - indicators are demanded to be "meaningful, contextualised and precise to support the evidence". A similar approach differentiates between societal quality, impact and valorisation, using contextually relevant indicators (Drooge et al 2010; SEP, 2010), but it is unclear in the working methods which indicators are recommended. Interestingly, in the guide by Drooge et al, there distinguishes between evidence of societal benefits that is available from retrospective analyses and evidence that will require a prospective study to collect.

Usage log data has the potential for interesting societal analyses,(Bollen, Biet-Arie, & Van de Sompel, 2006) but definition of usage and what it represents requires clarification before implementing in an evaluation. Further data and software accessibility, complexity of analyses, falsification and validity of data, privacy issues, and time-issues can deter the individual in using click-stream datamining in self-evaluation. In 2000 Wormell suggested text mining techniques to extract knowledge from literature concerning the topic 'Welfare' to thoroughly identify the topic's structure, developments in time intervals and a researchers contribution (Wormell, 2000). She indicated patterns and developments in the number of publications, term occurrences, similarity between the subject terms and formation of clusters among the subject segments to provide a comprehensive picture of trends influencing social policy and public opinion. This provided a useful pool of knowledge for individual researchers to use as a benchmark to validate their own innovation and contribution to societal benefits within this topic. However, the analysis work was extensive and had to been done on the behalf of the individual researcher and updated at regular intervals to ensure its currency.

Clearly, societal impact is harder to measure than scientific impact and there are (as of yet) no standardised indicators that can be used across all disciplines and institutions nor is there a method of evidence collection recommendable to the individual researcher. High scientific quality is not necessarily related to high societal quality, but perhaps most important for evaluation is the to acknowledge that societal benefits can take many years to become apparent and the routes through which research can effect behaviour or inform social policy are diffuse. We can agree that defining social benefits of research is challenging and measuring it appropriately even more so (Bornmann, 2012).

Sustainability

It is incorrectly assumed that the chance of a researcher's work being used declines with age as 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.

Non-valid information can still be useful for the growth of science and non-used publications can be caused by other factors than lack of validity as 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 publication (Archambault & Larivière, 2010; Costas, van Leeuwen, & van Raan, 2010b; Egghe & Rousseau, 2000). In addition, 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 aging but does not cause aging (Egghe & 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. It is therefore vital to present the researcher with a validated field "age" norm relevant to their specialty when evaluating sustainability.

Measures of sustainability have to cope with these diverse characteristics and fluctuations in usage by local groups. Cited half-life, immediacy index and their aggregated versions apply only to journals, not individual articles but are nonetheless widely used as performance benchmarks in individual evaluation. 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 more relevant indicators of sustainability at the micro-level.

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, PI, 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. Yet in Costas et als' "classification of durability" there is presented a simple percentile distribution of citations to documents normalised to field and document type. This index detects the possible effects that durability can have on the measurement of the performance of the individual, in an easily understandable form and is worthy of further empirical investigation (Costas, van Leeuwen, & van Raan, 2010b).

Demands to the calculation of indicators in individual self-assessment

Indicating scientific activity using bibliometrics is based on a mathematical framework that attempts to account for the quantity of publications and the effect, documented in forms of citation, they have had on the surrounding community. Without considering what the indicator expresses or its theoretical foundations, the indicator is purely instrumental and can be used inappropriately to distort, reduce, or enhance the elements of a researcher's CV that benefit from being distorted, reduced or enhanced. What the purpose of the self-evaluation is, what indicators do or do not measure and how to interpret the results has to be clear for the evaluators and the evaluand before any indicators are implemented.

This reviews shows that indicators that purport to measure the same aspects of a researcher's scientific activity produce different results because their mathematical foundations are different. Stochastic or deterministic mathematical models, that are the foundation of indicators, don't convey anything about the physical or social causes behind data production in the wide range of bibliographic and non-bibliographic recorded activities (De Bellis, 2009; Glänzel, 2006). For instance, fractional counting that adjust for the authors name rank in the by-line and number of authors credit contribution on one scale, while dividing the number of citations received by a paper by the square root of the number of co-authors to remove the dependence of co-authorship credit contribution on another (Carabone 2011). Accordingly the goodness of fit of the chosen mathematical model on the bibliometric data relative to researchers profiles within their field is vital as the fit balances a high or low production rate with the expected field norm for that academic position, gender and publishing history (Costas, Bordons, van Leeuwen, & van Raan, 2009; 2010a; 2010b). This is why inter- and intradisciplinary comparisons demand users of bibliometrics are aware of field specific publication and citation traditions and understand the influence these have on citation-based indicators (Alonso, Cabreriazo, Herrera-Viedma, & Herra, 2009; Iglesias & Pecharromán, 2007; Wagner et al., 2011).

To compare individual performance with peers, field normalization is recommended. Here the field is fixed as a reference to calculate normalizing factors by a multiplicative correction (Iglesias & Pecharromán, 2007), thus assuming that publication and citations are independent variables. In other words the effect of the publishing size on the citation count has been eliminated. Studies have shown that normalized indicators characterise the area but can be disadvantageous for the specific publication patterns of a researcher within his sub-field specialty (Bollen, Rodriguez, & Van, 2006; Ingwersen, Larsen, Rousseau, & Russell, 2001; van Leeuwen & Moed, 2002; Yan & Ding 2011). Further, normalization favours highly cited authors as impact increases in a power law relationship to the number of published papers (Iglesias & Pecharromán, 2007) which is why the law of the constant ratio is advantageous in comparing researchers' of low or average impact to their peers. Using this viewpoint of actual citations to works results in simple discipline to discipline citation ratios, e.g. where 1 citation in maths roughly corresponds to 15 in chemistry, thus acknowledging the complex reality of comparing researchers who work in increasingly multi- and interdisciplinary fields. It is also beneficial to account for the number of people and publications in different fields through the total number of citations produced by the people in those publications (Podlubny, 2005),

By combining indicators researchers can illustrate publication rate over time, document type-specific performance, presence in scientific communication (adjusted for field, seniority and gender) and provide an indication of the use and impact of their research in the scholarly community. However, using a series of indicators to capture such scientific activities has mathematical implications due to the structure of the data these indicators analyse. It is commonly known by bibliometricians that citation data is highly skewed and if the distribution is very skewed and far from a normal distribution, the mean and the standard deviation may be misleading measures (Bornmann & Werner,

2012; Lundberg, 2009). How should individual researchers handle this in self-evaluation, especially if correction is detrimental to their "scores"? By stabilizing the variance of the distribution of a skewed dataset so it exhibits a normal distribution, approximately standard normal variables can be managed in bibliometric analyses making analyses simpler and results comparable. Lundberg (2009) argues for the benefits of logarithmic transformation of citation rates to avoid using the geometric mean. Stability of indicators on small datasets, as will often be the case in individual evaluation, will be improved using transformed data but the transformation of data symmetry can significantly change the outcome of descriptive statistics. The benefits of this approach have to be examined critically before encouraging the individual to use them as overcompensating with mathematical formulas can lead to bad statistics, unwise comparisons and researcher's 'enflating' their CVs (De Bellis, 2009; Schreiber, Malesios, & Psarakis, 2012).

Demands to bibliometric indicators in self-assessment

It is obvious from the indicators presented in this review, that bibliometric self-evaluation goes beyond citation count and journal impact factor. Clearly, a single number will only give a rough approximation of an individual's multifaceted dissemination profile and it is recommended that indicators are combined in a well-designed method to facilitate a useful evaluation, as there are many indicators to choose from, each with their own strengths/weaknesses and researcher/field variables that can be redundant or counter-productive when used together (Bornmann, Mutz, & Daniel, 2008; Costas, van Leeuwen, & Bordons, 2010a; Franceschet, 2009; Jin, Liang, Rousseau, & Egghe, 2007; Retzer & Jurasinski, 2009; van Leeuwen, Visser, Moed, Nederhof, & Raan, 2003). As citation and publication data are used to inform dialogue with management on a departmental or institutional level, bibliometric evaluation demands the methodological strategy tailored to the aim of the evaluation.

If the assessment is to produce valid information useful to both the individual and the evaluation committee, a high level of attention to detail is demanded in the design of a replicable strategy and the consistency of interpretation. A bibliometric strategy has to employ understandable indicators that account for the individual's academic seniority and profile, discipline, publishing channels and scientific activities. This requires a complete data set of the researcher's oeuvre not just for statistical stability but to produce unbiased results, as possibilities and limits of indicators are dependent on the availability and quality of data.

Problems with data accessibility, English language bias in citation databases and missing publication and citation data limit performance analyses of measurable outcome and that can directly affect interpretations of the performance of the researcher, (Bach, 2011; Rousseau, 2006). Further, the combination of indicators have to fit:
the framework of disciplinary traditions and expectations (Batista, Campiteli, Kinouchi, & Martinez, 2006); the originality of the presented research or the further development of theories and methodologies; the presence of the researcher in national or international scholarly organizations; the involvement in projects with a socio-cultural relevance for the community; the dissemination in enlightenment literature and the application and utility of the work in practice (Hicks, 2004; Mostert, Ellenbroek, Meijer, van A., & Klasen, 2010; Must, Otsus, & Mustajoki, 2012; SEP, 2010).

The key challenge for self-evaluation then is its feasibity. Can the researcher complete it in regards to data collection, time and finances (Burnhill & Tubby Hille, 1994; Ingwersen, 2005)? What is or is not possible to evaluate must be clear as this can be both advantageous and detrimental to the researcher's CV to limit the evaluation (SEP, 2010). Assessment of the individual's production must go beyond

interpretation of patterns in bibliographic data to factor in differences in the granularity of measurements and assessment (Batista, Campiteli, Kinouchi, & Martinez, 2006; Wagner et al., 2011). This is a lot to demand of the individual who surely wants just to enrich his CV to his advantage. However, results of evaluations have been proven to contribute to both positive and negative culture changes in publishing activities of individuals, (Haslam & Laham, 2009; HEFCE, 2009; Hicks, 2004; Hicks, 2006; H. F. Moed, 2008) and with this is mind indicators must be verifiable at the individual level as, 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 & Jurasinski, 2009).

Methodological considerations

This review is limited to a subjective assessment of the characteristics of indicators at the individual level. We have not investigated empirically indicator applicability, validity, utility, objectivity, effect on the individuals publishing behaviour, cause and effect mechanisms inherent to the indicator, or inter-field variations of the indicators when implemented. These need to be analysed in future studies. Neither, have we considered the ethical implications of self-evaluation to strengthen and support an individual's CV. Further, input and process indicators were excluded from the review. Even though these have an important role for the execution of scientific activities, indicators of investment and expenditure fall outside the scope of the bibliometric assessment of publications and citations data.

Conclusions

The focus of this review is to judge the utility of indicators for researchers, in self-evaluation, to document scientific activities and publication performance on their CVs. The indicators are categorised as *output, outcome, quality, research infrastructure, impact, innovation and social benefits,* and *sustainability*. These are presented in tables to exemplify how this range of scientific activities can be collectively assessed and the advantages and limitations of each indicator are presented. This structure was chosen to emphasise that at the current time 1) certain scientific activities and publication performance are more easily evaluated using bibliometrics than others, 2) assessment of scientific activity and publication performance cannot be represented by a single indicator, 3) it is unwise to use citations as a proxy of research quality, 4) choice of indicators can have a direct positive or negative effect on the outcome of the evaluation of the individual and 5) the assessment can easily be biased towards for whom the results are for and by whom the assessment is conducted. The usability of indicators and the transparency of their mathematical composition are questioned. The types of 'quality' indicators can measure are presented.

A thorough self-evaluation requires the combination of quantitative and qualitative assessment methods. Which indicators and how these are combined to best express a researcher's 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 strategy of indicators for self-assessment, as well as locally relevant performance benchmarks, the researcher will reach a better understanding of the achievements of their published works and perhaps identify where this can be improved. Hopefully objective self-evaluation will contribute to an *informed* assessment, to research management at the institutional, faculty and departmental level, promote organisational learning and validate funding decisions. The success of the indicators are though 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. Until the information community addresses data completeness and accessibility, instead of inventing new indicators, measures of societal activities and performance evaluation in the "softer" sciences will lag behind.

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. Therefore, simple indicators are concluded to be better for individual self-evaluation as their requirements to bibliographic data are modest and calculations transparent. However, even though there is undoubtedly potential in self-evaluation to support a CV in an evaluation, extreme caution is called for as ethical issues have yet to be explored and a need for guidelines for Good Evaluation Practices is urgent.

Acknowledgements
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 Francoise Caillods (Ed.), *World Social Science Report 2010* (pp. 251-254). UNESCO publishing.

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. Resource document. Académie des sciences. http://www.academie-sciences.fr/activite/rapport/avis170111gb.pdf. Accessed 5 April 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., Biet-Arie, O., & Van de Sompel, H. (2006). Alternative metrics of journal impact based on usage data - the bX project. Powerpoint lecture presented at the *Meeting on Alternative Metrics of Publication Impact,* Humboldt University: Berlin, 20-21 February. www.dini.de/fileadmin/workshops/februar-2006/bollen.ppt Accessed 11 April 2013.

Bollen, J., Rodriguez, M., & Van, d. S. (2006). Journal status. *Scientometrics,* 69(3), 669-687.

Bollen, J., & Sompel van de, 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.

Bollen, J., Van, d. S., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. Resource document *PLoS ONE.* http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0006022. Accessed 5 April 2013.

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

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.

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.

Carabone, V. (2011) Fractional counting of Authorship to Quantify Scientific Research Output. *arXiv:1106.0114 [physics.soc-ph]*

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.

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

Chai, J., Hua, P., Rousseau, R., & Wan, J. (2008). The adapted pure h-index. *Proceedings of WIS 2008: Fourth International Conference on Webmetrics, Informetrics and Scientometrics & Ninth COLLNET Meeting,* Berlin. doi: www.collnet.de/Berlin-2008/ChaiWIS2008aph.pdf

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.,N., & 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.

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

Directorate-General for Research. (2008). *Assessing Europe's University-based Research.* Belgium: The European Commission.

Drooge van, L. and Besselaar van den, P. and Elsen, G.M.F. and Haas de, M. and Heuvel van den, J.J. and Maassen van den Brink, H. and Meulen van der, B. and Spaapen, J.B. and Westenbrink, R (2010). Evaluating the societal relevance of academic research: A guide. Resource document. ERIC: Evaluating Research in Context. http://depot.knaw.nl/9026/ Accessed 5 April 2013.

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

EFC. (2010). *Research evaluation guidelines: Why evaluate?* Resource document. EFC: European Foundation Center. http://www.efc.be/programmes_services/resources/Documents/Research_Evaluation_Guidelines_1_Why_Evaluate.pdf. Accessed 11 April 2013.

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.

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: http://dx.doi.org/10.1016/j.ipm.2007.05.003

Franceschet, M. (2009). A cluster analysis of scholar and bibliometric indicators. *Journal of the American Society for Information Science and Technology,* 60(10), 1650-1964.

Franceschet, M. (2010). The effect of scholar collaboration on impact and quality of academic papers. *Journal of Informetrics,* 4(4), 540-553.

Glänzel, W. (2003). Bibliometrics as a research field: A course on theory and application of bibliometric indicators. http://nsdl.niscair.res.in/bitstream/123456789/968/1/ Accessed 5 April 2013.

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.

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. Resource document. Harzing.com. http://www.harzing.com/pop_hindex.htm Accessed 5 April 2013.

Harzing, A. (2012). Publish or perish user's manual. Resource document. 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

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.* Resource document. Higher Education Funding Council for England. https://www.hefce.ac.uk/pubs/rereports/year/2009/bibliometricspilotlessons/ Accessed 11 April 2013.

Hessel, Niklas. (2013, February 15) Akademisk Doping. *Weekendavisen*, section 1, p.1.

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

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

Hicks, D., & Wang, J. (2009). Towards a bibliometric database for the social sciences and humanities. Resource document. Bepress. http://works.bepress.com/diana_hicks/18/ Accessed 5 April 2013.

Hicks, D. (2012). One size doesn't fit all: On the co-evolution of national evaluation systems and social science publishing. Confero, doi: 10.2284/confero13v1130117

Hirsch, J. (2005). An index to quantify an individual's scientific research output. *, 102 (46)* 16569-16572.

Hirst, G. (1978). Discipline impact factors: A method for determining core journal lists. *Journal of the American Society for Information Science,* doi: 10.1002/asi.4630290403.

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

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

Ingwersen, P. (2005). Scientometri: Videnskabspublicering og bibliometriske metoder. Ressource document. Biokemisk Forening. http://www.biokemi.org/biozoom/issues/506/articles/2162. Accessed 5 April 2013.

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.

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.

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

Leimu, R. (2005). What determines citation frequency of ecological papers? *Trends in Ecology and Evolution,* 20(1), 28-32.

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

Meulen van der, B., & Rip, A. (2000). Evaluation of societal quality of public sector research in the Netherlands**.** *Research Evaluation,* 8(1), 11-25.

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

Moed, H. F. (2005). *Citation analysis and research evaluation*. The Netherlands: Springer.

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

Moravcsik, M. (1986). Assessing the methodology for finding a methodology for assessment. *Social Studies of Science,* 16(3), 534-39.

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. Resourse document. Health

Research Policy Systems. http://www.biomedcentral.com/content/pdf/1478-4505-8-30.pdf Accessed 5 April 2013.

Must, Ã, Otsus, O., & Mustajoki, A. (2012). *Task 1.2. ACUMEN survey on peer review practices.* Resource document. ACUMEN. http://research-acumen.eu/wp-content/uploads/Del-1.3_27Dec2012.pdf

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), 229-30.

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

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. 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.

REF2014. (2012). Panel criteria and working methods. Resource document. Research Excellence Framework 2014. http://www.ref.ac.uk/pubs/2012-01/. Accessed 5 April 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. Resource document. Rosenberglab.net. www.rosenberglab.net/Pubs/Rosenberg2011_ImpactFactor.pdf . Accessed 5 April 2013.

Rousseau, R. (2006). New developments related to the hirsch index. Resource document. 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

Rymer, L. (2011). Measuring the impact of research - the context for impact development. Resource document. Group of Eight Backgrounder 23. http://www.go8.edu.au/__documents/go8-policy-analysis/2011/go8backgrounder23_measimpactresearch.pdf. Accessed 5 April 2013.

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., & Sandstrøm, 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.

Satyanarayana, K. (2010). Impact factor and other indicies to assess science, scientists and scientific journals. *Indian Journal Physiol Pharmacol, 54*(3), 197-212.

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

SEP. (2010). Standard Evaluation Protocol 2009-2015. Resource document. The Royal Netherlands Academy of Arts and Sciences.

http://www.knaw.nl/Content/Internet_KNAW/publicaties/pdf/20091052.pdf. Accessed 11 April 2011

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

Spaapen, J., Dijstelbloem, H., & Wamelink, F. (2007). *Evaluating research in context: A method for comprehensive assessment*. The Hague: Consultative Committee of Sector Councils for Research and Development (COS).

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., & Moed, H. F. (2002). Development and application of journal impact measures in the dutch science system. *Scientometrics, 53*(2), 249-266.

van Leeuwen, T. N., Visser, M., Moed, H., Nederhof, T., & Raan, A. V. (2003). The holy grail of science policy: Exploring nd 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., & van Eck, L. (2009). A Taxonomy of Bibliometric Performance Indicators Based on the Property of Consistency.Resource document ERIM Report. repub.eur.nl/res/pub/15182/ERS-2009-014-LIS.pdf. Accessed 5 April 2013.

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

Waltman, L., Eck, N. J., Leeuwen, T. N., Visser, M. S., & Raan, A. F. J. (2011). On the correlation between bibliometric indicators and peer review: Reply to Opthof and Leydesdorff. *Scientometrics,* doi: 10.1007/s11192-011-0425-7

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

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

Wormell, I., (2000) Critical aspects of the Danish Welfare State - as revealed by issue tracking. *Scientometrics*, 48(2), pp.237-250

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

Appendix 1. European Research Evaluation Agencies sourced in establishing; use of bibliometric indicators in practice, aim of assessment and definition of categories of research activities. The information used to build this table is publically available on the agency's website, accessed via Google, last updated December 2012.

Country/Agency	Primary unit of evaluation	Primary bibliometric analyses	Self-assessment	Peer review	Aim of assessment
Austria/ ERA	Research discipline at university institutions. Selected papers	P & quality ranking. Publishing frequency indicator for books, book chapters and conferences; network & citation analysis inc. centile relative counts.	no	yes	Assessment of quality, activity, application and recognition & esteem
Belgium/ ULB	Discipline specific research teams	No use of bibliometrics. Under discussion	no	yes	Improve performance, assess quality and achieve excellence.
Denmark/ Action plan for Research Evaluation	Funding instruments, areas of research & research programmes	Area adjusted publication volume and field normalized citations	yes	yes	Accumulate knowledge and increase visibility of research, including commercial and social impact
Finland/ AALTO	Dept., institutes, groups & academic seniorities	Unclear or no metrics	yes	yes	Quality, impact, esteem societal impact & potentials
Finland/ UH RAE	Research departments and institutes	P, FCSm and JCSm, top 10% highly cited publications and network analyses	yes	yes	Assessment exercise for university's own purposes of the quality of research
France/ AERES	Teams & centers	Production weighted per discipline with citation & network analysis	yes	yes	Funding based on an excellence ranking within the same field
Germany /CHE Ranking	University profile (selected papers)	Weighted & raw P per discipline; multi -dimensional ranking. CPP	yes	no	Benchmark performance and stimulate competition
Germany /Forschungsrating	Research units, university and non-university, (selected papers)	P & top 10% P. Citation count (raw and normalized),	yes	yes	Guidelines for: quality, impact/effectiveness, efficiency, promotion of young researchers, knowledge transfer.
Germany /Initiative for Excellence	University	Unclear. P, collaboration, JIF in selected areas.	no	yes	Identify excellence and allocate/attract funding. Encourage competition and gender equality.
Hungary /Maintainer Agreements	University	Unclear other than a suggested publication, citation and collaboration count.	yes	no.	Monitoring and assessment of education and research. Increase efficiency and improve quality
Italy /CIVR	University & departments, selected output	No use of bibliometrics	no	yes	Establish guidelines for research evaluation and funding
Netherlands /SEP	Institutes, groups of researchers and sets of academic seniorities	CPP compared to FCSm and JCSm. Centile publication ranking & most important books/chapters. Network analyses	yes	yes	Improve research performance, quality, management & dispersion of funds
Sweden / A new model for allocation of resources'	University	Area adjusted publication volume and field normalized citations	no	no	Allocation of resources, quality incentives & strategic independence based on bibliometric analyses and external funding
UK REF2014 (HEFCE)	Departments, institution, university	Citation analysis and impact assessment in economy, society, culture, public policy and services, production and environment.	yes	yes	Quality of research in higher education institutions

Output indicators and their dimensions

All indicators require verified publication data.

Indicator	Definition	Designed to indicate	Individual		Complexity		Comments
			Advantages	Limitations	Col.*	Cal.*	
P	Total counting. Each N author of a paper receives 1 credit.	Count of production used in formal communication	Potentially, all types of output can be included or selected in regards to theme of evaluation.	Does not measure importance, impact of papers, duration or volume of research work.	1	1	Counts vary across disciplines due to nature of work and conventions for research communication.
P_{isi}	Number of papers in ISI processed publications	Used in the calculation of impact compared to world subfield citation average based on ISI citation data.	Recognised benchmark for analyses and bibliometric research projects.	Includes only ISI defined normal articles, letters, notes, reviews and conference papers.	1	2	Provides a distorted or incomplete picture; more appropriate in some fields than others (Harzing, 2012).
P_{ts}	Publication in selected sources	Number of publications in selected sources defined important by the researcher's affiliated institution.	Reflects output in sources deemed locally important.	Provides only a snapshot of productivity	1	2	Provides a distorted or incomplete picture
Co-publications	Count or share of co-authored publications.	Collaboration on departmental, institutional, inter- or national level & identify networks.	Shows with whom researcher co-publishes and the intensity of co-publication	Usefulness is affected by how the identification of affiliation and partnerships is handled.	1	1	Identifies if collaboration is governed by immediate proximity.
Fractional counting on papers	Each of the N authors receives a score equal to 1/N	Shared authorship of papers gives less weight to collaborative works than non-collaborative ones.	Accounts for differences in publishing behaviour among fields of science and level of multi-authorship.	Favours secondary authors by allocating equal credit to all authors	1	2	Criticized for lack of fit between credit scores and contribution (Hagen, 2010)
Proportional or arithmetic counting	Author with rank R in by-line with N co-authors (R=1,..N) receives score N+1-R	Shared authorship of papers, weighting contribution of first author highest and last lowest.	Rewards level of contribution to a paper.	If authors adapt alphabetical ordering or take turns to be first or second author this indicator cannot be applied.	1	2	Can be normalized in such a way that the total score of all authors is equal to 1.
Geometric counting	Author with rank R with N co-authors re-ceives credit of 2N-R	Assumes that the rank of authors in the byline accurately reflects their contribution	The first few authors get most of the credit	Allotted authorship credit rapidly approximates asymptotic values as N increases.	1	2	Asymptopic values lose their validity on small sample size.
Harmonic counting	Ratio of credit allotted to ith and jth author is j:i regardless of total number of co-authors	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.,	Provides accurate representation of perceived quantitative norms of byline hierarchy.	Applies only in areas where unequal co-author contributions are the norm.	1	2	Tested in natural sciences
Noblesse oblige	Last author gets 0.5 credit, other N-1 authors receive 1/(2(n-1)) each	Indicates the importance of the last author for the project behind the paper.	Acknowledges that the last author contributes with resources and not data	There is no way to identify actual level of contribution apart from statements from the authors. (Bennett & Taylor, 2003)	1	2	This is one of many suggested counting schemes for noblesse oblige
FA First author counting	Only first of N authors of a paper receive a credit equal to 1.	Credit given to first author only	Simple method of crediting publication to the assumed main contributor.	Does not give an accurate picture of the relative contribution of the authors	1	1	Unfair when authors are ordered alphabetically or practice 'noblesse oblige' (Russell & Rousseau, 2002)
Weighted publication count	Applies a weighted score to the type of output.	A reliable distinction between different document types.	Accounts for importance of different publication types for communication within a field.	Has to be designed individual to field as no gold standard.	1	1	Enables comparisons of like with like.

* Col. = data collection, Cal. = calculation

Outcome indicators and their dimensions

All indicators require verified publication data and data from one or more citation databases. Some require an aggregate of "world" publication and citation data to calculate field normalisation scores.

Indicator	Definition	Designed to indicate	Individual		Complexity		Comments
			Advantages	Limitations	Col*	Cal*	
C + sc	Count of all citations to all or selected output, including self-citations	Indication of all usage for whole period of analysis	Reflects social side of research and the cumulative development of knowledge	Quality and timeliness of citation not considered	3	1	Self-citations affect the reliability & validity of the measure on small amounts of data in assessments (Costas & Bordons, 2007; Glänzel, Debackere, Thijs, & Schubert, 2006)
C	Number of citations recorded in CI†, minus self-citations	Recognised benchmark for analyses. Indication of usage by stakeholders for whole period of analysis	Reflects social side of research and the cumulative development of knowledge in CI processed publications	Quality and timeliness of citation not considered; Unclear what to exclude: cites of oneself, a co-author or institutional colleague.	2	2	Does not account for older articles being more cited and variation of citation rates between document types and fields.
Scimago Total Cites (STC)	STC is the number of citations received by articles in Scopus journals, during last 3 years	Indication of usage by stakeholders for whole period of analysis	All types of documents considered and different coverage in database than CI that could be beneficial to some fields.	Only citing information available on articles published after 1996.	2	1	Includes the journals and country scientific information contained in the Scopus® database
C-sc	Citation count, self-citations removed	Measure of usage for whole period of analysis	Reflects social side of research and the cumulative development of knowledge	Quality and timeliness of citation not considered; Unclear what to exclude: cites of oneself, a co-author or institutional colleague.	3	2	Does not account for older articles being more cited and variation of citation rates between document types and fields.
% SELFCIT	Number of self-citations divided by total citations	Share of citations to own publications	Reflects readership of work outside of author and group.	Unclear what to exclude: cites of oneself, a co-author or institutional colleague	3	2	Identifies unwarranted self-promotion
CPP	Sum of citations divided by number of publications.	Trend of how cites evolve over time	Enables comparisons of scientists of different ages and different type of publications	Tells nothing of the timeliness, origin or quality of the cite (positive or negative)	3	2	Citations can be hard to find, reward low productivity & penalize high productivity (Haslam & Laham, 2009) .
Ptop	Publications are grouped by type, age and subject, then ranked by citations.	Identify if publications are among the top 20, 10, 5, 1% most frequently cited papers in subject/subfield/world in a given publication year.	Indicates if publications are cited well but fail to produce really high impact or if researcher contributes to high impact publications but also has a pool of less well cited work.	Unlike mean based indicators, percentiles are not affected by skewed distribution	3	3	Percentiles are most suitable for normalisation of citation counts in terms of subject, document type and publication year (Bornmann & Werner, 2012)

Field top % citation reference value	Quota between count of publications in group, as above, and those with citations above n%.	World share of publications above citation threshold for n% most cited for same age, type and field	Percentiles can prevent a single, highly cited publication receiving an excessively heavy weighting	The degree to which top n% publications are over/under-represented differs across fields and over time (Waltman & Schreiber, 2012)	3	3	Accuracy of inter-field and inter-temporal comparisons decreases with level of representation.
E(Ptop)	Expected number of highly cited papers among the top 20, 10, 5, 1% in the subfield/world	Reference value: expected number of highly cited papers based on the number of papers published by the research unit.	Reflects deviations from the 80th, 90th, 95th, 98th, 99th percentile if tied values occur due to the discrete nature of the impact distribution.	Only Includes documents that have been cited at least once and is interpreted as normalised citations per cited paper not citations per paper	3	3	Expected scores are based on large data sets, their 'random' error is much smaller than that of the value CPP.
A/E(Ptop)	The ratio of the actual and expected presence in the top of the citation distribution.	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.	Indicates share of top impact publication.	Does not account for time delays between publication and citations	3	3	Can reveal if a high normalized score is due to a few highly cited papers or a general high level of citations.
Age of citations	Identifies how old citations are.	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.	Accounts for differences between delayed citations and sleeping beauties, and inter-field differences (van Raan, 2004)	Observed age of citations may not conform with theoretical distributions as the measure cannot cope with singularities from usage of literature on a micro level (De Bellis, 2009)	3	1	Usage and validity are not directly related and might merely reflect the availability of documents.
Number of significant papers	Papers with >y citations,	Gives idea of broad and sustained impact	y can be adjusted for seniority, field norm and publication types	Subjective.	3	1	Can randomly favour or disfavour individuals
Age and productivity (Costas, van Leeuwen, & Bordons, 2010a)	Mean number of documents by age and CPP (3 yr citation window) in 4 year age brackets, adjusted to field.	Effects of academic age on productivity and impact.	Identifies the age at which scientists produce their best research and the extent of the decline in their production	Mean impact declines with age regardless of quality of researcher's body of work.	2	3	If used independently, fosters practice of quantity over quality. Difficult to maintain high values of impact with increasing rates of production.
%Pnc	Number of non-cited publications divided by total number publications in same time period	Share of publications never cited after certain time period, excluding self-citations	Benchmark value: cited and non-cited publications reflect their underlying relevance for technological developments	Publications can be greatly used and of great influence, but never cited (MacRoberts & MacRoberts, 2010)	3	1	Authors cite only a fraction of their influences, many citations go to secondary sources, and that informal level of communication is not captured

* Col. = data collection, Cal. = calculation

†CI =Web of Science (CI) versions of the Science Citation Index, the Social Science Citation Index, Arts and Humanities Citation Index

Quality indicators and their dimensions

All indicators require verified publication data and data from one or more citation databases

Indicator	Definition	Designed to indicate	Individual		Complexity		Comments
			Advantages	Limitations	Col*	Cal*	
h-index (Hirsch, 2005)	Publications ranked in descending order by the times cited. H is the number of papers (N) in the list that have N or more citations.	Cumulative achievement	H is a simple but rough measurement of quality of work, when compared to JIF, citation & publication count (Alonso, Cabreriazo, Herrera-Viedma, & Herra, 2009)	Once a paper is in H-core, the number of citations it receives is disregarded. Loss of citation information means comparisons based on the h-index can be misleading (Schreiber, Malesios, & Psarakis, 2012)	3	2	Arbitrary cut off value for including or excluding publications from productive h-core.
g-index (Egghe, 2006)	Publications ranked in descending order by times cited. G is highest number g of papers that together received g2 or more citations	The distinction between and order of scientists (Egghe, 2006; Harzing, 2008)	Corrects h by weighting highly cited papers to make subsequent citations to highly cited papers count in calculation of the index.	Can be disproportionate to average publication rate. 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	3	3	Ignores the distribution of citations as based on arithmetic average. (Alonso, Cabreriazo, Herrera-Viedma, & Herra, 2009; Costas & Bordons, 2007)
b-index (Brown, 2009)	B is the integer value of the author's external citation rate (non-self-citations) to the power three quarters, multiplied by their h-index	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	Cut-off value for including or excluding publications in productive core is determined using a field-specific reference standard for scientific excellence (Bornmann, Mutza, & Daniel, 2007)	Assumes that relative self-citation rate is constant across an author's publications	3	4	The b index depends on the year in which it is determined, the period under consideration and the used database
Generalized h-index hf (Radicchi, Fortunatoa, & Castellanob, 2008)	Citations of each article normalized by average number of citations per article in the subject category of the article under observation	Allows comparison to peers by correcting individual articles' citation rates for field variation	Suitable for comparing scientists in different fields as rescales field variations and factors out bias of different publication rates	Scales number of citations and rank of papers by constants dependent on discipline, however constants are not available for all fields.	3	4	Calculation is not easy making it a nominal index and not a pragmatic one (Namazi & Fallahzadeh, 2010)
h-index sequences and matrices (Liang, 2006)	Calculates h-sequence by continually changing the time spans of the data. Constructs h-matrix based on a group of correlative h-sequences.	Singles out significant variations in individual scientists citation patterns across different research domains	Makes scientists of different scientific age comparable.	Difficult to determine the correct publication/citation window in construction of the matrix	3	4	Only tested on 11 well established physicists.

Hg-index (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009b)	Geometric mean of a scientist's h- and g-indicators, i.e. hg=$\sqrt{h \cdot g}$	Greater granularity in comparison between researchers with similar h- and g- indicators.	Accounts for influence of a big successful paper on g-index to achieve balance between the impact of the majority of the best papers of the author and very highly cited ones.	Combining H and G does not improve discriminatory power, hg has no direct meaning in terms of papers and citations of a scientist and can lead to hasty judgements (Franceschini & Maisano, 2011)	3	3	Simple to compute once the h- and g-indicators have been obtained.
hα (Eck & Waltman, 2008)	The value of hα is equal to N papers with at least α· hα citations each and the other n- Hα papers have fewer than ≤ α· hα citations each.	Cumulative achievement, advantageous for selective scientists.	Greater granularity in comparing scientists with same h is possible; α can be set to the practices in a specific field, allowing for fairer comparison between fields.	No agreement on the value of parameter α. The appropriate choice of α requires more study and is field dependent. Sensitivity of hα to α needs investigating.	3	4	Small α: ranks scientists based on number of papers with at least one citation (quantity measure: advantageous for scientist who publish a lot but are not very highly cited) Large α: measures number of citations of most cited paper (quality).
Gα (Eck & Waltman, 2008)	gα is the highest rank such that the first gα papers have, together, at least citations.	Based on same ideas as g-index, but allows for fractional papers and citations to measure performance at a more precise level.	gα-index puts more weight on the quality aspect of scientific performance than the hα-index.	No agreement on the value of parameter. The appropriate choice of Gα requires more study and is field dependent.	3	4	Empirical research is needed to find out whether in practical applications the gα index provides better results than g-index
Normalized h-index (Sidiropoulos, Katsaros, & Manolopoulos, 2007)	hn =h/Np, if h of its Np articles have received at least h citations each, and the rest (Np−h) articles received no more than h citations.	Normalizes h to compare scientists achievement based across fields	Accounts for the fact that scientists have different publication and citation habits in different fields.	The normalized h-index can only be used in parallel to h-index and as rewards less productive but highly cited authors	3	3	Using this parameter to judge someone still at the beginning of their career, with few publications, is prone to give paradoxical results.
H(2) index (Kosmulski, 2006)	The highest natural number such that the scientist's H(2) most cited papers received each at least H(2)2 citations.	Weights most productive papers but requires a much higher level of citation attraction to be included in index.	Precision/homograph problem reduced as only a small subset of the researcher's papers used to calculate H(2) index (Bornmann, Mutz, & Daniel, 2008; Jin, Liang, Rousseau, & Egghe, 2007)	Difficult to discriminate between scientists having different number of publications with quite different citation rates for relatively high H(2) indicators	3	3	Suffers from same inconsistency problems as h. (Waltman & van Eck, 2011)
A-index (Jin, 2006; Rousseau, 2006)	Average number of citations in h-core thus requires first the determination of h.	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, Malesios, & Psarakis, 2012)	a-index can increase even if h-index remains the same as citation counts increase (Alonso, Cabreriazo, Herrera-Viedma, & Herra, 2009)	a is h-dependent, has information redundancy with h, and when used together with h masks the real differences in excess citations of different researchers (Schreiber, Malesios, & Psarakis, 2012)	3	3	A-index involves division by h and punishes researchers with high h-index (Jin, Liang, Rousseau, & Egghe, 2007) ; sensitive to highly cited papers (Rousseau, 2006)

R-index (Jin, Liang, Rousseau, & Egghe, 2007)	Square root of the h and A index	Citation intensity and improves sensitivity and differentiability of A index	Adjusts for punishing the researcher with a high h index;	As above. R-index involves division by h and punishes researchers with high h-index; (Jin et al 2007);	3	3	Supplement to h. Easier to calculate than g index, but not as elegant.
Citation-weighted h-index (hw) (Egghe & Rousseau, 2008)	Hw is the square root of the total weighted citations (Sw) received by the highest number of articles that received Sw/h or more citations	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.	Improves sensitivity to the number of citations in h-core	Doesn't use h-table in calculation and is therefore not an acceptable h-type measure	3	4	Hw can be misleading and a contradiction of h (Maabreh & Alsmadi, 2012)
ħ-index (Miller, 2006)	Square root of half the total number of citations to all publications	Comprehensive measure of the overall structure of citations to papers	Includes papers h ignores ie. most highly cited articles and the body of articles with moderate citations	Difficult to establish the total citation count with high precision (Schreiber, 2010)	3	3	Is only roughly proportional to h.
m-index (Bornmann, Mutz, & Daniel, 2008)	Median number of citations received by papers in the h-core	Impact of papers in the h-core	To account for skewed distribution of citations, the median and not the arithmetic average is used to measure a central tendency.	Although median may be a better measure of central tendency it can be chronologically instable.	3	2	Reduces impact of heavily cited papers.
π-index (Vinkler, 2009)	π is one hundredth of the number of citations received by the top square root of the total number of papers ranked by decreasing number of citations.	Production and impact of scientist	Allows for comparative assessment of scientists active in similar subject fields. Sensitive to citedness of top papers and thus indicates impact of information on research.	Value depends on citation rate of papers in the elite set (top cited papers); the elite set is scaled by an arbitrary prefactor (Schreiber, 2010).	3	4	Can be calculated on a small number of papers. Unique index because it is defined in terms of the summed number of citations rather than the square root of the sum or the average (Schreiber, 2010).
Tapered h-index (hT) (Anderson, Hankin, & Killworth, 2008)	Using a Ferrers graph, the h-index is calculated as equal to the length of the side of the Durfee square assigning no credit to all points that fall outside.	Production and impact index that takes all citations into account, yet the contribution of the h-core is not changed.	Evaluates the complete production of the researcher, all citations giving to each of them a value equal to the inverse of the increment that is supposed to increase the h-index one unit.	Difficult to implement because of the computations needed to obtain the measure and the difficulty in obtaining accurate data from bibliographic databases (Alonso et al 2009).	3	5	Shows smooth increase in citations, not irregular jumps as in h-index. Conceptually complex (Anderson et al 2008).

Rational h-indicators hrat Index (Ruane & Tol, 2008)	hrat=(h+1) $-\frac{nc}{2.h+1}$ h is h index, nc is number of citations that are needed to make a h-index of h+1 and 2.	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).	Increases in smaller steps than h-index providing greater distinction in ranking of individuals	The relative influence of the interpolation will be stronger for smaller values of the indicators therefore utilize the generalized indicators when comparing many data sets with very small values of h.	3	5	Interpolated indicators have the advantage that one does not have to wait so long to see one's index growing.
Rational g-index grat, (Schreiber, 2008a; Tol, 2008)	Interpolates between g and g+1 based as above on the piecewise linearly interpolated citation curve.	Indicates the distance to a higher g-index	It is not a complementary index requiring first the determination of h, but rather follows from a self-consistent definition (Schreiber, 2010) .	Limits as for hrat.	3	5	As every citation increases interpolated g, the index is sensitive to self-citations (Schreiber 2008a)
e-index (Zhang, 2009)	E is the number of excess citations (more-than-h citations received by each paper in the h core)	Complements the h-index for the ignored excess citations	The combination *h,e* provides complete citation information.	E value can only be calculated if h is given.	3	2	Complements h especially for evaluating highly cited scientists or for precisely comparing the scientific output of a group of scientists having an identical h-index.
f-index (Tol, 2009)	Fractional counting and ranking scheme of papers:cites, where the average is calculated as the harmonic mean	Attempts to give weight/value to citations. Highest number of articles that received *f* or more citations on average.	An additional citation to a not-so-often cited paper counts more than an additional citation to an often-cited paper.	Both f & t indicators are maximum if every paper is cited the same number of times, but the f-index deviates much faster from this maximum than the t-index.	3	4	More discriminatory power than the h- and g-indicators. Because of the non-linearity of the harmonic mean, the f-index is more sensitive to small differences between researchers
t-index (Tol, 2009)	Fractional counting and ranking scheme of papers:cites, where the average is calculated as the geometric mean	Attempts to give weight/value to citations. Highest number of articles that received t or more citations on average	Using geometric mean doesn't place much weight on the distribution of citations.	Sensitivity to small differences between researchers is stronger with harmonic mean (f-index) than geometric mean.	3	4	It is not sufficient to determine the function and value of citations using indicators; their cognitive background should also be taken into consideration.
Hmx-index (Sanderson, 2008)	Rank academics by their maximum *h* (hmx) measured across WOS, Scopus and GS.	Ranking of the academics using all citation databases together.	Accounts for missing citations, lack of correlation between databases and disparities in h across databases.	Assumes that the differences in h across the databases are due to false negative errors and that these were negligible.	3	2	Although hmx provides a better estimate of h than any single database, a close examination of the overlaps of citations and publications between the databases will provide a better estimate.

w-index (Wu, 2008)	w is the highest number of papers have at least 10w citations each	The integrated impact of a researcher's excellent papers.	More accurately reflects the influence of a scientist's top papers	H dependent. Tendency to describe quantity of the productive core	3	2	w-index of 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.
Index of Quality and Productivity (Antonakis & Lalive, 2008)	Ratio actual citations to estimated citations and total papers (corrected for subject)	Quality reference value; judges the global number of citations a scholar's work would receive if it were of average quality in its field.	Corrects citation count for scholarly productivity, author's academic age, and field-specific citation habits with reference to estimated citation rate.	Tested in natural sciences, medicine and psychology and dependent on WOS field specific journal impact factors.	3	3	Correlates better with expert ratings of greatness than h index. Allows comparison as brings papers in low cited fields on same scale as papers in highly cited fields.
x-index (Claro & Costa, 2011)	x is a researcher's absolute score divided by a reference score	Indication of research level. Describes quantity and quality of the productive core and allows for comparison with peers.	Accounts for multi-and interdisciplinary research by using the journals the researcher publishes in as reference and not field classification	x is based on (5 year) Impact Factor which has well-documented limitations; x is also vulnerable to scale issues	3	4	Using a measure based on citation counts would permit a more meaningful assessment of scientific quality
H per decade (Hpd-index) (Kosmulski, 2009)	Hpd is highest number of papers that have at least hpd citations per decade each and other papers have less than hpd + 1 citations per decade each.	Compare the scientific output of scientists in different ages. Seniority-independent Hirsch-type index.	In contrast with h-index, which steadily increases in time, hpd of a mature scientist is nearly constant over many years, and hpd of an inactive scientist slowly declines.	Hpd uses scaling factor of 10 to improve granularity between researchers is as an arbitrary number, which randomly favors or disfavors individuals.	3	4	hpd can be further modified for multi-authored papers where the individual cites per year of each paper is divided by the number of co-authors to produce the contribution of single co-author.
Q^2 –index (Cabrerizoa, Alonso, Herrera-Viedmac, & Herrerac, 2012)	Q^2 is the geometric mean of h-index and the median number of citations received by papers in the h-core	Relates two different dimensions in a researcher's productive core: the number and impact of papers	Combines robustness of h-index' measurement of papers in core with m-index correction of the distribution of citations to papers.	h- and m-indicators have to be obtained before calculation of q^2	3	3	Geometric mean is not influenced by extremely higher values, and obtains a value which fuses the information provided by the aggregated values in a balanced way.

Research Infrastructure indicators and their dimensions

All indicators require verified publication data and data from one or more citation databases

Indicator	Definition	Designed to indicate	Individual		Complexity		Comments
			Advantages	Limitations	Col*	Cal*	
Number of co-authors	Count of authors per paper	Indicates cooperation and growth of cooperation at inter- and national level;	Measure volume of work by teams of authors at individual level	Whole or fractional counts of authorship produce different results	1	1	How affiliation is listed can be problematic and can affect aggregation;
Co-citations	Number of times 2 papers are cited simultaneously in same article	Thematic networks and influence and impact of researcher.	Cluster analysis shows related subjects, communities and evolvement of field over time.	Highly selective analysis of science as they describe only part of the process of assembling knowledge	3	1	Limited to scientific publications in citation indicators.
Fractional counting on citations	Gives an author of an m-authored paper only credit of c/m if the paper received c citations	Designed to remove the dependence of co-authorship (Egghe, 2008)	Gives less weight to collaborative works and leads to proper normalization of indicators and fairer comparisons	Regards credit as a single unit that can be distributed evenly, making share dependent on number of authors.	3	2	Comparison to field norm unwise as citations to the publications may not be representative of the field but biased towards the highly or poorly cited.
hi-index (Batista, Campiteli, Kinouchi, & Martinez, 2006)	Divides h-index by the mean number of researchers in the h-core publications.	Indicates number of papers with at least h citations scientist would have written if worked alone.	Accounts for differences in co-authorship patterns, disciplinary differences and self-citations (Schreiber, 2008a)	Might decrease when a paper with many authors advances into the h-core by attracting additional citations and reduces size of the h-core.	3	3	The average is sensitive to extreme values and disfavours people with some papers with a large number of co-authors
POP variation individual H-index (Harzing, 2008)	Divides number of citations by number of authors for that paper, then calculates the h-index of the normalised citation counts	Accounts for co-authorship effects	Gives an approximation of the per-author impact, which is what the original h-index set out to provide.	Normalisation by mean number of authors of publications in the h-core leads to reduction of the index. This is a fractionalised count of citations and publications (Schreiber, 2008a)	3	3	(Egghe, 2008) also considered multiple authors by computing g and h indicators using a fractional crediting system.
n-index (Namazi & Fallahzadeh, 2010)	Researcher's h-index divided by the highest h-index of the journals of his/her major field of study	Enables comparison of researchers working in different fields:	Can surmount the problem of unequal citations in different fields	Still awaiting validation.	2	2	Calculation based on Scopus definition of h and SCImago Journal and Country Rank website for journal information

H_m-index (Schreiber, 2008b)	Uses inverse number of authors to yield a reduced or effective rank. Hm is the reduced number of papers that have been cited hm or more times	Softens influence of authors in multi-authored papers	Does not push articles out of the h-core; each paper is fully counted allowing for a straightforward aggregation of data sets.	Precision problem is enhanced, as additional papers enter into the hm-core.	3	4	Uses fractional paper counts instead of reduced citation counts
Alternative H index (Batista et al., 2006)	Alternative h is h-index divided by mean number of authors in the h publications	Indicates the number of papers a researcher would have written along his/her career if worked alone.	Rewards scientists whose papers are entirely produced by themselves from the authors that work groups that publish a larger amount of papers.	Mean is sensitive to extreme values and could penalize authors with papers with a large number of authors.	3	2	Valid quantification of output across disciplines allowing for comparison.
Pure h-index (Hp) (Wan, Hua, & Rousseau, 2007)	Hp is the square root of h divided by normalised number of authors and credit to their relative rank on the by-line of the h-core articles	Corrects individual h-scores for number of co-authors	Reduces effect of collaboration in multi-authored, highly cited paper.	Results vary dependent on method of distributing credit to authors- fractional count, arithmetic to determine h,	3	3	More refined approach is pure R-index. Takes the number of collaborators, possibly the rank in the byline and the actual number of citations into account.
Adapted pure H-index (h_{ap}) (Chai, Hua, Rousseau, & Wan, 2008)	H is interpolated rank value between papers (fractionally counted) and citations (counted as square root of equivalent number of authors).	Finer granularity of individual h-scores for number of co-authors by using a new h-core.	Alters h-core to be less biased than Hp with respect to authors with many multi-authored papers	Precision an issues and difficult to calculate.	3	5	Lead to a more moderate correction of authorship than h_i as divides citation count by the square root of author count rather than full author count (Rosenberg, 2011)
Cognitive orientation	Analysis by aggregating papers according to scientific subfields the individual publishes or is cited in.	Identify how frequently a scientist publishes or is cited in various fields; indicates visibility/usage in the main subfields and peripheral subfields.	Can easily be related to the position a researcher holds in the community	More applicable in some fields than others as often journal based and limited to CI† definition of scientific fields	3	1	Useful to identify future areas for collaboration and production.
Visual representation techniques	Variety of techniques of multidimensional analysis to construct maps	Based on bibliographic data graphical representations are generated of publishing, collaboration, citations, growth and activity in research field.	Maps of relational networks depict structure of research with greater clarity than in statistical tables.	Data loss: not all data contained in a multidimensional system in two dimensions can be represented.	3	1	Requires software and instruction but can provide a comprehensive picture of the development of a researcher's work.

* Col. = data collection, Cal. = calculation

†CI =Web of Science (CI) versions of the Science Citation Index, the Social Science Citation Index, Arts and Humanities Citation Index

Impact indicators and their dimensions

All indicators require verified publication data and data from one or more citation databases. Some require an aggregate of "world" publication and citation data to calculate field normalisation scores

Indicator	Definition	Designed to indicate	Individual		Complexity		Comments
			Advantages	Limitations	Col*	Cal*	
ISI JIF (SIF) Synchronous IF	Number of citations a publication has received during a single citing year to documents from previous 2 publication years	Average number of citations a publication in a specific journal has received limited to ISI document types and subject fields.	Readily available. The "mix" of different publication years makes SIF robust indicator of permanent impact	Measure of journal popularity not scientific impact (Bollen, Rodriguez, & Van, 2006) Not designed for indication of individual performance.	2	1	Does not allow for different citation window to benefit field; hides variation in article citation rates as citations are results of skewed distribution.
Diachronous IF (Ingwersen, Larsen, Rousseau, & Russell, 2001)	A ratio calculation of citations from two or more citing years to documents issued in a fixed publication year	Reflects actual and development of impact over time of a set of papers.	Can be calculated for one-off publications, such as books containing contributions of different authors, or conference proceedings	Demands more resources than simply using impact factors from JCR, because it has to be based on manual collection of data.	3	2	Better represents the researcher in evaluation than SIF.
Weighted PageRank rating of journal status (Bollen, Rodriguez, & Van, 2006)	Assigns a numerical weighting to each element of hyperlinked set of documents.	Indicates relative importance of journal within a journal citation network	Takes into account the popularity and prestige factor of status, avoids assigning high ranks to popular but irrelevant journals	Assumes links are trust votes and ranks journals based on these links interconnecting them.	2	5	Assumes prestige is not only a matter of the number of citations, but who is actually citing.
Y Factor (Bollen, Rodriguez, & Van, 2006)	Y is JIF multiplied by PageRank	Scientific impact defined as a combination of popularity and prestige	Accounts for ISI JIF reliance on citation frequencies (popularity) and the Weighted PageRank reliance on prestige values	Has not yet been fully justified, but performed well in physics, computer science (Bollen et al 2008:Satyanarayana 2010)	2	2	Reduces effect of review articles/journals in ranking and promotes original articles
Scimago Journal Rank (SJR)	Citation PageRank of a journal divided by the number of articles published by the journal, in a 3 year citation period	Average per article PageRank based on Scopus citation data	Assigns different values to citations depending on the importance of the journals where they come from	Scopus is limited to the time period after 1996 for which citation analysis is available	2	1	Open access journals included in indicator
EigenFactor	Ratio total weighted citations to journal in a certain year to documents from previous 5 years	Journal's total importance to the scientific community	Includes citations from non-standard items and a longer citations window.	Based on journals listed in JCR; journals that publish less than 12 articles per year averaged over 5 years are not included, nor journals that do not cite other journals listed in the JCR	2	1	*Eigenfactor* journal categories differ from the ISI categories; journals can only belong to one category based on citation patterns

Article influence score (AI)	EigenFactor score divided by *i*-th entry in the normalized article vector	Measure of average per-article citation influence of the journal	Comparable to ISI JIF	Both EigenFactor and AI are redundant indicators as add little to easily understandable JIF, total citations and 5 year impact indicator (Chang, McAleer, & Oxley, 2010)	2	1	Large disciplinary differences that persist in the Article Influence Score limit its utility for comparing journals across different fields (Arendt, 2010)
Co-authorship network analysis (Yan & Ding, 2011)	Weighted PageRank algorithm considering citation & co-authorship network topology	Individual author impact within related author community	Focuses on the random surfing aspect and develops it into citation ratios.	PR algorithm, only the top 10%-20% of overall authors in the co-authorship network can produce useful data.	2	5	Success of indicator is field dependent as rate of co-authorship varies
Normalised journal impact	Journal impact divided by citation average in subfields covered by the journal	Mean impact value of all the normalized citation counts for publications in a specific journal	Accounts for differences in reference practices in sub-fields and type, age and distribution of documents	Difficult to calculate normalised measure of multi-disciplinary journals	2	2	Enables cross-comparisons among disciplines and not biased in favour of review journals
Journal to field impact score (JFIS) (van Leeuwen & Moed, 2002)	Compares citations to one journal to world average of citations to journals within same field for 5 year period	Journal to fields citation score that indicates relative impact of a journal	Accounts for journal subject area and document type, allowing for comparisons between subject areas.	A problem with normalization to document type is that in some journals/fields the amount can be so low that it hardly constitutes a meaningful standard for comparison.	3	2	Lengthened time period and identification of specific document type improves usefulness of measure.
Discipline Impact Factor (DIF) (Hirst, 1978)	DIF is the number of citations to a journal by the citing set divided by the number of citable items published in the journal over time.	Number of times a journal is cited by the core literature of a single subfield rather than a complete set of ISI journals.	Gives a good approximation of core journals as a performance benchmark	Requires at least 3 iterations of the calculation to identify the core literature and stabalize the indicator; Can be affected by continued citations to older articles	2	3	Index loses detail as dependent on ISI Journal Citation Reports i.e. it is affected by JCR field coverage and minimum cites inclusion criterion.
Median impact factor (IF med)	IF med is the median value of all journal Impact Factors in the subject category.	The aggregate Impact Factor for a subject category	Accounts for the number of citations to all journals in the category and the number of articles from all journals in the category.	The number of journals that make up categories and the number of articles in these journals influence the calculations of these ratios.	2	2	Not designed to replace the JIF, but is a complementary indicator.
Normalised journal position (NJP) (Bordons & Barrigon, 1992)	Ordinal position of each journal in JCR category, ranked by JIF, divided by number of journals in that category.	Compare reputation of journals across fields	Allows for inter-field comparisons as it is a normalized indicator.	NJP is confounded by editorial decisions. All manuscripts have same rank position & the position is the result of successful publication decisions.	2	2	The citation counts of the published manuscripts determine the position of the journal (Bornmann, Mutz, Marx, Schier, & Daniel, 2011)

Item oriented field normalized citation score average ($\bar{c}f$) (Lundberg, 2009)	Citations to individual publications divided by world average of citations to publications of the same type, year and subject area	Item orientated field normalised citation score.	Normalisation is on the level of individual publication giving each publication equal weight in the final field score value. Accounts for the prevailing skewness of citation distributions	Value of field normalised citation score can be unproportionately affected by highly cited publications in a moderately cited field.	3	4	More appropriate for some document types than others; there are differences in average availability of citation data, citation rates, and document types used in research.
Field citation score (FCS)	Publications sorted by type, age and subject. Mean value of citations within group is field reference value	Represents the number of citations expected for a paper of the same type, published in all journals within a specific field in the same year, and document type.	Is an international reference to compare relative impact of publications to those published in the group of journals that constitute a field	Classification of journals into journal categories is less appropriate for researchers in multidisciplinary areas	2	3	ISI CI field categories are inadequate for some disciplines, providing a distorted picture
Field Citation Score Mean (FCSm)	Mean citation rate of all papers published by unit of evaluation in all subfields in which he or she is active	Weighted average for comparison of impact in different subfields	Accounts for impact level of an units journal set.	Often based on subject classifications in ISI and ISI world average where subfields defined by CI subject categories	2	3	Most suitable indicator of international position.
JSCS or JRV Journal citation score (journal reference value)	Publications are grouped after type. Mean value of citations to all publications within group is calculated	Worlds average of citations to publications according to type and age.	Journal-based worldwide average impact as an international reference level for the university/institute/department/ group/researcher etc.	Affected by rate of citation or time delay between publication and citation, dependent on field.	2	3	Expanding the size of the group can be counterproductive
Normalised Journal Citation Score (JSCm)	Mean citation rate of all articles published in the journals in which the individual has published.	Reference value accounting for type of paper and years in which papers were published.	Weighted average, weights determined by number and type of papers published in each journal.	Low impact publications published in low impact journals may get a similar score to high impact publications in high impact journals	2	3	More accurate for activity in subfields than FSCm especially for developing and interdisciplinary fields.
JCSM/FCSm (Costas, Bordons, van Leeuwen, & van Raan, 2009)	Journal citation score mean divided by field score mean**.**	Journal based worldwide average impact mean for an individual researcher compared to average citation score of the subfields	Indicates if the researcher publishes in journals with high or low impact within the field.	Based on ISI data, low impact sources are often not included. Valuable information can also be obtained by retrieving impact data from non-CI publications.	3	2	Favours senior researchers as minimum publication value if 50 is recommended for informative analysis.
Crown Indicator CPP/FCSm	Sum of citations divided by sum of world average	Individual performance compared to world citation average to publications of same document types, ages, and subfields.	Sum of citations before normalization makes indicator resistant to effect of highly cited papers in low-cited	Limited to same document type as world citation average is based on.	3	3	Calculation benefits older articles in highly cited fields (Moed, 2005)
Prediction of article impact (Levitt & Thelwall, 2011)	Weighted sum of article citation and impact factor of the journal in which the article was published.	Predictor of long term citation	Aims to include new publications in analysis of an individual's research.	Indicator tested on only one subject category with a short publication window and may not apply to other subjects	2	4	Comparisons between the weighted sum indicator and the indicators from which it is derived (sum of citation and IF) need to be conducted with care.

P_{tj}	Count of number of publications published in selected journals in a time span.	Performance of articles in journals important to (sub)field or institution.	Reflects potential impact of articles in sources defined locally as important.	Does not take the size of the analyzed unit into account.	1	2	More interesting than mere publication count.
CPP/JCSm	Impact of individual's articles compared to average citation rate of individuals journal set.	Indicates if the individual's performance is above or below the average citation rate of the journal set.	Not affected by few publications that have a high/low citation count compared to world average.	Can be manipulated by publishing in averagely cited journals with a below average journal impact indicator (Moed, 2005)	3	2	Citation rates are normalised as: the average citation rate of the researcher compared to average citation rate for field
JCSm/FCSm (Gaemers, 2007)	Journal citation score mean divided by field citation score mean	Relative impact level of the journals compared to their subfields	Normalised values are free from influences by distribution and document type effects.	The CPP/JCSm, CPP/FCSm and JCSm/FCSm indicators are not independent. The value of each one follows directly from the values of the other indicators.	2	3	An unambiguous classification of articles in journals is impossible and different weighting schemes may lead to very different ratings in the evaluation
C/FCSm (van Leeuwen, Visser, Moed, Nederhof, & Raan, 2003)	Total citation count divided by world mean citation rate of all publications in the same field (from same year of publication).	Applied impact score of each article/set of articles to the mean field average in which the researcher has published	Accepted as reliable measure for visibility in natural sciences. Highlights diversity of publication performance.	Unreliable due to non-para-digmatic nature of different fields, the heterogeneity of publication behaviours and insufficient coverage in citation databases.	3	2	Inadequate coverage in social and humanist sciences in citation indexes effects validity of indicators.
Logarithm based citation z-score (Lundberg, 2009)	Log. number of citations a publication has re-ceived to the mean & standard deviation of log. citation rates for all publications of same type, age and subject.	Accounts for citation rate variability of different fields and skewed distribution of citations over publications on an item level.	Normalizes citation impact level of individual production to allow better control over the variability of citation rates across research fields.	If the distribution of citation values is very skewed and far from a normal distribution, the mean and the standard deviation may be misleading measures.	3	5	Approaches normal distribution already within low aggregation levels.
Usage Impact Factor (UIF) (Bollen & Sompel van de, 2008)	Number of full text downloades in a year to articles published in the journal in the previous two years divided by the number of articles published by the journal in the preceding two years.	Average local usage rates for the articles published in a journal	On the basis of detailed usage data, subsets of the scholarly community can be analysed (students, researcher, lecturers & public)	Scalability of the approach (infrastructure, privacy & sample size) and quality of data should be considered.	4	5	Usage precedes citation, thereby serving as an earlier indicator of scholarly impact.

Innovation and Social Benefits indicators and their dimensions

Requires data from citation databases, the internet, internal databases and verified activity data.

Indicator	Description	Designed to indicate	Individual		Complexity		Comments
			Potentials	Limitations	Col*	Cal*	
Knowledge exchange (Mostert, Ellenbroek, Meijer, van A., & Klasen, 2010)	Weighted count of keynote speeches, activity in agencies & organisations, public forums, committees, conferences & co-operation with companies.	Knowledge production, knowledge exchange, knowledge use and earning capacity	Can justify/promote research programme or individual scientist's work	No well-defined bibliometric indicators recommended	1	1	Based on normalised peer reviewed science citation impact analysis
Dissemination in public sphere (Mostert, Ellenbroek, Meijer, van A., & Klasen, 2010)	Count of contributions to, inc.: tv & radio pro-grams, newspapers, non-peer reviewed journals, text books, public & professional websites and news forums	Impact and use in public sphere (knowledge transfer)	Useful addition to evaluation of scientific dissemination activities in the academic environment;	Many indicators and no gold standard method of weighting relative to departmental norm or expected performance in discipline	1	1	Societal quality is dependent on different activities than scientific quality and is not a consequence of scientific quality.
Knowledge use (Mostert, Ellenbroek, Meijer, van A., & Klasen, 2010)	Count of use of output in schoolbooks, curriculum, protocols, guidelines, policies and in new products	Impact on learning in stakeholders environment.	Analysis of citations and references in guidelines, policies, protocols to indicate links (use) with stakeholders.	Has to be adjusted to the mission and objectives of the scientist and department/discipline	5	1	Focuses on research group level
Patent applications (Okubu, 1997)	Count of patent applications	Innovation	Resources invested in R&D activities and role of scientist in development of new techniques.	Patent application varies from field to field.	1	1	Quality or significance of patents is not on an equal level;
Citations in patents (Okubu, 1997)	Count and source assessment of citations in patents	Impact on or use in new innovations	Depicts state of a given art, newness and significance of innovation; length of time between publication of paper and patent application.	Cites might be legally or competitively motivated and not of innovative or scientific nature. Indicates impact of technology rather than science	5	1	Requires access to specialized database and cooperation of several specialists to verify results (Quomiam, Hassanaly, Baldit, Rostaing, & Dou, 1993)

Scientific proximity (Okubu, 1997)	Relative number of citations of papers in patents applied for in specific sector	Intensity of an industrial or technological activity	Interaction between science and technology	Credibility of any utilisation of such data for analytical and statistical purposes.	5	2	Patents serve a legal purpose, and authors demonstrate their technological links and conceal the essentials of their content
Usage log data (Bollen, Biet-Arie, & Van de Sompel, 2006)	Log data from webportals collected – date/time of request, request type, article identifier.	User activity that expresses interest or preference	Allows analyses of immediacy, representativeness and structural aspects of prestige and impact in the scholarly community	Privacy and legal issues in data-recording, verification and falsification issues and usage definition	5	3	Eliminates time-lag of citations (published in literature and included in citations databases)
Tool to measure societal relevance (Niederkrotenthaler, Dorner, & Maier, 2011)	Questionnaire used as the (self-assessment) application form and the assessment form for the reviewer	Aims at evaluating the the level of the effect of the publication, or at the level of its original aim	Accounts for knowledge gain, application &increase in awareness; efforts to translate research results into societal action; identification of stakeholders and interaction with them.	Only developed and evaluated in a focus group in the biomedical sciences	1	1	Tool requires further development, specification and validation.

Sustainability indicators and their dimensions

Requires verified publication data and data from citation databases.

Indicator	Description	Designed to measure	Individual		Complexity		Comments
			Advantages	Limitations	Col*	Cal*	
Citation age c(t) (Egghe & Rousseau, 2000)	c(t) is the difference between the date of publication of a researcher's work and the age of citations referring to it.	The age of citations referring to a researcher's work.	The entire distribution of the citation ages of a set of citing publications provides insight into the level of obsolescence or sustainability.	Possibility of measuring aging in a meaningful way is questionable by means of citation counting as this doesn't account for role of literature growth, availability of literature and disciplinary variety	3	3	Usage and validity are not necessarily related
Aging rate a(t) (Egghe & Rousseau, 2000)	a(t) is the difference between ct and c(t+1)	Aging rate of a publication.	For individual documents stochastic models are preferable as they allow for translation of diverse factors influencing aging into parameters that can be estimated from empirical data with a specified margin of error	A corrective factor is required if citation rates are to be adjusted for changes in the size of citing population and discipline (De Bellis, 2009; Dubos, 2011)	3	4	There are many models to study aging, the simplest is study of the exponential decay of the distribution of citations to a set of documents
Contemporary h-index h^c (Sidiropoulos, Katsaros, & Manolopoulos, 2007)	An article is assigned a decaying weight depending on its age	Currency of articles in h-core.	Accounts for active versus inactive researchers	The weighting is parametrized gamma=4 and delta=1, making this metric identical to hpd, except measured on a four year cycle rather than a decade. (Rosenberg, 2011)	3	4	An old article gradually loses its "value", even if it still gets citations thus newer articles are prioritized in the count.
Trend H index h^t (Sidiropoulos, Katsaros, & Manolopoulos, 2007)	Each citation of an article is assigned an exponentially decaying weight, which is expressed as a function of the "age" of the citation.	Age of article and age of citation.	Identifies pioneering articles that set out new line of research and still cited frequently.	The weighting is parametrized and for gamma = 1 and delta = 0, this metric is the same as the h-index.	3	4	Estimates impact of researchers work in a particular time instance i.e. whether articles still get citations by looking at the age of the cites.
Dynamic H-type index (Rousseau & Ye, 2008)	Built on 3 time dependent elements: $R(T) \cdot vh(T)$ where R(T) is the R-index computed at time T and vh is the h-velocity	Accounts for the size and contents of the h-core, the number of citations received and the h-velocity.	Detects situations where two scientists have the same h index and the same number of citations in the h core but that one has no change in his h index while another scientist's h index is on the rise.	H dependent. To define vh it is better to find a fitting for hrat(t) - and not for h(t)- as this function is more similar to a continuous function than the standard h-index.	3	4	For evaluation purposes self-citations should be removed (Alonso et al 2009).

M-quotient (Hirsch, 2005)	M is h-index divided by years since first publication	H type index, accounting for length of scientific career	Allows for comparisons between academics with different lengths of academic careers, as h is approximately proportional to career length.	m stabilizes later in career; small changes in h can lead to large changes in m; first paper not always an appropriate starting point.	3	2	m discriminates against part time researchers/career interruptions (Harzing, 2008)
AR-index (Jin, Liang, Rousseau, & Egghe, 2007)	AR is the square root of the sum of the average number of citations per year of articles included in the h-core.	Accounts for citation intensity and the age of publications in the core.	AR is necessary to evaluate performance changes.	Divides the received citation counts by the raw age of the publication. Thus the decay of a publication is very steep and insensitive to disciplinary differences. (Järvelin & Person, 2008)	3	2	AR index increases and decreases over time (Alonso et al 2009); Complements h. Jin et al do not consider AR convincing as a ranking metric in research evaluation.
Discounted Cumulated Impact (DCI) (Ahlgrena & Järvelin, 2010; Järvelin & Person, 2008)	Sum of weighted count of citations over time to a set of documents divided by the logarithm of the impact in past time intervals	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.	Gives more weight to highly cited publications as these are assumed to be quality works.	Difference caused by weighting: some authors gain impact while some others lose.	3	5	Rewards an author for receiving new citations even if the publication is old.
Price index – PI (Price, 1970)	PI = (n1/n2)*100 where n1, is the number of cited references with a relative age of less than 5 years, n2 is the total number of references.	Percentage references to documents, not older than 5 years, at the time of publication of the citing sources	Accounts for the differing levels of immediacy characteristic of the structurally diverse modes of knowledge production occurring in the different sciences	Does not reflect the age structure in slowly ageing literature (De Bellis, 2009)	3	2	In the calculation of PI it is unclear whether the year of publication, is year zero or year one. Moreover, it is unclear whether or not this year is included. (Egghe & Rousseau, 1995)
Immediacy index	Ratio number of citations a journal receives in a given year to the number of articles it issues during the same year.	Speed at which an average article in a journal is cited in the year it is published	Discounts the advantage of large journals over small ones.	Frequently issued journals may have an advantage because an article published early in the year has a better chance of being cited than one published later in the year.	2	2	Different types of journals influence the immediacy index, such as length of publishing history, prestige and atypical references.

Aggregate Immediacy Index (AII)	AII cites to all items published in journals in a particular subject category in one year divided by the number or articles/reviews published in those same journals in the same year	How quickly articles in a subject are cited	Useful context for evaluating how a journal compares to other journals publishing within the same discipline.	Metric can be limited by field coverage of citation database.	2	2	For comparing journals specializing in cutting-edge research, the immediacy index can provide a useful perspective.
Cited half-life (CHL) & Aggregate Cited Half-Life (ACHL)	CHL is the number of years, going back from the current year, that account for 50% of the total citations received by the cited journal in the current year	A benchmark of the age of cited articles in a single journal	ACHL is an indication of the turnover rate of the body of work on a subject and is calculated the same way as CHL.	A lower or higher cited half-life does not imply any particular value for a journal.	2	2	It is possible to measure the impact factor of the journals in which a particular person has published articles however misuse in evaluating individuals can occur as there is a wide variation from article to article within a single journal
Classification of durability (Costas, van Leeuwen, & Bordons, 2010; 2010b; 2011)	Percentile distribution of citations that a document receives each year, accounting for all document types and research fields.	Durability of scientific literature on distribution of citations over time among different fields	Aids study of individuals from general perspective using composite indicators. Discrim-inates between normal, flash in the pan and delayed publications.	Minimum 5 yr citation history threshold for reliable results and empirically investigated in WOS using journal subject categories.	2	3	Can be applied to large sets of documents or documents published in different years; Documents can be classified in more than one field and can be updated yearly/monthly
Age-weighted citation rate (AWCR, AW & per-author AWCR) (Harzing, 2012b)	Age-weighted citation rate, is the number of citations to a given paper divided by the age of that paper	AWCR measures the number of citations to an entire body of work, adjusted for the age of each individual paper	Using the sum over all papers instead, represents the impact of the total body of work allowing younger, less cited papers to contribute to the AWCR	Field norm has to be decided to account for field characteristics such as expected age of citations, “sleeping beauties”, and delayed recognition.	2	3	The AW-index is defined as the square root of the AWCR. It approximates the h-index if the mean citation rate remains constant over the years. The per-author age-weighted citation rate is similar to the plain AWCR, but is normalized to the number of authors for each paper.

Appendix 2 - 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 Informetrics,* 3(4), 273-289.

Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009b). Hg-index: A new index to characterize the scientific output of researchers based on the h- and g-indicators. *Scientometrics,* doi: 10.1007/s11192-009-0047-5.

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

Arendt, J. (2010) Are Article Influence Scores Comparible across Scientific Fields? *Issues in Science and Technology Librarianship.* Doi: 10.5062/F4FQ9TJW

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

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

Bollen, J., Biet-Arie, O., & Van de Sompel, H. (2006). Alternative metrics of journal impact based on usage data - the bX project. Powerpoint lecture presented at the *Meeting on Alternative Metrics of Publication Impact,* Humboldt University: Berlin, 20-21 February. www.dini.de/fileadmin/workshops/februar-2006/bollen.ppt Accessed 11 April 2013.

Bollen, J., Rodriguez, M., & Van, d. S. (2006). Journal status. *Scientometrics,* 69(3), 669-687.

Bollen, J., & Sompel van de, 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. (2007). The b-index as a measure of scientific excellence: A promising supplement to the h-index. *International Journal of Scientometrics, Informetrics and Bibliometrics.* 11(1) paper 6.

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.

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. *Proceedings of WIS 2008: Fourth International Conference on Webmetrics, Informetrics and Scientometrics & Ninth COLLNET Meeting,* Berlin. doi: www.collnet.de/Berlin-2008/ChaiWIS2008aph.pdf

Chang, C., McAleer, M. & Oxley, L. (2010) Journal Impact Factor Versus Eigenfactor and Article Influence. Resource document. University of Canterbury Research Repository. http://hdl.handle.net/10092/5387. Accessed 15 May 2013.

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 Sociaet 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.,N., & 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.

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

Dubos, G. (2011) Stochastic Modelling of Responsiveness, Schedule Risk and Obsolescence of Space Systems, and iMplications for Design Choices. Resource Document. Georgia Institute of Technology:PhD Thesis. https://smartech.gatech.edu/bitstream/handle/1853/43656/dubos_gregory_f_201105_phd.pdf. Accessed 15 May 2013.

Eck, N. V., & Waltman, L. (2008). Generalizing the g- and h-indicators. *ECON Papers,* doi: 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. (2008). Mathematical theory of the h- and g-index in case of fractional counting of authorship. *Journal of the American Society for Information Science and Technology,* doi:10.1002/asi.20845

Egghe, L. & Rousseau, R. (1995) Averaging and globalising quotients of informetric and scientometric data. *Journal of Information Science.* 22(3), pp.165-70.

Egghe, L., & Rousseau, R. (2000). Aging, obsolescence, impact, growth, and utilization: Definitions and relations. *Journal of the American Society for Information Science and Technology,* 51(11), 1004-17.

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

Franceschini, F. & Maisano, D. (2011) Criticsm of the hg-index. *Scientometrics.* 86(2), pp. 339-346.

Gaemers, K. (2007). *Evaluation report: Institute for biodiversity and ecosystem dynamics Universiteit van amsterdam.* Resource document. Universiteit van Amsterdam: Institute for Biodiversity and Ecosystem Dynamics (IBED). *Assessment-report-of-ibed-including-the-bibliometric-analysis.pdf. Accessed 11 April 2013*

Glänzel, W., Debackere, K., Thijs, B., & Schubert, A. (2006). A concise review on the role of author self-citations in information science, bibliometrics and science policy. *Scientometrics,* doi: 10.1007/s11192-006-0098-9

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. Resource document. Harzing.com. http://www.harzing.com/pop_hindex.htm Accessed 5 April 2013.

Harzing, A (2012). Document Categories in the ISI Web of Knowledge: Misunderstanding the Social Sciences? Resource document. Harzing.com. http://www.harzing.com/download/isi.pdf. Accessed 15 May 2013.

Harzing, A. (2012b). Publish or perish user's manual. Resource document. 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

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.

Hirst, G. (1978). Discipline impact factors: A method for determining core journal lists. *Journal of the American Society for Information Science,* doi: 10.1002/asi.4630290403.

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

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), pp.4-6.

Kosmulski, M. (2009). New seniority-independent hirsch-type index. *Journal of Informetrics,* 3(4), 341-347.

Leimu, R. (2005). What determines citation frequency of ecological papers? *Trends in Ecology and Evolution,* 20(1), 28-32.

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

Maabreh, M. & Alsmadi, I. (2012) A Survey of Impact and Citation Indicators: Limitations and Issues. *International Journal of Advanced Science and Technology.* 40, pp.35-54.

MacRoberts, M.H, & MacRoberts, B.R. (2010) Problems of Citation Analysis: A Study of Uncited and Seldom-Cited Influences. *Journal of the American Society of Information Science and Technology.* doi: 10.1002/asi.21228

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

Moed, H. F. (2005). *Citation analysis and research evaluation*. The Netherlands: Springer.

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

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. Resource document. Health Research Policy Systems. http://www.biomedcentral.com/content/pdf/1478-4505-8-30.pdf Accessed 5 April 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), 229-30.

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.

Price, D. 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.

Quomiam, L., Hassanaly, P., Baldit, P., Rostaing, H. & Dou, H. (1993) Bibliometric Analysis of patents for R & D management. *Research Management.* 3(1), pp. 13-18.

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.

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

Rousseau, R. (2006). New developments related to the hirsch index. Resource document. 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

Russell, J., & Rousseau, R. (2002). Bibliometrics and institutional evaluation. In Arvantis, R. (Ed.), *Encyclopedia of Life Support Systems (EOLSS).* Part 19-3: Science and Technology Policy. Oxford:EOLSS

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.

Satyanarayana, K. (2010). Impact factor and other indicies to assess science, scientists and scientific journals. *Indian Journal Physiol Pharmacol,* 54(3), 197-212.

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. *arXiv:0802.1820[Physics.Soc-Ph].*

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,* doi: 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., & Moed, H. F. (2002). Development and application of journal impact measures in the dutch science system. *Scientometrics,* 53(2), 249-266.

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

van Raan, A. (2004) Sleeping Beauties in science. *Scientometrics*, 59(3), 461-466.

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

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

Waltman, L & Schreiber, M. (2012) On the calculation of percentile-based bibliometric indicators. *Journal of the American Society for Information Science and Technology,* doi: 10.1002/asi.22775

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

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

ACUMEN
academic careers understood through measurement and norms

FP7 Grant Agreement	266632

Deliverable No and Title	D5.8 Part 2 – Bibliometric Indicators of Young Authors in Astrophysics: Can Later Stars be Predicted?
Dissemination Level	PU (public)
Work Package	WP5 Bibliometric Indicators
Version	1.0
Release Date	15-04-2014
Author(s)	Frank Havemann Birger Larsen

Scientometrics

Bibliometric Indicators of Young Authors in Astrophysics: Can Later Stars be Predicted?

--Manuscript Draft--

Manuscript Number:	
Full Title:	Bibliometric Indicators of Young Authors in Astrophysics: Can Later Stars be Predicted?
Article Type:	Manuscript
Keywords:	bibliometric indicators; research evaluation
Corresponding Author:	Frank Havemann GERMANY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	
Corresponding Author's Secondary Institution:	
First Author:	Frank Havemann
First Author Secondary Information:	
Order of Authors:	Frank Havemann
	Birger Larsen, PhD
Order of Authors Secondary Information:	
Abstract:	We test 16 bibliometric indicators with respect to their validity at the level of the individual researcher by estimating their power to predict later successful researchers. We compare the indicators of a sample of astrophysics researchers who later co-authored highly cited papers before their first landmark paper with the distributions of these indicators over a random control group of young authors in astronomy and astrophysics. We find that field and citation-window normalisation substantially improves the predicting power of citation indicators. The two indicators of total influence based on citation numbers normalised with expected citation numbers are the only indicators which show differences between later stars and random authors significant on a 1% level. Indicators of paper output are not very useful to predict later stars. The famous h-index makes no difference at all between later stars and the random control group.

Scientometrics manuscript No.
(will be inserted by the editor)

Bibliometric Indicators of Young Authors in Astrophysics: Can Later Stars be Predicted?

Frank Havemann · Birger Larsen

Received: date / Accepted: date

Abstract We test 16 bibliometric indicators with respect to their validity at the level of the individual researcher by estimating their power to predict later successful researchers. We compare the indicators of a sample of astrophysics researchers who later co-authored highly cited papers before their first landmark paper with the distributions of these indicators over a random control group of young authors in astronomy and astrophysics. We find that field and citation-window normalisation substantially improves the predicting power of citation indicators. The two indicators of total influence based on citation numbers normalised with expected citation numbers are the only indicators which show differences between later stars and random authors significant on a 1 % level. Indicators of paper output are not very useful to predict later stars. The famous h-index makes no difference at all between later stars and the random control group.

Keywords bibliometric indicators · research evaluation · astrophysics

1 Introduction

Any indicator should actually indicate what it is made for. If an indicator is used for evaluation it should not provide an incentive for an unwanted behaviour. In scholarly publishing we know salami and multiple publications, unjustified assignment of co-authorship, and different practices of tactical ci-

Frank Havemann
Institut für Bibliotheks- und Informationswissenschaft, Humboldt-Universität zu Berlin, D 10099 Berlin, Dorotheenstr. 26, Germany
Tel.: +49-30-20934228
E-mail: frank.havemann@ibi.hu-berlin.de

Birger Larsen
Department of Communication, Aalborg University, Copenhagen, Denmark

tation behaviour. Bibliometricians should strive to develop valid research indicators which have no unwanted adverse effects (Kreiman and Maunsell 2011).

Most bibliometric indicators are not developed for the evaluation of individual researchers (Costas, van Leeuwen, and Bordons 2010, p. 1565), however individuals are increasingly being evaluated using such indicators. We test selected indicators with respect to their validity at the level of the individual researcher by estimating their power to predict later successful researchers. For this reason, we compare bibliometric indicators of a sample of astrophysics researchers who later co-authored highly cited papers (later stars, for short) before their first landmark paper with the distributions of these indicators over a random control group of young authors in astronomy and astrophysics.

Results obtained with some standard basic indicators have been presented on a poster at ISSI 2013.[1] Here we extend the study to more sophisticated measures with the aim to find the best indicators for predicting later stars. We imagine that later stars apply for a job in an astrophysical research institute five years after their first paper in a journal indexed in Web of Science (WoS). Do they perform better bibliometrically than the average of applicants with the same period of publishing?

2 Data and method

2.1 Sampling of authors

We inspected 64 astronomy and astrophysics journals to find researchers who started publishing after 1990 and had published for a period of at least five years in WoS journals. We excluded those who had more than 50 co-authors on average because evaluating those big-science authors cannot be supported by bibliometrics. We draw a random sample of 331 authors mainly publishing in this field and affiliated longer in Europe then elsewhere. The latter criterion contradicts with the international character of astrophysics research but makes the sample more homogenous with respect to the educational and cultural background of the researchers.

To find authors with highly cited papers, for each journal considered we ranked papers with more than four citations per year and less than ten authors according to their citations per year. We excluded papers with ten or more authors because we want to have later stars whose contributions to the successful papers are not too small. From the top 20 percent of these paper rank-lists we extracted all European authors of highly cited papers. We obtained 362 candidates who published their first highly cited paper at least five years after their first paper in one the 64 journals.

We ranked these later-star candidates according to their number of highly cited papers. We went through this list and checked whether the authors had really five years or more to wait for the breakthrough paper if all their papers

[1] *14th International Society of Scientometrics and Informetrics Conference* in Vienna, Austria, 15th to 20th July 2013 (Havemann and Larsen 2013)

in WoS-journals are taken into account. We chose the first 40 authors to keep the effort manageable. For all WoS-papers of the 40 later stars and of the 331 random authors (downloaded at Humboldt-University, Berlin) all citing papers were determined by CWTS, Leiden. All bibliometric indicators presented below are based on papers and their citations within the first five years of the author. To compare only authors with similar collaboration behaviour we restricted both samples to authors with less than four and more than one co-author on average ending up with 30 later stars and 179 random authors.

We further restricted both samples to authors starting before 1999 because there is only one star starting later (in 2002) but many random authors (more than 100). By this restriction to 29 stars and 74 authors in the control group we take into account that the citation behaviour of astrophysicists has changed remarkably during the last 25 years. The numbers of references have increased. The median of reference numbers of the 448 papers published in the 1986 volume of the *Monthly Notices of the Royal Astronomical Society* was 24. Till the year 2010 the median of reference numbers has doubled (calculated with 2,006 papers, data source: WoS).[2] Longer lists of references induce higher citation numbers of papers. Thus, both samples still have a time variance of expected citation numbers. This time variance increases the overlap between the citation-indicator distributions of the samples when citation numbers are not normalised. In other aspects the union of our samples is surely more homogenous than many real groups of applicants (career duration, collaboration behaviour, geographical background).

An alternative data source for astrophysics publications and their citations is the *Astrophysics Data System* (ADS)[3] delivered jointly by the US National Aeronautics and Space Administration (NASA) and the Smithsonian Astrophysical Observatory (Henneken, Kurtz, and Accomazzi 2011). ADS includes also non-refereed publications. Any user can obtain a whole slew of bibliometric indicators for any set of selected publications.

2.2 Statistics

For each bibliometric indicator considered, we test whether both samples behave like random samples drawn from the same population by applying a one-sided Wilcoxon rank sum test with continuity correction. We test the null hypothesis that for both samples we have the same probability of drawing an author with a larger value in the other sample. The alternative hypothesis is that indicator values of later stars exceed the values of random authors.[4]

We have also tested the hypothesis that for both samples we have the same probability of drawing an author with a larger value of the *collaborative coefficient* (Ajiferuke, Burrell, and Tague 1988, cf. also our Table 1, p. 5) in

[2] cf. Henneken, Kurtz, and Accomazzi (2011, p. 5)

[3] `http://adsabs.harvard.edu`

[4] cf. the Wikipedia article `http://en.wikipedia.org/wiki/Mann-Whitney-Wilcoxon_test`

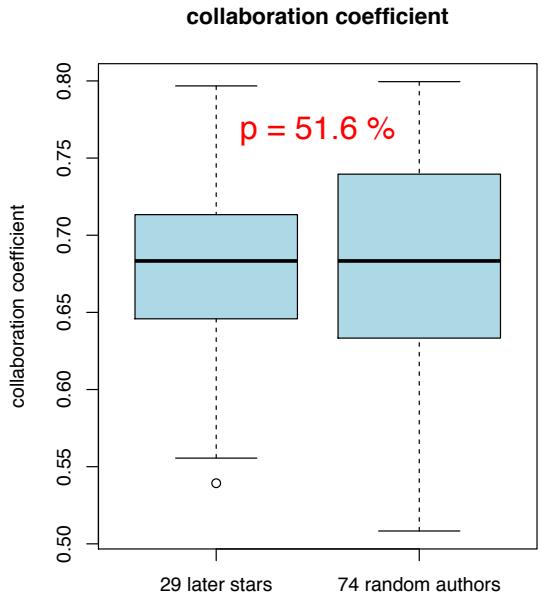


Fig. 1 The authors in the two samples have similar distributions of collaboration behaviour.

the other sample. In both samples we have a similar collaboration behaviour (cf. Figure 1). If we would refuse the null hypothesis we would fail in about one half of possible cases (test probability $p = .516$). This result ensures that differences between both groups are not due to different typical team sizes.

All work was done using the free open-source statistics software **R** (which includes a graphics package).[5]

2.3 Selection of indicators

The indicators analysed here are listed together with their mathematical definitions in Table 1. In Appendix A.1 we discuss the definition of each of these indicators.

We have calculated and tested two simple output indicators and nine indicators of influence. Beside pure numbers of papers and their citations within the first five publishing years of the authors we use fractionally counted papers and citations as the input for indicators of productivity and of influence. The use of fractional counting in evaluation penalises unjustified assignment of co-authorship to friends.

If we compare papers published in fields with different citation behaviour any citation indicator should be field normalised with expected citation numbers. Here we consider only one field but—as mentioned above—the citation

[5] `http://www.r-project.org` (**R**-scripts for indicator calculation and sample data can be obtained from the first author of this paper.)

behaviour of astrophysicists has changed dramatically within the last decades. That means, distributions of unnormalised citation indicators of the two samples overlap partly due to the changing citation behaviour.

Another wanted effect of normalising with expected citation numbers is that we account for different citation windows of papers. Thus, citations to papers published in the beginning of a period obtain a lower weight than those to papers published in the last year. The estimation of expected citation numbers of papers is described in Appendix A.2.

Another method to deal with varying citation behaviour is to determine each paper's percentile in the citation distribution of a control sample of papers. Bornmann, Leydesdorff, and Wang (2013) compare five approaches to this promising method. Percentile ranking avoids the use of arithmetic means of heavily skewed citation distributions. We minimise the influence of skewness by calculating expected citation numbers by a linear regression over all years considered (s. Appendix A.2). We have to leave a test of the percentile method with our samples to further work due to a lack of citation data of control samples.

Table 1 List of author indicators: a_i is the number of authors of paper i; c_i is the number of citations of paper i; $\mathrm{E}(c_i)$ is the expected number of citations of paper i (cf. Appendix A.2); we assume that papers of an author are ordered according to c_i and denote the paper's rank with r; the effective rank is defined as $r_{\mathrm{eff}}(r) = \sum_i^r 1/a_i$.

name	definition
productivity:	
nr. of papers	$\sum_i 1 = n$
fractional score	$\sum_i 1/a_i = f$
total influence:	
nr. of citations	$\sum_i c_i$
norm. nr. cit.	$\sum_i c_i/\mathrm{E}(c_i)$
j-index	$\sum_i \sqrt{c_i}$
fract. citations	$\sum_i c_i/a_i$
fract. norm. cit.	$\sum_i c_i/(\mathrm{E}(c_i)a_i)$
typical infl.:	
mean cit. nr.	$\sum_i c_i/n$
mean fract. cit.	$\sum_i (c_i/a_i)/n$
med. fract. cit.	$\mathrm{median}(c_i/a_i)$
max. fract. cit.	$\max(c_i/a_i)$
h-type indices:	
Hirsch index	$\max(r \vert c_r \geq r)$
g-index	$\max(r \vert \sum_i^r c_i \geq r^2)$
fract. h-type:	
h_{m}-index	$\max(r_{\mathrm{eff}} \vert c_{r(r_{\mathrm{eff}})} \geq r_{\mathrm{eff}})$
g_{f}-index	$\max(r \vert \sum_i^r c_i/a_i \geq r^2)$
g_{m}-index	$\max(r_{\mathrm{eff}} \vert \sum_i^{r(r_{\mathrm{eff}})} c_i/a_i \geq r_{\mathrm{eff}}^2)$
collaboration:	
collab. coeff.	$1 - f/n$

Recently, several authors tested a third approach to field normalisation of citation numbers. Here data on the citing side are normalised. Waltman and van Eck (2013, s. also references of this paper) discuss three variants of this method. Also this approach cannot be tested with the data we have at hand. We could test the simplest variant where each citation of a paper is divided by the number of all references of the citing paper (Zhou and Leydesdorff 2011; Pepe and Kurtz 2012). Waltman and van Eck (2013) and also Radicchi and Castellano (2012) found that this fractional counting of references does not properly normalise for field and subfield differences. A further drawback of this variant is that citation numbers are not corrected for the age of the cited paper. We therefore did not test it.

In addition to the eleven indicators of productivity and of influence we calculated the widely used Hirsch or h-index (Hirsch 2005), a number combining influence and output performance in an uncontrolled and arbitrary manner, and four variants of it which have been introduced to avoid disadvantages of the Hirsch index.

We did not consider any indicator based on the number of highly cited papers because this contradicts our sampling procedure: we selected *later* stars who have *no* highly cited paper in their first five years of publishing.

3 Results

Medians of all 16 indicators of both samples are given in Table 2. In the next to last column of Table 2 we list the failure probability p of rejecting the null hypothesis that both samples behave like random samples drawn from the same population. In the last column we give the rank R according to p. For all but the two indicators on least ranks (Hirsch index and *median of fractional citation numbers*) the stars' sample has a higher median than the random sample.

The boxplots in Appendix A.3 allow a comparison of indicator distributions for both samples. The figures are ordered according to the ranking R. That means that p-values increase from the first to the last boxplot. The boxplots have a logarithmic scale because all indicator distributions are highly skewed. All citation indicators have zero values for some uncited authors in the control sample. Therefore we display the logarithm of indicator values + 1.

The two indicators based on normalised citation numbers are the most useful among the 16 indicators considered (s. Figure 3). With respect to *normalised numbers of citations* and to *fractional normalised citations* both samples behave not like random samples from the same population. In both cases, rejecting the null hypothesis has a failure probability below 1 %.

The distributions of eight further indicators differ at least on a 5 % significance level (s. Figures 4–7). For the remaining six indicators there is no significant difference between distributions of later stars and of authors in the control group (s. Figures 8–10). The Hirsch-index has very similar distributions for both samples ($p = 21$ %, rank 15, s. Figure 10).

Table 2 Median indicators of samples, test probability p, and rank R (according to p)

indicator	stars	random	p	R
productivity:				
nr. of papers	8	6	.076	12
fractional score	2.67	1.86	.095	13
total influence:				
nr. of citations	36	22.5	.028	6
norm. nr. cit.	6.03	3.83	.003	1
j-index	11.86	8.76	.031	9
fract. citations	10.00	6.57	.030	7
fract. norm. cit.	1.82	1.10	.008	2
typical infl.:				
mean cit. nr.	5.25	4.00	.117	14
mean fract. cit.	1.23	0.99	.062	11
med. fract. cit.	0.50	0.67	.260	16
max. fract. cit.	4.67	3.00	.030	8
h-type indices:				
Hirsch index	3	3	.210	15
g-index	5	4	.037	10
fract. h-type:				
h_{m}-index	1.32	1.00	.020	3
g_{f}-index	3	2	.024	4
g_{m}-index	2.38	1.68	.025	5
collaboration:				
collab. coeff.	.683	.683	.516	17

4 Discussion

Our results underline the necessity to correct citation indicators for the age of the cited papers and also for varying citation behaviour.[6] The two indicators of total influence based on citation numbers normalised with expected citation numbers are the only indicators among a total of 16 which show significant differences between later stars and random authors on a 1 % level. Thus, normalised citation indicators of total influence can indeed help to predict later successful authors. Despite this relative good performance of normalised citation indicators of total influence we cannot recommend to use them as the only basis for an evaluation of young authors in astrophysics and in similar fields of natural sciences. Normalisation at the field level cannot correct for a variability in citation numbers between different topics. Opthof (2011) analysed the citation density in different topics of cardiovascular research papers and concluded that even normalised citation indicators "should not be used for quality assessment of individual scientists" (cf. his abstract).[7] In each case, bibliometrics can only support evaluation and cannot replace individual peer review.

[6] It would be interesting—from a theoretical point of view—to determine the influence of each of both corrections separately.

[7] Topics in physics as in astrophysics also differ substantially in citation density (Radicchi and Castellano 2011; Pepe and Kurtz 2012).

None of the two output indicators have a significant difference below the 5 % level.[8] Thus, it is very unlikely to discover a later star in astrophysics by comparing her productivity with the productivity of a random author (Figures 8 and 9). The Hirsch index makes no difference at all ($p = 21\,\%$, Figure 10). This is in agreement with conclusions drawn by Lehmann, Jackson, and Lautrup (2006) and also by Kosmulski (2012) who analysed small samples of mature scientists and found that the number of publications "is rather useless" as a tool of assessment and that also the h-index is not really helpful. In contrast to these findings, Pudovkin, Kretschmer, Stegmann, and Garfield (2012) found that h-index and number of papers are indicators which differ most significantly between group leaders and other scientists at a medical research institution. This can surely be explained by real output differences of elder and younger researchers but maybe partly also by the assumption that group leaders have more often been working at the institute over the whole analysed 5-years period than other researchers.

We could have analysed the generalised h-index proposed by Radicchi, Fortunato, and Castellano (2008) who use normalised citation and paper numbers. We did not because h performs much worse than indicators of total influence.

The g-index proposed by Egghe (2006) to improve the h-index performs indeed better than the original ($p = 3.7\,\%$, Figure 7). The same holds for the analysed three h-type indices which are based on fractional counting. They have been introduced by Egghe (2008) and by Schreiber (2008c, 2009) to account for varying collaboration behaviour.

There is no significant difference between the two samples when we compare citation indicators which are designed to reflect the mean influence of an author's papers. We calculated three of them: the arithmetic mean of citation numbers ($p = 11.7\,\%$, Figure 9), fractionally counted citations per paper ($p = 6.2\,\%$, Figure 8), and the median of the fractionally counted citations ($p = 26\,\%$, Figure 10). We wondered whether for a later star a large maximum of (fractional) citations is more typical than a large value of any measure of central tendency of citation numbers. The answer is yes. The maximum of fractional citations is a better indicator of typical influence ($p = 3\,\%$, Figure 6). We could have analysed normalised indicators of typical influence, too. We did not because indicators of typical influence do not perform better than those of total influence.

We do not exclude self-citations when calculating citation indicators. There are arguments for their exclusion in evaluative bibliometrics but we assume that it would be difficult for young authors to massively cite their own papers within their first five years of publishing.

We expect that weighting (fractional) paper numbers with a measure of journal reputation would improve the predictive power of output indicators. We did not test this because the only journal-reputation indicator available for us was the *journal impact factor* which is not useful here—albeit often used for

[8] This is in accordance with the result obtained by Neufeld, Huber, and Wegner (2013, cf. p. 9) when comparing successful with non-successful applicants of a funding programme for young researchers.

weighting paper numbers (Seglen 1997; Lozano, Larivière, and Gingras 2012, s. also the references of these papers).

Analysing 85 researchers in oncology Hönekopp and Khan (2012) found that "a linear combination of past productivity and the average paper's citation" is a better predictor of future publication success than any of the single indicators they had studied. We did not consider combinations of indicators of productivity and of mean influence because the simpler indicators of total influence also reflect productivity—as far as the produced papers have been cited. Neglecting uncited papers is a wanted effect that is also quoted in favour of the h-index.

Hornbostel, Böhmer, Klingsporn, Neufeld, and von Ins (2009) found only small differences in numbers of publications and citations between approved and rejected applicants to a German funding programm for young researchers. In an earlier study, Nederhof and van Raan (1987) compared 19 PhD graduates in physics with best degrees to 119 other graduates with lower grade. They considered the total number of papers before and after graduation and their total and average (short time) impact. The 19 best graduates performed significantly better but, interestingly, the impact of their papers declined and reached the level of the control-group papers a few years after graduation. The authors speculate about the reason of this phenomenon and suggest that better students could have been engaged for hot and therefore highly cited research projects. They conclude, that maybe "the quality of the research project, and not the quality of the particular graduate is the most important determinant of both productivity and impact figures" (Nederhof and van Raan 1987, p. 348). This hypothesis could also hold for the young astrophysicists analysed by us. Its confirmation would further diminish the weight of bibliometric indicators in the evaluation of young researchers.

Acknowledgements We thank Jesper Schneider for helpful discussions of an early draft and Paul Wouters at CWTS in Leiden for providing citation data. The analysis was done for the purposes of the ACUMEN project, financed by the European Commission, cf. `http://research-acumen.eu/`.

References

Ajiferuke, I., Q. Burrell, and J. Tague (1988). Collaborative coefficient – a single measure of the degree of collaboration in research. *Scientometrics 14*, 421–433.

Bornmann, L., L. Leydesdorff, and J. Wang (2013, October). Which percentile-based approach should be preferred for calculating normalized citation impact values? An empirical comparison of five approaches including a newly developed citation-rank approach (P100). *Journal of Informetrics 7*(4), 933–944. s. a. `http://arxiv.org/abs/1306.4454`.

Costas, R., T. N. van Leeuwen, and M. Bordons (2010, August). A bibliometric classificatory approach for the study and assessment of research performance at the individual level: The effects of age on productivity

and impact. *Journal of the American Society for Information Science and Technology 61*(8), 1564–1581.

Egghe, L. (2006). An improvement of the h-index: the g-index. *ISSI Newsletter 2*(2), 8–9.

Egghe, L. (2008). Mathematical theory of the *h*- and *g*-index in case of fractional counting of authorship. *Journal of the American Society for Information Science and Technology 59*(10), 1608–1616.

Havemann, F. and B. Larsen (2013). Bibliometric Indicators of Young Authors in Astrophysics: Can Later Stars be Predicted? In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, and H. Moed (Eds.), *PROCEEDINGS OF ISSI 2013 Vienna*, Volume 2, pp. 1881–1883.

Henneken, E. A., M. J. Kurtz, and A. Accomazzi (2011). The ADS in the Information Age—Impact on Discovery. arXiv preprint arXiv:1106.5644.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences 102*(46), 16569–16572. `http://arxiv.org/abs/physics/0508025`.

Hönekopp, J. and J. Khan (2012). Future publication success in science is better predicted by traditional measures than by the h index. *Scientometrics 90*(3), 843–853.

Hornbostel, S., S. Böhmer, B. Klingsporn, J. Neufeld, and M. von Ins (2009, April). Funding of young scientist and scientific excellence. *Scientometrics 79*(1), 171–190.

Kosmulski, M. (2012, July). Calibration against a reference set: A quantitative approach to assessment of the methods of assessment of scientific output. *Journal of Informetrics 6*(3), 451–456.

Kreiman, G. and J. H. R. Maunsell (2011). Nine criteria for a measure of scientific output. *Frontiers in Computational Neuroscience 5*, article nr. 48 (6 pages).

Lehmann, S., A. D. Jackson, and B. E. Lautrup (2006, December). Measures for measures. *Nature 444*(7122), 1003–1004.

Lehmann, S., A. D. Jackson, and B. E. Lautrup (2008, August). A quantitative analysis of indicators of scientific performance. *Scientometrics 76*(2), 369–390.

Levene, M., T. Fenner, and J. Bar-Ilan (2012). A bibliometric index based on the complete list of cited publications. *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics* (16), 1–6. s.a. arXiv:1304.6945.

Lozano, G. A., V. Larivière, and Y. Gingras (2012). The weakening relationship between the impact factor and papers' citations in the digital age. *Journal of the American Society for Information Science and Technology 63*(11), 2140–2145.

Lundberg, J. (2007, April). Lifting the crown—citation *z*-score. *Journal of Informetrics 1*(2), 145–154.

Marchant, T. (2009, June). Score-based bibliometric rankings of authors. *Journal of the American Society for Information Science and Technology 60*(6), 1132–1137.

Nederhof, A. J. and A. F. J. van Raan (1987, May). Peer review and bibliometric indicators of scientific performance: A comparison of cum laude doctorates with ordinary doctorates in physics. *Scientometrics 11*(5-6), 333–350.

Neufeld, J., N. Huber, and A. Wegner (2013, January). Peer review-based selection decisions in individual research funding, applicants' publication strategies and performance: The case of the ERC starting grants. *Research Evaluation 22*(4), 237–247.

Opthof, T. (2011, June). Differences in citation frequency of clinical and basic science papers in cardiovascular research. *Medical & Biological Engineering & Computing 49*(6), 613–621.

Opthof, T. and L. Leydesdorff (2010, July). Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance. *Journal of Informetrics 4*(3), 423–430.

Pepe, A. and M. J. Kurtz (2012, November). A measure of total research impact independent of time and discipline. *PLoS ONE 7*(11), e46428.

Pudovkin, A., H. Kretschmer, J. Stegmann, and E. Garfield (2012). Research evaluation. Part I: productivity and citedness of a German medical research institution. *Scientometrics 93*(1), 3–16.

Radicchi, F. and C. Castellano (2011, April). Rescaling citations of publications in physics. *Physical Review E 83*(4), 046116.

Radicchi, F. and C. Castellano (2012, January). Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts. *Journal of Informetrics 6*(1), 121–130.

Radicchi, F., S. Fortunato, and C. Castellano (2008, November). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences 105*(45), 17268–17272.

Schreiber, M. (2008a, July). A modification of the h-index: The h_m-index accounts for multi-authored manuscripts. *Journal of Informetrics 2*(3), 211–216.

Schreiber, M. (2008b). The influence of self-citation corrections on Egghe's g index. *Scientometrics 76*(1), 187–200. Cf. also arXiv:0707.4577.

Schreiber, M. (2008c). To share the fame in a fair way, h_{m} modifies h for multi-authored manuscripts. *New Journal of Physics 10*(4), 040201.

Schreiber, M. (2009). Fractionalized counting of publications for the g-index. *Journal of the American Society for Information Science and Technology 60*(10), 2145–2150.

Schubert, A. and T. Braun (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics 9*(5), 281–291.

Seglen, P. O. (1997, 2). Why the impact factor of journals should not be used for evaluating research. *BMJ: British Medical Journal 314*(7079), 498–513.

van Eck, N. J. and L. Waltman (2008, October). Generalizing the h- and g-indices. *Journal of Informetrics 2*(4), 263–271.

Waltman, L. and N. J. van Eck (2012). The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology 63*(2), 406–415.

Waltman, L. and N. J. van Eck (2013, October). A systematic empirical comparison of different approaches for normalizing citation impact indicators. *Journal of Informetrics 7*(4), 833–849.

Waltman, L., N. J. van Eck, T. N. van Leeuwen, M. S. Visser, and A. F. J. van Raan (2011, February). Towards a new crown indicator: an empirical analysis. *Scientometrics 87*(3), 467–481.

Zhou, P. and L. Leydesdorff (2011). Fractional counting of citations in research evaluation: A cross-and interdisciplinary assessment of the Tsinghua University in Beijing. *Journal of Informetrics 5*(3), 360–368.

A Appendix

A.1 Descriptions of indicators

A.1.1 Productivity indicators

Number of papers: This elementary indicator of productivity belongs to a bygone era when co-authorship was the exception and not the rule. It has the unwanted adverse effects of multiple publishing of the same results and of honorary authorships.

Fractional score: Each paper i is divided into a_i fractions where a_i is the number of authors. These fractions are summed up for the papers of the evaluated author. We use the simplest variant where all fractions of a paper are equal: $f = \sum_i 1/a_i$. This indicator penalises honorary authorships and takes into account that larger teams can be more productive.

A.1.2 Total influence

All indicators of total influence tend to increase with the author's number of papers. That means, they are also indicating productivity.

Number of citations: Each citation of a paper indicates that it has influenced the citing author(s). The sum $\sum_i c_i$ of raw numbers c_i of citations of an author's papers is highly field dependent. The paper's number of citations c_i depends on the age of a paper at the time of evaluation. Highly cited papers have surely some quality but less cited ones can also be of high quality.

Normalised numbers of citations: We normalise each paper's number of citations c_i by an expected number of citations $\mathrm{E}(c_i)$ which takes into account the paper's age and the citation behaviour in astrophysics during the first five (calendar) years in the paper's lifetime (cf. Appendix A.2). After normalising each paper's citation number we sum the ratios of observed and expected citation numbers:

$$\sum_{i=1}^{n} \frac{c_i}{\mathrm{E(c_i)}}.$$

Some bibliometricians do not calculate the sum of ratios but the ratio of sums $\sum_i c_i / \sum_i \mathrm{E}(c_i)$ (Schubert and Braun 1986). This procedure is thought to evaluate the whole oeuvre of an author but has been criticised recently for being not "consistent" (Opthof and Leydesdorff 2010; Waltman, van Eck, van Leeuwen, Visser, and van Raan 2011).[9]

The j-index: The j-index is the sum of the square roots of citation numbers of the author's papers

$$\sum_{i=1}^{n} \sqrt{c_i}.$$

It was proposed by Levene, Fenner, and Bar-Ilan (2012) to downgrade the influence of highly cited papers in the sum of citation numbers.

Fractional citations: Analogously to the fractional score described above we distribute citations of each paper equally to its authors:

$$\sum_{i=1}^{n} \frac{c_i}{a_i}.$$

Fractional normalised citations: The normalised numbers of citations can also be distributed among the authors involved (Radicchi and Castellano 2011):

$$\sum_{i=1}^{n} \frac{c_i}{\mathrm{E}(c_i) a_i}.$$

A.1.3 Typical influence

Mean citation number: The arithmetic mean of citations of an author's papers

$$\frac{1}{n} \sum_{i=1}^{n} c_i.$$

is the simplest indicator of influence which does not tend to increase with the author's productivity.

Mean fractional citations: The arithmetic mean of fractionally counted citations of an author's papers:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{c_i}{a_i}.$$

Median of fractional citations: The median of fractionally counted citations of an author's papers median(c_i/a_i) is considered because citation distributions are skewed.

Maximum of fractional citations: We wondered whether for a later star a large maximum of (fractional) citations max(c_i/a_i) is more typical than a large value of any measure of central tendency of citation numbers (Lehmann, Jackson, and Lautrup 2008, cf. p. 375).

[9] The h-index is also not consistent (Marchant 2009; Waltman and van Eck 2012).

A.1.4 Indices of h-type

Hirsch index: The h-index was introduced by Hirsch (2005) "to quantify an individual's scientific research output." It is defined as the maximum rank r in a rank list of an author's papers according to their citation numbers c_i which is less than or equal to the citation number c_r of the paper with rank r: $h = \max(r|c_r \geq r)$. The h-index has been criticised for its arbitrariness (van Eck and Waltman 2008). It is arbitrary because in the definition Hirsch "assumes an equality between incommensurable quantities" (Lehmann, Jackson, and Lautrup 2008, p. 377), namely a rank and a citation number. Hirsch himself stated that his index depends on field-specific citation and collaboration behaviour (Hirsch 2005, p. 16571).

Egghe's g-index: Egghe (2006) criticised the h-index for being insensitive to the citation frequency of an author's highly cited papers. His g-index can be defined as the maximum rank r which is less than or equal to the mean citation number $(\sum_i^r c_i)/r$ of papers till rank r (Schreiber 2008b). This condition is equivalent to $\sum_i^r c_i \geq r^2$. That means, g can also be defined as

$$g = \max(r|\sum_{i=1}^{r} c_i \geq r^2).$$

A.1.5 Fractional indices of h-type

Schreiber's h_{m}-index: Fractional counting of papers or of citations could be applied to define an h-index which takes multi-authorship into account (Egghe 2008; Schreiber 2008c). Schreiber (2008a) argued that fractionally counted citations could remove highly cited papers from the h-core if they have a lot of authors. This led him to define the h_{m}-index as the maximal effective rank $r_{\mathrm{eff}}(r) = \sum_i^r 1/a_i$ which is less than or equal to the number of citations c_r:

$$h_{\mathrm{m}} = \max(r_{\mathrm{eff}}|c_{r(r_{\mathrm{eff}})} \geq r_{\mathrm{eff}}).$$

Egghe's g_{f}-index: Egghe (2008) proposed to define a fractional g-index g_{f} as

$$g_{\mathrm{f}} = \max(r|\sum_{i=1}^{r} \frac{c_i}{a_i} \geq r^2).$$

Here the citations are counted fractionally.

Schreiber's g_{m}-index: Schreiber (2009) proposed a fractional g-index g_{m} where both, papers and citations, are counted fractionally:

$$g_{\mathrm{m}} = \max(r_{\mathrm{eff}}|\sum_{i=1}^{r(r_{\mathrm{eff}})} \frac{c_i}{a_i} \geq r_{\mathrm{eff}}^2).$$

A.2 Expected citation numbers

Usually, for field normalisation expected citation numbers of papers are calculated as arithmetic means of citation numbers of all papers (of the same document type) published in all journals of the field in the same year. There are two main technical problems with this method, the rough delineation of fields and the skewness of citation distributions.

We do not evaluate single authors but only want to show the influence of field normalisation on distributions of citation indicators of authors. Therefore we can use a random sample of papers (for which we have already the citation data) instead of all papers in the field. This

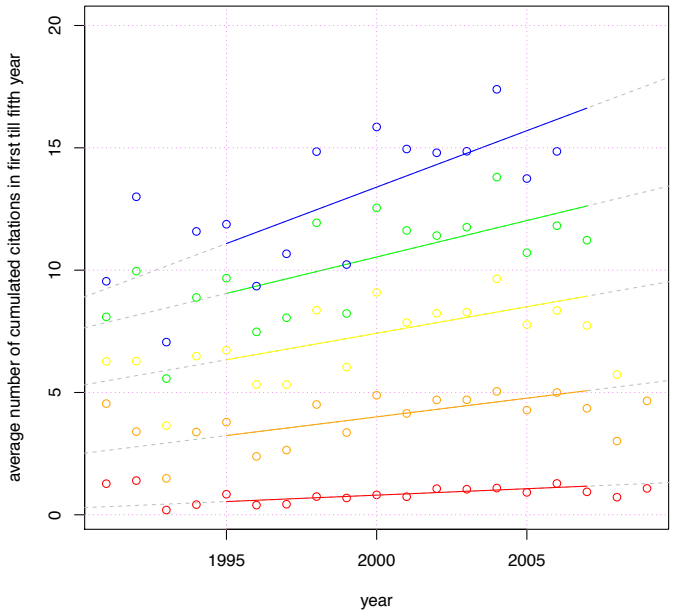


Fig. 2 Linear regressions and averages of citation numbers of papers of random authors in astrophysics after the first (the publication) year (red), the second year (orange), the third year (yellow), the fourth year (green), and the fifth year (blue).

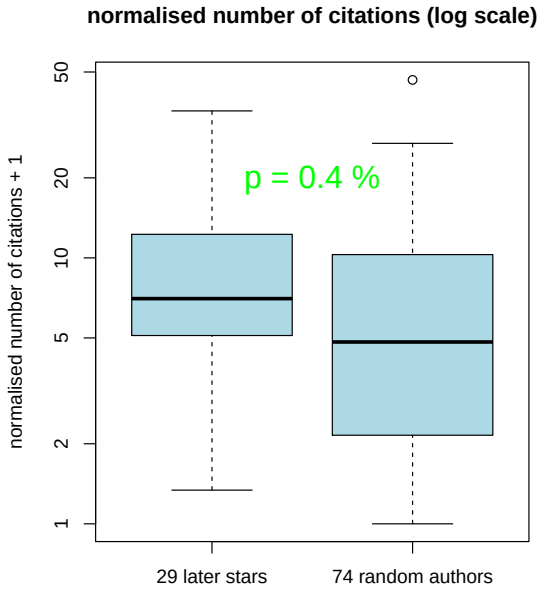
sample contains papers published in the years 1991–2009 by all 331 random authors of our initial control sample. We only consider those 2342 papers with at most 20 authors. Figure 2 shows the average cumulated citation numbers in the publication year, one year later, two years later etc. Due to the skewness of citation distributions these arithmetic means fluctuate. Therefore we made a linear regression for each of the five time series of citation numbers of papers (not of the averages) but restricted the analysis to the years 1995–2007 (coloured part of the regression lines) where we have more than 100 papers in each year. The interpolated citation numbers obtained by linear regression are used as expected citation numbers $\mathrm{E}(c_i)$ of papers published in the corresponding years.

From these data we estimate a doubling of citation numbers in astrophysics in the two decades around the millennium.

Calculating expected citation numbers as field averages is problematic because the arithmetic mean is not a good measure for the central tendency of skewed citation distributions. Lundberg (2007) therefore proposed to determine expected citation numbers as geometric means of citation numbers of papers in the field. Because papers can have zero citations he adds 1 to be able to calculate the geometric mean. This can be justified by saying that publishing a paper is the first citation of the published results.

A.3 Boxplots of indicators

On the next pages you find boxplots of distributions of all 16 indicators both of the sample of 29 later stars and of the control sample of 74 random young astrophysicists.



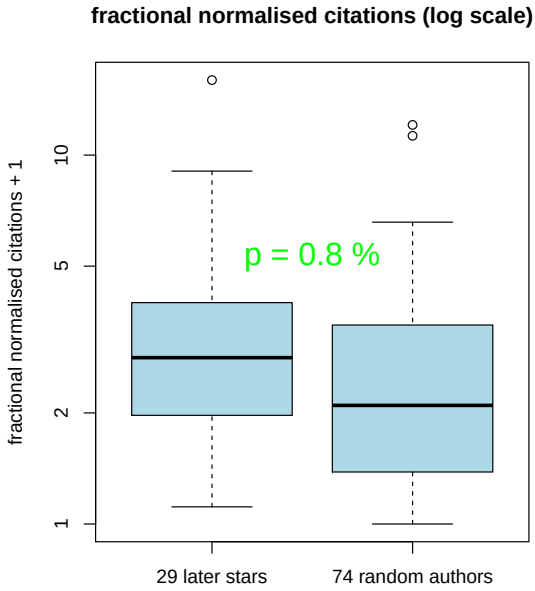


Fig. 3 The two indicators with best p-values: $p < 1\,\%$

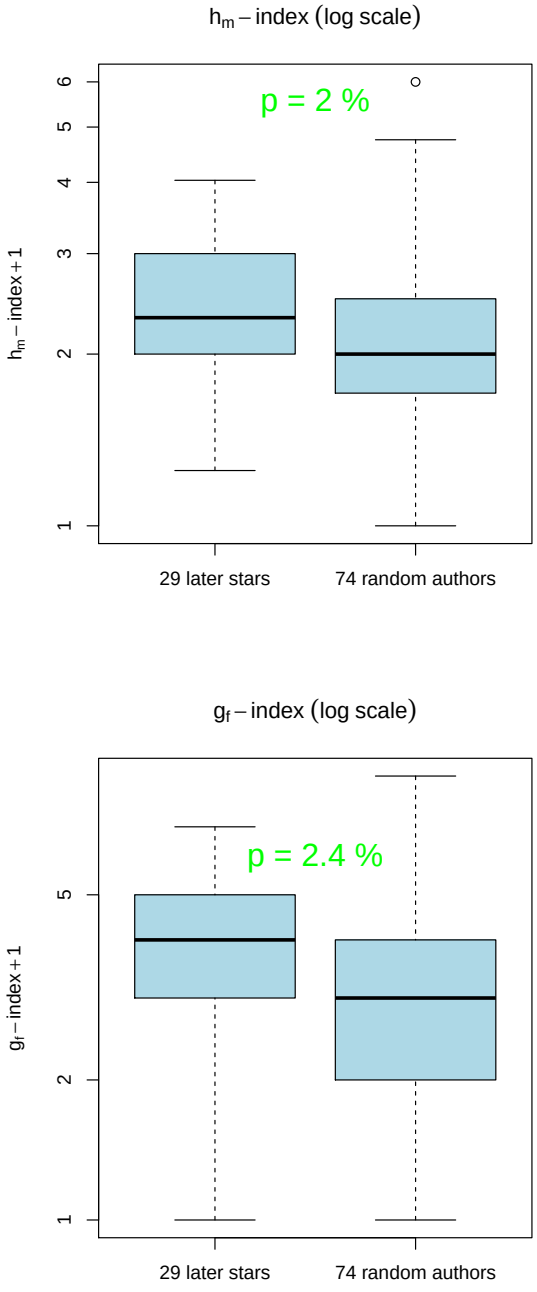
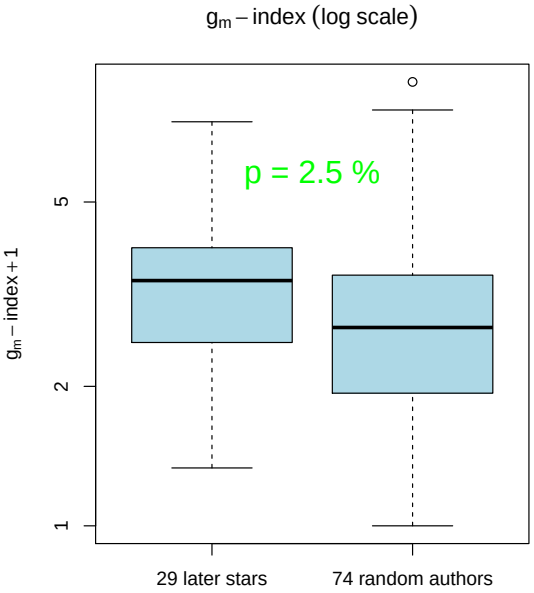


Fig. 4 The indicators on rank 3 and 4 according to p-values: $p < 5\,\%$



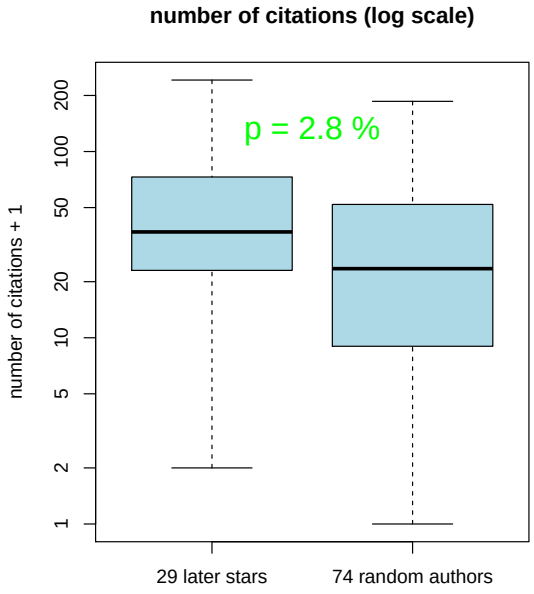


Fig. 5 The indicators on rank 5 and 6 according to p-values: $p < 5\,\%$

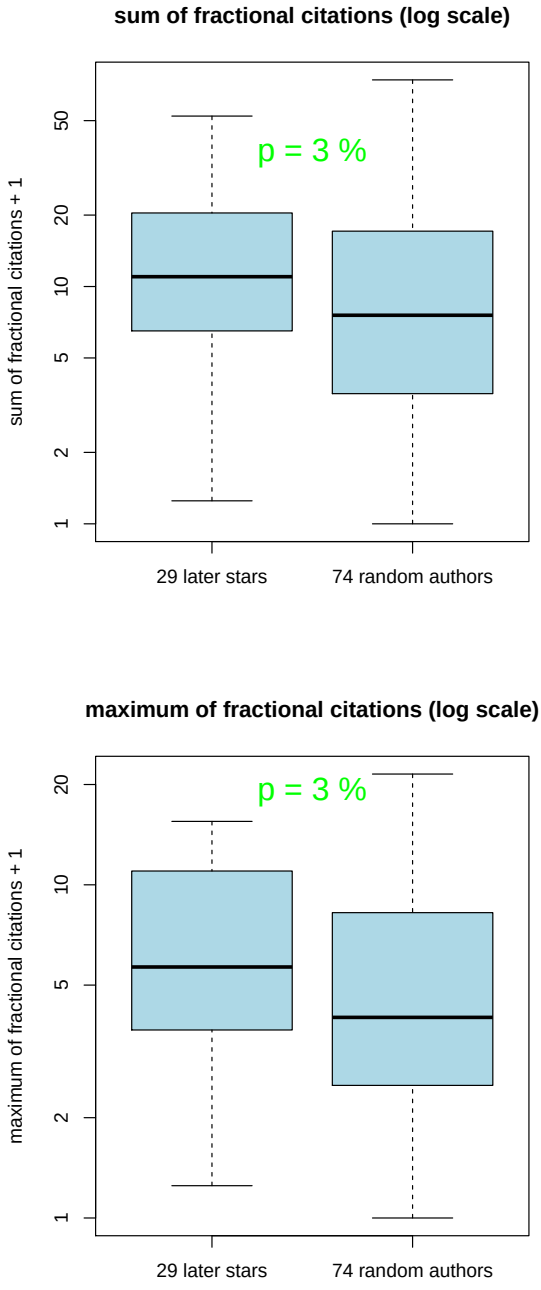
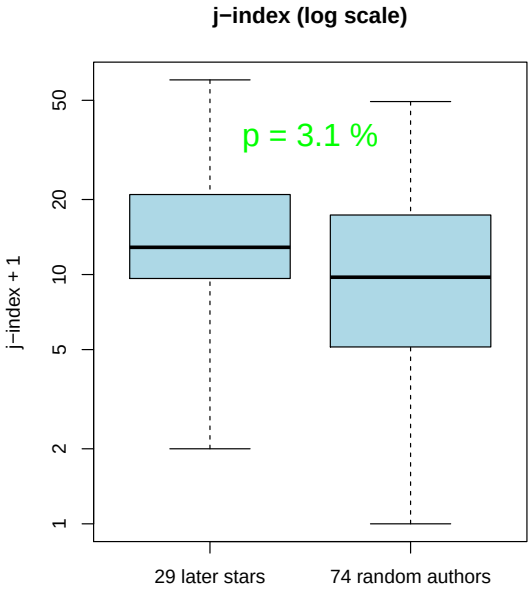


Fig. 6 The indicators on rank 7 and 8 according to p-values: $p < 5\,\%$



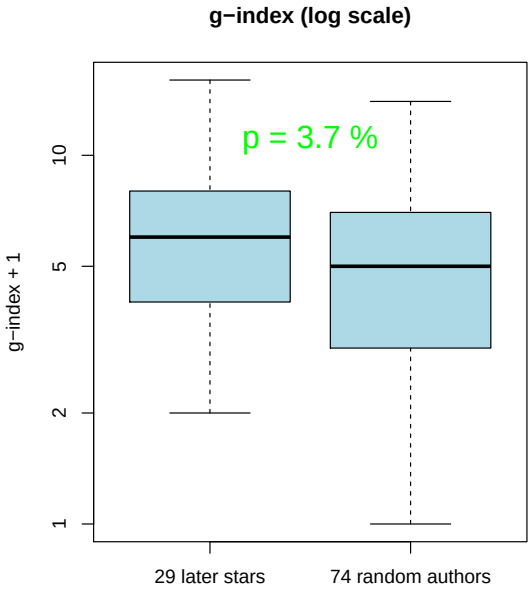
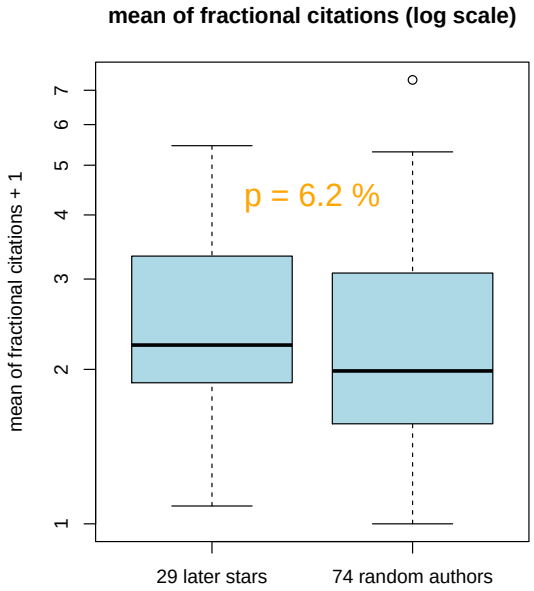


Fig. 7 The indicators on rank 9 and 10 according to p-values: $p < 5\,\%$



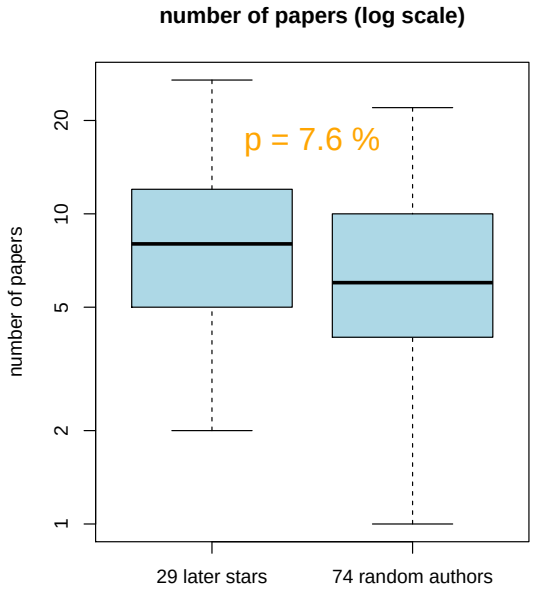
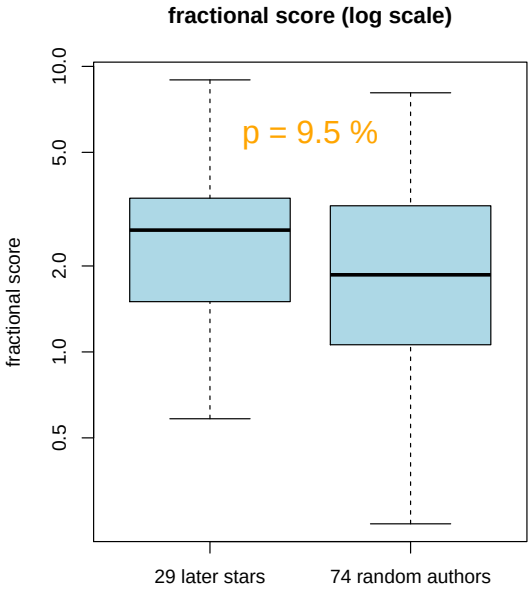


Fig. 8 The indicators on rank 11 and 12 according to p-values: $p < 10\,\%$



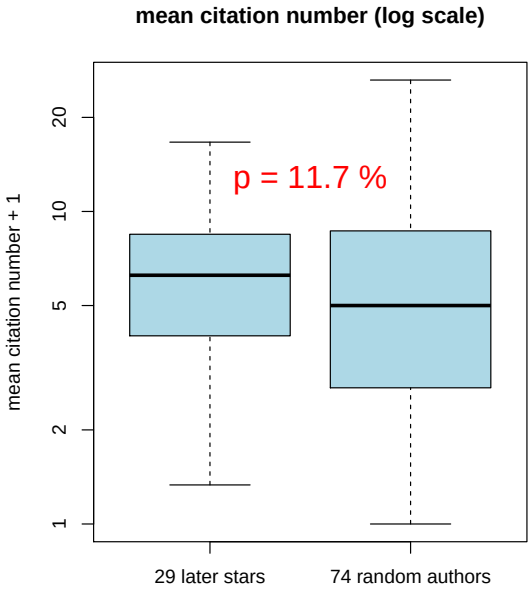


Fig. 9 The indicators on rank 13 and 14 according to p-values

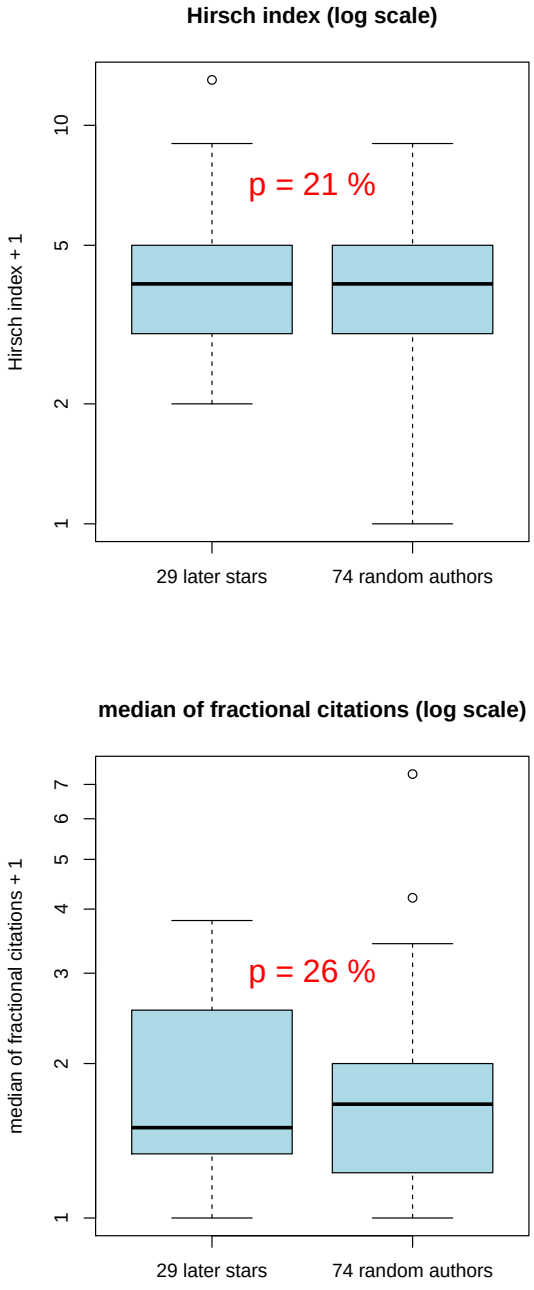


Fig. 10 The indicators on rank 15 and 16 according to p-values

http://www.springer.com/journal/11192

Scientometrics
An International Journal for all Quantitative Aspects of the Science of Science, Communication in Science and Science Policy
Editor-in-Chief: Glänzel, W.
ISSN: 0138-9130 (print version)
ISSN: 1588-2861 (electronic version)
Journal no. 11192

ACUMEN

academic careers understood through measurement and norms

FP7 Grant Agreement	266632
Deliverable No and Title	D5.8 Part 3 - Selection of Samples
Dissemination Level	PU (public)
Work Package	WP5. Bibliometric Indicators
Version	1.0
Release Date	06-08-2013
Author(s)	Lorna Wildgaard Birger Larsen Jesper W Schneider
Project Website	http://research-acumen.eu/

European Commission
7th Framework Programme
SP4 - Capacities
Science in Society 2010
Grant Agreement: 266632

EUROPEAN COMMISSION

SEVENTH FRAMEWORK PROGRAMME

PART A

Part A. Preparing for the analysis. Sampling strategy and methodological considerations in developing bibliometric indicators of the performance and impact of individuals for use in the ACUMEN portfolio.

Work Package 5: New Bibliometric indicators
June 28th, 2013

Project partners: Department of Information Studies, Royal School of Library and Information Science; Department of Library and Information Science, Humboldt University Berlin

Executive Summary:

Based on the samples from the four research fields used in the other WPs we have identified 793 researchers with online publication lists. Publication data from these researchers were gathered and combined with demographic data from the survey. Bibliometric analyses of these publications were undertaken in WoS and Google Scholar using a set of indicators designed for assessment at the individual level. The sample of 64 indicators were previously identified in the review of 114 bibliometric indicators, D5.8 Part 1 as presented in Madrid in January 2013. The set of 64 indicators has been reduced to 40 using a number of selection criteria.

We decided to use (construct) a decision-tree (which in a reworked form could go into the portfolio) as the guiding principle when choosing and comparing indicators. Our basic pragmatic assumption is that since indicators are already provided on many curriculum vitaes (CV's), though there are great variations across fields, simplicity and the ease with which such indicators can be obtained and/or compiled, are the basis for our analyses and later recommendations. We observed that what sets the ACUMEN portfolio apart from the current use of indicators on CV's, is the portfolios potential to give the researcher guidelines to aid interpretation of the indicators and set them in a narrative the enriches the cv.

The main tasks therefore are 1) to characterize types of indicators; 2) to examine (within the dataset) to what extent easily obtainable indicators correlate with more sophisticated indicators, as the latter would be close to impossible for individuals to obtain and provide in a CV; 3) subsequently provide an annotated guideline for the use of individual indicators in relation to their CV's, with special focus on gender, current career position, research field, as well pitfalls/deficiencies (important here is that the perspective is the researcher); 4) an ethical perspective on the use of individual metrics (for example, ecological fallacies concerning journal indicators being used at the individual level etc.), and finally we will also provide a guideline including the ethical perspective for evaluators (aka their point of view).

It is essential that our suggestions as to which type of indicators to use (and not use), are supported with guidelines - more explicit than ”read the fine print” – on their interpretation and limitations, and how to present such indicators on a CV.

Introduction

The ACUMEN portfolio is more than just a registry of CVs and publication lists. The portfolio aims to help the researcher document their activities and connect these activities with their results and the effect of these on research spaces. In this sense the portfolio enables the researcher to express the full richness of what they do. The idea is that through bibliometrics, bibliographic information can be linked to these research activities and their reception in the scientific and public communities. This is challenging as these activities and their effects are in the form of different types of publications, uses, values, applications, relationships, and roles in inspiring creativity and innovation; these in turn are only measurable by the researcher dependent on the completeness of their record and accessibility. Figure 1 illustrates interconnections in the research zone and thus the challenges we face in fitting indicators to at the level of the individual. So apart from recommending bibliometric indicators, WP5 aims to develop standards and guidelines for implementation and interpretation, to do help the researcher do meaningful bibliometric self-evaluation. But ultimately success is dependent on a fair amount of effort on the part of the researcher, which is why simplicity is the key.

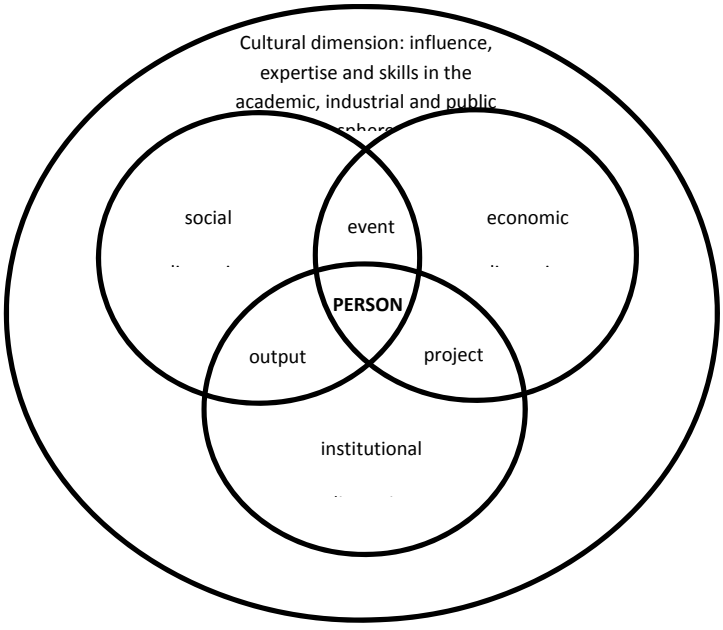


Fig.1. Visualising the research zone

The informed use of bibliometrics will make it possible for the researcher to disseminate their academic identity. Disseminating an identity is philosophically, socially and culturally challenging. To ease this, WP5 suggests that only the researcher who owns the CV can edit and append the created document and the bibliometric analyses. The identity researchers present through their ACUMEN portfolio are their

academic profiles that the consumer or those who have permission to view the CV should validate, not ACUMEN. Hence, guidelines will also be tailored to the consumer to guide interpretation of bibliometrically enriched CVs to allow contextual judgements of performance, and the use of bibliometrics at the individual level.

Clearly trust is an issue just as ethics are an issue. Self-evaluation presents the researcher with the opportunity to exploit the procedures for their own personal gain at the detriment to science (Cheung, 2008; Lawrence, 2008). The challenge for the bibliometrics is to improve the representativeness of research output evaluations at the individual level. Where it is not the ACUMEN portfolios' task to validate the bibliographic and bibliometric information the researcher provides on his portfolio CV, it is our task to provide appropriate bibliometrics that are designed for micro-level analysis, that are transparent in their application, and understandable so their use and limitations are clear. We must consider if the effort it takes the researcher to do the analyses and contextualise the scientific activities reported on the CV is worth it, as ethically speaking, how reliable is the outcome?

Reliability is trust-based and a different parameter conditioned on the point of view: from the evaluators' point of view the main issue is if individual level bibliometric evaluation is at all ethically defensible while from the individual researcher's point of view, the issues could be more related to self-promotion. A core problem is that self-evaluation is subjective (Potočnik, 2005) and it is a common fear that instead of monitoring the research process, bibliometrics will be used in evaluations to monitor the researcher (Collini 2012; Bach 2011; Cheung 2008). Hopefully encapsulating bibliometrics in a narrative will avoid fitting the indicators to the natural sciences' traditions of writing, publishing in journals and linking these publications to citations represented in WOS, (Campbell 2008; Laloë & Mosseri, 2009; Bornmann, L. et al, 2008). It should also reduce the pressure to publish, preferably in journals with a high impact factor included in citation databases, rather than journals that fit the writing talent of the author and content of the paper. This approach can result in competitive and aggressive researchers being rewarded over modest or irregular publishers (Cheung, 2008).

Accordingly, the guidelines and contextualisation of results help researchers enrich the information on their CVs and consumers understand the listed information, and this is where the ACUMEN portfolio stands apart from other CV providers with bibliometric applications. Common for existing providers is the lack of "fine print" describing the limits of bibliometrics and their interpretations, or the fine print being so distant from the CV that it is intelligible, such as HEP Inspire where the bibliometric results are presented as a box of statistics at the end of a publication list. ACUMEN supports a short narrative, that briefly and explicitly presents the meaning of such statistics for the consumer. When used correctly the informed use and informed interpretation of bibliometrics can bring objectivity into the process of individual evaluation (Bornmann et al, 2008). This avoids promoting "ready to use" amateur indicators where the validity of the use of these measures can affect the validity of self-evaluation (Lundberg, 2009). As both the researcher and evaluator are bound by professional codes of conduct that ensure professional reliability and accountability we assume this applies in an evaluation. To avoid the researcher or evaluator relying on the parsimony principle 'one indicator is better than two', such as the h-index (Zitt, 2008), we suggest developing a pallet of robust and valid indicators to recommend to the researcher. The indicators

must be easy to use and understand. Our codex is an accompliment to these indicators to regulate ethical principles and rules of behaviour for bibliometric self-evaluation.

Aim

Our aim is to recommend bibliometric indicators, traditional and new, researchers can use themselves to enrich their CVS. When combined with the other ACUMEN members' expertise, a portfolio of validated qualitative and quantitative measures will be available for the researcher to document not only their publication activities, but also contextualise these activities in narratives that showcase their expertise and influence in the context of their demographic information, specialty and academic seniority. The aim of the bibliometric indices is to document the core activities of output and reception to their work. This is nothing new. However, investigated as a form of self-evaluation, new complex aspects are introduced, such as access to data, ethics and the dependency of the success-rate of indicators dependent on complicated mathematics, software or complete datasets. The beauty of our study is that it is tested on real life data, that is flawed, incomplete and under-representative of certain academic groups and gender. But such is demographic of the scientific community and thus our dataset is highly representative of how science is practiced.

It is important to remember that bibliometric indicators are not limited to publication and citation counts, or limited to traditionally measureable forms of scientific communication in scientific journals. They are used in combination with qualitative and quanitative indicators recommended in other work packages, to document all a researcher's activity. Thus, the combined indicators also support the researcher's creativity and work with perhaps low-prestige but highly relevant problems that are "published", in the broadest sense of the word, as a lot of communication is on the web, through popular media channels or in interactive installations. The following case study exemplifies our aim with enriching the CV with bibliometric indicators.

The publication list for Researcher A is presented as it appeared on the website. The font or layout has not been changed. Only part of it is shown here.

(This list is presented chronologically and includes all editions of books and compendiums. The list includes reviews, chronicles, popular science articles and textbooks.)

1. **Researcher A. (1979): *XXX*, Speciale i biologi ved Kbh. Universitet**
2. **Researcher A. (1980): ”Article 1”**
3. **Researcher A. (1981): “Article 2”, s. 96-151 i *Niche: Nordisk tidsskrift for kritisk biologi*. Årg. 2 nr. 2.**
4. **Researcher A. (1982): “Article 3”. s. 95-143 i *Psyke & Logos*, nr. 1, 1982.**
5. **Researcher A. (1982): “Article 4”, *Biofag*, nr. 6, dec.**
6. **Researcher A. (1985): “Article 5 s. 60-72 i *Biofag*, nr. 2. april.**
7. **Researcher A. (1985): “Article 6”. s. 422-426 i *Højskolebladet*, nr. 27.**
8. **Researcher A. (1985): “Article 7”, *Ingeniøren*, 11.okt.**
9. **Researcher A. (1985): “Book chapter 1” s. 25-46 i *Informationssamfundet* red. Thomas Söderqvist, Forlaget Philosophia.**
10. **Researcher A. (1985): “Article 8”, s. 40-47 i *Naturkampen* nr. 38.**

Short Narrative: addition to researcher A's curriculum vitae

Bibliometrics

Output

My output is defined as the 112 published works from 1993-2013. This total is compared to three reference groups, comparison values resourced April 2013. The reference group on the Local Level consists of the median number of publications of associate professors at my institution; likewise the National Level consists of associate professors in my field at from the University of Copenhagen, Aalborg and Roskilde, while the Expert Reference group consists of the publications of leading scholars in my field.

1993-2013 my output level is 112 publications; w.r.t the local level it is 32 (range 5-76); w.r.t. the national level 62 (range 28-214); w.r.t. the expert level 129 (31-414).

Generally, I do not co-author works. 93/112 works are single authored. I have been most comfortable working in repeated small collaborations; these works are authored by teams of 2 to 5 scholars and a single workshop paper by 8 scholars. In terms of number of papers I do certainly better than the median person on a local and national level and in terms of the expert group I am in the top 10, rank 10/21. Fifty-five of my works, in 80 publications, have been published in 6 languages and are included in 362 academic library holdings.

Citations

It is interesting to know where my works are being cited. Even though citations to books and national language works are under-represented in citation indices, one can roughly see that I have influence in: cybersemiotics, computer science, business and economics, linguistics, engineering, social sciences, library and Information Science as well as Philosophy. Citations to my works and those of the Expert reference group have been sourced in Google Scholar and Web of Science.

Parameter	Myself	Expert (median scores)
Npapers	112	129
Year of first publication	1993	1977
Works per year	5.6	3.5
H index	16	11
M quotient	0.8	0.47

More recently, the use of the h-index (the number of papers that have received more citations than their rank in

Sampling strategy

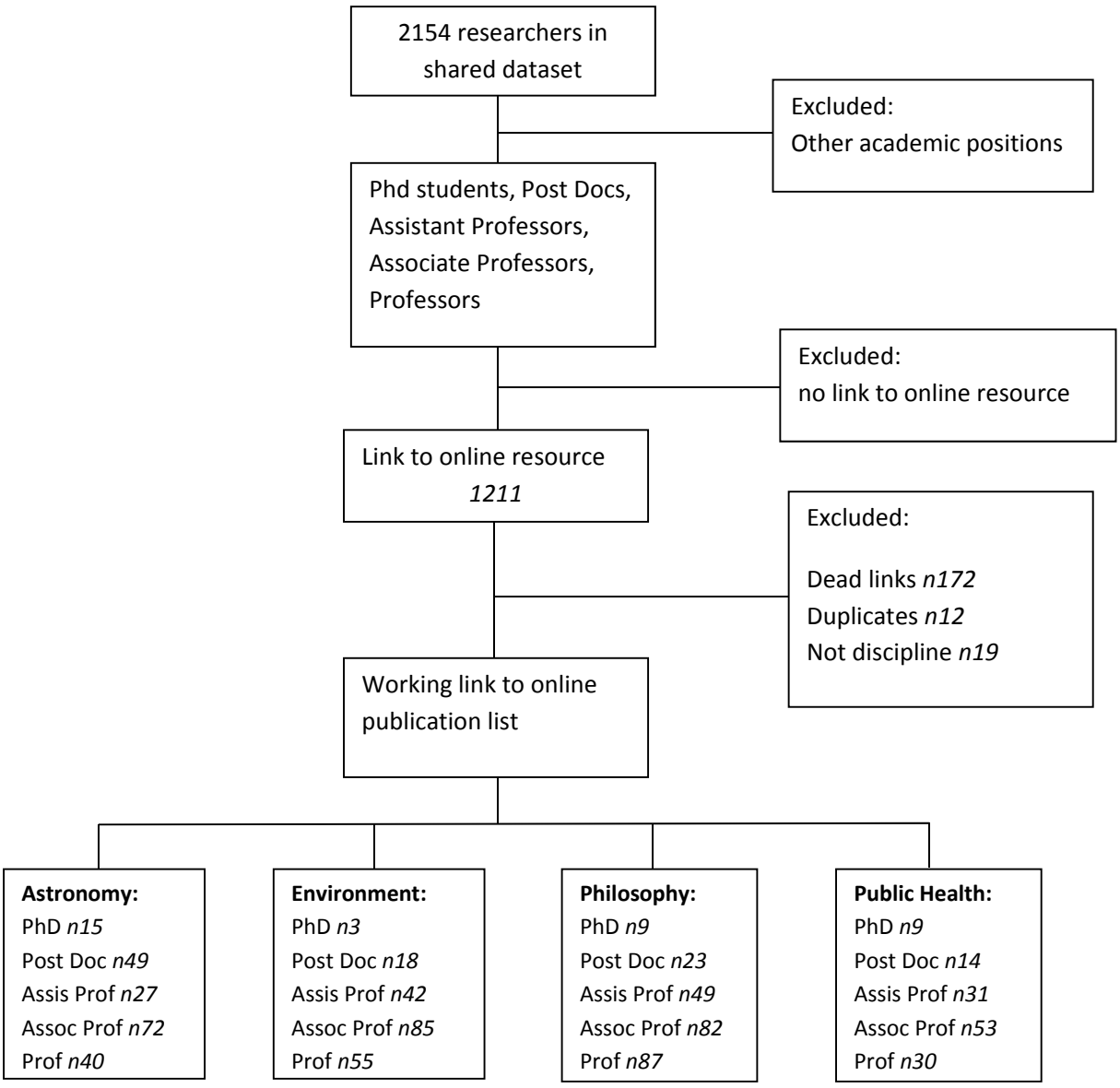
The sample of publication lists used for the bibliometric analyses were sourced from the shared dataset of 2,154 academic profiles collected by WP2. The shared dataset includes 4 subject areas (astronomy & astrophysics, public environmental and occupational health, environmental engineering, and philosophy (including the history and philosophy of science)) and 15 European countries (Bulgaria, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Israel, Italy, The Netherlands, Poland, Slovenia, Spain, and the United Kingdom). Details of the method and rationale of how the shared dataset was collected can be found in the Progress Report (2): ACUMEN Web Presence Survey Results (WP2, 2012).

Briefly, WP2 formed the shared dataset by extracting automatically a list of emails from published research papers indexed in the Thomsen Reuters Web of Science (WOS) during 2005-2011 in the four studied fields, which are based on WOS subject categories, for each European country. Because of the low coverage of Philosophy in WOS the Scopus citation index was also sourced to get sufficient email addresses for this field. A large scale survey in selected scientific fields and EU countries was conducted, resulting in information on online presence from 2,154 respondents. This information included URLs, online CVs, PDFs, PPT files publication lists, links to repositories, journals, individual websites, group websites and group publication lists as well as demographic data (gender, affiliation, discipline/specialty, and academic status).

We originally intended to use the entire sample of *n*2154 researchers as our aim was to identify how much variation exists or is estimated to exist in the population in relation to the performance of the indicators. However, not all these respondents had an online presence. Therefore the dataset was reduced further by only including the researchers who provided a link or links to any form of online material, figure 2. From this set we extracted only the researchers who had the academic status of PhD Student, Post Doc, Assistant Professor, Associate Professor or Professor resulting in a set of *n*1211 researchers. The professional titles were limited to these five seniorities to ensure we could investigate potential correlations or trends in academic life cycles and bibliometrics. Finally, all links were followed to verify how many actually led to a publication list. This led to a further reduction of the dataset as the following were excluded: dead links, duplicates, links to materials that were not an individuals' publication list or CV including a list of publications, not one of our identified 5 academic status' or subjects that fall outside our four disciplines. Our resulting sample is 793 publication lists, appendix 1 & 2.

Cleaning the base data, collecting publication and citation data, and validating bibliogaphical information is a time craving process, but is resulting in god data of a high quality with which we can contextualize the bibliometric results and counts to. We collected enough baseline data to capture an entire iteration (or cycle) of the researcher's life cycle. An iteration should account for the different types of variation seen within these process, such as cycles, trends, volume ranges, cycle time ranges etc.

Fig. 2. Flowchart of sampling strategy



Characteristics of sample

Gender and disciplinary representation

In our sample of 793 researchers, 182 are women, 23%. This is under the expected European percent for women in science, 30% and 44% dependent on field as reported in the SHE figures for 2012: http://ec.europa.eu/research/science-society/document_library/pdf_06/she-figures-2012_en.pdf.

Table 1. Gender ratio and disciplinary representation (women:men)

	Astronomy	**Environment**	**Philosophy**	**Public Health**	***Seniority ratio***
Ph.D.	1:4	0:3	1:2	1:3	*1:2*
Post Doc.	1:3	1:2	1:6	1:1	*1:3*
Assis. Prof.	1:3	1:3	1:5	1:3	*1:4*
Assoc. Prof.	1:5	1:5	1:3	1:2	*1:4*
Professor	1:19	1:6	1:5	1:2	*1:5*
Disciplinary ratio	*1:5*	*1:4*	*1:4*	*1:2*	

Academic posts and disciplinary representation

The prime objective of the indicators, are their stability and performance on different academic seniorities. For bibliometrics, this means their usability and ease to calculate small amounts of citation and publication data (as in phd students with 3 years publishing history) to large amount of data (professors with publishing histories spanning decades). The distribution of researchers across academic seniorities and disciplines is unequal, skewed in favour of senior researchers.

Table 2. Academic posts and disciplinary representation

	Astronomy	**Environment**	**Philosophy**	**Public Health**	***Seniority Total***
Ph.D.	15	3	9	9	*36*
Post Doc.	49	18	23	14	*104*
Assis. Prof.	27	42	49	31	*149*
Assoc. Prof.	72	85	82	53	*292*
Professor	40	55	87	30	*212*
Disciplinary Total	*203*	*203*	*250*	*137*	*793*

Disciplinary and linguistic representation

This demographic represents the disciplinary and linguistic representation of the departments to which the academics in our sample are affiliated. Linguistic hereditary of the research centres in the sample are more indicative of disciplinary publication and citation traditions than the researcher's nationality or the

centre's geographical location. Figure x illustrates how the sample is weighted towards the Romance (Italian, Spanish, French and Algerian), Germanic (German, Dutch, Yiddish and Swiss), and Anglo-Saxon (English, American and Australian) research and writing traditions. The corresponding table shows that at a disciplinary level the distribution is weighted differently dependent on the discipline. The categories are based on the indo-european family of languages, appendix 3.

Fig. 3. Linguistic representation of research centres in the entire sample

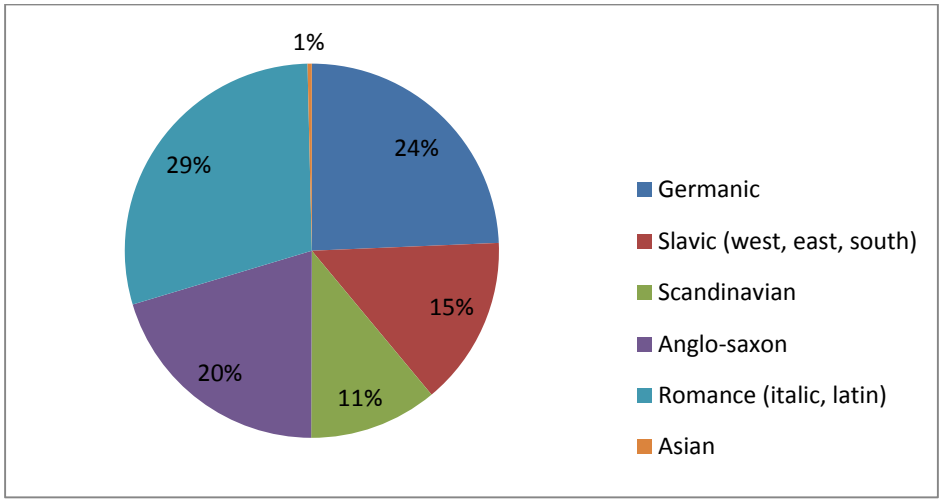


Table 3. Disciplinary and linguistic distribution

	Anglo-Saxon	Asian	Germanic	Romance	Scandinavian	Slavic	*Total*
Astronomy	37	3	59	62	7	35	*203*
Environment	25		32	60	33	53	*203*
Philosophy	71		56	83	20	20	*250*
Public Health	28		46	27	28	8	*137*

Limitations

Gender bias

Our sample has a strong male bias, the overall ratio of men to women is 3:1, which is though the same ratio as is the original shared data set. However, the gender distribution at the disciplinary level differs in two of the fields compared to the shared dataset. In the shared dataset the ratio men to women in Astonomy is 1:4, our sample represents 1:5, and in Environment there are 1:3 women, our data shows 1:4. However, it is a fact that women are outnumbered by men in math, science and engineering fields, which are two of our four selected disciplines. Our data includes relatively few women in high-level faculty positions, which is also supported in the literature (RAISE, 2013). A study, detailed in the journal *Psychological Science* (Murphy et al, 2007) claims to bring a new feature of gender bias to light that is important to remember when we contextualize our counts of scientific activity, write the guidelines and the indicators included in the ACUMEN portfolio. The feature is that women are less likely to participate in science and engineering settings in which they are outnumbered by men. These "situational cues" have an important meaning and effect on the careers of women, and these cues are the cultural and social factors that discourage women from a career in science. This includes socialization in which girls are taught, directly and indirectly, to steer clear of studies and jobs typically pursued by boys and men. **In addition, past research has revealed an unconscious bias at universities where evaluators rate resumes and journal articles lower on average for women than men**[1]**.** The responsibilities of family caretaking still fall disproportionately on women and so women often choose the stay-at-home-mum position or their household responsibilities make it nearly impossible for them to meet the long hours required for a high-level faculty position. Conversely, our sample also shows traces of the effect of female dominated fields on men, Public Health Policy, where the academic playing field is more evenly distributed, perhaps this could be attributed to the male sense of not belonging.

Ultimately, this means that our analyses of effects on gender are limited and we will as a result be focusing on academic status and research field. "Gender" will be supplementary analyses where the amount of data allows sensible investigations.

Sampling bias

We used the shared dataset as it has been an aim of ACUMEN since the kick off meeting in 2011 to connect the work packages through a shared dataset with real world parameters. In this way the findings of the work packages compliment and supplement each other in a way that the respondents and their bibliographic data are investigated through interviews, surveys, institutional documents, web presence and bibliometrics. For our work package this has meant that a sample has been drawn from the shared dataset and is as such defined as "convenience" sampling, i.e. a type of nonprobability sampling which involves the sample being drawn from that part of the population which is close to hand. Using such a sample means we cannot make scientific generalizations to the total population. This type of sampling is however useful for pilot testing and power analyses. Power analyses are used to calculate the minimum sample size required to detect an effect and accordingly determine how significant our results have to be, to be considered statistically significant *even though we cannot test the significance of our results*. As we

[1] A overview of sources is too extensive to list. Please refer to, amongst others, the Boston University Recruitment Guide lines and corresponding reference list, available at: http://www.bu.edu/apfd/recruitment/fsm/assumption_awareness/

have a convenience sample, several important matters must be considered in the design of the bibliometric analyses:

- the sample is weighted in favour of senior researchers.
- the academic seniorities are unevenly distributed across the disciplines.
- the disciplines are represented unevenly, range 137 to 250 researchers. This affects the types of analysis we can implement, the statistics we can use and the strengths of the conclusions we can draw.
- can the purpose of our analyses be adequately answered using a convenience sample?, ie characterize types of indicators, examine the correlation between simple and sophisticated indicators, provide guidelines for application of indicators on CVs and the ethical perspectives on the use of individual metrics.
- at the present time we are unaware of any controls within our analyses which can lessen the impact of a our convenience sample, thereby ensuring the results will be more representative of the population. But, how can we be sure that our convenience sample is responding or behaving differently than a random sample from the same population?

Sources used in data collection

A copy of each publication list was saved, as the internet is dynamic and we are well aware that the links that are working today could be dead tomorrow. Further a publication list is a living document that is updated and thus our base data can potentially change. We used sources of citation data that are readably available to researchers in all disciplines. Four students from RSLIS were employed to extract the data in June 2013. Multiple IP addresses were generated to solve the aggressive blocking policy of Google Scholar. The process for finding and exporting publication data from WOS and GS are described in detail in the Work Task description, appendix 4.

Publication lists, bibliographic and citation data were thus sourced in Web of Science (WOS) and Google Scholar (GS) with the aim to compare the alignment and performance of a multi-disciplinary structured citation index and a scholarly web search engine, where full text information is collected and presented through a web-crawler. Performance is defined as usefulness at the individual (disciplinary) level and the effect the choice of database has on the size of the researcher's indices. It was a tactical choice to use multidisciplinary databases rather than disciplinary specific databases such as the Astrophysics Data System (ADS) or High Energy Physics Literature Database (Inspire). Common for these systems are that they provide ready to use indices and to some extent "fine print" that define the function of bibliometric indicator and how to interpret them. However, none provide clear guidelines for implementation and their limitations and none attempt to contextualise the results. Instead the indices are presented as statistics beside a profile of the researcher. Likewise there are publication databases that attempt full discipline

coverage, such as the Philosophers Index or ECON lit. Although more representative of a discipline's literature than WOS, citations are not indexed and we do not have the necessary knowledge of a researcher's subject speciality and hence preference of database. Would the public health researchers in our sample prefer we sourced their publications in Pubmed, as all medical publications that are worth anything can be found there, or in Cinahl, as the research is nationally oriented and practice-based? Likewise how can we guess if an environmental scientist regards Inspec as *the* database rather than the Energy Citation Database (ECD)? Rather, the disciplinary specific indexes will be used in our case studies as we are very aware of the importance of these databases and it is important to address their role in the ACUMEN portfolio. In the case studies we show how good the coverage of subject-specific databases are compared to WOS and GS, the quality of the data and how difficult it is for the researcher to extract publication and citation information from these sources.

We did though experience some practical problems with our choice of citation sources, due to the amount of data we extracted. These problems are described below, but are considered not be an issue at the individual level, as extracting citation information for one publication list at a time is vastly different than extracting 793 publication lists.

Google Scholar

Data is difficult and time consuming to extract en masse from GS. Hence we used Harzing's Publish or Perish version 4.0.12 (POP) software[2] to identify publications and retrieve, and to a limited extent analyse, academic citations in GS. We are aware that GS offer a personal citation service "My Citations" where the researcher can create a profile in GS that automatically harvest relevant publication and citation data. This service is easy to use but the generated bibliometrics are limited to h index, total citations, citations over time and i10 index. We are instead recommending the researcher uses POP to search GS even though it requires effort to keep the amount of citations up-to-date, remove duplicates and publications that are not written by the researcher. Another thought behind this choice is that by researchers actively updating their publication and citation lists, they will build an understanding for what bibliometric results are built on, and not blindly trust ready to use indices presented out of context. Unlike GS, POP support this rationale by presenting a range of indices that attempt to cover basic assessment considerations such as adjusting for writing collaboratives and length of publication history (amongst others number of citations, cites per year, cites per paper, h, g, hc, hl, AWCR, AW, e, and hm-index). Publication data can be easily sorted in POP and citation results can be easily exported into Excel. At the individual level the amount of data cleaning would be, in comparison to our study, minimal.

In February 2013 GS reduced the maximum number of results per page from 100 to 20. This means that Publish or Perish now has to retrieve up to 5 times as many result pages per query in order to show the full results and has following effect on data extraction:

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

- More page requests mean that POP hits the maximum number of requests that Google Scholar allows per hour sooner.
- If the number of page requests exceeds the maximum that Google Scholar allows, the IP address will be temporarily blocked by Google Scholar. This block can last for up to 24 hours.
- To avoid hitting the maximum allowable request limit, POP uses an adaptive request rate limiter. This limits the number of requests that are sent to Google Scholar within a given period, both short-term (during the last 60 seconds) and medium term (during the last hour).
- It is no longer possible to limit to research field: Google Scholar has redesigned its interface and integrated the advanced search page in its general search page. In doing so it removed the option to select specific subject areas. As a result subject filtering is now no longer possible, neither in Google Scholar, nor in Publish or Perish.
- By default, Google Scholar matches the name and initials anywhere in the list of authors, so CT Kulik would also be matched by P Kulik, CT Williamson. To match an author's initials only in combination with her or his own surname, use "quotes" around the author's name: "CT Kulik" will not match P Kulik, CT Williamson, but it will match CT Kulik and CTM Kulik, or any other name that contains both CT and Kulik. To exclude unwanted author names, these have to be found by sorting through the results list and entering them in the Exclude these names field. For example, to exclude CLC Kulik from the earlier example, enter "CLC Kulik" in the Exclude these names field. However for both au id #9 (B Jansen) & #11 (S Ward) the result lists numbered over 1000 even after excluding unwanted names and the only option left is to manually remove publications not written by the researcher.

To achieve the required reduction in requests, Publish or Perish delays subsequent requests for a variable amount of time (up to 1 minute). The higher the recent request rate, the longer the delays.

This meant that for our study the amount of data collection per session was limited and the speed of data extraction was slow. The alternative is being blocked by Google Scholar for up to 24 hours. As we are performing queries that yield many results (several hundred or more at the professor level) and issue a large number of queries in short succession, the request rate limiter will insert progressively longer delays to keep the overall request rate within acceptable limits and warn us of an upcoming block from GS. To avoid being block or having to stop collection to stay within a required rate, we created 100 IP addresses which we switched between when we received a warning.

Extended citation analysis of GS data

A drawback of using POP for analyzing a great quantity of citation information is that it does not support export of details of citing sources. It links instead directly to the list of citing sources in GS. This lack of detail hampers our analysis of the foundations of the indicators. We are investigating the possibility of using the Online Citation Service[3] (OCS) software to retrieve details of citing sources, with the kind

[3] http://dbs.uni-leipzig.de/ocs/

permission of the developers, Professor Erhard Rahm and Professor Stefan Endrullis from Leipzig University. Apart from the traditional search by author name and venue, OCS allows the upload of a list of publications and returns the results for this. However, OCS has recently been affected by the GS interface changes and aggressive blocking policy. Knowing this, the advantages of the OCS have to be revisited and other options discussed before we implement any extended analysis of data.

Web of Science

WOS is a highly valuable resource for researchers to discover prior work in their research areas, as the scope extends across multiple publisher's lines. The use of WOS in the evaluation of academic performance through the counting of individuals' publications and citations, weighted often by Journal Citation Reports (JCR) as a proxy indicator of the quality of the publications, is more contentious in the bibliometric community.

This contention arises in part from the peer review process and publishing quota that has to be met before a journal is accepted. Critics of the database suggest that these barriers have resulted in a strong bias in favour of "long-established, commercial publishers (disciplines), and against recently-started publications, independent journals, and conferences" (Clarke & Pucihar 2012). Moreover, the declared policy of WOS is that only current and forthcoming issues are considered in the evaluation. Back issues are not accepted (TS 2013a) i.e. recognition of worth is not retrospective. The result of the WoS approach is that major journals of relevance to some disciplines could be missing, or have been taken up only from recent dates and without any retrospectivity. This means that for some senior researchers, the proportion of their publications that are indexed by WoS is as low.

A further consideration is that journals are deleted from Web of Science throughout the year (TS 2013b). This represents historical revisionism, with publications and citations being effectively cleansed from the record (Clarke 2008). Also publications and citation-counts are not cumulative, because they change not only upwards, as new documents are published, but also downwards, as venues are deleted. Studies have also shown database bias towards international English language journals, and certain document types, primarily articles and the citation culture in article-based disciplines.

Table 4. Overall ISI coverage by main field*

EXCELLENT (> 80%)	VERY GOOD (60-80%)	GOOD(40-60%)	MODERATE (<40 %)
Biochem & Mol Biol	Appl Phys & Chem	Mathematics	Other Soc Sci
Biol Sci – Humans	Biol Sci – Anim & Plants	Economics	Humanities & Arts
Chemistry	Psychol & Psychiat	Engineering	
Clin Medicine	Geosciences		
Phys & Astron	Soc Sci ~ Medicine		

*table reference: (Moed 2007)

In a preliminary randomised study of 20 researchers we confirmed the common conception that WOS under-represents the "softer" sciences and non-article based disciplines and searches in GS result in a lot

of noise and clean-up. We found that WOS underrepresents Philosophers, books and national language/small publications and Google Scholar requires patience and tenacity to search, Table 5.

Table 5. Disciplinary representation in GS and WOS

Author id	Discipline	Seniority	N publications	Found GS	Citations GS	Found WOS	Citations WOS
1	Astronomy	Prof	257	233	3614	148	7302
2	Astronomy	Assoc Prof	42	54	257	28	171
3	Astronomy	Assis Prof	89	143	1407	46	907
4	Astronomy	Post Doc	251	262	291	54	138
5	Astronomy	Phd	10	15	67	7	36
6	Environment	Prof	84	167	1459	41	282
7	Environment	Assoc Prof	63	74	3927	46	2066
8	Environment	Assis Prof	30	30	398	33	426
9	Environment	Post Doc	25	-	-	5	21
10	Environment	Phd	12	20	34	3	13
11	Health	Prof	415	-	-	441	8245
12	Health	Assoc Prof	90	200	3472	0	0
13	Health	Assis Prof	151	95	407	21	151
14	Health	Post Doc	49			13	327
15	Health	Phd	24	17	138	19	211
16	Philosophy	Prof	41	22	43	13	12
17	Philosophy	Assoc Prof	36	27	36	4	0
18	Philosophy	Assis Prof	18	35	91	7	57
19	Philosophy	Post Doc	8	10	11	0	0
20	Philosophy	Phd	4	3	0	1	0

The overlap between citations and publications sourced in Web of Science and Google Scholar was not investigated, as this is not an issue for us. We are calculating indicators separately in each database and contextualising the results as we would not expect the researcher to attempt an indicator using combined data from both sources where the citation data is cleaned for duplicates to calculate a fully representative citation count. In the process of collecting data for the analyses we have main broad observations that GS is finding citations from national language publications, books and book chapters, and local journals published in English language as well as citations from sources indexed in WOS. The question is if is there a pattern in the type of publications we don't find and if this is problematic for what we want to do? What is the effect if we miss something highly cited or many minor publications?

We accept there is an overlap, and acknowledge that the researcher would wish to write the highest resulting indicator on the CV. However in the bibliometric analysis we did compare the difference between results in GS and WOS and find that the score only varies by ± 1 dependent on the discipline. We are aware of potential ethical and validity problems here which is why in the guidelines we stipulate the researcher reports which database was used to calculate the indicator and we offer alternative indicators that account for database bias, such as the hmx - index (the median h of h-indices calculated in WOS, GS and Scopus).

In summary, disciplinary (under)representation in WOS has been well documented in the literature (Clarke, 2008; Salisbury 2009). However there appears to be an agreement, that even though other databases such as GS or Scopus cover a wider range of materials, WOS has much more complete coverage, with more articles indexed and more current citations. As with bibliometric analysis in any single database publication counts are of limited value and citation analysis should always be in context as the future of research **assessment exercises lies in the intelligent combination of metrics and peer review** (Moed 2007). This observation forms the ACUMEN portfolio, and sets it apart from any other CV enrichment application currently available.

Final observations in preparation for the bibliometric analyses.

The exploratory study of 20 researchers also provided useful information in guiding the data-collection and analysis. The results are listed below:

1. A publication list is not a publication list! It is a link to a webpage with selected publications, a short narrative, a link to a database a list in pdf format or a list on a website separated into article types, chronological, and each type accompanied by a short narrative.

2. Some authors publish more than one publication list, an institutional list and a full list on a their personal website fx author id #3, table x, gave 4 publication lists: ADS (89 references), ArXiv (59 references), SPIRES (dead link), Citebase (not a publication list).

3. Some lists are more complete than others. Some include only peer reviewed, published articles while others include everything: rapid responses, popular articles, encyclopedia, conference papers, letters, articles, book chapters and works in preparation.

4. Publication lists are not as a rule up to date. During data-collection we should expect to find more publications by an author than listed on the publication list.

5. Publications by authors with common names, such as au id# 9 & 11, are bordering on the impossible to verify in GS using Publish or Perish. We expect the sample to be reduced.

6. Au id #12 writes national language articles and publishes in books. Even though #12 is an accomplished author he or she is not represented in WOS. Further the publication list is written in Italian, and GS includes both Italian and English translations of the works. Even though this

increases the publication list two-fold, we consider translated and original papers as two different works, attracting different readers and different citations.

Method of Bibliometric analysis

Characterization of types of indicators.

The indicators tested in our study were previously identified in our comprehensive literature review of 114 bibliometric indicators used in individual evaluation, D5.8 Part 1. In the review we categorised the indicators into the main type of impact they *purport* to measure, be it outcome, output, quality, impact, sustainability, innovation & social benefits or research infrastructure. The mathematical foundation of each indicator was rated on scale of 1 to 5, where 1 is simple counting and 5 is extremely advanced math. Likewise we studied how difficult it would be for the individual to access and collect the information needed to calculate the indicator. This rough complexity rating reduced the set from 114 to 64 indicators that were considered potentially useful for self-evaluation.

In preparation for the analyses of the indicators, we sorted and filtered the indicators investigating in detail their applicability at the individual level. This resulted in separating the set into 37 indicators and 16 potentially useful reference standards, appendix 5. The applicability of this set was discussed during a meeting of WP5 in May 2013. Using the decision tree, below, we identified and categorised the indicators, discussed their function in the light of previous findings and disciplinary considerations as well as the potentials for correlative analyses. Disagreements were discussed until consensus was reached.

Is the indicator relevant for our 4 disciplines?

No. Exclude indicator from study.
Yes. Continue to next question.

Can the indicator be calculated in WOS and GS?

No. Exclude indicator from study.
Yes. Continue to next question.

Is the data needed to calculate the indicator available to the individual in WOS or GS?

No. Exclude indicator from study.
Yes, see appendix 6. Continue to next question.

Is there information redundancy between the indicators?
No. Continue to next question.
Yes. Does this overlap need investigating before we can responsibly exclude one of the indices from the set? Yes. Include the indicator in the study. No. Exclude the indicator from the study

This resulted in 40 indicators that were then categorised as "simple", *n*27, or "sophisticated", *n*13. We wish to compare and correlate the performance of simple and sophisticated indicators. A research question that developed during our discussions is if, at the individual level, simple perhaps rougher indicators perform just as useful as the sophisticated (professional) refined indicators. The sophisticated indicators tend to be more complicated in design and calculation. Finally, the indicators were sorted into the ACUMEN sub-portfolio they best represent, Table 6.

Table 6. Bibliometric indicators included in the analysis; their description, the type of impact they purport to measure, complexity and sub-portfolio categorization.

ID	Indicator	Description	Type of impact	Complexity	*Sub-portfolio
1	P	Count of production used in formal communication	Output	Simple	Output
2	Pisi, Pgs	Publications indexed in WOS or GS	Output	Simple	Output
3	Pts	Publications in sources defined as important by researcher's affiliated institution or specialty	Output	Simple	Expertise
4	Co-publications	Collaboration on a group, departmental, institutional, national or international level	Output	Simple	Output
5	Categorised publication type	Distinction between document types	Output	Simple	Output
6	C +sc	Citations including self-citations	Outcome	Simple	Influence
7	CPP	Citations per paper	Outcome	Simple	Influence
8	Number of significant papers	Top cited papers	Outcome	Simple	Influence
9	Ptop	Publications among the top 20, 10, 5 or 1% most frequently cited papers in subject/field/world in a given year	Outcome	Sophisticated	Influence
10	Age and productivity	Effects of academic age on productitivty and impact	Outcome	Sophisticated	Output
11	%Pnc	Share of publications that are not cited. Identify trends in type, subject etc	Outcome	Simple	Output
12	Number of different co-authors	Growth of co-operation at group, departmental, institutional, national or international level.	Research Infrastructure	Simple	Expertise
13	Hi-index	Accounts for co-authorship effects	Research Infrastructure	Simple	Influence
14	POP variation individual H index	Accounts for co-authorship effects	Research Infrastructure	Simple	Influence
15	n-index	Accounts for co-authorship effects	Research Infrastructure	Simple	Influence
16	Alternative h index	Accounts for co-authorship effects	Research Infrastructure	Simple (same as hi index)	Influence
17	Hp	Accounts for co-authorship effects	Research Infrastructure	Sophisticated	Influence
18	Diachronous IF	Development of impact over time of a set of papers	Impact	Simple	Influence
19	Y Factor	Scientific impact defined as a combination of popularity and prestige	Impact	Sophisticated	Expertise
20	NJI	Normalised journal impact	Impact	Sophisticated	Influence
21	JFIS	Journal to field impact score	Impact	Sophisticated	Influence
22	DIF	Discipline impact factor	Impact	Sophisticated	Influence

23	IFmed	Median impact factor	Impact	Sophisticated	Influence
24	NJP	Normalised journal position	Impact	Sophisticated	Influence
25	FCS	Field Citation Score, number of citations expected for a paper of the same type within a field and year.	Impact	Sophisticated	Influence
26	CPP/JCSm	Normalised citation score (CS/NCS)	Impact	Sophisticated	Influence
27	H	Cumulative achievement	Quality	Simple	Expertise
28	hmx	Median h across multiple databases	Quality	Simple	Expertise
29	g	Cumulative acheivement, includes more information than h	Quality	Simple	Expertise
30	H(2)	Weights most productive papers, but requires more citations to be included in index	Quality	Simple	Expertise
31	A-index	Magnitude of citations to a researcher's papers	Quality	Simple	Influence
32	R-index	Improves sensitivity of A	Quality	Simple	Influence
33	h-index	Structure of citations to papers	Quality	Simple	Influence
34	M-quotient	Adjusts h for length of career	Quality	Simple	Influence
35	E index	Includes ignored excess citations in h index	Quality	Simple	Influence
36	Citation Age	The age of citations referring to a researchers work	Sustainability	Simple	Influence
37	Aggregate Immediacy index	How quickly papers in a subject are cited	Sustainability	Sophisticated	Influence
38	AWCR, AW & per author AWCR	Age weighted citation weight	Sustainability	Simple	Influence
39	WorldCat	Inclusion in academic libraries internationally	Innovation and social benefits	Simple	Expertise
40	National and local Library Catalogues	Inclusion in national library catalogues and bibliographies that include press coverage	Innovation and social benefits	Simple	Expertise

*As we learn more about the indices during the tests, we expect to find that some measure activity better in another sub-portfolio than that they were originally assigned.

To fully understand how complicated even simple indices can be and ensure that this is the final list of indicators for the analyses, we examined the independence or dependence of the indicators on other indices and if their interpretation is dependent on the use reference standards and weighting systems, appendix 7. No unexpected complications were discovered and no further indicators were excluded.

Table 7. Analysis of independence

ID	Indicator	Independent	Dependent on another index	Dependent on reference standard
1	P	✓		
2	Pisi, Pgs	✓		
3	Pts			Authority list
4	Co-publications	✓		
5	Categorised publication type	✓		
6	C +sc	✓		
7	CPP	✓		
8	Number of significant papers	✓		
9	Ptop			Authority list
10	Age and productivity		CPP	
11	%Pnc	✓		
12	Number of different co-authors	✓		
13	Hi-index		H dependent	
14	POP variation individual H index		H, authors per paper	
15	n-index		H, journal h,	
16	Alternative h index		H, authors per paper	
17	Hp		H, authors per paper	
18	Diachronous IF	✓		
19	Y Factor		ISI JIF	
20	NJI			Citation average in subfield
21	JFIS			5 year field journal average
22	DIF		Number of citable items in journal over time	
23	IFmed			Median IF of journals in subject category
24	NJP			JCR category ranked by JIF
25	FCS			Field citation score

26	CPP/JCSm			Average citation rate of individuals in journal set
27	H	✓		
28	hmx	✓		
29	g		H dependent	
30	H(2)		H dependent	
31	A-index		H, A dependent	
32	R-index	✓		
33	ħ-index		H dependent	
34	M-quotient		H dependent	
35	E index		H dependent	
36	Citation Age	✓		
37	Aggregate Immediacy index	✓		
38	AWCR, AW & author AWCR	✓		
39	WorldCat	✓		
40	National and local Library Catalogues	✓		

Method of analysis

The forty indicators will enable the following analyses that will help us include stable and recommended indices in the portfolio:

1. The success of simple contra sophisticated indicators.
2. Correlation between simple and sophisticated indicators.
3. Correlation between the four disciplines and the indicators.
4. Correlation between the five seniorities and the indicators.
5. Correlation between gender and the indicators (where data allows sensible analyses)
6. Correlation between (gender) seniority, field and indicator.
7. Correlation between (gender) seniority, field and indicator categorised as simple or sophisticated.
8. The differences in performance between indicators of the same type of impact.
9. The effect of discipline on the success of the indicators.
10. The effect of seniority on the success of the indicators.
11. The effect of gender on the success of the indicators, (if data allows sensible analyses).
12. The effect of data quantity on the indicators.

Methodological considerations

Simple vs sophisticated

Lessons learnt from the test-case narrative taught us that simple indicators can give a lot of information which in turn can be demanding to contextualize. We wish to understand if they perform just as well as the sophisticated indicators which more or less indicate the same thing and to understand the correlation between them and how useful they are for the discipline and the seniority. This is why these sophisticated indicators appear on the list, even though they would be too intricate and demanding for the researcher to calculate. The indices in the impact category are all apart from one "sophisticated" and traditional disciplinary benchmarks. This problematic was already identified in the review, because good measures of impact are dependent on a high level of aggregation to be comparable to global performance standards. We are interested in if other indicators such as CPP are as informative as these and could used as a proxy for impact.

Indicators that account for co-authorship effects

The hi, POP variation, N, alternative h and hp overlap and are information redundant if used together. We will rank these and discuss which are the most disciplinary representative at the individual level. The usefulness of identifying individual contribution depends on the field. Of course bibliometrically it is interesting to provide a metric that accounts for the number of papers researchers would have written if they had worked alone or support intra- or interdisciplinary analysis. But from the researchers point of view it is debatable if this is important. If it is a disciplinary tradition to multi-author papers, fractionalising the contribution would be detrimental to the individual and we would not recommend the author to use fractionalisation schemes. However, if researchers in a multi-authoring discipline choose to write alone, it is important to provide the fractionalisation counting tools to emphasize their efforts.

Indicators of quality

The information redundancy between the h, g, n H(2), A, R, ħ, M-quotient and e-indies will be investigated and if the indices favour an academic seniority or field. Further the use of the h-index (or g-index) as a benchmark in different areas, for different seniorities or gender will be investigated, such as h-

index of author compared to h index of seniority (within specialty). Further, we wish to investigate if CPP gives a better representation of impacts of quality than h-index. Compared to h, CPP is more intuitive as all citations and papers are included in the calculation rather than a "core" of papers. As h is acknowledged for its simplicity and is known in the research community, the guidelines for both the evaluator and the researcher the main pitfalls of the h-index will be listed, emphasizing how comparison across fields is unwise.

Indicators of impact

Clearly there are more sophisticated indicators of impact in our study than simple ones. Note though, that these are designed for a higher level of aggregation than the individual. However, researchers will undoubtedly want to draw attention to how successful they are within their field especially if they have published in journals with high impact factor and their papers have received a lot of citations throughout their career. We will test Y, NJI, JFIS, DIF, IFmed, NJP and SPP/JCSm to understand how they correlate with more simple impact indicators, and if these simple indicators can be aggregated to be used as local bench marks, Table 8.

Table 8. Local benchmarks developed from simple indicators

Reference Standard	**Indicator**
Production of colleagues of same academic seniority within department or institution	P
Production of same academic seniority within field, national or international level	P
Production of experts in specialty	P
Citations to colleagues of same academic seniority within department or institution	C + sc
Citations to same academic seniority within field, national or international level	C + sc
Citations to experts in specialty	C + sc
H index at local, national or international level	H
M quotient at local, national or international level	M-quotient

The case narrative taught us that simple indicators can be aggregated to useful local performance benchmarks. However indicators that are simple at an individual level become complex and time consuming when used on a higher level of aggregation. The time and effort needed in calculation must be clear in the guidelines as this affects the practicality and usefulness of the standard, however relevant it may be. Other possible benchmarks, where the amount of data allows for sensible comparisons, could be in disciplinary databases, such as the individual's visibility and representation in ADS, Inspire, Inspec, Biomed, PubMed, or the Philosophers Index. The challenge for us, is to find an easy method the researcher can reproduce, to find out which are the most highly cited papers in regards to a researchers specialty and not ISI defined subject category. This will be extremely difficult in areas where citation

activity is not high and we need to analyse how publication types, years and citations correlate with sophisticated field-citation indicators.

Indicators of sustainability

Together with the indicator Age and Productivity, with is purported to primarily measure outcome, we will test which of the indicators in this category best reflect the researcher's currency.

Indicators of innovation and social benefits

The success and informativeness of the indicators of innovation and social benefits are dependent on the completeness of the information on the researcher's CV and are also highly dependent on culture, politics and economics of the country and/or domain the researcher is active. A self-evaluation questionnaire covering the issues of knowledge exchange, earning capacity, use in the public sphere, patent applications and the effects of publication is currently being tested in the HEFCE evaluations in the UK (Neiderkrontenhaler et al 2011; Wildgaard et al 2013). This form of evaluation falls outside our framework of bibliometrics. We recommend Neiderkrontenhalers questionnaire as useful in developing a checklist or guideline for reporting innovation and social benefit. In the narrative case study, we found WorldCat and the Danish bibliography accessed through bibliotek.dk useful sources for indicating incorporation of published works in public libraries and appearance in the media. Being in a public library catalogue is used as a proxy for dissemination in the social sphere and appearance in the media is also assumed to be a measure of societal impact. The disciplinary usefulness of similar national library catalogues will be investigated.

Next steps

Status June 2013: data is still being collected and analysed.

The bibliometric analyses, results, conclusions and recommendations will be presented in the final report (D5.8). The thorough methodological preparations and preliminary studies described in this document have enabled us to design analyses targeted to our potential users within the four disciplines that will result in useful information. Further, we can already now sketch a structure for the guidelines that will accompany the recommended bibliometric indicators:

For Researchers: Guidelines for using bibliometric indicators on your CV

- Coverage in databases. How to choose where to extract data?
- Gender
- Academic status
- Discipline
- Suggestions to benchmarks that are relevant to you
- Pitfalls
- Deficiencies
- Presentation techniques
- Good self-evaluation practice

For Evaluators: Guideline for Evaluators

- Interpreting bibliometric self-evaluation
- Ethics of self-evaluation

Reference list

Bach, J. (2011). *On The Proper Use of Bibliometrics to Evaluate Individual Researchers.* Rapport de l'Académie des sciences - 17 janvier 2011: Report presented on 17 January 2011 to the Minister of Higher Education and Research.

Bornmann, L., Mutz, R., Neuhaus, C., & Daniel, H.-D. (2008). Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics, 8*, pp. 93-102.

Clarke, R (2008) An Exploratory Study of Information Systems Researcher Impact. *Commun. AIS 22, 1 (January 2008),* PrePrint at http://www.rogerclarke.com/SOS/Cit-CAIS.html

Clarke, R., & Pucihar, A., (2012) The Web of Science Revisited: Is it a Tenable Source for the Information Systems Discipline or for eCommerce Researchers? Accessed 9. June 2013 http://www.rogerclarke.com/SOS/WoSRev.html

Campbell, P. (2008). Escape from the Impact Factor. *Ethics in Science and Environmental Politics,* 8, pp. 5-7.

Cheung, W. (2008). The Economics of Post-Doc Publishing. *Ethics in Science and Environmental Politics,* 8, pp. 41-44.

Collini, S. (2012). Bibliometry. In *What Are Universities For?* London: Penguin.

Laloë, F., & Mosseri, R. (2009). Bibliometric Evaluation of Individual Researchers: not even right....not even wrong! *Europhysics News, 5*, pp. 26-29.

Lawrence, P. (2008). Lost in Publication: how measurement harms science. *Ethics in Science and Environmental Politics*, 8, pp. 9-11.

Lundberg, J. (2009). Lifting the crown: citation z-score. *Journal of Informetrics,* 1(2), pp. 145-154.

Moed, H. F. (2007) *The use of bibliometric indicators in research evaluation and policy*. Power Point lecture presented at: Colloque de l'Académie des sciences "Évolution des publications

scientifiques - Le regard des chercheurs" des 14-15 mai 2007. Retrieved June 10, 2013 from : http://www.academie-sciences.fr/video/v140507.htm

Murphy, M., Steele, C.M., & Gross, J.J., (2007) Signaling Threat: How Situational Cues Affect Women in Math, Science, and Engineering Settings. *Psychological Science*, 18(10), pp. 879-885.

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.

Potočnik, J. (2005). *The European Charter for Researchers: The Code of Conduct for the Recruitment of Researchers.* Brussels: European Commission:Directorate-General for Research.

RAISE. (2013). *Recognizing the acheivements of Women in Science, technology, eningeering, maths and medicine*, at http://www.raiseproject.org/index.php [accessed 10 June 2013]
Salisbury, L (2009) Web of Science and Scopus: A Comparative review of Content and Searching Capabilities*. The Charleston Advisor* (July), pp.5-18.

TS (2013a) Journal Submission Process' Thomson Reuters, at http://ip-science.thomsonreuters.com/mjl/selection/ [accessed 10 June 2013]

TS (2013b) *'The Thomson Reuters Journal Selection Process' Thomson Reuters,* May 2012, at http://thomsonreuters.com/products_services/science/free/essays/journal_selection_process/ [accessed 8 June 2013]

Wildgaard, L., Schneider, J., & Larsen, B. (2013). Quantitative Evaluation of the Individual Researcher: a review of the characteristics of 114 bibliometric indicators. *Manuscript submitted for publication*.

WP2. (2012) *Progress Report (2): ACUMEN Web Presence Survey Results*
Work Package 2: Assessing the Institutional Web presence of researchers . University of Wolverhampton, Statistical Cybermetrics Research Group.
Retrieved June 10, 2013 from:
https://www.dropbox.com/home/Project%20Managment/WP2%20work%20folder.

Zitt, M. B. (2008). Challenges for scientometric indicators: data demining, knowledge-flow measurements and diversity issues. *Ethics in Science and Environmental Politics,* 8, pp. 49-60.

List of Appendices

1. Sample, corrected for working links and duplicates removed, p.26

2. Researchers excluded from sample, p.27

3. Seniority, Disciplinary and geographical distribution, p.28

4. Work task protocol, p.29

5. Reduction of 64 indicators for analysis to 37 plus 16 reference standards, p.41

6. Identification of the information needed to calculate the indicators and reference standards, p.50

7. The dependence of indicators on other indicators, reference standards and weighting systems, p.55

Appendix 1: Sample corrected for working links and duplicates

We have a sample of researchers, *n*1211, who provided links to a publication list. I have been through all the links to remove duplicates, researchers who do not belong in the discipline, deadlinks and links to material other than personal publication list, eg. blogs, group websites and information about areas of research. This has resulted in a sample of 776 researchers with working links to publication list(s), distributed as follows:

In Astronomy we have 203 researchers, 17% women

Astronomy	**Phd**	**Post Doc**	**Assis. Prof**	**Assoc. Prof**	**Prof**
ACUMEN shared data set	57	142	66	144	86
Provide link to web material	18	71	37	93	63
Working link to publication list	15	49	27	72	40
Men/women with working link	12/3	37/12	20/7	61/11	38/2

In Environmental Science we have 203 researchers, 23% women

Environment	**Phd**	**Post Doc**	**Assis. Prof**	**Assoc. Prof**	**Prof**
ACUMEN shared data set	31	65	92	200	126
Provide link to web material	8	29	64	135	83
Working link to publication list	3	18	42	85	55
Men/women with working link	3/0	12/6	33/9	71/14	47/8

In Philosophy, we have 250 researchers, 19% women

Philosophy	**Phd**	**Post Doc**	**Assis. Prof**	**Assoc. Prof**	**Prof**
ACUMEN shared data set	25	47	85	147	151
Provide link to web material	14	34	67	124	129
Working link to publication list	9	23	49	82	87
Men/women with working link	6/3	20/3	41/8	64/18	72/15

In Public Health we have 137 researchers, 39% women

Health	**Phd**	**Post Doc**	**Assis. Prof**	**Assoc. Prof**	**Prof**
ACUMEN shared data set	48	54	82	194	97
Provide link to web material	17	21	49	97	58
Working link to publication list	9	14	31	53	30
Men/women with working link	2/7	7/7	36/13	36/17	20/10

Overall in our sample of 793 researchers, 182 are women, 23%. This is under the expected European percent for women in science, 30% and 44% dependent on field as reported in the SHE figures for 2012: http://ec.europa.eu/research/science-society/document_library/pdf_06/she-figures-2012_en.pdf.

Appendix 2: Researchers excluded from sample

Astronomy	**Phd**	**Post Doc**	**Assis. Prof**	**Assoc. Prof**	**Prof**	***Total***
Dead link	2	16	1	6	13	*38*
Not Discipline				1		*1*
Duplicate			1			*1*
Not publication list	1	6	8	15	10	*40*
Not correct seniority						

Environment	**Phd**	**Post Doc**	**Assis. Prof**	**Assoc. Prof**	**Prof**	***Total***
Dead link	2	6	7	25	11	*49*
Not Discipline		1			2	*3*
Duplicate						
Not publication list	2	4	15	25	15	*61*
Not correct seniority	1					*1*

Philosophy	**Phd**	**Post Doc**	**Assis. Prof**	**Assoc. Prof**	**Prof**	***Total***
Dead link	2	5	9	12	17	*45*
Not Discipline		1	1	4	2	*8*
Duplicate	1	1		5	3	*10*
Not publication list	2	4	8	21	20	*55*
Not correct seniority						

Public Health	**Phd**	**Post Doc**	**Assis. Prof**	**Assoc. Prof**	**Prof**	***Total***
Dead link	3	2	8	17	10	*40*
Not Discipline	2		1	1	2	*6*
Duplicate		1				*1*
Not publication list	3	4	9	26	16	*58*
Not correct seniority						

Appendix 3: Seniority, disciplinary and geographical distribution

	AU	BG	CH	CN	CZ	DE	DK	DZ	EE	ES	FI	FR	HU	IL	IN	IT	NL	NO	PL	RU	SK	UK	USA
Astro Phd					1	3										2	4		2			2	1
Astro Post Doc			1		1	16	1			5	1		3	1			6				1	10	3
Astro Assis Prof	1				2	4				2		1	1	1		9			1			5	
Astro Assoc Prof					2	3	2		2	9		11	3	3	2	12	8		4	1	2	6	1
Astro Prof		1		1	1	2	2			3	1	5	4	6		3	1		3			7	
Total Astro.	***1***	***1***	***1***	***1***	***7***	***28***	***5***		***2***	***19***	***2***	***17***	***11***	***11***	***2***	***26***	***19***		***10***	***1***	***3***	***31***	***5***
Enviro Phd					1											1	1						
Enviro Post Doc					1	1	5			2	2		1			1			1			4	
Enviro Assis Prof		1			4	1	2		1	5			3			11	8		3		1	2	
Enviro Assoc Prof					5	1	13		7	5	3	3	4	7		15	4	1	4		1	12	
Enviro Prof					3	4	3		2	5	4	3	7	4		9	1		3			7	
Total Enviro.		***1***			***14***	***7***	***23***		***10***	***17***	***9***	***6***	***15***	***11***		***37***	***14***	***1***	***11***		***2***	***25***	
Phil Phd						1				2	1		1	1								3	
Phil Post Doc					2	6	1			1	3	2		1		1	2					4	
Phil Assis Prof					3	8	1		1	3	1	5		1		6	6		1		1	11	1
Phil Assoc Prof						5	8		1	12	2	6	1	1		16	4		5		1	18	2
Phil Prof					1	10	3			7		6		3		16	7				2	31	1
Total Phil.					***6***	***30***	***13***		***2***	***25***	***7***	***19***	***2***	***7***		***39***	***19***		***6***		***4***	***67***	***4***
P. Health Phd							2			1		1					4					1	
P. Health Post Doc						5	2					1				1						5	
P. Health Assis Prof						4	2		1	1		1	1			4	9					8	
P. Health Assoc Prof						4	11	1	1	1	3	1	3	3		7	7		2			9	
P. Health Prof						7	8			3						4	3					5	
Total P. Health.						***20***	***25***	***1***	***2***	***6***	***3***	***4***	***4***	***3***		***16***	***23***		***2***			***28***	
Overall	**1**	**2**	**1**	**1**	**27**	**85**	**66**	**1**	**16**	**67**	**21**	**46**	**32**	**32**	**2**	**118**	**75**	**1**	**29**	**1**	**9**	**151**	**9**

Appendix 4:

Work guideline: Extracting publications from Google Scholar and Web of Science.

Contents

ACUMEN Project description: What is ACUMEN? 35
Your Job: A brief outline and how to save your work 36
Method of Data Collection: Google Scholar, through Publish or Perish 38
How to export from POP to excel 38
Tips to searching 39
Example of a step-by-step search strategy 40
Example of an author that is impossible to verify 41
Method of Data Collection: Web of Science 42
How to export to from WOS Excel 45
Before the next search 46

Contact Information

Lorna Wildgaard (project leader, Copenhagen) pnm664@iva.ku.dk
tlf: 32341460

Jesper W Schneider (project leader, Århus) jws@cfa.au.dk

Birger Larsen (project advisor, Copenhagen) ftm448@iva.ku.dk

Send your email address to Lorna to join the project's Dropbox folder to share files, experiences and store completed work.

ACUMEN Project description: What is ACUMEN?

ACUMEN stands for Academic Careers Understood through Measurements and Norms. ACUMEN is a European research collaboration aimed at understanding the ways in which researchers are evaluated by their peers and by institutions, and at assessing how the science system can be improved and enhanced. This FP7 project is a cooperation among nine European research institutes with Professor Paul Wouters (CWTS – Leiden University) as principal investigator.

The aim? To use the ACUMEN member's combined expertise to produce a portfolio of both traditional indicators and new (useful) qualitative indices and quantitative web-based and bibliometric measures. These measures will be presented to the researcher as an online enriched CV, which documents their research activities as well as supporting assessments of their expertise, output and influence *in the context* of their demographic information and career path narratives. This visualization tool will support the core creativity of research in all disciplines and not steer the aim of research as publishing in high JIF journals rather than work with low-prestige but relevant problems. Hence the indicators are not limited to publication and citation counts, or limited to traditionally measureable forms of scientific communication in journals as a lot of communication now-a-days is on the web or through popular media channels or interactive installations.

The philosophy behind the project is to address the gap between creating research, evaluating research and promoting excellence. There is a problem in current systems of research evaluation and this problem is complicated. Researchers are people who are being evaluated between narrow frameworks and limited technology. In these systems the societal role of their research is secondary and the methods of evaluation, such as peer review can be biased, subjective, give power to scientific elite and enforce the gender power structure. To understand the effect of evaluation, we need to be aware of differences between disciplines, gender and culture. Thus to obtain a consistency between the mission of the researcher and the mission of evaluation ACUMEN will also be developing guidelines for Good Evaluation Practice, in the hope that evaluation will be implemented in such a way that does not undermine the authority of the researcher in their process of quality, and support their craftsmanship without giving them all the freedom or taking freedom away.

What difference will ACUMEN make? ACUMEN is investigating how evaluation plays out in diversity of labour force and gender. This questions the neutrality of evaluation and how straightforward it is. In cooperation with the European Commission, ACUMEN will contribute to policies and that get research evaluation on a better track. The goal is still to promote excellence and tools that can solve societal problems but keep space for creativity. The connection of analysis of the individuals career with evaluation and the interaction between evaluation process and career advancement will be strengthened. The measures created will enrich CVs and point to activities in systematic way that is acceptable to evaluators. The ACUMEN Portfolio is the link between knowledge evaluation and how this is embedded in research careers evaluation.

Your Job: A brief outline and how to save your work

Please send your email to pnm664@iva.ku.dk (Lorna) and you will be invited to join the Dropbox Folder: ACUMEN Data Extraction. In the Dropbox folder "ACUMEN Data Extraction" you will find a folder for each of the four disciplines. There is also a "Troubleshooting" folder where you will find tips on how to search Web of Science and Google Scholar. Feel free to up load your own tips to share with your project colleagues.

You will be allocated a master excel sheet containing a list of authors and links to their online publication list(s). **All text in the excel sheets is to be written in English**. The only information you alter is this sheet is the following:

Part 1

1.1) Follow the link to the author's publication list.
1.2) Verify that the link is working. Mark in the Excel sheet, in the cell "link", if the link is:
working and a researcher within the discipline you have been assigned (w),
dead (d),
not a publication list (n),
not the academic seniority you have been assigned (not seniority)
if the researcher does not belong to the discipline (nd), or
if the researcher appears on the list more than once (duplicate)
1.3) If the link leads to a publication list (w) of a researcher within the discipline you have been assigned (duplicates removed), copy/past the whole line of author information into the 2nd sheet, labeled "working links".
1.4) Save this excel sheet in the Dropbox folder, ACUMEN Data Extraction, under the correct discipline, under the correct academic seniority as so:
Discipline_academic seniority_workinglinks_your initials
1.5) Save a copy of the publication list in the corresponding folder in our Dropbox.
Save it as "Author surname_Bib ID number_your initials" for example "Druckmullerova_8_LEW"
What format to save in?
-If the publicationlist can be easily exported, export into an excel file, test file or word document (whatever is easiest).
-If the publication list is a PDF, save as PDF where as,
-if the publication list is a list on a website that requires the references are copy/pasted one by one, take a screen shot and save that. Ensure you have all the bibliographical information.

Part 2

2.1) Using the sheet "working links" as your master, start with the first author on the list. Follow the link and keep it open while you find the authors publications in Web of Science and Google Scholar. 2.2) Add 3 more cells in the header of the "working links" at the end of the author information: "number of publications on list", "number of publications GS", "number of publications WOS".
2.3) If the author has links to more than one list, you'll have to compare the lists for duplicates. Assess what the author writes about, the institutions they are affiliated to and the age range of the

publications. This will help you verify the publications found in Web of Science and Google Scholar.
2.4) Note how many publications the author has listed, and write the amount in the cell “number of publications on list”

Part 3

3.1) For each author create a new Excel folder “Discipline_seniority_author name_yourinitials” with 3 sheets – name the first “author name_GS”, the second “authorname_WOS”, and the third “authorname_duplicates”.
3.2) Search Google Scholar (GS) using Publish or Perish version 4 or newer, for publications by the author and export to the sheet “author name_GS”.
3.3) Search Web of Science (WOS) for publications by the author and export to the sheet “authorname_WOS”.
3.4) Some researcher’s names are so common that they generate an enormous amount of results in GS and it is accordingly impossible to verify authorship. Mark in the authors excel sheet (“Discipline_seniority_author name_impossible_yourinitials”) that they were impossible and save this sheet to the Dropbox folder ACUMEN Data Extraction, Impossible
3.4) Copy and paste the GS list into the third sheet “authorname_duplicates”. Highlight the list with a colour. Copy and paste the WOS list into the same sheet. Make sure the titles are in the same column. Mark the entire list and sort after title alphabetically. The colour makes it easy to see the duplicate publications, both between WOS and GS, and GS and GS.

If you make changes to the files you have saved in the Dropbox folder, please save with a revised number, such as
Public Health_Professor_JSmith_LW02

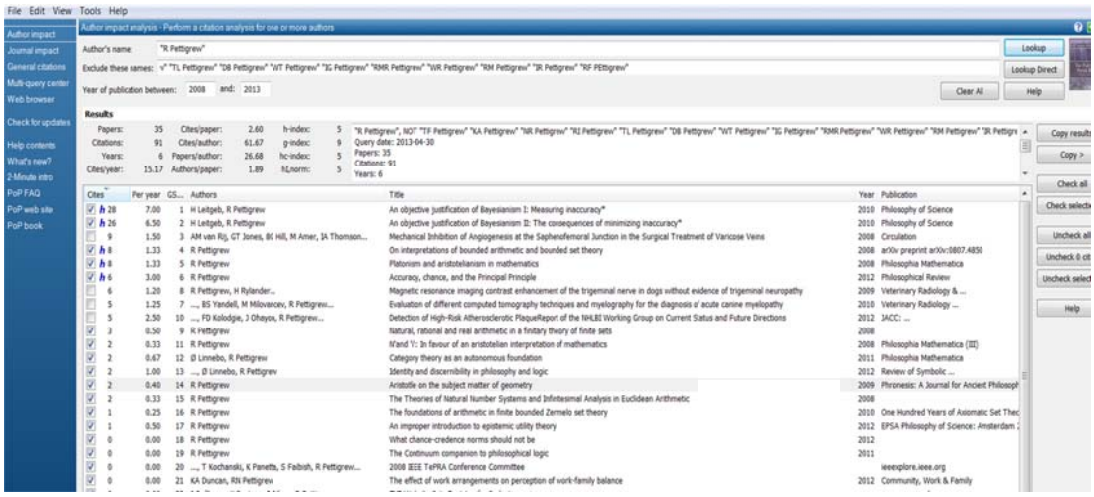
For both GS and WOS:
if the researcher has no publications please write in their corresponding excel sheet and write “No publications”.

Method of Data Collection: Google Scholar, through Publish or Perish

Download and install Publish and Perish: http://www.harzing.com/pop.htm
Search using the **Author Impact** function. The Author impact analysis page allows you to perform a quick analysis of the impact of an author's publications. The Author impact analysis page contains the following panes:
-Author query pane
-Results pane

How to perform an Author impact analysis
To perform a basic impact analysis:

1. Enter the author's name in the Author's name field;
2. Click Lookup or press the Enter key.
3. The program will now contact Google Scholar to obtain the citations, process the list, and calculate the Citation metrics, which are then displayed in the Results pane. The full list of results is also available for inspection or modifications and can be exported in a variety of formats.



From the researcher's publication list see how does the researcher writes their name in the author byline. Use this form to search the databases. Fx The author name below has the following forms, so you will have to search them all. Write them with "quotes" with OR in between each name.

"Piotr A Dybczynski" OR "PA Dybczynski" OR "Dybczynski, P"

How to export from POP to excel:

Step 1
Copy>
copy statistics for excel with header
Open excel arc
ctrl v

Step 2
Copy>
copy results for excel with header
Open excel arc
ctrl v

Tips to searching

1. Always use "quotes" around the author's name, e.g. "A Harzing".
2. PoP is not case dependent, "A HARZING" gives the same result as "a harzing"
3. The order of search terms does not matter. "A Harzing" will give the same result as "Harzing A".
4. Use an author's initials rather than their full given name as not all journals publish author names in full.
5. If an author has consistently published with only one initial, you can exclude namesakes using 2nd and 3rd initials by using wildcards in the "exclude these names" field, e.g. when searching for "G Sewell", you can exclude "G* Sewell" "G** Sewell".
6. You cannot use "*G Sewell" to exclude "WG Sewell" or "AG Sewell". You need to manually exclude these authors by listing them in the "exclude these names" field. To exclude certain author names, enter them in the Exclude these names field. For example, to exclude CLC Kulik from the earlier example, enter "CLC Kulik" in the Exclude these names field. You can enter more than one exclusion in Exclude these names: "CL Kulik" "CLC Kulik" would exclude both these combinations from the search.
7. If an author has published under two different names (e.g. maiden name and married name) use OR between search terms for a combined search "WG Sewell" OR "W Sewell"
8. If an author has mostly published with two initials, but has incidental publications with one initial, a combined search with initials and full given name (e.g. "CT Kulik" OR "Carol Kulik") will usually capture all of their publications.
9. Do not try to use the AND keyword in an author search. Google Scholar does not recognize this keyword and will treat it as a normal search word. Instead, just enter multiple author names; this will behave as an "and" search by default.
10. If you are looking for an author whose name contains accented letters, then it might help if you include several variations of the name, both with and without accents, and also with the accented letters missing. For example, to search for someone with the surname Veríssimo (note the accent on the first 'i'), use the following names in the Author field: "Veríssimo" OR "Verissimo" OR "Verssimo"
11. If the list of results is fairly limited, you can manually include or exclude citations from the analysis by checking or clearing the boxes in the Results list.

Limiting year

Before limiting the year range, always check whether an author has highly cited publications without a year listing. If you know that a certain author only published after (or before) a certain year, you can enter the start or end years in the Year of publication between ... and ... fields. You can also use these fields if you want to analyse the author's publications from a given period.

(De)Selecting and merging results

You can deselect publications not published by the target author. Simply remove the tick mark in the first column by clicking on it.

You can (de)select more than one publication at once by first selecting the relevant publications and then clicking the "(un)check selection" button.

If the results contain duplicate entries, you can merge them by dragging and dropping the duplicate entries onto the master record.

Selecting relevant publications for unchecking or merging can be made easier by first sorting the results by Cites, Authors, Title, Year, Publication, or Publisher. Sorting is done simply by clicking on the corresponding column heading. Click twice to reverse the sort order.

Here are some shortcuts:

1. The Check all button places check marks in all boxes;
2. The Uncheck all button clears all boxes;
3. When you use the keyboard to travel up and down in the Results list, pressing the space bar toggles the check mark on and off on the selected line.
4. You can also select a consecutive range of items in the list (left-click on the first item, then hold either Shift key and left-click on the last item) and use the Check selection/Uncheck selection buttons to check/uncheck all selected items and recalculate the citation statistics.

Example of a step-by-step search strategy

Search for the target academic's name with his/her first initial and surname in quotes, e.g. "a harzing". Please note that Google Scholar matches the surname and initials anywhere in the initials+surname combination, so "C Kulik" would be matched by CT Kulik, CLC Kulik, but also by PC Kulik.

It is generally better to use fewer initials and then exclude the ones you don't want (see next point) instead of using more initials, because many citations (or authors) are sloppy with the initials they use. With too many initials in the Author's name field you run the risk of missing a substantial number of relevant articles.

To exclude certain names, enter them in the Exclude these names field. For example, to exclude CLC Kulik from the previous example, enter "CLC Kulik" in the Exclude these names field (and keep "C Kulik" in the Author's name field). You can enter more than one exclusion in Exclude these names: "CL Kulik" "CLC Kulik" would exclude both these combinations from the search.

If the result includes publications not published by the target academic, deselect those publications (remove the tick mark in the first column by clicking on it). If the list is long, it might be easier to deselect all publications first and then only select the relevant publications. Please note that any titles with less than 5 citations usually have very little or no impact on the h-index, but might influence the g-index. Hence, if you are faced with a very long list and are only interested in the h-index, you might consider deselecting all and only reviewing titles with 5 or more citations.

Selecting relevant publications might be easier by sorting the results by Cites, Authors, Title, Year, Publication, or Publisher. Sorting is done simply by clicking on the corresponding column heading.

Example of an author that is impossible to verify

Common names are time consuming, but it is still quicker to use POP than export by hand. I found that for common names general search is quicker than author search. Write the name of the author in quotes in the author field and then in the “None of the words” field write the author names you wish to exclude, again in quotes around each name.

Author’s name: B Jansen

None of the words: "BJ jansen" "BAJ Jansen" "BG Jansen" "KMb Jansen" "bsh Jansen" "bjp Jansen" "bes Jansen" "bmp Jansen" "bh jansen" "bd jansen" "hb jansen" "be jansen" "bjm jansen" "gb jansen" "br jansen" "rb jansen" "brj Jansen""hwb Jansen" "bd jansen" "ba jansen" "jb jansen" "bgm jansen" "bc jansen" "mb jansen" "bjm jansen" "lb jansen" "bjh jansen" "bd jansen" "pb jansen" "bp jansen" "jansen-schulz"

Year of publication: 2001-2013

The search time still returns over 1000 references. Also I’m being warned that Google will block me. When you find such an author, mark in your dataset that he/she impossible. Copy the all the author’s information from your master excel arc into the ACUMEN data extraction dropbox folder_impossibles.

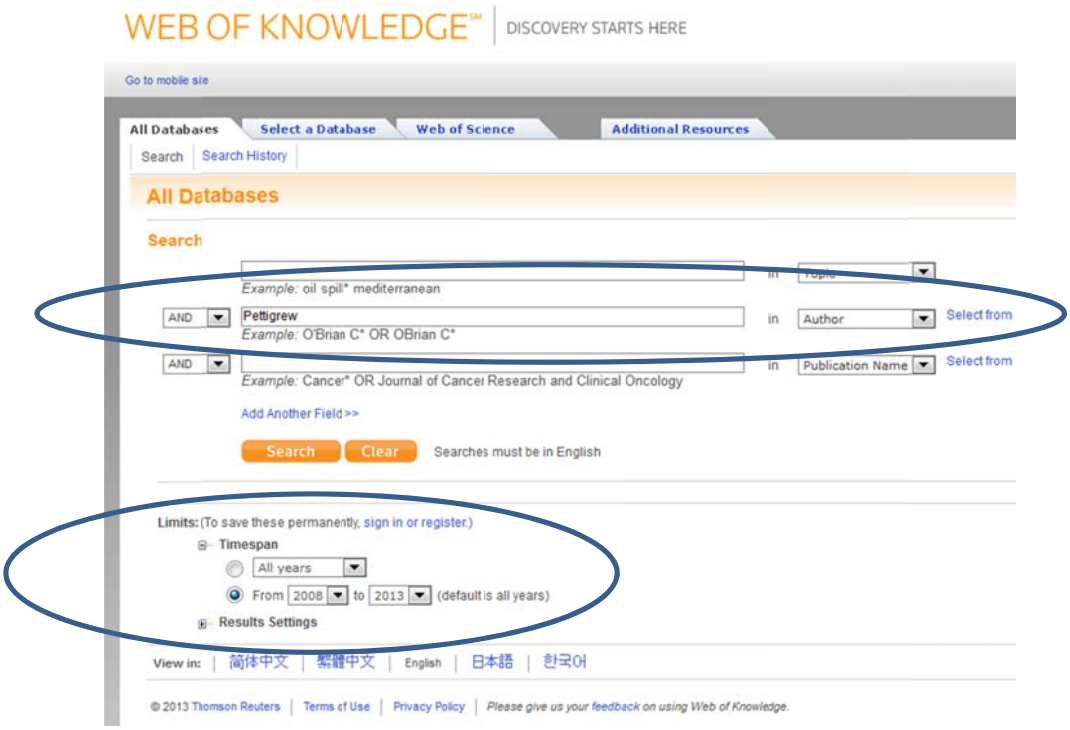
Searching and making the results accurate is time-consuming as in February 2013 Google Scholar reduced the maximum number of results per page from 100 to 20. This means that Publish or Perish now has to retrieve up to 5 times as many result pages per query in order to show the full results and has following effect on data extraction:

- More page requests mean that POP hits the maximum number of requests that Google Scholar allows per hour sooner.
- If the number of page requests exceeds the maximum that Google Scholar allows, our IP address will be temporarily blocked by Google Scholar. This block can last for up to 24 hours.
- To avoid hitting the maximum allowable request limit, POP uses an adaptive request rate limiter. This limits the number of requests that are sent to Google Scholar within a given period, both short-term (during the last 60 seconds) and medium term (during the last hour).
- To achieve the required reduction in requests, Publish or Perish delays subsequent requests for a variable amount of time (up to 1 minute). The higher the recent request rate, the longer the delays.

This means for us that the amount of data collection per session is limited and the speed of data extraction is slower than before. The alternative is being blocked by Google Scholar for up to 24 hours. As we are performing queries that yield many results (several hundred or more) and issue a number of queries in short succession, the request rate limiter will insert progressively longer delays to keep the overall request rate within acceptable limits. To avoid this, spread the queries over the day.

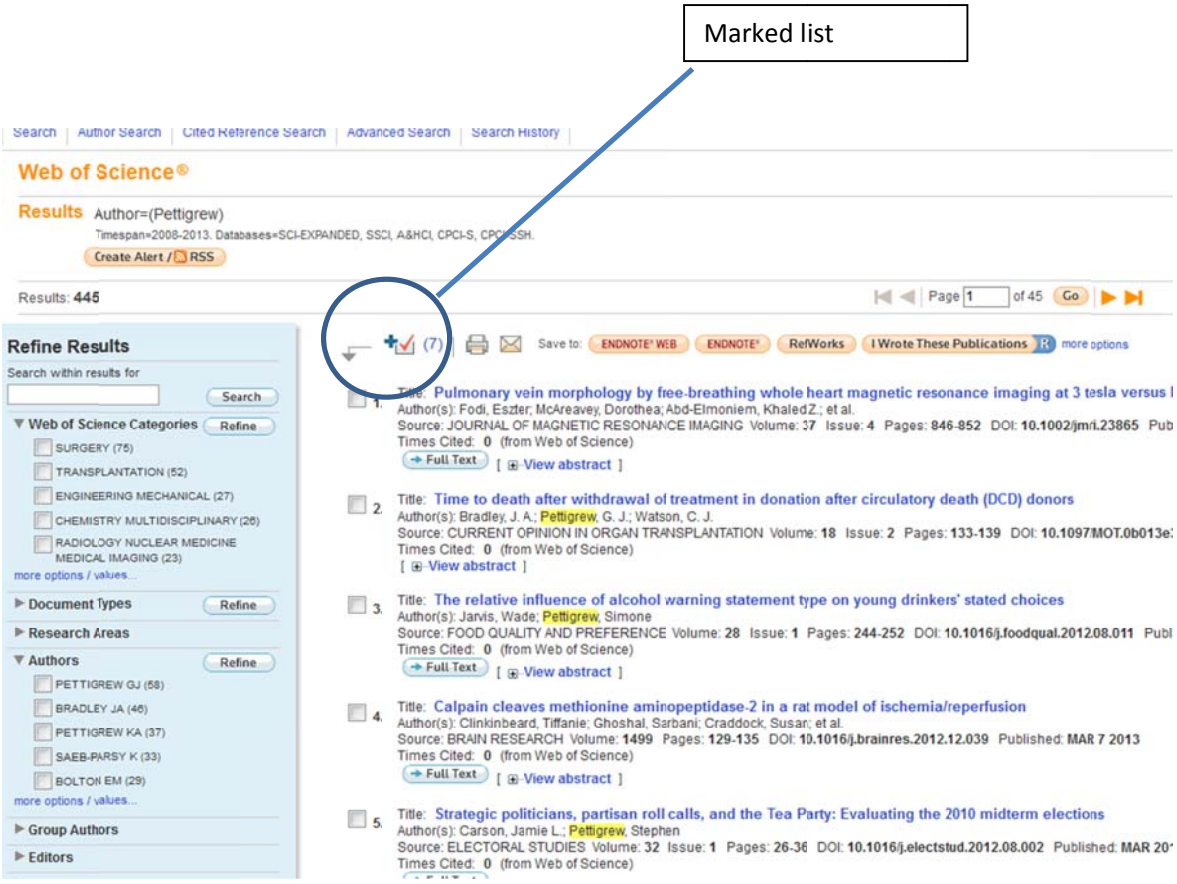
Method of Data Collection: Web of Science

Open Web of Science (a citation database that is part of Web of Knowledge).



Enter the researcher's surname and possible initials in the search box. Limit the field to "author". Limit the search, under Timespan, from the earliest publication year reported on the author's publication list.

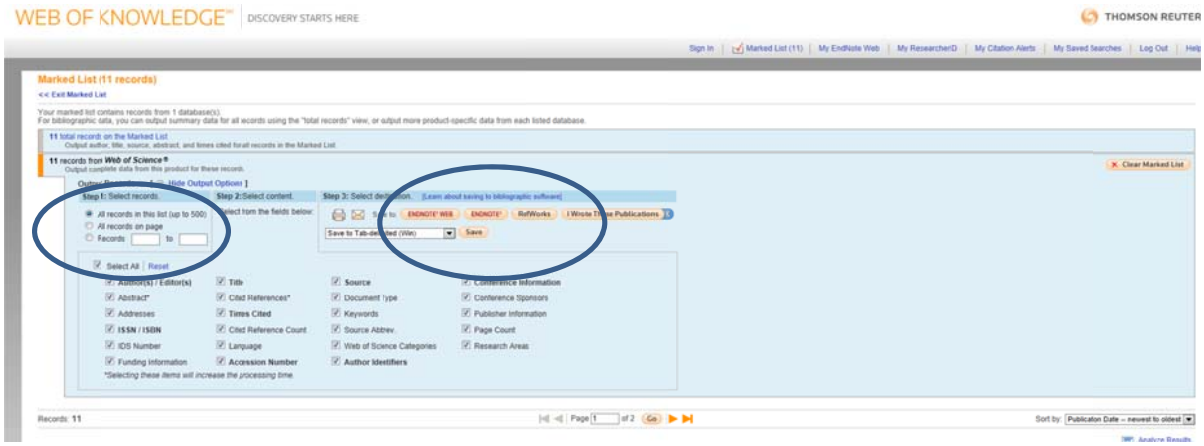
Press search.



First limit to author name: In the column refine results click on Author, and more options. Click the surname and initial option(s) that are relevant and click refine to just include these variants.

Repeat for Web of Science categories. If there are just a few categories click on those you wish to exclude and then click on “exclude”. If there are many options, select the relevant categories and “refine”. Think broadly when using the categories and narrow the search slowly, continuously checking the results list. Philosophy can for example also be included in the mathematics, social studies, or management category.

When you are satisfied with the list, click the boxes beside the references to add the articles to your marked list. You find the marked list at the top of the search. Click the plus to add to your list. Click on the number in parenthesis to enter your marked list.


WEB OF KNOWLEDGE
DISCOVERY STARTS HERE
THOMSON REUTER
Marked List (11 records)
<< Exit Marked List
Clear Marked List
Step 1: Select records
Step 2:Select content
Step 3:
All records in this list (up to 500)
All records on page
Records
to
Select All
Reset
Author(s) / Editor(s)
Title
Source
Conference Information
Abstract*
Cited References*
Document Type
Conference Sponsors
Addresses
Times Cited
Keywords
Publisher Information
ISSN / ISBN
Cited Reference Count
Source Abbrev.
Page Count
IDS Number
Language
Web of Science Categories
Research Areas
Funding Information
Accession Number
Author Identifiers
*Selecting these items will increase the processing time
Save
Records: 11
Page 1 of 2
Go
Sort by:
Publication Date -- newest to oldest

Step 1: In the marked list check “All records in this list” and “Select All”

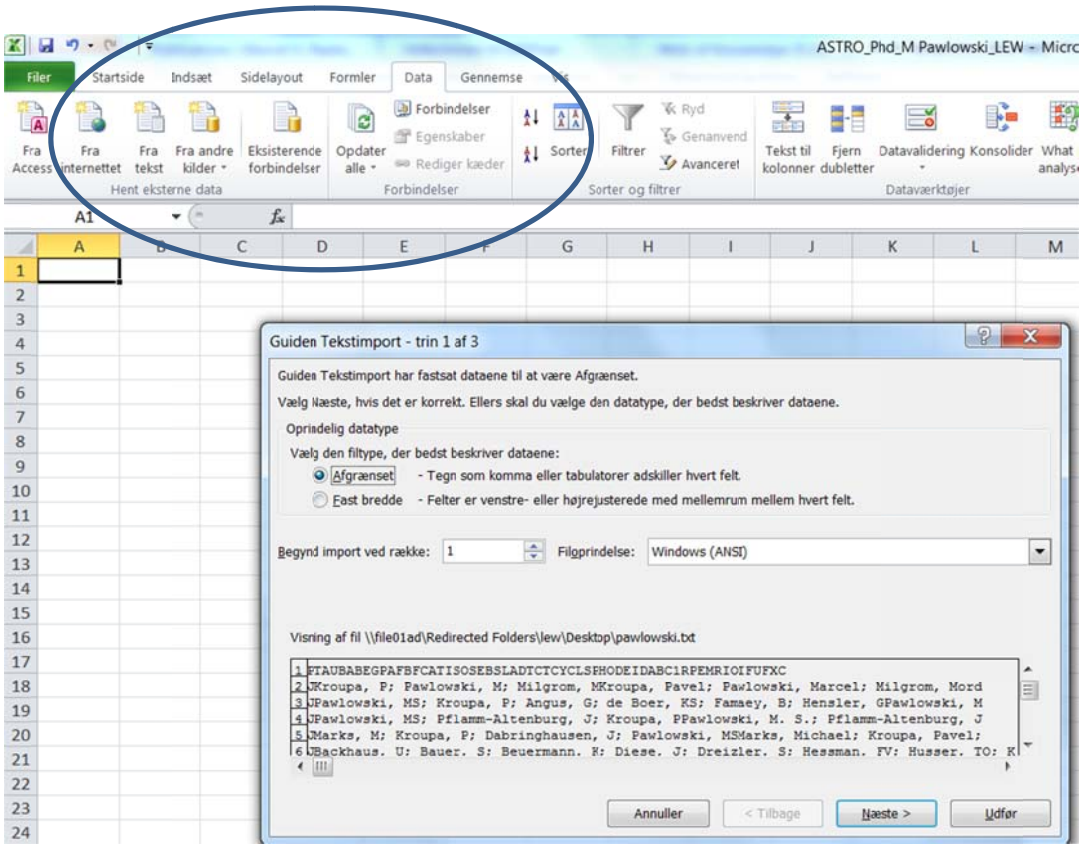
Click on Step 2: Selected destination and save as Tab de-limited Win or Mac dependent on your computer.

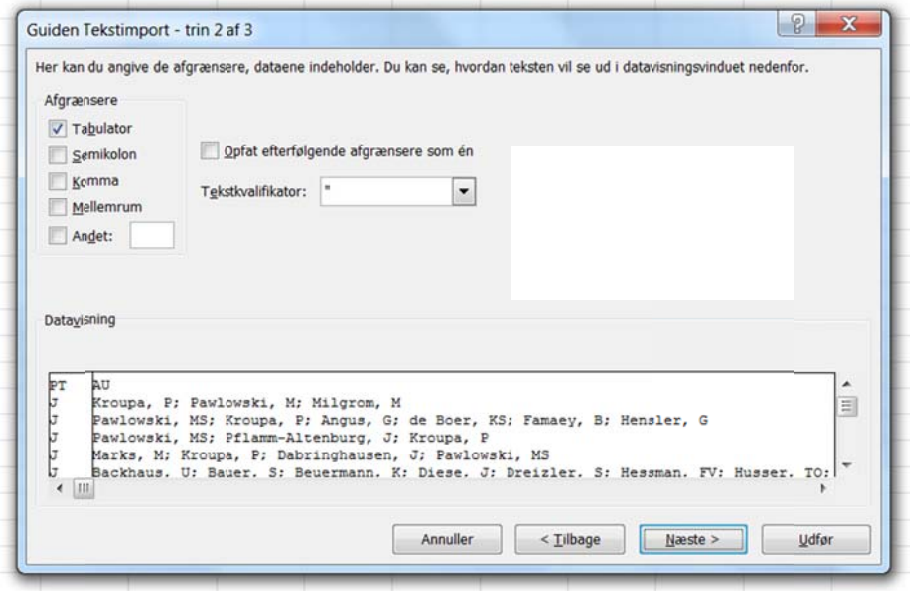
Save

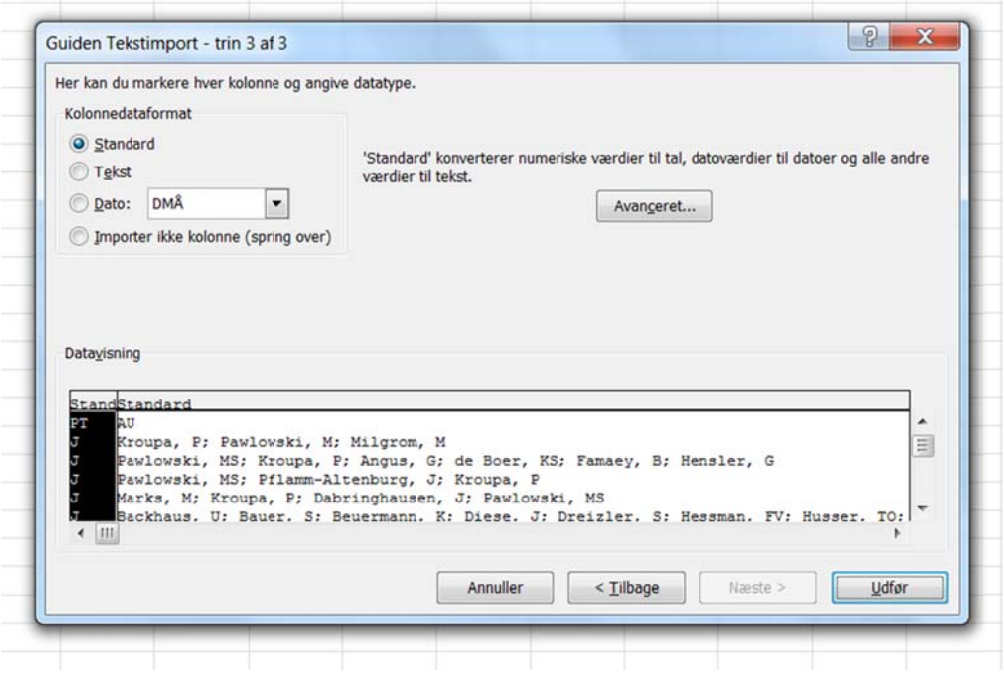
How to export to from WOS Excel

Save the file on your computer.

Open Excel and choose the “Data” tab from the navigation menu. Click on “from text”. Choose the text file from the pop-up menu and “import”. The Text Import Guide pops up. Follow the guide to import the text into the cells of the Excel sheet.







Before the next search

Before you do a new search in WOS, remember to clear your marked list.

After you have typed in the next author name, check the year limits are correct.

Indicators of Output: Published or unpublished countable works

ID nr.	Indicator	Description	WOS	GS	Astro.	Enviro.Sci	Phil.	Health	Comments
1	**P**	Count of production used in formal communication			✓	✓	✓	✓	From authors CV
2	**P_{isi}**	Used in the calculation of impact compared to world subfield citation average based on ISI citation data.	✓	(✓)	✓	✓	✓	✓	Also in GS.
3	**P_{ts}**	Number of publications in selected sources defined important by the researcher's affiliated institution.			✓	✓	✓	✓	Exemplify with BFI for Denmark, evt. other countries authorized lists
4	**Co-publications**	Collaboration on departmental, institutional, inter- or national level & identify networks.	✓	✓	✓	✓	✓	✓	More relevant in some fields than others
5	**Fractional counting on papers**	Shared authorship of papers gives less weight to collaborative works than non-collaborative ones.							Fractional counting is not beneficial from the individual's viewpoint. No one would want to reduce their score.
6	**Proportional or arithmetic counting**	Shared authorship of papers, weighting contribution of first author highest and last lowest.							Ditto
7	**Geometric counting**	Assumes that the rank of authors in the byline accurately reflects their contribution							Ditto
8	**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.,							Ditto
9	**Noblesse oblige**	Indicates the importance of the last author for the project behind the paper.							Ditto
10	**FA** First author counting	Credit given to first author only							ditto
11	**Weighted publication count**	A reliable distinction between different document types.			✓	✓	✓	✓	Which weights should be applied there are no standards. A table summary of type of work would be interesting- If the author does it themselves, a high level of detail is achievable, if we do it in GS/WOS it would be limited.

Remember: The researcher has to be able to do these indicators themselves.

ID nr.	Indicator	Description	WOS	GS	Astro.	Environ. Sci.	Phil.	Health	Comments
1	**C + sc**	Indication of all usage for whole period of analysis	✓	✓	✓	✓	✓	✓	
2	**C**	Recognised benchmark for analyses. Indication of usage by stakeholders for whole period of analysis							Do self-citations include cites from co-writers? This could be messy
3	**Scimago Total Cites (STC)**	Indication of usage by stakeholders for whole period of analysis							Citing info only available from after 1996. Access to Scopus can be limited because of the cost
4	**C-sc**	Measure of usage for whole period of analysis							
5	**% SELFCIT**	Share of citations to own publications							
6	**CPP**	Trend of how cites evolve over time	✓		✓	✓	✓	✓	Very rough measure
7	**Ptop**	Identify if publications are among the top 20, 10, 5, 1% most frequently cited papers in subject/subfield/world in a given publication year.	✓	✓	✓	✓	✓	✓	Percentiles not affected by skewed distribution. Requires reference standard
8	**Field top % citation reference value**	World share of publications above citation threshold for n% most cited for same age, type and field	✓	✓	✓	✓	✓	✓	Ditto
9	**E(Ptop)**	Reference value: expected number of highly cited papers based on the number of papers published by the research unit.							More interesting on department level
10	**A/E(Ptop)**	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.							Ditto
11	**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.	✓	✓	✓	✓	✓	✓	
12	**Number of significant papers**	Gives idea of broad and sustained impact	✓	✓	✓	✓	✓	✓	Logical measure if individuals define own reference standard and compare to that
13	**Age and productivity**	Effects of academic age on productivity and impact.							Other effects could be more interesting such as effect of grant on productivity.
14	**%Pnc**	Share of publications never cited after certain time period, excluding self-citations	✓	✓	✓	✓	✓	✓	Useful in reflection and justifying why something is not cited fx according to type; encyclopeadia, preface or schism between language & subject

Indicators of Research Infrastructure: Collaboration and to which extent these are citing the work

ID nr.	Indicator	Description	WOS	GS	Astro.	Environ. Sci.	Phil.	Health	Comments
1	**Number of co-authors**	Indicates cooperation and growth of cooperation at inter- and national level;	✓	✓	✓	✓	✓	✓	General interest to see if author works in groups, alone, repeated collaborations
2	**Co-citations**	Thematic networks and influence and impact of researcher.							Not interesting for CV
3	**Fractional counting on citations**	Designed to remove the dependence of co-authorship (Egghe, 2008)							Not interesting for individual to reduce citation count
4	**hi-index**	Indicates number of papers with at least h citations scientist would have written if worked alone.	✓	✓	✓	✓		✓	Useful in subjects with extreme co-authorship such as Astronomy. Not too much work for author as limited to h core
5	**POP variation individual H-index**	Accounts for co-authorship effects							Above is easier even though granularity is lost.
6	**n-index**	Enables comparison of researchers working in different fields:							Based on a journal's h (how will researcher get that, comparison between fields is interesting to evaluator not author)
7	**Alternative H index**	Indicates the number of papers a researcher would have written along his/her career if worked alone.							Same as hi-index
8	**Pure h-index (Hp)**	Corrects individual h-scores for number of co-authors							Based on fractional counting and place in author by-line. A lot of work.
9	**Cognitive orientation**	Identify how frequently a scientist publishes or is cited in various fields; indicates visibility/usage in the main subfields and peripheral subfields.	✓	✓	✓	✓	✓	✓	Interesting to see where work is published and cited (used). Graphically good addition to CV, easy to read.
10	**Visual representation techniques**	Based on bibliographic data graphical representations are generated of publishing, collaboration, citations, growth and activity in research field.			✓	✓	✓	✓	Sure. But which graphics/tools should be used in which fields?

Indicators of Impact: Visibility in the field. (Highlighted were excluded in review, as these are impact of journal and not author).

Even though these are indicators of journal performance, we have to establish a field norm. A field norm is used as comparison in the other categories (fx sustainability, quality) and general yardstick measure of what is expected. If the researcher can document he is performing better than a field standard he will want to do that. Thus, the portfolio has to either present field norms that are up to date or present methods for the researcher to define his own standard.

ID nr.	Indicator	Description	WOS	GS	Astro.	Enviro. Sci.	Phil.	Health	Comments
1	**ISI JIF (SIF) Synchronous IF**	Average number of citations a publication in a specific journal has received limited to ISI document types and subject fields.							Limited usefulness, but calculable by the individual. Measure of journal popularity and not designed for individual performance
2	**Diachronous IF**	Reflects actual and development of impact over time of a set of papers.	✓		✓	✓			Possible in WOS, time consuming with GS. Better represents impact of researcher than ISI JIF.
3	Y Factor	Scientific impact defined as a combination of popularity and prestige							Based on JIF. Measure of journal impact.
4	Scimago Journal Rank (SJR)	Average per article PageRank based on Scopus citation data							Not GS or WOS
5	EigenFactor	Journal's total importance to the scientific community							Not GS or WOS
6	Article influence score (AI)	Measure of average per-article citation influence of the journal							Not GS or WOS
7	Normalised journal impact	Mean impact value of all the normalized citation counts for publications in a specific journal							Measure of journal impact
8	Journal to field impact score (JFIS)	Journal to fields citation score that indicates relative impact of a journal							Measure of journal impact
9	Discipline Impact Factor (DIF) (Hirst, 1978)	Number of times a journal is cited by the core literature of a single subfield rather than a complete set of ISI journals.							Index loses detail as dependent on ISI Journal Citation Reports i.e. it is affected by JCR field coverage and minimum cites inclusion criterion.
10	**Median impact factor (IF med)**	The aggregate Impact Factor for a subject category. Median value of all journal Impact Factors in the subject category.							Author can specify journals from websites (if report IF). Aggregate impact factor for a subject category. Compliments JIF
11	**Normalised journal**	Compare reputation of journals across fields							Based on JCR, used in across

	position (NJP)								field comparisons. Not relevant for individual
12	**Field citation score (FCS)**	Represents the number of citations expected for a paper of the same type, published in all journals within a specific field in the same year, and document type.	✓		✓	✓			ISI CI field categories are inadequate for some disciplines, providing a distorted picture
13	**Field Citation Score Mean (FCSm)**	Weighted average for comparison of impact in different subfields							Indicator on a higher level of aggregation than individual
14	**JSCS or JRV Journal citation score (journal reference value)**	Worlds average of citations to publications according to type and age. Journal-based worldwide average impact as an international reference level for the university/institute/department/group/researcher etc.							How can the individual do this?
15	**Normalised Journal Citation Score (JSCm)**	Reference value accounting for type of paper and years in which papers were published. Mean citation rate of all articles published in the journals in which the individual has published.							More accurate for activity in subfields than FSCm especially for developing and interdisciplinary fields.
16	**JCSM/FCSm**	Journal based worldwide average impact mean for an individual researcher compared to average citation score of the subfields							Favours senior researchers as minimum publication value if 50 is recommended for informative analysis. Dependent on calculation of JCS and FCS
17	**Crown Indicator CPP/FCSm**	Individual performance compared to world citation average to publications of same document types, ages, and subfields.							Limited to same document type as world citation average is based on. Dependent on calculation of FSCm.
18	**P_{tj}**	Performance of articles in journals important to (sub)field or institution.	✓	✓	✓	✓	✓	✓	
19	**CPP/JCSm**	Indicates if the individual's performance is above or below the average citation rate of the journal set.	✓	✓	✓	✓			We can't expect the individual to calculate the score of the journal set. These would have to be available standards, hence relation to individual is limited. Also limited in philosophy and public health (national interest)
20	**JCSm/FCSm**	Relative impact level of the journals compared to their subfields.							Measure of journal impact
21	**C/FCSm**	Applied impact score of each article/set of articles to the mean field average in which the researcher has published							Dependent on calculation of FCSm

Indicators of quality: Level and performance of research

ID nr.	Indicator	Description	WOS	GS	Astro.	Enviro.Sci.	Phil.	Health	Comments
1	**h-index**	Cumulative achievement	✓	✓	✓	✓	✓	✓	Anbefale reference standard w.r.t specialty. Guidelines how to establish on local (peers in dept), national and expert level if necessary (leaders in field).
2	**g-index**	The distinction between and order of scientists (Egghe, 2006; Harzing, 2008)	✓	✓	✓	✓	✓	✓	
3	**Hg-index**	Greater granularity in comparison between researchers with similar h- and g- indices.			✓	✓	✓	✓	
4	**Normalized h-index**	Normalizes h to compare scientists achievement based across fields							Not relevant for this study
5	**H(2) index**	Weights most productive papers but requires a much higher level of citation attraction to be included in index.	✓	✓	✓	✓	✓	✓	Weight most productive papers but requires higher citation level.
6	**A-index**	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	✓	✓	✓	✓	✓	✓	Average number of citations in H core, to imply that some papers are more highly cited than others. Has information redundancy with h.
7	**R-index**	Citation intensity and improves sensitivity and differentiability of A index	✓	✓	✓	✓	✓	✓	Square root of H and A index. Pretty much the same as g, but easier to calculate
8	**ħ-index**	Comprehensive measure of the overall structure of citations to papers	✓	✓	✓	✓	✓	✓	Includes citations to all papers (square root of half of the total number of citations to all publications)
9	**m-index**	Impact of papers in the h-core (median nr of citations to papers in h core)	✓	✓	✓	✓	✓	✓	To demanding to be used as reference standard, as detailed citation data required. M quotient better.
10	**M-quotient**	Adjusts for length of career	✓	✓	✓	✓	✓	✓	simple
11	**e-index**	Complements the h-index for the ignored excess citations	✓	✓	✓	✓	✓	✓	Can only be used with h, as e accounts for the "more than h" citations, thus providing complete citation information
12	**Hmx-index**	Ranking of the academics using all citation databases together.	✓	✓	✓	✓	✓	✓	Maximun h across WOS, GS and Scopus (can compare with WOS, GS and database of choice fx ADS in astrology?)
13	**w-index**	The integrated impact of a researcher's excellent papers.							Not as recognisible as H and just like h the cut off point is arbitrary.

14	**Index of Quality and Productivity**	Quality reference value; judges the global number of citations a scholar's work would receive if it were of average quality in its field.	✓	✓	✓	✓	✓	✓	Could be interesting but requires reference standards to field and academic seniority. I'll look at it again to see if it is researcher tool or a system/evaluator tool.
15	**Q^2**	Relates two different dimensions in a researcher's productive core: the number and impact of papers	✓	✓	✓	✓	✓	✓	Dependent on calculation of m index and h index.

Indicators of Innovation & Social Benefits: Contribution to society's social, economic and cultural capital

ID nr	Indicator	Description	WOS	GS	Astro.	Enviro. Sci.	Phil.	Health	Comments
1	**Knowledge exchange**	Knowledge production, knowledge exchange, knowledge use and earning capacity			✓	✓	✓	✓	Information from CV as this is weighted count of keynote speeches, activity in agencies & organisations, public forums, committees, conferences & co-operation with companies. How to weight?
2	**Dissemination in public sphere**	Impact and use in public sphere (knowledge transfer)			✓	✓	✓	✓	Often not reported on CV. Count of contributions to, inc.: tv & radio programs, newspapers, non-peer reviewed journals, text books, public & professional websites and news forums.
3	**Patent applications**	Innovation			✓	✓			Count of patent applications. Quality or significance of patents is not on an equal level; Citations in patents is more interesting. How can researcher get these, and what are reasons to cite – influence, legal or political?
4	**Tool to measure societal relevance**	Aims at evaluating the the level of the effect of the publication, or at the level of its original aim			✓	✓	✓	✓	Questionnaire used as the (self-assessment) application form and the assessment form for the reviewer (Niederkrotenthaler, Dorner, & Maier, 2011)

Indicators of Sustainability: Use or decline in use

ID nr.	Indicator	Description	WOS	GS	Astro.	Enviro. Sci.	Phil.	Health	Comments
1	**Citation age c(t)**	The age of citations referring to a researcher's work.	✓	✓	✓	✓	✓	✓	
3	**AR-index**	AR is the square root of the sum of the average number of citations per year of articles included in the h-core. Accounts for citation intensity and age of publications in H core	✓	✓	✓	✓	✓	✓	do not consider AR convincing as a ranking metric in research evaluation as the decay of a publication is very steep and insensitive to disciplinary differences
4	**Price index – PI (Price, 1970)**	Percentage references to documents, not older than 5 years, at the time of publication of the citing sources							Interesting bibliometrically, but not interesting for researcher
5	**Immediacy index**	Speed at which an average article in a journal is cited in the year it is published							
6	**Aggregate Immediacy Index (AII)**	How quickly articles in a subject are cited	✓	✓	✓	✓	✓	✓	If we can define a subject area and journals this could be an useful metric
7	**Cited half-life (CHL) & Aggregate Cited Half-Life (ACHL)**	A benchmark of the age of cited articles in a single journal							
8	**Classification of durability**	Durability of scientific literature on distribution of citations over time among different fields	✓		✓	✓			Only tested in WOS using journal subject categories
9	**Age-weighted citation rate (AWCR, AW & per-author AWCR) ***	AWCR measures the number of citations to an entire body of work, adjusted for the age of each individual paper	✓	✓	✓	✓	✓	✓	Field norm has to be decided to account for field characteristics such as expected age of citations, "sleeping beauties", and delayed recognition.

*The AW-index is defined as the square root of the AWCR. It approximates the h-index if the mean citation rate remains constant over the years. The per-author age-weighted citation rate is similar to the plain AWCR, but is normalized to the number of authors for each paper.

Appendix 6.

Identification of the data needed to calculate the indicators and reference standards

	Elements needed to calculate metric											
Output	Author name	Author byline	Full CV	affiliation	country	Publication list	Article id	Authority list	Ref. standard(s)	Weighting standard	Citation database	WOS only
P	✓					✓						
P_{isi}	✓					✓	✓					✓
P_{ts}	✓					✓	✓	✓				
Co-publications	✓	✓		✓	✓	✓						
Weighted publication count	✓					✓				✓		

	Elements needed to calculate metric											
Outcome	Author name	Author byline	Full CV	affiliation	country	Publication list	Article id	Authority list	Ref. standard(s)	Weighting standard	Citation database	WOS only
CPP	✓					✓					✓	
Ptop	✓					✓			✓		✓	
Age of citations	✓					✓					✓	
Number of significant papers	✓					✓			✓		✓	
%Pnc	✓					✓					✓	

	Elements needed to calculate metric											
Research Infrastructure	Author name	Author byline	Full CV	affiliation	country	Publication list	Article id	Authority list	Ref. standard(s)	Weighting standard	Citation database	WOS only
Numbers of co-authors	✓	✓				✓						
Hi-index	✓	✓				✓	✓				✓	
Cognitive orientation	✓					✓	✓					✓
Visual representation techniques	✓	✓		✓	✓	✓	✓		✓		✓	

	Elements needed to calculate metric											
Impact	Author name	Author byline	Full CV	affiliation	country	Publication list	Article id	Authority list	Ref. standard(s)	Weighting standard	Citation database	WOS only
Diachronous IF	✓					✓	✓				✓	
P_{tj}	✓					✓	✓					
CPP/JCSm	✓					✓	✓				✓	

	Elements needed to calculate metric											
Quality	Author name	Author byline	Full CV	affiliation	country	Publication list	Article id	Authority list	Ref. standard(s)	Weighting standard	Citation database	WOS only
h-index	✓					✓					✓	
g-index	✓					✓					✓	
H(2) index	✓					✓					✓	
A-index	✓					✓					✓	
R-index	✓					✓					✓	
ħ-index	✓					✓					✓	
m-index	✓					✓					✓	
M-quotient	✓					✓					✓	
e-index	✓					✓					✓	
Hmx-index	✓					✓					✓	
Q^2	✓					✓					✓	

Innovation & social benefits	Author name	Author byline	Full CV	affiliation	country	Publication list	Article id	Authority list	Ref. standard(s)	Weighting standard	Citation database	WOS only
Knowledge exchange			✓							✓		
Dissemination in public sphere			✓			✓						
Patent applications			✓								Evt. Patent citation database	
Tool to measure societal relevance			✓									
Library holdings (academic/community library)	✓					✓					✓ (WorldCat)	

	Elements needed to calculate metric											
Sustainability	Author name	Author byline	Full CV	affiliation	country	Publication list	Article id	Authority list	Ref. standard(s)	Weighting standard	Citation database	WOS only
Citation age c(t)	✓					✓	✓				✓	
AR-index	✓					✓	✓				✓	
Classification of durability	✓					✓		✓			✓	✓
Age-weighted citation rate (AWCR, AW & per-author AWCR)	✓					✓	✓		✓		✓	

Reference standards individual can calculate	Author name	Author byline	affiliation	country	Publication list	Journal list w.r.t. subject(s)	Article id	Authority list	Weighting standard	Citation database	WOS only
ISI JIF synchronous IF					✓	✓					✓
Y factor					✓	✓					✓
Field citation score (FCS)/(FCSm)					✓	✓			✓		✓
JSCS or JRV Journal citation score (journal reference value)					✓	✓				✓	
Normalised Journal Citation Score (JSCm)					✓				✓	✓	
C/FCSm					✓	✓				✓	
production of colleagues of same academic seniority at dept-/institution,	✓		✓		✓						
Production of same academic seniority within field, national level/international	✓	✓	✓	✓	✓						
Production of expert reference group	✓	✓	✓	✓	✓						
citations to colleagues of same academic seniority at dept-/institution,	✓		✓		✓					✓	
Citations/ median citations to same academic seniority within field, national level/international	✓	✓	✓	✓	✓					✓	
Citations/median citations to expert reference group	✓	✓	✓	✓	✓					✓	
H index at local/national/expert level	✓	✓	✓	✓	✓					✓	✓
M quotient at local/national/expert level	✓	✓	✓	✓	✓					✓	✓
Index of Quality and Productivity	✓				✓	✓				✓	
Aggregate Immediacy Index (AII)						✓				✓	

All reference standards are time consuming to calculate.

Appendix 7

Overview of the dependence of indicators on other indicators, reference standards and weighting systems.

37 indicators of individual performance. An overview				
Metric	**independent**	**Dependent on calculation of another index**	**Dependent on calculation of reference standard**	**Comments**
P	✓			
P_{isi}	✓			
P_{ts}			✓	
Co-publications	✓			
Weighted publication count			✓	
CPP	✓			
Ptop			✓	
Age of citations	✓			
Number of significant papers	✓			
%Pnc			✓	
Numbers of co-authors	✓			
Hi-index		✓ (h)		Supplement to h
Cognitive orientation	✓			
Visual representation techniques	✓			
Diachronous IF	✓			
P_{tj}	✓			
CPP/JCSm			✓	
h-index	✓			
g-index	✓			
H(2) index		✓ (h)		
A-index		✓ (h)		Supplement to h
R-index		✓ (h) (a)		Supplement to h
ħ-index	✓			
m-index		✓ (h)		Supplement to h
M-quotient		✓ (h)		
e-index		✓ (h)		Supplement to h
Hmx-index		✓ (h)		

Q2		✓ (h)		
Knowledge exchange	✓			
Dissemination in public sphere	✓			
Patent applications	✓			
Tool to measure societal relevance	✓			
Library holdings	✓			
Citation age c(t)	✓			
AR-index		✓ (h)		Supplements h
Classification of durability	✓			
Age-weighted citation rate (AWCR, AW & per-author AWCR)			✓	

16 Reference standards, suggested methods that can be calculated by the individual.				
Metric	**independent**	**Dependent on calculation of another metric**	**Dependent on weighting**	**Comments**
ISI JIF synchronous IF	✓			
Y factor		✓ (isi jif)		
Field citation score (FCS)/(FSCm)			✓	If both FCS and JSCS are calculated, then JSCSm/FCSm (impact mean for an individual researcher compared to average citation score of the subfields)
JSCS or JRV Journal citation score (journal reference value)	✓			Simpler than FCS, but a rougher measure
Normalised Journal Citation Score (JSCm)			✓	
C/FCSm		✓ (FCSm)		
production of colleagues of same academic seniority at dept-/institution,	✓			
Production of same academic seniority within field, national level/international	✓			
Production of expert reference group	✓			
citations to colleagues of same academic seniority at dept-/institution,	✓			
Citations/ median citations to same academic seniority within field, national level/international	✓			
Citations/median citations to expert reference group	✓			
H index at local/national/expert level	✓			
M quotient at local/national/expert level		✓ (h)		
Index of Quality and Productivity		✓ estimated rate w.r.t. citation count, productivity, academis age, field citation habits		
Aggregate Immediacy Index (AII)	✓			

Part B

Part B. Data-collection

Work Package 5: New Bibliometric indicators August 6th, 2013
Project partners: Department of Information Studies, Royal School of Library and Information Science; Department of Library and Information Science, Humboldt University Berlin

Abstract

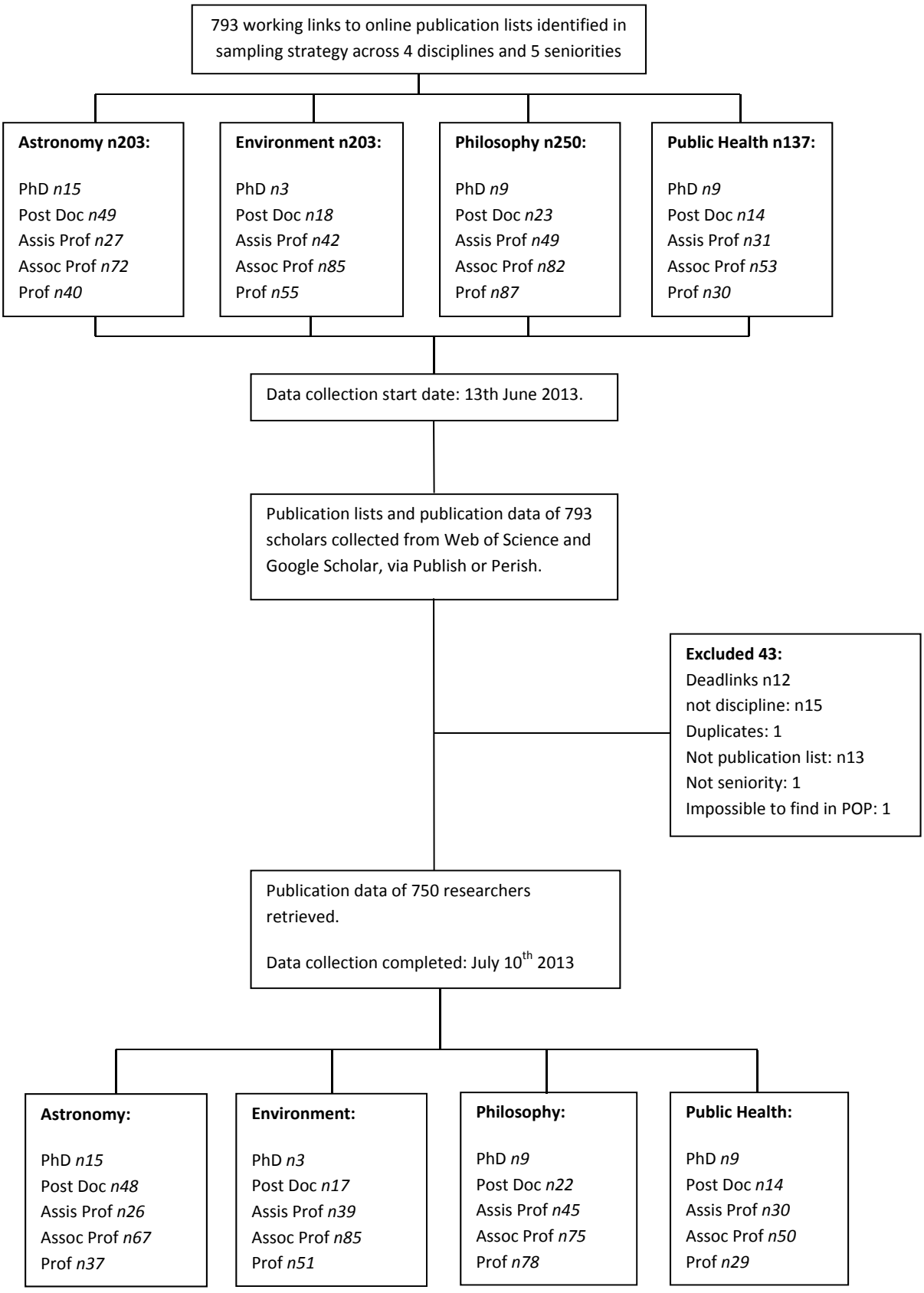
This report summarizes observations from the collection of publication data of the 793 scholars identified in WP5 sampling strategy dated 28th of June 2013: "Progress Report (draft to final report): Preparing for the analysis. Sampling strategy and methodological considerations in developing bibliometric indicators of the performance and impact of individuals for use in the ACUMEN portfolio".
The scholars' publication lists were collected. Individual scholar's lists of publications were then sourced in Web of Science and Google Scholar, using Publish or Perish. The information on 750 scholars was successfully collected and an overview of this sample of scholars is presented in this report. This final WP5 sample is available for all consortium members to use and can be found in the ACUMEN dropbox. To evaluate bibliometrically the scholar's performance in WOS, UT codes where collected and sent to CWTS where simple and sophisticated bibliometric indicators are currently being calculated, (a UT code is a unique article identifier used by Thomson Reuters that appears in databases in their Web of Knowledge service). The scholar's performance in GS will be evaluated using Publish and Perish's standard bibliometric indicators. Each scholar's POP statistics were collected. Observations from the data collection that could have importance for the design ACUMEN portfolio are presented in this report.

Data-collection

793 working links to online publication lists across 4 disciplines and 5 seniorities were identified in the sampling strategy[4]. The publication lists of these 793 scholars were collected from the scholar's homepage and publication data was searched for in Web of Science and in Google Scholar, via Harzings Publish or Perish. Forty-three scholars were excluded due to: the scholar's specialty falling outside the four disciplines investigated in preparation for the ACUMEN portfolio (15), no available publication list (13), deadlinks (12), duplicates (1), scholar impossible to identify (1) and the scholar's academic seniority is not considered in our study (1). This resulted in a dataset of 750 scholars: 193 in Astronomy, 195 in Environmental Studies, 229 in Philosophy and 133 in Public Health, Fig, 1. Data collection commenced on the 13th of June 2013 and was completed by the 10th of July 2013.

[4] WP5 (June 2013) Progress Report (draft to final report): Preparing for the analysis. Sampling strategy and methodological considerations in developing bibliometric indicators of the performance and impact of individuals for use in the ACUMEN portfolio.

Fig. 1. Flowchart of data-collection



Gender distribution in the Sample

In the sample of 750 researchers 584 are men and 165 are women, Table 1. Women make up 22% of the overall sample, a reduction of 1% from the potential sample identified in the sampling strategy but still reflecting the European ratio of men to women in science, 3:1[5]. Overall the data shows the trend that in the junior categories the ratio men to women is 2:1: phd students, post doc and assistant professor, while in the senior categories, associate professor and professor, the ratio is 4:1. This trend reflects the 2012 SHE figures of gender in research, confirming that our sample patterns the share of women employed in academia across Europe. Gender imbalance increases with age and women represent only 20% of Grade A academic staff, who are associate professors and professors[6].

It is important to understand however if the exclusion of the 43 scholars has consequences for the ration men to women within disciplines and academic seniorities. The ratio men to women in the astronomy, environment and public health disciplines remain unchanged. The majority of the exclusions, 21/43, were in philosophy. This was partly due to a large amount of dead links and partly due to scholars identified as not belonging to the discipline. The title “Doctor of Philosophy” does not necessarily relate to a scholar working as a philosopher or being affiliated with the history of science. In the context of academic degrees, the term "philosophy" does not refer solely to the field of philosophy, but is used in a broader sense in accordance with its original Greek meaning (love of wisdom) and thus is awarded to scholars in other specialties. This first became clear during data collection as the publication lists and publishing patterns of the scholar did not correlate with the other scholars in this discipline. The inclusion of these “false-positive” scholars in the dataset is a result of the automatic data-harvesting by the software used by WP2 to collect the original shared dataset from Web of Science. Manual filtering, that is reading the CVs and publication lists and consulting institutional webpages, was the only way to decide if the scholar’s specialty belonged to Philosophy or the History & Philosophy of Science.

Table 1. Distribution of seniorities and gender across the disciplines in the sample

	PhD	Post Doc	Assis Prof	Assoc Prof	Prof	Total
Astronomy	15	48	26	67	37	193
Gender M/F	*12:3*	*37:11*	*20:6*	*58:9*	*35:2*	*162:31*
Environment	3	17	39	85	51	195
Gender M/F	*3:0*	*11:6*	*30:9*	*72:13*	*44:7*	*160:35*
Philosophy	9	22	45	75	78	229
Gender M/F	*6:3*	*20:2*	*37:8*	*57:18*	*63:15*	*183:46*
Public Health	9	14	31	50	29	133
Gender M/F	*2:7*	*7:7*	*18:13*	*34:16*	*19:10*	*79:53*
Total	36	101	140	277	195	750
Discipline M/F	*23:13*	*75:26*	*105:36*	*221:56*	*161:34*	*585:165*

[5] Directorate-General for Research and Innovation, Unit B6 (2012) SHE Figures 2012: Gender in Research and Innovation. European Commision: Brussells.
Retrieved from:
http://ec.europa.eu/research/science-society/document_library/pdf_06/she-figures-2012_en.pdf
[6] SHE figures 2012.

The reduction has however had an overall positive effect on demographic of the philosophy category as the ratio men to women has decreased. By comparing the potential sample with the collected data, ratio men to women in the the phd category remains the same at 2:1, the post doc category has increased from 6:1 to 10:1, the assistant professor category has decreased from 5:1 to 4:1, the associate professor category from 4:1 to 3:1 and the professor category is also improved from 5:1 to 4:1.

Observations from the data-collection

Forty-three scholars were excluded during the data-collection: 10 from astronomy, 8 from environment, 21 from philosophy and 5 from public health. In appendix 1 we illustrate, in tables, from which discipline and seniority these scholars have been excluded and what caused the exclusion.

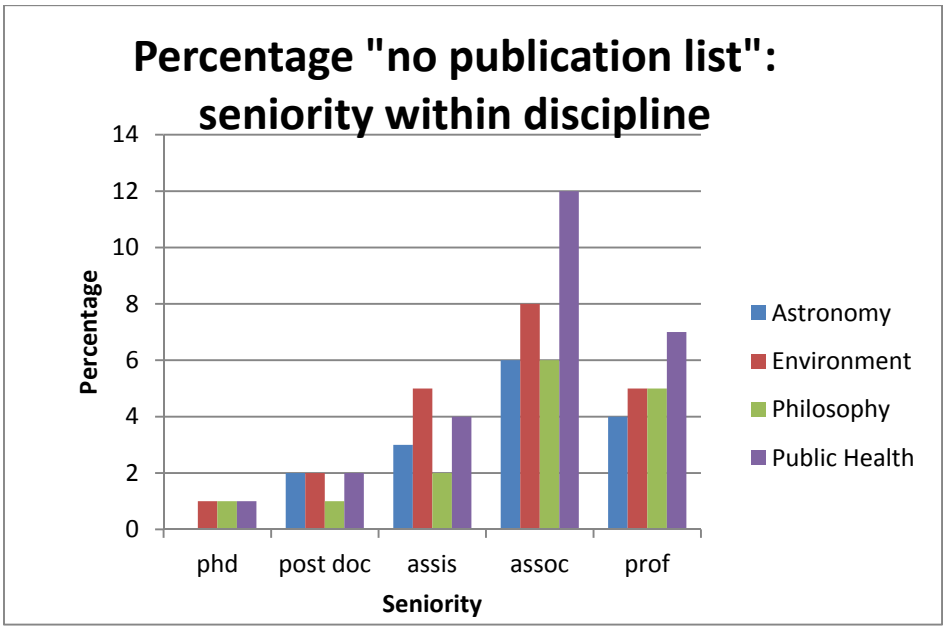
Our disciplinary samples are different sizes which mean direct comparisons of the causes of exclusions are not possible. Percentages are then used in the following analysis to indicate trends in online behaviour that lead to the exclusion. The total number of excluded scholars and included scholars within each discipline were added together and used as the denominator in the percentage calculations in Table 2 and figures 2 & 3.

Table 2. Percentage exclusion per discipline

	% Dead links	% Not discipline	% Duplicates	% No publication list
Astronomy	14	0.3	0.3	16
Environment	17	1	0	20
Philosophy	13	6	3	15
Public Health	17	2	0.4	25

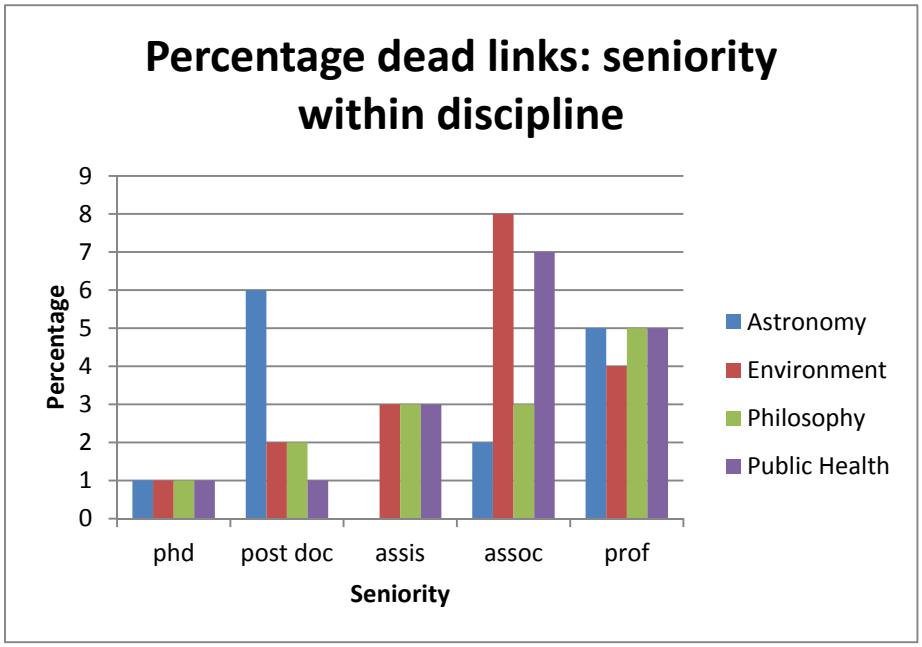
Noticeably the greatest reason for exclusion is that the scholar's online presence does not include a publication list. Often scholars write about their specialty, projects, activities and achievements to promote interest in themselves and their field of study but omit the publication list. This appears to be more prominent in public health and environment where the norm seems to be to link to a repository like Pubmed, Inspire or ADS. In these cases the "publication list" is a link to an author search in the chosen repository. For example scholar number 523, who is a professor in public health, links directly to his publications in PubMed with the simple author search: Reis S[Author]. This retrieves 523 references. These works are authored by Ries S, Reis SE, Ries SR, Ries Si etc. After exhaustive sorting we found that his real number of publications is only 62. We have interpreted this to mean that some scholars are either unaware of name ambiguity problems, of how databases "think" or are uncritical of numbers pulled from databases. This could really be a problem, even for simple indicators as we had expected the scholar at least would know their number of publications and would question such an inflated number. Perhaps the ACUMEN portfolio will have to encourage scholars to use Google Citation or a similar system, or stipulate scholars have an ORCID id to be a part of the portfolio so that they can claim all their real publications and calculate impact indicators more easily. The data indicates that in our sample the more senior the scholar is, the more likely the publication list was missing from their web profile, fig. 2.

Fig. 2 Percentage "no publication lists" to seniority within discipline



Dead links are the second major cause of exclusion and are fairly evenly distributed across the disciplines. The internet is a dynamic resource with information being added and removed constantly and the dead links, in our sample, do not appear to be more prominent in one discipline over another, which would indicate disciplinary issues with site maintenance. It is though worth stressing that the sample we present here is a snapshot of the internet and a different sample could be produced if the collection process was repeated at a later date.

Fig. 3 Percentage dead links to seniority within discipline



Scholars appear to leave homepages or profiles incomplete when a new type of online profile tool becomes available or they move institutions. This has had a direct effect on our access to the publication lists of the scholars in our sample, especially senior scholars. In the short time since defining the sampling strategy and collecting the data links to publication lists have died, persons have moved institutes, been promoted and sites closed down or are under construction meaning that publication data could not collected and verified. This was especially noticeable in Public Health and Environment, whose scholars have a very active web presence often with 3 or more e-profiles available with varying degrees of currency on for example Linked In, blogs, Google Citations, Inspire, Scopus ID, PURE, CURIS, ORCID, Mendeley, Facebook, Microsoft Academic Search, Academia.eu, Impact Story, institutional homepages, project websites, etc., but this means that sites are neglected or expired when a new profile is created, and often under construction during our data collection window.

We observed that Astrophysicists enjoy using online dissemination tools the most and take enormous pride in personalizing homepages with all manner of interactive communication techniques, animations and outlinks to other interesting pages on the internet. This was however challenging in the data collection process as publication lists were “hidden” in solar systems or split up under different project pages or types of publication. The ACUMEN portfolio will have to encourage personalization to attract these scholars but also be simple enough so the information is easily findable by consumers of ACUMEN CVs. Further some astrophysicists, as well as environmental scientists and public health scholars, already include metrics on their CVs. The use varies from the very competent who contextualize the metrics in great detail to scholars who list the impact factor of the journals they publish in, please see the examples in appendix 3.
In ADS[7] ready-to-use metrics are available, as they are in the database Inspire[8], with little or no

[7] http://adsabs.harvard.edu/tools/metrics/
[8] http://inspirehep.net/author/G.Aad.1/

guidance to responsible use and interpretation of these statistics. The metrics are presented as a list of numbers leaving the interpretation open for the consumer. The inclusion of metrics on CVs in our sample indicates that scholars in three of our disciplines are interested in bibliometrics enriching their publication lists but this interest is noticeably absent in the fourth discipline, Philosophy. This will be the strength of the ACUMEN portfolio and how it differs from other resources that solicit CVs using bibliometrics. ACUMEN presents the scholar with metrics that are not only beneficial to the hard sciences, but relevant to the individual scholar, their seniority and their specialty, and gives the scholar tools to contextualize the metrics and present them to the consumer in a narrative that explains what the numbers mean and how the resulting "impact" has been interpreted.

The performance of WOS and Publish or Perish (POP) during data collection

The students collecting the data were asked to keep a log book of their experiences searching WOS and POP. Two students did this and their log books can be found in appendix 2. The notes are written in a mixture of Danish and English, and are copy/pasted without grammatical correction from the students' log books. The notes have however been anonymized and categorized into disciplines and seniorities. The main observations are reported in the next sections.

Publication lists

Publication lists are rarely complete and more often than not out of date. In the data collection our method was to search from the date of the first reported publication on the list to 2013, regardless if the publication list did not report publications up to this year.

Google Scholar includes publication types such as reports, comments and teaching materials that give a different publication/activity profile of the scholar than the profile in WOS which is limited to primarily to journal articles, reviews and conference papers. Scholars boost their publication lists or activity by including publications by colleagues in their project group while junior scholars' link to lists by their department peers to increase their visibility and show their network. These publications were not included in our publication data.

Not all the publications found in WOS have a UT number, which means there will be a slight discrepancy between the descriptive statistics based on the actual number of publications found in WOS and the bibliometric results based on the WOS UT numbers, such as P, CPP.

Name ambiguity

As expected, finding a scholar with a common name such as "Fan" or "Li" and identifying their real publications was in some cases impossible in POP, for example:

Author name: ”ab logan”
NOT ”ba logan” ”bb logan” ”bc logan” ”cb logan” ”db logan” ”bd logan” ”cb logan” ”bc logan” ”db logan” ”bd logan” ”eb logan” ”be logan” ”fb logan” ”bf logan” ”gb Logan” ”bg logan” ”hb logan” ”bh logan” ”ib logan” ”bi logan” ”jb logan” ”bj logan” ”kb Logan” ”bk logan” ”lb logan” ”bl logan” ”bm logan” ”mb logan” ”nb logan” ”bn logan” ”ob Logan” ”bo logan” ”pb logan” ”bp logan” ”qb logan” ”bq logan” ”rb logan” ”br logan” ”sb Logan” ”bs logan” ”tb logan” ”bt logan” ”ub logan” ”bu logan” ”vb logan” ”bv logan” ”bx Logan” ”xb logan” ”yb logan” ”by logan” ”zb logan” ”bz logan” "ahb logan" "elb logan" "lb Logan-fain"

It is not possible to limit to discipline and POP stops the search when the one thousand publications limit has been reached, eliminating what it considers to be less relevant publications than the ones returned. In terms of citations, these are usually articles with few (or no) citations. The omission may or may not be significant: most high-level citation metrics such as the h-Index and g-index are fairly robust and are unlikely to be affected. However, as we were looking for specific results, then these might be missing from the results list. It was not possible to search publications individually and group them to generate the bibliometric statistics. In these cases, POP ready to use bibliometrics are not useful as they do not reflect the true publication profile of the author and will give invalid information.

Homonyms are also a problem in POP, the students found that it was not uncommon for two or more authors to share the same surname and initial and be active within the same discipline. It was difficult to attribute the correct publications to the author. In POP tenacity and creativity is required to identify the scholar eg. the scholar Dvorak spells his name differently when publishing in English than when publishing in Hungarian was searched in POP:

"peter dvorak" or "petr dvorak" or "p dvorak" or "petra dvořáka" or " p Dvořáka"
Eksklude: "pa dvorak" "pj dvorak" "pf dvorak" "lp dvorak"
Likewise, scholars use a formal name for scientific articles and books “Samuel Clark” and an informal name on popular science documents, blogs, reviews, newspaper articles, etc, “Sam Clark”. This is an important distinction to be aware of when searching for publications on the internet.

In self-evaluation name ambiguity should not be a problem as scholars will know the alternative names they used on their publications however this must not be assumed as we have already reported in this paper scholars’ unquestioning acceptance of search results.

National language challenges
Researchers publish in their national languages which made it challenging to correctly couple the author to publications, especially in POP. In these cases the method was to firstly find the publications in WOS, as here the English language publications are prominently indexed, and use the abstrcts and indexing terms to understand the subject area. Using the researcher’s publication list as a master, the publications in GS were compared to the publication list, WOS list and key title words translated using google translate. In this way works with the same author name and not on the publication list, were identified and foreign language publications attributed correctly. This was a painstakingly slow process, but by doing so, non-english language publications were systematically collected and hence well-represented in the sample. We thus ensured that national language publications were not excluded due to our lack of knowledge of foreign languages.

Disciplines
Many scholars in the sample work with multi-disciplinary specialties and publish in a wide range of different formats and academic journals. Designing useful benchmarks for the scholars to contextualize their performance to will be challenging. For example statisiticians in Public Health publish in the traditions of the medical specialty they are working with, and surgeons publish, for example, at a very higher rate than practitioners of emergency medicine. The same trait is apparent

in Philosophy, where cosmic-philosophers publishing styles mimic Astrophysicists with a high amount of multi-author publications whereas philosophers of economics appear to single author papers and publish more books than their cosmic-philosopher fellows.

Recommendations

- Emphasize the importance of storing the online CV, publication list and online profile in one place and keeping it up-to-date. As a consumer it is difficult to gather a complete picture of the scholar when information is separated into personal homepages, institute homepages, pdfs and various profile tools.
- We cannot expect the researcher to sort through two or more citation indices and remove duplicate citations to get a complete citation record. We do however encourage the researcher to explore different indices to understand their coverage in them and be critical of what the ready to use metrics reported in these sources represent. The optimum would be if the scholar presented indicators on their ACUMEN CV, such as amount of citations per paper, h index, extracted from more than one database and present the range.
- Describe name ambiguity problems and how these affect the usefulness of citation indices and ready-to-use metrics. Ensure the scholar has room to write all the names he or she publishes or has published under. Name forms will make it easier for the consumer of the CVs to track activities and validate information. Research funders, research organisations, publishers, integrators etc. will find this useful.
- Require the scholar to have an ORCID id or Google Citation profile to ensure the scholar can easily claim his publications.
- Ensure easy import of publications into the portfolio. It will take effort to start an ACUMEN CV. The portfolio must support import of exisiting publicationslists in RIS, Bibtex, refman format, scopus ID, Google Mycitations, WOS, Mendeley and excel etc. Possible support in a “search and link” wizard? Search and link metadata on books, manuscript submissons, patents etc.
- Enable the researcher to set up an alert/search profile that can pull publications into the CV after the researcher has accepted the publication as theirs and not a duplicate.
- Develop guides to calculation and interpretation of metrics, both for the scholar AND the consumer.
- The portfolio must include a description of the problems with the representability of reference standards at the individual and specialty level. We must provide guidelines to how the scholar can establish local standards that reflect their specialty as a field, acknowledging their multi-disciplinary character.
- Personalisation of the ACUMEN CV will encourage use.
- Ensure that scholars can link to their peers ACUMEN CVs, like Linked In.
- A guide to how to present indices on the CV.

Next steps

Data-analysis will continue with a description and trend analysis of the simple statistics from POP and later a correlation analysis of the simple and sophisticated indicators from CWTS, based on the WOS data. These analyses will enable us to decide if the indicators we recommend for the ACUMEN

portfolio are a strong model of the disciplines and help us to identify which indicators are missing. Reference standards will be investigated as we are already aware of the difficulty the scholar will have in calculating useful peer comparisons. We will exemplify using performance standards supplied by CWTS that are based on a large level of aggregation and compare them with pseudo-h indices of the scholar's peers and percentile citations at the article level. Are these simple indices a useful predictor of impact within a community?

We will be looking at indicators and gender, academic posts and disciplinary representation. Perhaps the indicators and data we have identified are data-driven and not researcher-driven. What consequences will this have for the usefulness of the metrics in the portfolio?

Appendices

1. Composition of disciplinary sample before and after data-collection……11

2. Log book from data-collection……13

3. Excerpts of a CVs using bibliometrics……30

Appendix 1: Composition of disciplines before and after data-collection

Astronomy

Composition of discipline identified in sampling strategy

Astronomy	Phd	Post Doc	Assis. Prof	Assoc. Prof	Prof	*Total*
Dead link	2	16	1	6	13	*38*
Not Discipline				1		*1*
Duplicate			1			*1*
Not publication list	1	6	8	15	10	*40*
Not correct seniority						

Composition of discipline after data collection

Astronomy	Phd	Post Doc	Assis. Prof	Assoc. Prof	Prof	*Total*
Dead link	2	17	1	7	14	*41*
Not Discipline				1		*1*
Duplicate			1			*1*
Not publication list	1	6	9	18	12	*46*
Not correct seniority						

The set is reduced from 203 to 193 scholars, a reduction of 5%

Environment

Composition of discipline identified in sampling strategy

Environment	Phd	Post Doc	Assis. Prof	Assoc. Prof	Prof	*Total*
Dead link	2	6	7	25	11	*49*
Not Discipline		1			2	*3*
Duplicate						
Not publication list	2	4	15	25	15	*61*
Not correct seniority	1					*1*

Composition of discipline after data collection

Environment	Phd	Post Doc	Assis. Prof	Assoc. Prof	Prof	*Total*
Dead link	2	6	9	25	12	*54*
Not Discipline		1			3	*4*
Duplicate						
Not publication list	2	5	16	26	16	*65*
Not correct seniority	1					*1*
Impossible to find in POP					1	*1*

The set is reduced from 203 to 195 scholars, a reduction of 4%

Philosophy

Composition of discipline identified in sampling strategy

Philosophy	Phd	Post Doc	Assis. Prof	Assoc. Prof	Prof	*Total*
Dead link	2	5	9	12	17	*45*
Not Discipline		1	1	4	2	*8*
Duplicate	1	1		5	3	*10*
Not publication list	2	4	8	21	20	*55*
Not correct seniority						

Composition of discipline after data collection

Philosophy	Phd	Post Doc	Assis. Prof	Assoc. Prof	Prof	*Total*
Dead link	2	6	10	12	18	*49*
Not Discipline		1	3	10	8	*22*
Duplicate	1	1		6	3	*11*
Not publication list	2	4	8	21	20	*55*
Not correct seniority						

The set is reduced from 250 to 229 scholars, a reduction of 8%.

Public Health

Composition of discipline identified in sampling strategy

Public Health	Phd	Post Doc	Assis. Prof	Assoc. Prof	Prof	*Total*
Dead link	3	2	8	17	10	*40*
Not Discipline	2		1	1	2	*6*
Duplicate		1				*1*
Not publication list	3	4	9	26	16	*58*
Not correct seniority						

Composition of discipline after data collection

Public Health	Phd	Post Doc	Assis. Prof	Assoc. Prof	Prof	*Total*
Dead link	3	2	8	17	11	*41*
Not Discipline	2		1	1	2	*6*
Duplicate		1				*1*
Not publication list	3	4	9	29	16	*61*
Not correct seniority						

The set is reduced from 137 to 132 plus one scholar moved from environment to public health (n133), a reduction of 3%.

Appendix 2: Log book from the data-collection.

These are observations by the students collecting the publication data in Web of Science and Google Scholar, via Publish or Perish. The students were asked to note any problems or challenges they had collecting data in these two indices. They were also encouraged to write down their thoughts about the performance or "usefulness" of WOS and POP in searching for a scholar's publications.
The notes are written in a mixture of Danish and English, and are copied without grammatical correction from the students' log books. The notes have however been categorized into disciplines and seniorities.

Astronomy & Astrophysics

phd-students

Forfatteren akos kereszturi har udgivet artikler siden 1994, hvilket kunne indikere at han måske ikke er Post.doc. Kereszturi har 35 publikationer på sin publikationsliste, men noterer også en del Populærvidenskabelig formidling, der formodentlig vil dukke op i gs. Han går meget op i bred formidling af Fysik, hvilket kan forklare det høje antal af publikationer i gs - måske er 294 dog lige lovligt højt. De er alle Inden for astrofysik og jeg åbnede de dokumenter, jeg var i tvivl om og de var af akos kerezturi.

Michael weidinger kan være et problem i gs, da der er en anden fysiker ved navn matthias weidinger, der Udgiver fra university of Wurzburg, der også udgiver inden for astronomi og astrofysik. Det bliver svært at Skelne de to fra hinanden i gs.

Gs: "erik bartoš" ville udelukke "me bartoš" men det viste sig at være gs, der havde taget hans titel med Som fornavn, altså var det ham.

Assistant professsors

Mange dubletter i publish or perish

Daphne weihs, no. 77, er biomediciner og arbejder ikke med astronomi eller lignende.

Msg_max_results
“Warning: results limit reached.
The query returned <n> results, which is the maximum that google scholar Allows. This may affect the query coverage. Click help for more information.
Indicates that your query returned the maximum number of results that google scholar Allows (1000; sometimes a few less). Your query may have more matches, but the remainder Are not available. As a result, some potential matches may be omitted from the list of results. Generally speaking, the missing results are deemed by google scholar to be less relevant than the ones that were returned. In terms of citations, these are usually articles with few (or No) citations.
The omission may or may not be significant: most high-level citation metrics such as the h-Index and g-index are fairly robust and are unlikely to be affected. However, if you are looking for one or more specific results, then these might be missing from the results list.

Professors

Professor li
Google scholar search: stopped after 1000 posts retreived, the search is not representative of his Work. Search query:
"cheng li" from 2001 to 2013: all
Query date: 2013-06-27
Papers: 47
Citations: 3491
Years: 13

Professor varga
Google scholar search: stopped after 1000 posts retreived, the search is not representative of his Work. Search query:
"p varga" from 1966 to 2013: all
Query date: 2013-06-27
Papers: 1000
Citations: 16203
Years: 48
The search “peter varga” resulted in 34 posts, mostly hungarian, but they all belong to our professor. Hungarian posts verified by title opslag in google translate/and on his cv (which is out of date)

Environmental Science, studies & engineering

Assistant professors

255 Freni, g
Fik firdoblet sine publikaitoner i gs. Udover at det skyldes ikke-engelsk sproget litteratur var Der også en del praksis-orienteret materiale (rapporter osv.)

Associate professors

280 rajta i
Ikke inden for environmental, udgiver inden for fysik.

281 gendel y
Linker til en anden persons cv. Hans cv er ikke til at finde på siden, men ved at google kommer det frem at Han er ph.d studerende og den persons cv, han linker til, er hans vejleder. Fik først sin ph.d i 2011 og er Derfor tvivlsomt assoc_prof.
Http://www.google.dk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=10&ved=0cgmqfjaj&url=http%3a%2f%2fwww.neaman.org.il%2fneaman2011%2fuserdata%2fsendfile.asp%3fdbid%3d1%26lngid%3d2%26gid%3d2344&ei=9zjpudpmoczzsgakk4dicg&usg=afqjcngnfabqimgdoss6mrglx0rguvtjiq&bvm=bv.48572450,d.yms

Philosophy and the History & Philosophy of Science

Post doctoral students

26/6 gramelsberger, gabriele (417)
Ingen navnesammenfald eller anden støj

26/6 lessmann, ortrud (418)
Navnesammenfald

Lessmann, olivier
Umiddelbart let at adskille, da deres fagområde var meget forskelligt

26/6 novotny, daniel d. (419)
Navnesammenfald
Novotny, duan
Novotny, david
Søgning på ”novotny dd” fik sorteret det meste af støjen fra.

26/6 dicken, paul (420)
Navnesammenfald
Dicken, peter
Forskelle i fagområde gjorde adskillese let

26/6 malmqvist, erik (421)
Navnesammenfald
Malmqvist, ebba
Fagområderne var meget tæt på hinanden men hun var klart praktiker, hvor han er meget teoretisk Orienteret. Dette lettede sorteringen en del.

26/6 frega, roberto (422)
Navnesammenfald
Frega, romeo
Fagområde var forskelligt, så det var let at sortere

26/6 marvan, tomas (423)
Ingen navnesammenfald eller anden støj

26/6 eronen, markus (424)
Op til flere navnesammenfald, men ved at søge på ”eronen mi” kom kun relevante dokumenter Frem, der kan godt være nogen der ikke er kommet med, men dem jeg fandt var højt relevante.

26/6 gerken, mikkel (425) (impossible)
Mange navnesammenfald indenfor mange forskellige fagområder, oprydningsarbejdet især i gs
Viste sig meget tidskrævende

26/6 herran, néstor (426)
Enkelte navnesammefald, men da han har et meget snævert fokus for sit fagområde var det let at Sortere.

26/6 shultziner, doron (427)
Ingen navnesammenfald eller anden støj.

26/6 hennig, boris (428)
Masser af navnesammenfald og enkelte fagområdesammenfald, især i gs vil det måske blive Nødvendigt at tjekke resultaterne efter, da nogle af dem jeg bedømte som relevante godt kan have Været af en navnefælle.

26/6 backman, jussi (429) (impossible)
En meget høj grad af navnesammenfald også på eget universitet, fagområdesammenfald er ikke så Udtalt, men mængden af støj fra navnefæller gør det til et kæmpearbejde at sortere i det.

26/6 roinila, markku (430)
Et navnesammenfald med en amerikansk forsker der skrev om det finsk-svenske Immigrationsmindretal i nordamerika. Let at skille fra hinanden.

26/6 milne, richard (431) (impossible)
Høj frekvens af navnesammenfald også i beslægtede fagområder.

26/6 buczek, pawel (432)
Navnesammenfald
Buczek, piotr
Fagområder er forskellige nok til at kunne sortere

26/6 vagelpohl, uwe (433)
Umiddelbart ingen navnesammenfald eller anden støj

26/6 pieters, wolter (434)
Et navnesammenfald indenfor nært beslægtet fagområde
”Pieters, willem” Sortering lidt besværlig i gs da jeg ikke forstår hollandsk, men det gik forholdsvis smertefrit

26/6 lönnqvist, jan-erik (435)
Navnesammenfald med en kemiker

26/6 stokes, patrick (436)
Masser af navnesammenfald, ville måske være værd at gennemse igen

26/6 evers, daan (437)
Meget høj frekvens af navnesammenfald, svært at indkredse i gs da han blev frasorteret i Forbindelse med at jeg prøvede at udelukke diverse ekstra initialer.
Burde eventuelt gennemgås igen

26/6 sanchez leon, alberto (438)
Få navnesammenfald, men dem der var lå også tæt på i fagområde, især i gs var det svært at Afkode hvilke dokumenter der hørte til.
Burde eventuelt gennemgås igen

Assistant professors

Dvorak 454 (impossible)
Gs:
Forfatternavne: "peter dvorak" or "petr dvorak" or "p dvorak" or "petra dvořáka" or " p Dvořáka"
Ekskludering: "pa dvorak" "pj dvorak" "pf dvorak" "lp dvorak"
Fandt petre dvoraka på forfatterens egen side hvor jeg gik et skridt tilbage fra den engelske Side. Der er 940 poster efter ovenstående søgning. Dvorak kan åbenbart staves på mange måder, og umiddelbart ud fra hvad jeg har kunne se Kan peter dvoraks navn også staves på flere måder, så hvorledes jeg ellers kunne ekskludere Ved jeg ikke.

Roy 455
Gs
Forfatternavne: "oliver roy" or "o roy"
Ekskludering: "oc roy" "jo roy" "ofa roy" "mo roy" "op roy" "po roy"
312 poster
Rangerede efter publication og gennemgik Dem som var relevante stod oftest sammen med andre relevante pga. Publikationen.

Ridge 458 (impossible)
Gs

Forfatternavne: "steve ridge" or "s ridge"
Ekskludering: "sgm ridge" "sa ridge" "se ridge" "sgk ridge"
Ingen relevante resultater – kan det passe?

Simon 461 (impossible)
Gs
Forfatternavne: "fabrizio simon" or "f simon"
Ekskludering: "af simon" "fb simon" "fa simon" "fjg simón" "fx simon" "fg simon" "fr simon" "mf simon" "jf simon" "fjg simon" "lf simon" "df simon" "hf simon" "fp simon" "bf simon" "fm simon" "f simon-ritz" "f simon nieto"
Tilsyneladende er der mange der hedder f simon, så der kom over 1000 poster selvom jeg Ekskluderede en del efternavne. Så den er impossible.

Wilkinson 464 (impossible)
Wos
Author=(angus j wilkinson) or author=(wilkinson aj) or author=(wilkinson a)
Refined by: authors=(wilkinson a or wilkinson aj) and [excluding] Web of science categories=(biochemistry molecular biology or Health policy services or surgery or medicine research Experimental or transplantation or psychology or Infectious diseases or pathology or microbiology or Pediatrics or biochemical research methods or immunology Or cell biology or medicine general internal or genetics Heredity or hematology or physiology or zoology or Psychology multidisciplinary or nursing or behavioral Sciences or tropical medicine or psychology experimental Or psychiatry or psychology biological or clinical Neurology) and research areas=(engineering or materials Science or physics or metallurgy metallurgical Engineering) and authors=(wilkinson aj)

Timespan=1991-2013. Databases=sci-expanded, cpci-s.

Gs Author name: "angus j wilkinson" or "aj wilkinson"
561 poster

Schäfer 471
Wos Refinede med de universiteter han har arbejdet ved – gav 12 poster ud af de originale 151. Spørgsmålet er om der er noget materiale som ikke står registreret under universitetet som Er skrevet af schäfer

Author=(mike s schaefer) or author=(schaefer ms) or author=(schaefer m)
Refined by: organizations-enhanced=(free university of berlin or University of hamburg) Timespan=2002-2013. Databases=ssci, a&hci, cpci-ssh.

Clark 472

Han hedder samuel clark men der står på hans egen side at han hedder sam clark. Fandt først hans Udgivelser efter kun at søge på samuel og ikke sam. Prøvede at søge på hans andre artikler på title i Wos, men fandt ingenting. Så han er kun katalogiseret som samuel, ihvertfald i wos.
Gs
"samuel clark" or "sam clark" "sj clark" "sl clark" "sr clark" "sa clark" "se clark" "st clark" "js clark"
Fandt 4

Moreno munõz 476
Wos
Ved brug af munoz i søgningen fandt jeg ingenting på hans navn. Ved brug af kun moreno Kom der over tusind poster, men ved at kigge i categories var der intet der havde med hans Område at gøre. Så jeg skrev 0 resultater

Associate professors

Chapman 489 (impossible)
Gs
"siobhan chapman" or "s chapman" NOT "sc chapman" "cs chapman" "rs chapman" "ds chapman" "sj chapman" "ms chapman" "ds Chapman" "ls chapman" "sw chapman" "fs chapman" "sk chapman" "sb chapman" "as Chapman" "ks chapman" "bs chapman" "st chapman" "st chapman" "ss chapman" "ls Chapman" "sr chapman" "sg chapman" "es chapman" "sp chapman" "js chapman" "ps Chapman" "ns chapman" "sd chapman" "sg chapman" "st chapman"
1993-2013
Over 1000 poster

Gonzales 498 (impossible)
Wos
Hun linker selv til en researcherid.com side, hvor hun har 91 udgivelser. Når jeg taster Hendes author id nummer ind i wos får jeg kun 10 poster. Ved søgning på hendes navn Dukker der langt flere frem, men ved afgrænsning i hvilken organisation det kommer fra (university of navarra) kommer der 11 frem. 2 af dem er nye hvor en af dem er en af Hendes. Hvor den sidste er henne er et godt spørgsmål. Men jeg får altså kun 11 resultater.

Gs

"ana marta gonzalez" or "am gonzalez"
NOT "am gonzales-angulo" "am gonzales-paramas" "am gonzales-vadillo" "am gonzalez-Rodriguez" "jm alvarez-suarez" "am gonzalez soca" "am gonzalez gonzalez" "jm alvarez-Suarez" "am gonzalez-angulo" "am gonzalez-cameno"
Stadigvæk over 1000 poster.

Obrien 499 (impossible?)
Hverken i gs eller wos fandt jeg nogle poster.

Christensen 505
Wos
Author=(anne-marie soendergaard christensen) or author=(anne-marie sondergaard Christensen) or author=(anne marie soendergaard christensen) or author=(anne Marie sondergaard christensen) or author=(christensen ans) or Author=(christensen as)
Timespan=2006-2013. Databases=sci-expanded, ssci, a&hci, cpci-s, Cpci-ssh.
Ingen poster
Gs
Kun 6 poster

Kuna 529 (impossible)
Gs
Der kom error 13 ved min søgning. De resultater der kom frem var ikke relevante.

Logan 535 (impossible
Gs
Afgrænsning: ”ab logan” NOT ”ba logan” ”bb logan” ”bc logan” ”cb logan” ”db logan” ”bd logan” ”cb logan” ”bc logan” ”db logan” ”bd logan” ”eb logan” ”be logan” ”fb logan” ”bf logan” ”gb Logan” ”bg logan” ”hb logan” ”bh logan” ”ib logan” ”bi logan” ”jb logan” ”bj logan” ”kb Logan” ”bk logan” ”lb logan” ”bl logan” ”bm logan” ”mb logan” ”nb logan” ”bn logan” ”ob Logan” ”bo logan” ”pb logan” ”bp logan” ”qb logan” ”bq logan” ”rb logan” ”br logan” ”sb Logan” ”bs logan” ”tb logan” ”bt logan” ”ub logan” ”bu logan” ”vb logan” ”bv logan” ”bx Logan” ”xb logan” ”yb logan” ”by logan” ”zb logan” ”bz logan” "ahb logan" "elb logan" "lb Logan-fain"
Der kommer stadig over 1000 poster. Når jeg afgrænser kommer de alligevel frem. Så jeg Kan ikke se hvad jeg kan gøre anderledes.

Professor

27/6 borgato, maria teresa (570)
Ingen navnesammenfald eller anden støj

29/6 osborne, catherine (571) (impossible)
Navnesammenfald indenfor samme fagområde, især et problem i gs, da jeg kom i tvivl om jeg Markerede den rigtge forfatter eller ej.

29/6 klein-braslavy, sara (572)
Ingen navnesammenfald eller anden støj

29/6 lam, alice (573) (impossible)
Navnesammenfald også indenfor beslægtede fagområder

29/6 lorch, marjorie perlman (574)
Enkelte navnesammenfald, men adskillese af fagområder og hendes fokus på et meget snævert Emne gjorde det let at sortere.

29/6 galavotti, maria carla (575)
Umiddelbart ingen navnesammenfald eller anden støj

29/6 enslin, penny (576)
Meget få navnesammenfald hovedsageligt i gs, ingen umiddelbare fagområdeoverlap

29/6 unterhalter, elaine (577)
Ingen navnesammenfald eller anden støj

29/6 galeotti, anna elisabetta (578)
Ingen umiddelbare navensammenfald eller anden støj

29/6 griffiths, morwenna (579) (impossible)
Navnesammenfald indenfor nært beslægtede fagområder
29/6 frewer, lynn j (580)
Navnesammenfald udenfor fagområde
Frewer, lorna
Skrev om fredsbevarende styrker og militær udstationering
Utroligt mange resultater i især gs, kan måske skyldes dubletter

29/6 chemla, karine (581)
Ingen navnesammenfald eller anden støj

30/6 verbrugge, rineke (582)
Enkelte navnesammenfald, forholdsvis let at sortere da der ikke var nært beslægtede fagområder

30/6 garcia-encinas, maria jose (583)
Ingen navnesammenfald eller anden støj

30/6 campos boralevi, lea (584)
Ingen navnesammenfald eller anden støj

30/6 fernandez, angel nepomuceno (585)
Navnesammenfald med beslægtet fagområde, det var dog stadig muligt at sortere dem fra Hinanden.

1/7 chaline, jean (586)
Der var et meget stort antal af ekstra poster i gs, om det er dubletter eller fordi der er flere Indenfor samme felt er jeg ikke helt sikker på, men jeg inkluderede alle der holdt sig indenfor Emnet.

1/7 malo, antinio (587) (impossible)
Der var umiddelbart for mange navnesammenfald til at kunne lave en meningsfyldt sortering uden At bruge mange timer på det.

1/7 d'agostino, marcello (588) (impossible)
Mange navnesammenfald, men ikke i nært beslægtede fagområder
Gs var umiddelbart et utroligt stort sorteringsarbejde

1/7 buzzoni, marco (589) (impossible)
Linket til hans egen litteraturliste var dødt

Public health and Public Health Policy

Assistant professors

17/6 bode, christina (703)
Havde mange navnesammenfald i både wos og gs indenfor beslægtede felter
Bode, christoph
Bode, carole
Løsning Wos: Søge på fuldt fornavn og se hvilke categories der var tilknyttet søgeresultatet, Derefter bruge dem Udelukke institutioner og universiteter som forskeren ikke er eller har været Tilknyttet (organizations, enhanced -> exclude i more options) Gennemgå titler for at se om de stemmer overens med forskningsspecialisering.
Gs
Søge på fuldt eller delvis fornavn, Ekskludere initialer per vejledning
Fejlkilder
Har måske ekskluderet dokumenter hvor hun står med kun første initial (bode, c)
Wos
Har måske ekskluderet conference dokumenter ved at ekskludere bestemte Organisationer

18/6 booth, alison (704) (impossible)
Hun har selv andet initial m. Fremgår ikke af hendes universitetshjemmeside
Booth, am ifølge wos Mange navnesammenfald i både wos og gs
Booth, andy m.
Booth, alexander
Booth, al
Booth, ao
Løsning Wos: Søg på fulde fornavn plus initial: booth alison m
Gs
Det var umuligt umiddelbart at få et brugbart resultat.

18/6 williams, john r (705) (impossible)
Der findes så mange john r. Williams at det var umuligt at lave en søgning der umiddelbart gav Gode resultater.

18/6 huhtala, heini (706)
Ud fra stikprøver fandt jeg ingen navnesammenfald og stikprøver viste også samme lokalitet. Wos categories for datasættet ligger alle sammen indenfor medicinske eller beslægtede kategorier
Gs data var for omfangsrig til mere end en overfladisk gennemgang, det ser dog ud til ligesom i wos At falde indenfor det medicinske felt eller beslægtede felter.

18/6 gardner, benjamin (707) (impossible)
Ved at begrænse på både organizations-expanded og wos categories kom jeg frem til de resultater Der er i regnearket i forhold til wos.
I gs var det noget nær umuligt at begrænse søgningen således at man ramte den rigtige forfatter. Jeg har inkluderet de resultater jeg kom frem til men en større oprydning er nødvendig er min Bedømmelse.

18/6 spilková, jana (708)
Navnesammenfald
Spilkova, jirina Ansat ved samme universitet og har udgivet i nogenlunde samme periode. Har i både gs og wos sorteret ud fra at de skrev om forskellige fagområder

18/6 andreucetti, daniele (709)
Ingen navnesammenfald eller andre problemer i hverken wos eller gs.
I gs var der en del titler på italiensk, men ud fra hvad jeg kunne dechiffrere, så var de alle relevante.

18/6 van solinge, hanna (710)
Ingen problemer med navnesammenfald eller lignende
I gs var der to artikler på spansk. Umiddelbart kunne jeg med mine spanskkundskaber ikke Bedømme deres relevans, men det virkede til at den ene ihvertfald havde noget med familier og Gamle at gøre, de er derfor ikke blevet udeladt fra datasættet.

Associate professors

19/6 hakkaart-van roijen, leona (711)
Ingen problemer med afgrænsninger i hverken gs eller wos

19/6 baron-epel, orna (712)
Ingen navnesammenfald eller anden åbenlys støj i hverken gs eller wos

19/6 johnsen, søren p. (713)
Umiddelbart ingen navnesammenfald i wos
Gs
Dokumenterne virkede umiddelbart relevante pånær en enkelt post der var skrevet i Kyrillisk, jeg kunne ikke bedømme indholdet, men den er inkluderet i datasættet.

19/6 reis, shmuel (714) (impossible)

Ufatteligt mange navnesammenfald.
Både reis, s****. Mange forskellige fornavne til afternavnet reis.
Wos medtog også forfattere med sammensatte navne af typen reis-s****. F.eks. Reis-silva.

19/6 jensen, jesper ole (715) (impossible)
Mange navnesammefald. Prøvede at afgrænse i wos med ”countries/territories” og valgte ”denmark”. Forsker på dtu med navnet jensen, jens oluf dominerede stadig listen. I gs er der alt for meget støj til at få et meningsfyldt resultat umiddelbart.

19/6 nielsen, claus vinther (716)
Navnesammenfald med forskere indenfor andre felter. Andre forskere var indenfor videnskabelige felter der var markant anderledes

19/6 toft, gunnar (717)
Ingen problemer med fremfinding, ingen navnesammenfald.

19/6 hesse, morten (718)
Mange navnesammenfald
Wos: En begrænsning til ”countries/territories” hvor jeg valgte ”denmark” gav kun Artikler af morten hesse så vidt som jeg kunne bedømme
Gs: Blev nødt til at begrænse søgningen til ”hesse morten” da at medtage ”hesse m”
Gav over 1000 hits.

20/6 ramlau-hansen, cecilia (719)
Ingen navnesammenfald eller andre problemer

20/6 støvring, henrik/stovring, henrik (720)
Ingen navnesammenfald eller andet støj

20/6 muth, christiane (721) (impossible)
Har ingen egentlig egen publikationsliste, det var nødvendigt at søge på hvor mange af hendes Instituts udgivelser hun var (med)forfatter på.
Har i ”muth christiane_721_mabr.pdf” markeret navnet ”muth” da listen indbefatter 734 Hvoraf hun kun optræder på 53 af dem.
I wos begrænsedes søgningen til kun at indbefatte det universitet hun er tilknyttet I gs var det umuligt at få et brugbart resultat da der var navnesammenfald indenfor både Ubeslægtede og beslægtede forskningområder.

20/6 hougaard, karen sørig (722)

Ingen navnesammenfald eller andet støj
20/6 vehtari, aki (723)
Ingen navnesammenfald eller anden støj

20/6 kabai, péter (724)
Ingen navnesammenfald eller anden støj

20/6 bødker, réne (725)
Ingen navnesammenfald eller anden støj

20/6 ansel, pat (726) (impossible)
Navnesammenfald og forskningsområdesammenfald
Ansell, peter

20/6 chin a paw, mai (727)
Ingen navnesammefald eller anden støj. Det var dog nødvendigt at søge på både ”chin a paw, m” og ”chinapaw, m” da hun optræder under Begge navne.

20/6 de bruyne, martine (728) (impossible)

Professors

U vogel 774 (impossible)
Publikationsliste er samling af to forfatteres.
Gs: Afgrænsning: "uf vogel" and "ub vogel" and "ur vogel" Gav 243 resultater
Wos:
Au=(vogel u*) and (sh=(physical sciences or life sciences biomedicine) or Wc=(multidisciplinary sciences))
Refined by: authors=(vogel u) and organizations-enhanced=(university of Wurzburg or natl reference ctr meningococci or hannover medical School) and [excluding] publication years=(1989 or 1990)
Timespan=all years. Databases=sci-expanded, a&hci, ssci, cpci-ssh, cpci-s.

Mj prince 775
Gs
Frasortering ved at kigge dem alle sammen igennem. Alt det der har med disability og canada har jeg frasorteret

A katalinic 776
Wos:
Au=("katalinic a") and (sh=(physical sciences or social sciences or life

Sciences biomedicine) or wc=(social sciences, interdisciplinary or multidisciplinary Sciences))
Refined by: [excluding] web of science categories=(dentistry oral surgery Medicine or food science technology or computer science artificial Intelligence or telecommunications or computer science information Systems)
Timespan=all years. Databases=sci-expanded, a&hci, ssci, cpci-ssh, cpci-s.

Ad grant 777 (impossible)
Wos:
846 resultater før refining med organizations-enhanced=(london school of hygiene tropical medicine). Efter 126. Men om hun har arbejdet andre steder ved jeg ikke.
Author=(grant ad) or author=(grant a)
Refined by: [excluding] web of science categories=(physics particles fields or Computer science theory methods or environmental sciences or Engineering electrical electronic or astronomy astrophysics or food Science technology or history or nuclear science technology or Instruments instrumentation or telecommunications or marine Freshwater biology or computer science information systems or Agriculture dairy animal science or economics or education Scientific disciplines or fisheries or engineering environmental or Oceanography or business or meteorology atmospheric sciences or Computer science interdisciplinary applications or veterinary Sciences or imaging science photographic technology or dentistry Oral surgery medicine or political science or substance abuse or Zoology or engineering civil or literature british isles or sport Sciences or linguistics or chemistry applied or management or Language linguistics or materials science multidisciplinary) and Authors=(grant a or grant ad) and organizations-enhanced=(london school of Hygiene tropical medicine)
Timespan=1990-2013. Databases=sci-expanded, ssci, a&hci, cpci-s, cpci-ssh.

H montgomery 784 (impossible)
Gs
Problemer med eksludering af forkerte forfatternavne
"he montgomery" "jh montgomery" "hl montgomery" "gh montgomery" "hdb montgomery" "he
Montgomery-downs" "rh montgomery jr" "wh montgomery" "ah montgomer" "dh montgomery"
"hj montgomery" "mh montgomer" "ah montgomery" "rh montgomery" "ch montgomery" "sh

Montgomery" "mh montgomery" "h montgomery-massingberd" "jl montgomery"
Nogle af disse endte op alligevel på listen
Fs violante 786
Har kun artikler fra 2004 til 2008 i sin publikationsliste. Tog alle andre år med også, da det er usandsynligt at
Han i løbet af de år er blevet professor. Derfor søgte jeg på alle år i wos og

Appendix 3: Excerpts of a CVs using bibliometrics

Excerpt 1: from Public Health

PUBBLICAZIONI del Prof. Giuseppe Verlato

1) LAVORI SU RIVISTE INDICIZZATE SUI CURRENT CONTENTS
Impact Factor (I.F.) complessivo = 129.071

1982-1993 (I.F. complessivo = 8.378)

1.1) Cevese A, **Verlato G** (1985) Haemodynamic effects of withdrawal of efferent cervical vagal stimulation on anaesthetized dogs - Relative importance of chronotropic and non-chronotropic mechanisms. *Journal of the Autonomic Nervous System*, 14: 125-136 (I.F. 92 = 0.953)

1.2) Borgdorff P, **Verlato G**, Cevese A (1987) Cardiac alpha-1 adrenoceptors are not involved in heart rate control of the anaesthetized dog. *Pflugers Archiv European Journal of Physiology*, 410: 495-500 (I.F. 92=3.115)

1.3) Cevese A, **Verlato G**, Cerutti G (1989) 'Non-chronotropic' mechanisms on withdrawal of efferent vagal stimulation in anesthetized dogs. *Journal of the Autonomic Nervous System*, 28: 155-166 (I.F. 92=0.953)

1.4) Braggion C, Cornacchia M, Miano A, Schena F, **Verlato G**, Mastella G (1989) Exercise tolerance and effects of training in young patients with cystic fibrosis and mild airway obstruction. *Pediatric Pulmonology*, 7: 145-152 (I.F. 92=1.395)

1.5) **Verlato G**, Borgdorff P (1990) Endogenous adenosine enhances vagal negative chronotropic effect during hypoxia in the anaesthetised rabbit. *Cardiovascular Research*, 24: 532-539 (I.F. 92=1.476)

1.6) Colletti V, Fiorino FG, **Verlato G**, Montresor GC (1991) Reduced active protection to the cochlea during physical exercise. *Acta Otolaryngologica* (Stockholm), 111: 234-239 (I.F. 92=0.486)

1994 (I.F. complessivo = 2.759)

1.7) **Verlato G**, Poltronieri R. (1994) Usefulness and limitations of the positive protosystolic peak of coronary arterial blood flow as an index of epicardial arterial compliance. *Cardioscience*, 5: 87-94 (I.F. 94=0.648)

1.8) de Marco R, **Verlato G**, Zanolin E, Bugiani M, Drane JW (1994) Nonresponse bias in EC Respiratory Health Survey in Italy. *European Respiratory Journal*, 7: 2139-2145 (I.F. 94=2.111)

Bibliometric indicators

CWTS - Leiden:

The impact ratio used by the Centre for Science and Technology Studies in Leiden is defined as the number of citations to a person's papers over a certain period divided by the number of citations expected on the basis of a comparison set (average number of citations of all papers from the same years in either the same journals or in all journals in the field).
The CWTS works with windows as follows. E.g. 1980 – 1995 means publications that appeared between start of 1980 and the end of 1995, and citations to those until the end of 1995. The field is 'astronomy and astrophysics'.

1980 – 1995: My impact ratio w.r.t. journals is **3.28**; w.r.t. the field it is **2.89**.

1991 – 2001: My impact ratio w.r.t. journals is **2.64**; w.r.t. the field it is **1.80**.

1994 – 2003: My impact ratio w.r.t. journals is **2.67**; w.r.t. the field it is **2.04**.

Henk Spruit's "curve of growth" method.

For a large set of astronomers their total number of citations for first-author papers is listed as a function of years since their Ph.D. This follows a curve with a universal shape, which is then fitted for each individual with an amplitude scaling factor α.

Up to 1998: My $\alpha = \mathbf{3.07 \pm 0.19}$.

Dave Burstein's compilation.

Burstein has made a compilation of paper and citation counts of 4617 astronomers.

1981.0 – 1997.5: The compilations lists me with an average of **47.16** citations per paper. This is **1.75** times the average value and **2.58** times the median. Of the sample, **90.3%** of astronomers have fewer citations per paper then I do.

NASA ADS:

These are impact ratio's analogous to those of the CWTS derived by myself with the NASA Astronomy Database System.

1970 – 2002: My impact ratio w.r.t. journals **2.54**; w.r.t. the field (the major astronomical journals) it is **2.28**.

In 2004 I performed a comparative study of bibliographic indices of astronomers at institutes in the Netherlands, using the NASA Astrophysics Data System (ADS). The results of that exercise are available through my homepage or directly at http://www.intra.astro.rug.nl/~vdkruit/jea3/homepage/ads.pdf.

In the table below my own scores are compared to the distribution among the 98 active astronomical researchers in the Netherlands. The data are as of summer 2004 and concern only papers in refereed journals according to the definition of ADS. In "normalised" scores the number publications or of citations to these have been divided by the number of authors of the publications involved.

Publication scores **(refereed papers only!)** up to 2004 for myself and the distribution among Dutch astronomers; for comparison my current values (January 2010).

parameter		median	quartile	90%-tile	current
number of papers	74	60	102	175	86
citations to papers	3894	2051	3203	5196	5245
citations per paper	**52.6**	24.7	33.1	43.8	70.0
year of first publication	1970	1980	1972	1966	1970
papers per year	2.2	2.8	4.8	6.8	2.2
citations to papers per year	114.5	76.0	126.4	178.9	131.1
expected number of citations per paper	24.4	23.9	24.5	24.6	31.5
impact ratio	**2.16**	1.11	1.45	2.08	2.22
h-index	**39**	25	32	38	40
normalised number of papers	50	22	33	52	54
normalised citations to papers	**2454**	420	1025	1636	3192
normalised citations per normalised paper	**49.1**	22.0	30.8	49.1	59.1
normalised number of papers per year	1.5	1.0	1.4	2.0	1.4
normalised citations to papers per year	**72.2**	19.7	36.0	56.9	79.9
average number of authors on papers	**1.5**	3.3	2.4	2.0	1.6
normalised h-index	**32**	12	18	22	33

The **bold** parameters indicate where I score within the 90%-tile.

In terms of number of papers I do certainly better than the median person, and in terms of number of citations I am in the upper quartile. The number of papers per year and number of citations per year are heavily influenced by the fact that I have mostly published papers with small numbers of authors. In addition, up till recently I have generally not been co-author on papers of my Ph.D. students. As a result, my average is 1.5 authors per paper[1] and in that distribution (which has a median of 3.3) I am well into the (lower) ninety-percentile. This has been my style of doing research; I have been most comfortable working in small collaborations, often with a single colleague or graduate student.

In terms of normalised properties my scores in numbers of publications or citations (both absolute and per year) are excellent. My score in the number of normalised citations and those per paper or per year are actually above the ninety-percentile.

My number of citations per paper is very high (well above the ninety-percentile). Also the impact ratio (actual number of citations per paper relative to the expected number based on the length of one's career) is over 2 and above the ninety-percentile of the distribution of that parameter.

More recently the use of the h-index (the number of papers that have recieved more citations than their rank in a list sorted according to number of citations; if the h-index is n then the person –or group– has n papers that have received n citations or more) is replacing the impact ratio. The disadvantage of the index is that it grows with the lengthd of ones career and it does not allow for the number of authors (many papers with many authors gives more chance for a high h-index).

In connection with the review of the Netherlands Research School for Astronomy (NOVA) Peter Kamphuis and I produced a comparative study of citations for a number of samples of as-

[1]This has been estimated by dividing the number of papers by the number of normalised papers.

tronomers in February 2010 (see www.astro.rug.nl/~vdkruit/jea3/homepage/NOVA_ADS.pdf). It compared various measures of citation scores for staff members of NOVA research institutes at the Duitch universities, faculty of the 10 top-institutes in the U.S.A., a random sample of astronomers from the AAS (American Astronomical Society) membership list and one from the IAU (International Astronomical Union) membership list, and the 'key-researchers' of NOVA (leading astronomers in the Netherlands). We also derived the h-index for first author papers and normalised citations (normalised means the number of publications or citations divided by the number of authors on each paper).

In the table I reproduce these results together with my individual scores. These data have been derived from ADS in January 2010.

parameter	NOVA	TopUS	AAS	IAU	NOVA-kr	PCK
number of people in sample	79	177	172	193	26	1
number of papers	90	94	26	58	123	124
number of first-author papers	21	22	8	19	23	62
normalised number of papers	26	25	7	20	33	73
number of normalised first-author papers	11	10	3	9	13	50
number of citations	3325	4175	544	1042	4558	5255
number of first-author citation	795	971	112	256	1166	3618
number of normalised citations	704	929	86	271	1213	2571
h-index	31	34	12	17	39	40
first-author h-index	13	14	4	8	14	30
normalised h-index	14	16	5	9	16	33
first-author normalised h-index	9	10	3	6	10	27
citations per paper	36	43	21	18	38	42
citations per first-author paper	39	39	13	15	45	58
normalised citations per normalised paper	33	35	17	14	37	35
papers per year	3.9	3.3	1.6	1.9	5.6	3.1
citations per year	131	141	34	34	229	131
normalised papers per year	1.1	0.9	0.4	0.7	1.5	1.8
normalised citations per year	34	33	6	10	57	64
first-author papers per year	0.9	0.9	0.5	0.6	1.2	1.6
first-author citations per year	40	33	8	9	59	90
number of publishing years	25	30	18	30	22	40

I compare very favorably with astronomers in top-US institutes and with the best colleagues in the Netherlands (NOVA key-researchers), especially when the papers are normalised by the number of authors or have been restricted to first-author papers.

ACUMEN
academic careers understood through measurement and norms

FP7 Grant Agreement	266632
Deliverable No and Title	D5.8 Part 4 – Consequences of Indicators. Effects on the users.
Dissemination Level	PU (public)
Work Package	WP5 Bibliometric Indicators
Version	1.0
Release Date	28-04-2013
Author(s)	Lorna Wildgaard Birger Larsen Jesper Schneider
Project Website	http://research-acumen.eu/

European Commission
7th Framework Programme
SP4 - Capacities
Science in Society 2010
Grant Agreement: 266632

EUROPEAN COMMISSION

SEVENTH FRAMEWORK PROGRAMME

Guidelines for using bibliometrics in the ACUMEN portfolio: considerations, development and TOC.

Work Package 5: New Bibliometric indicators
Project partners: Department of Information Studies, Royal School of Library and Information Science; Department of Library and Information Science, Humboldt University Berlin

Motivation

As funding and evaluation are presented to the researcher as part of the same package, control of the assessment of 'quality' and 'impact' (in their many guises) in an evaluation can be improved by actively involving the individual researcher. However, encouraging researchers to document their activities with bibliometrics means it is important to understand the ethical implications of this type of self-evaluation. At the present time, researchers are bibliometrically evaluated with standardized indicators by regulatory bodies for universities despite differing disciplinary and institutional infrastructures (Bach, 2011; Toncich, 2006). The results of evaluation rounds are used beyond judging merit: to monitor performance, productivity and inform financial or managerial decisions (Collini, 2012). Uniform schemes are implemented to achieve this, but the uniformity of evaluation schemes does not allow contextual judgments of individual performance (Bornmann & Marx, 2013). They also present researchers with the opportunity to exploit the procedures for their own personal gain at the detriment to science (Cheung, 2008; Lawrence, 2008).

The challenge is how to improve the representativeness of research output evaluations at the individual level. The gap between creating research, evaluating research and promoting excellence needs to be addressed as this is the problem in current systems of research evaluation. This problem is complicated. Researchers are people who are being evaluated between narrow frameworks and limited technology. In these systems the societal role of their research is secondary and the methods of evaluation, bibliometrics included, can be biased, subjective, give power to scientific elite and enforce the gender power structure.

Aim

The ACUMEN portfolio encourages researchers to use bibliometrics themselves to contextualise the scientific activities reported on their CVs to improve the representativeness of the evaluation. To obtain a consistency between the mission of the researcher and the mission of evaluation, ACUMEN needs to develop Guidelines for Good Evaluation Practice (GGEP). GGEP will support self-evaluation and evaluation by the consumers of the researcher's CV, one that does not undermine the authority of the researcher in their scientific processes. The GGEP together with the interactive structure of the portfolio will endorse the craftsmanship of the researcher without giving them all the freedom or taking freedom away. The purpose of this paper is to develop the behavioural code of conduct for the application and interpretation of bibliometric self-evaluation that can be included in the GGEP.

The key questions to be answered to develop a useful codex of behaviour:

1. What do we already know about ethical issues attributed bibliometric evaluation at the individual level from both the perspective of the evaluator AND the perspective of the individual researcher?
2. Based on what we know, is the current state of individual level bibliometric evaluation ethically correct?
3. Accordingly, which ethical issues, need to be addressed in individual self-evaluation from the viewpoint of both the evaluator and the researcher?

Current evaluation practices

By reviewing the literature, we found that ethical issues are different conditioned on the point of view: from the evaluators' point of view the main issue is if individual level bibliometric self-evaluation is at all ethically defensible while from the individual researcher's point of view, the issues seem to be more related to self-promotion. A core problem is that evaluation is considered to have a large degree of subjectivity and, in the case of peer review, a cliquish nature (Potočnik, 2005). Bibliometrics has been suggested as a form of objective evaluation to supplement the subjective peer review process. However, instead of monitoring the research process bibliometric evaluation is suspected to monitor the researcher (Collini 2012; Bach 2011; Cheung 2008).

We assessed the evaluation procedures of 14 European Research Evaluation Agencies and found that in practice individual bibliometrics rely heavily on publication counts, collaboration patterns and ranking of excellence adjusted to disciplinary representation in Thomsen Reuters Web of Science, D5.8 Part 1, unfortunately this endorses two well-known ethical issues:

1. that evaluation has been designed to fit the natural sciences' traditions of writing, publishing in journals and linking these publications to citations represented in Web of Science, (Campbell 2008; Laloë & Mosseri, 2009; Bornmann, L. et al, 2008) and,
2. that there is a pressure to publish in journals with a high impact factor included in citation databases or authority lists, rather than journals that fit the writing talent of the author and content of the paper. This approach has been criticized for rewarding competitive and aggressive researchers over modest or irregular publishers (Cheung, 2008).

Further, it appears that quantitative methods of assessing of individual performance and the discrepancies between the criteria used in performance assessment do not make sense when regarding the broader socio-economic function of scientific and scholarly research (Collini, 2012; Cheung 2008). The uninformed use of quantitative measures at the individual level and the lack of indicators of types of scientific activities other than article publication undermine the representativeness and hence validity of the evaluation (Bach, 2011).

Nevertheless, evaluation is a part of the researcher's and the institution's everyday life, and it is a balancing act between conducting informative evaluations and monitoring behaviours. The individual researcher will probably never welcome the prospect of a qualitative or quantitative evaluation, even though consumers of research do enjoy hearing just as much about failures as they do sucess – please refer to the media frenzy around the alleged dishonesty of the neuroscientist Penkowa in Denmark. But it doesn't have to be this sensational. Sune Auken, leader of the PhD school at the Faculty of the Humanities, has recently reflected on the differences between the humanist and the hard sciences, and how in evaluation and in subsequent funding, humanists can be treated as failed scientists[1]. Thus evaluation measures must be designed specifically to account for the different perspectives of quality and influence in the humanities and in as well as the hard sciences. The use and interpretation of the h index in awarding funds is a case in point[2]. Åström found that it is just assumed that reviewers know and understand the differences between fields and the effects these have on bibliometric statistics. But this assumption is not in any way regulated or monitored. The resulting small resources invested in

[1] Auken, Sune (2013, March 8) Measuring the Spirit? Bibliometrics and the Humanities. Powerpoint lecture presented in Fest Salen at the Royal School of Library and Information Science, Copenhagen.

[2] Åström, F, (2013, March 8). Questions concerning funding agencies suggesting that grant applicants include their h-index (or similar citation indices) in their CV- when there are grant programmes that gather applications from different research fields. Powerpoint lecture presented in Fest Salen at the Royal School of Library and Information Science, Copenhagen.

humanist research mean that the effort to measure performance may not be worth it both time-wise and financially. These three examples show, that when failures come to light, negativity can make a complete discipline feel inadequate or the quality of evaluation judgments can be based on assumptions, which could result in the necessity of a revised self-image of the researcher in an evaluation. Self-image is the core concept of the ACUMEN portfolio. The portfolio creates a space for researchers to promote their self-image by enabling the researcher to document their activities with substantiating evidence before presenting this to potential consumers. Hence, the evaluation in the ACUMEN portfolio is seen as a bidirectional activity, as researchers evaluate themselves before being evaluated by consumers.

In summary, to reduce the chance of violating standard codes of scholarly conduct and behaviour in professional scientific research self-evaluation, both the calculation and the interpretation of the indicators must be transparent to stop misuse and misinterpretation that in turn could cause fabricated self-images and damaged reputations - by researchers themselves and by consumers. Guidelines alone cannot ensure the correct use of bibliometrics, but can promote the informed use and informed interpretation of the indicators that bring objectivity into the process of self-evaluation and will not unduly expose the researcher (Bornmann et al, 2008). This approach will avoid promoting "ready to use" amateur indicators where the validity of the use of these measures can affect the validity of self-evaluation (Lundberg, 2009). Informed indicators will enrich CVs with and point to activities in systematic way that is acceptable to consumers. Evaluation of the individual researcher is the cornerstone of the scientific and scholarly workforce and shapes the quality and relevance of knowledge production in science, technology and innovation. The bibliometric indicators recommended in the ACUMEN portfolio must be simple and effective to make it worth the researcher's time and effort. Simplicity means though that not all the activities and efforts of the individual to communicate this research can be measured bibliometrically, but they should still be reported in the CV because this does not mean that what cannot be measured is not important.

In the next sections, we consider ethical issues in bibliometric self-evaluation to understand the construction and effects of an evaluation on the researcher and the interpretations by the consumer. The contents of the behavioural codex builds on this study.

The motives of self-evaluation: self-improvement or self-protection?

Self-evaluation motives affect the behaviour of the evaluand[3] and the consequences of the evaluation. When the motive is self-improvement the individual may determine in self-evaluation how failure occurred, consider their shortcomings and identify corrective actions to be taken in the future in order to grow (Tyser et al, 2012). When the concern is self-protection, the individual uses the self-evaluation to positively judge their performance and ability in order to maintain or increase self-esteem, thereby excusing or omitting failure (Crocker et al 2003; Tyser et al, 2012). Which motive the individual pursues is dependent on the circumstances and how malleable the evaluation is. Self-improvement and self-protection arise in many situations and can come into conflict. In self-protection the individual may ignore useful negative feedback, whereas self-improvement would require attention to this information, even though it could be damaging to the researcher's self-esteem.

Are researchers able to document their performance through self-evaluation?

Slife (2008), concluded that the individual is not the one best able to document their performance as they would write things that are significant to them but not significant to the consumer of these documents. The consumer or evaluator on the other hand are in a position to communicate the kinds

[3] Definition: the person under evaluation

of information an advisory board for example needs to determine the potential of the candidate. This assumes though that the individual can trust the peer system to provide fair and honest evaluations. We, WP5, do not share that assumption, which is why the ACUMEN portfolio encourages contextualizing the results of performance measures in a narrative or "dialogue" between evaluand and evaluator.

However, by enriching the CV with a narrative, researchers must be aware that they are at the same time presenting for appraisal their core personality traits, as the indicators are presented as comparisons to peers and as a snap-shot of the researcher's self-image. The empirical and conceptual personality traits that are commonly appraised in self-evaluations of work satisfaction and career success[4] are: self-esteem (seeing oneself as successful and worthy), self-efficacy (trust in ones capability to perform in many contexts) and the internal locus of control (believing in one's ability to control one's environment), while career success is defined as "…work related outcomes or achievements one has accumulated as a result of one's work experiences" (Stumpp et al, 2010). Career success contains both subjective aspects, e.g. attitudes to work and career, and objective aspects, e.g. awards, ascendency, and invited talks. The objective aspects are particularly interesting in documenting performance, because they can be measured bibliometrically, and in turn become explicit indicators of success that can be directly extracted from the CV by the consumer.

Specifically junior researchers capitalize on their personality traits and capitalize on outcomes later in life. This was evidenced by (Judge & Hurst 2007) who found positive relationships between core personality traits and academic achievement, socio-economic status and income. Self-esteem was found to affect the overall self-evaluation, promoting both positive and negative self-reflection narratives, by (Vallacher et al, 2002). Using a validated instrument Stumpp et al (2010) continued the work of Judge & Hurst, and found that people with high-core traits in self-evaluation focus on the career goals they have achieved instead of goals they have not realized so far. The tendency is that academics, with high core traits, have taken more actions to attain their goals and therefor achieve their goals. The result is that evaluators (consumers) judge individuals with explicitly presented high core traits more favourably than others.

How to reduce the uncertainty of self-evaluation?

By providing relevant information uncertainty is reduced (Misra 1973). Given that the results of the bibliometric analyses are of personal significance to the individual, it is anticipated that the individual will seek and utilize whatever information is available that will increase their subjective validity. Misra reports that using evidence of consistency and evidence supplied through social consensus contribute to the stability of the self-evaluation. Thus, if the individual provides substantiating, consistent evidence that informs the CV, the more stable it is. If however, only meager and unreliable information are provided, the less valid or more uncertain the self-evaluation is assumed to be.

Social comparison is a process whereby information from others is used in order to make stable attributions about the individual. In bibliometric selv-evaluation the performance of relevant others is used to inform social comparison. In the event of a sharp discrepancy between the individual's performance and the performance of others, the individual will be more susceptible to influence and the self-evaluation will become unstable due to lack of self-confidence (Misra, 1973). Misra's investigation of the instability of self-evaluation, though only using 13 female junior academics at UCLA, showed that the subjects who were told they were mediocre performers showed less interest in

[4] An extensive overview of CSE literature can be found in (Stump et al, 2010).

exposing themselves for self-evaluation. They avoided future interaction in groups who were more successful than them and questioned their own abilities. Subjects who were informed they had high abilities readily exposed their knowledge and partook in evaluations.

We may speculate that for the bibliometric self-evaluation, the individual will choose not to report the results of the indicators if they are exposed as low-achievers compared to their peers. Using social comparison indicators can though provide positive self-enhancement possibilities. The indicators can verify the belief researchers have in their abilities, and the more researchers feel they have something to contribute, the more active and vocal they are in their scientific communities, and this will be reflected in their CVS. A similar strategy is to document the researcher's influence on others using citation indicators which satisfy a need in its own right. Documenting influence also reduces the individual's uncertainty in their abilities.

Does self-worth affect self-evaluation?

The pressure to publish means that researchers see their self-worth as contingent on publication success, which unfortunately is easy to measure bibliometrically and easy to misinterpret if the publication count is not set in context of the researcher's gender, seniority, specialty, affiliation and discipline. Researchers in self-evaluation can be tempted to self-regulate their publishing success or failures to maintain positive self-views of themselves (Nicholls & Stukas, 2011), as in bibliometric self-evaluation the researcher is exposed to the effects of social-comparisons with peers, some known, when they develop local benchmarks. It stands to reason that upward academic comparisons are threatening. Bibliometrics expose the researcher, as they are contextualized by upper social comparisons in academic fields that require somewhat constant external validation (Crocker et al, 2003; Nicholls & Stukas 2011). Being out performed and further having to document it is detrimental to the researcher's self-definition and is theorized to be more extreme when the social comparisons are acquaintances or colleagues rather than strangers (Crocker et al, 2003). Crocker et al identified areas in which university students may develop contingencies of self-worth such as achieving academically, competing well with others, getting approval from others and attempting to protect their self-image in these areas. This is why the bibliometric indicators should not stand alone. They are supplementary to other information in the CV that includes both qualitative and quantitative indicators and techniques to maintain positive self-image.

Are there gender differences in self-evaluation

Many women believe that discrimination limits their opportunities, especially in relation to promotion. There is an unconscious bias at universities where evaluators rate CVs and journal articles lower on average for women than men[5]. Not surprising then that there are relatively few women employed in high-level faculty positions, though masculinity lessens for lower-level positions (RAISE, 2013; Koenig 2011). In self-evaluation, female researchers reflect gender stereotypes. Predominantly "communal" qualities, such as being nice or compassionate, are associated with women, and predominantly "agentic" qualities, such as being assertive or competitive, are associated with men (Koenig 2011; Cai, 2007). It is the agentic qualities that are believed to be essential to success and are the qualities that are prominent on a CV - as the results of being competitive or assertive are measurable, e.g. winning awards, initiating projects, where in contrast the researcher is not awarded a grant or published because they are "nice" or "compassionate".

[5] A overview of sources is too extensive to list. Please refer to, amongst others, the Boston University Recruitment Guide lines and corresponding reference list, available at: http://www.bu.edu/apfd/recruitment/fsm/assumption_awareness/

What are the cultural differences in self-evaluation?

Like gender differences in self-evaluation, cultural differences are less prominent in communal qualities, than they are in the agentic ones. There are the classic east versus west differences which are well covered in the literature, but also inter- and intra-European differences as well as subcultural differences, which have received less attention. Examples follow:

People from East Asian countries evaluate themselves in an excessively less positive manner than those in the West (Cai et al, 2007). The results of Cai et als study of junior researchers at the East China Normal University compared to their peers at the University of Washington point to this being due to cultural differences in modesty, not self-esteem, for example the Confucian tradition emphasizes modesty, difference and self-effacement. A similar culture is in the Scandinavian countries, the 10 rules in the Law of Jante[6], where children are encouraged from an early age not to brag about themselves. The law de-emphasizes individual effort and places all emphasis on the collective, while discouraging those who stand out as achievers: "*You are not to think you're anyone special or that you're better than us"*. Much has been written on the problematic nature of cultural differences in self-evaluation, and is too extensive to be listed here[7]. Topics worth considering in the construction of behavioural guidelines are how cross-cultural differences affect self-enhancement (Kurman, 2002; Takata, 2003); variance in measures of self-esteem across academic life-span (we have not succeeded in finding literature on this topic) and the effect of age, gender, ethnic groupings and variances in self-esteem (Cheng, C.H.K., & Watkins, D, 2000; Yin & Fan, 2003). However, agreement appears to be that self-evaluation is interpreted differently by different (sub)cultures. As a result these cultural ambiguities around presentation of self, especially in a bibliometrically enriched CV, demand that indicators, interpretations and the purpose of the self-evaluation is clear and standardised.

Conclusions

The informed use of bibliometrics will result in data that substantiates the claims and activities listed on the researcher's CV. This will document the work of researcher in a systematic way that is acceptable to consumers. However, there are important issues to consider in recommending the use of bibliometric self-evaluation. These are the reliability of the individual's calculation and interpretation of the indicators, and how bibliometric evaluation can affect the researcher's self-worth. Accordingly, the use of indicators must be voluntary and not a requirement.

A behavioural codex has been designed to inform the use and interpretation of bibliometrics used in self-evaluation. Using metrics can be complicated and time consuming, even simple indicators produce a lot of information. Guides to how to do bibliometric analyses in common citation databases must be available in the portfolio, either in the form of step-by-step instructions or links to online tutorials. We suggest a collaboration with the DLR LInCS programme sponsored project MyRi (Measuring your research Impact): http://www.ndlr.ie/myri/.

Further, the portfolio must provide the researcher with the tools to sort and filter all this information and present it in a short and useful narrative. As part of this, the methods of calculation and interpretation must be standardised and readily available to both the owner of the CV and the consumer. The consumer must be made aware that numbers are just numbers and must be set in the context of the individual's academic seniority, specialty, gender and culture. As experts, we the developers of the portfolio must before-hand take into account the diverse problems and difficulties

[6] The 11 principles or commandments that form the "Jante's Shield" of the Scandinavian people can be found in: Sandemose, Aksel (1933). En flyktning krysser sitt spor. Oslo: Aschehoug (Repr. 2005). ISBN 82-03-18914-8

7 Please refer to: Russon, C., & Russon, K. (2000). The annotated Bibliography of International Programme Evaluation. USA: Kluwer Academic Publishers. ISBN: 0-7923-8426-1

that could arise in the bibliometric analysis and in the interpretation of the metrics by the researcher in the narrative and by the consumer. Most importantly, bibliometric analyses cannot stand alone.

Limitations

Whether the findings in these studies presented here apply to all academic seniorities, disciplines, cultures or other ethnic groups is an unanswered question.

Suggested TOC behavioural codex for researchers and consumers using bibliometric self-evaluation. The TOC is built on the literature cited in the background study.

Observe good self-evaluation practice: This codex is developed to regulate ethical principles and rules of behaviour for bibliometric self-evaluation.

1. A short statement about professional codes of conduct.

Both the researcher and evaluator are bound by professional codes of conduct that ensure professional reliability and accountability. This conduct applies in a self-evaluation.
The bibliometric analyses and CV are subject to the researcher's integrity. Integrity is defined as: a person with integrity takes responsibility for their own successes or failures, and accepts the consequences of actions taken, never accepting or seeking undue credit for the accomplishments of others. ACUMEN provides the space and the guidelines for self-evaluation, researchers have the sole responsibility for the content of their CVs. Do not use bibliometrics if this gives you a negative self-image or you are uncertain of its benefits for you.

The calculation of indices can lead to many errors as evidenced by their variability in the databases. a researcher should calculate his own indices (in the disciplines where the databases are available) before submitting them for validation by persons in charge of indices at the level of a research institution or academic establishment. This opportunity is not always available.

ACUMEN endorses the idea of a unique identifier associated with each researcher, to verify publications attributed to the author and assist in assessment of the validity of the metrics.

When establishing local benchmarks, maintain the anonymity of the relevant others. Do not distribute the data you have collected about the performance of your peers unless you have anonymized it.

2. The limitations of bibliometric indicators

Bibliometrics do not stand alone. They are supplementary to other quantitative and qualitative indicators and must be contextualized to other information on your CV, your academic history and your ambitions. The ability to apply bibliometrics and its importance in the overall assessment of research varies between disciplines.

Bibliometric indices have no intrinsic value. They can only be understood relative to the distribution of index values for a particular field and by taking into account the age of the researchers concerned.

In some fields it is not the tradition to cite extensively the work that your scholarship and research is building upon – yet this is the whole principle of the citation analysis system. Do not use bibliometrics to compare performance between disciplines. Your score may be low in relation to the broad discipline or subject category, but high in relation to your particular seniority or speciality's publication production and received citations. Always contextualise the metric data just as you would the results of your research.

Citations come from the users of your work and can be complementary or critical. Negative citations, critical of a work, are counted as valid.

Self-citations can be legitimate citations of your own work that use to show how your research is developing. There is a practice of manipulating citations - over citing yourself or co-authors to boost your citation record. Always state if self-citations are included. It you included self-citations in your count, include self-citations in the counts of the peers you compare yourself to.

To avoid the researcher or evaluator relying on the parsimony principle 'one indicator is better than two', such as the h-index, the ACUMEN portfolio suggests a pallet of robust and valid indicators which are easy to use and understand.

The procedure, criteria and indicators used in bibliometric self-evaluation, as well as their adaptions to specific fields or sub-fields, are different at the national, university and department level. One indicator does not fit all.
Recommended disciplinary indicators. The indicators recommended by the ACUMEN portfolio are gender, academic seniority and disciplinary dependent. The operation of the indicator in self-evaluation is standardised.

3. A practical guide to bibliometric self-evaluation

The main source datasets – databases holding research and citations to it - are those of Thomson Reuters (Web of Science, Journal Citation Reports and other products), Elsevier (Scopus and other products) and Google Scholar plus subject-specialist options in some fields. Each collects the citation information from the articles in a select range of publications only – the overlap between the content of these sources has been shown to be quite modest in particular studies. So using just one source is providing a partial view of both research and citations to it. Where citation is common, the data sources often do not index the publications where research in a field is typically published – local publications, non-English, monographs, conference and working papers are poorly indexed. Learn more at http://www.ndlr.ie/myri/ and use the Open Access tutorials and work sheets that support bibliometrics training and awareness. Her you can also find reviews of some of the citations sources in the overview presented on the next page.

Subscription services are marked in red, and free software in green.

What analysis do you want to do?	Web of Science	Google Scholar/ Publish or Perish	Scopus	CWTS Journal Indicators	Journal Citation Reports	Eigen Factor	SCImago	Essential Science Indicators	ORCID	Impact story (online impact)
Article analysis										
Author Analysis										
Journal Analysis										
Journal Rankings										
Institution ranking										
Country Ranking										
Citing pattern analysis in discipline										
See top people, top places and trends										
Acquire a single perpetual ID										

Web of Science: http://thomsonreuters.com/web-of-science/
Publish or Perish: http://www.harzing.com/pop.htm
Scopus: http://www.info.sciverse.com/scopus
CWTS journal indicators: http://www.journalindicators.com/
EigenFactor: http://www.eigenfactor.org/
SCImago: http://www.scimagojr.com/
Essential Science Indicators: http://thomsonreuters.com/essential-science-indicators/
ORCID: http://orcid.org/
Impact Story: http://impactstory.org/

4. How to calculate each indicator recommended in the ACUMEN portfolio.
An example: Citation Count to one document or all documents.

The raw count of how many citations have been received by your document or set of documents over time. Do not remove self-citations. Always write the name of the database you used to source your citations.

Suggested sources: Web of Science, Scopus, Google Scholar with Publish or Perish

4a. Guide to interpreting each indicator in the portfolio.

An example: Simple Citation analysis

Compare the citations of your article in a given journal to the mean number of citations within same journal over a given period. This will add value to articles that are frequently cited in low impact journals.
Suggested cource: Journal Citation Reports, Eigen Factor, Scimago Journal and country rank

Compare the number of citations to your article to the citation data of another article published at the same time and in the same field. This will indicate the performance or use of your work.
Suggested sources: Web of Science, Scopus, Google Scholar with Publish or Perish

Examine the quality of the citations: knowing which articles (or types of articles) have cited a given article (or person) not only can reveal who has appreciated the work but also be used to assess its

interdisciplinarity, longevity,scope and timeliness.
Suggested sources: Web of Science, Scopus, Google Scholar with Publish or Perish

Examine the age of the citations: Knowing the age of the citations can show how current the "use" of your document is. Dividing the citations in to specific time periods, typically 5 year periods, show the growth of citations over time.
Suggested sources: Web of Science, Scopus, Google Scholar with Publish or Perish

5. Guide to local benchmarks.
An example: National-speciality citation benchmark
Establish a peer group by identifying researchers of the same academic seniority as yourself, in your country, working within the same specialty as yourself. Investigate, researcher by researcher, the amount of citations their documents receive. Use the same database you used to create your own citation count. Compare the median number of citations to your documents to the median number of citations to documents by the peer-group. The median is used, as citation counts are highly skewed and the mean can misrepresent performance as it can be affected by extreme high/low citation counts. Comparing performance to peers of the same academic seniority within your specialty will indicate to the consumer how your citation count ranks in regard to significant others.

Suggested sources: Web of Science, Scopus, Google Scholar with Publish or Perish.

5a. Guide to global benchmarks
An example: Field top 5% citation threshold value
The top 5% threshold value is the minimum number of citations essential to make a publication one of the 5% most cited publications of the same age, of the same publication type within the same field. Other top reference values, as top 1% and top 10% are also used, and calculated in the same way as top 5%.
All publications are divided into groups where the items have the same document type, age and subject area. The publications in the group are counted and sorted according to the number of citations in descending order. The number of citations needed to belong to the top 5% share of publications, i.e. the 95thpercentile limit, is equal to the top 5% threshold value.

The index is not suitable for junior researchers. Subject areas defined in Web of Science do not necessarily reflect sub-specialties. It takes extensive work to establish a global benchmark.

Suggested sources: Web of Science

6. Presenting the metrics
As soon as you contextualise your metrics in a narrative, your academic character and personality will become public. Be aware how your personality affects the evaluation: your self-efficacy, your modesty, your self-esteem and your ego.

Use the metrics to substantiate how you are achieving your goals, to account for failures and to reduce the chance of the consumer misinterpreting your metrics. Documenting your efforts to do this could justify a sporadic publishing or collaboration strategy or rapid changes in affiliation that could be read by the consumer as disloyalty.

Measures of career success are easy to document such as number of awards or invited talks in relation to promotion and grants received. Include the subjective aspects of success in the narrative to

contextualize your publication count if for example your publication count may be at the lower end of the comparison group for this specialty, you came late to academia or you took a break to prioritize other activities.

Substantiate your metrics with social comparisons or local benchmarks. A list of numbers presented as statistics is just noise. Don't expect the consumer to take the effort to interpret and contextualize these numbers.

Bibliometric evaluation should be associated to a close examination of your work, in particular to evaluate its originality, an element that cannot be assessed through a bibliometric study.

Use the ACUMEN case studies for tips and inspiration on how to contextualize your metrics in a narrative

7. Specifically for consumers

Bibliometric indices should be used and interpreted differently depending on the purpose of the evaluation, such as recruitment, promotion, grants and distinctions.
All the CVs in the ACUMEN portfolio suffer from self-presentation bias that can have both positive and negative outcomes for the researcher. Cultural differences in presentation may be due to modesty not limited self-esteem or lack of belief in competences. Bibliometrics have limited value for assessing junior researchers at the start of their academic career. There are also disciplinary differences in publication and citation traditions. These pattern variations must not be ignored in an evaluation as these differences affect the calculation and results of the metrics. It is important to be aware that some researchers might chose to steer their activity in such away as to get articles accepted in journals with a high impact factor rather than engaging in original and creative research.

The data used to calculate the indicators and individuals use and interpretation of the metrics must be validated by the consumer. ACUMEN takes no responsibility for the information presented in the CVs.

Bibliometric self-evaluation is of no value unless a number of prerequisites are met:

- The self-evaluation focusses on the articles/papers and not the journals.
- There must not be cross-disciplinary comparisons, such as comparing h-index across fields.
- It is inappropriate to use the Journal Impact Factor of a journal title to evaluate the quality of an individual researcher's output.
- It is important to consider bibliometric data against the specific distribution of values of the researcher's field and also to take into account the rate of career progression.
- The metrics must be justified in a narrative.

References

Bach, J. (2011). *On The Proper Use of Bibliometrics to Evaluate Individual Researchers.* Rapport de l'Académie des sciences - 17 janvier 2011: Report presented on 17 January 2011 to the Minister of Higher Education and Research.

Bornmann, L., & Marx, W. (15th. February 2013). Standards for the application of bibliometrics in the evaluation of individual researchers working in the natural sciences. Retrieved 25th. February 2013 from ArXiv.org: http://arxiv.org/abs/1302.3697v1

Bornmann, L., Mutz, R., Neuhaus, C., & Daniel, H.-D. (2008). Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics, 8*, pp. 93-102.

Cai, H., Brown, J.D., Deng, C., & Oakes, M.A. (2007) Self-esteem and culture: Differences in cognitive self-evaluations or affective self-regard? *Asian Journal of Social Psychology.* 10, pp. 162-170.

Campbell, P. (2008). Escape from the Impact Factor. *Ethics in Science and Environmental Politics, 8*, pp. 5-7.

Cheung, W. (2008). The Economics of Post-Doc Publishing. *Ethics in Science and Environmental Politics, 8*, pp. 41-44.

Collini, S. (2012). Bibliometry. In *What Are Universities For?* London: Penguin.

Crocker, J., Luhtanen, R.K., Cooper, M.L., & Bouvrette, S. (2003) Contingencies of self-worth in college students: Theory and measurement. *Journal of Personality and Social Psychology*, 85, pp.894-908

Ingwersen, P. (2005). Scientometri: Videnskabspublicering og bibliometriske metoder. *Biozoom, 2*.

Judge, T.A., & Hurst, C. (2007) Capitalizing on one's advantages: Role of core self-evaluations. *Journal of Applied Psychology*, 92, pp.1212-1227

KAOW-ARSOM. (2012, 7 25). Conditions for a Good Evaluation. Retrieved 1 29, 2013, from *The Guidelines Project: Providing Instruments for evaluating development Research:* http://www.guidelines.kaowarsom.be/en

Koenig, A.M., Eagly, A.H., Mitchell, A.A. & Ristikari, T. (2011). Are Leader Stereotypes Masculine? A Meta-Analysis of Three Research Paradigms. *Psychological Bulletin,* 137(4), pp.616-642.

Kurman, J. (2002). Measured cross-cultural differences in self-enhancement and the sensitivity of the self-enhancement measure to the modesty response. *Cross Cultural Research*, 36 (1), pp.73-95.

Laloë, F., & Mosseri, R. (2009). Bibliometric Evaluation of Individual Researchers: not even right....not even wrong! *Europhysics News, 5*, pp. 26-29.

Lawrence, P. (2008). Lost in Publication: how measurement harms science. *Ethics in Science and Environmental Politics, 8*, pp. 9-11.

Lundberg, J. (2009). Lifting the crown: citation z-score. *Journal of Informetrics, 1*(2), pp. 145-154.

Misra, S. (1973) Instability in self-evaluation, conformity and affiliation. *Journal of Psychology*, 41(3), pp.361-375

Murphy, M., Steele, C.M., & Gross, J.J., (2007) Signaling Threat: How Situational Cues Affect Women in Math, Science, and Engineering Settings. *Psychological Science*, 18(10), pp. 879-885.

Nicholls, E. & Stukas, A.A. (2011) Narcissism and the Self-Evaluation Maintenance Model: Effects of Social Comparison Threats on Relationship Closeness. *The Journal of Social Psychology*, 151 (2), pp.201-212

Potočnik, J. (2005). *The European Charter for Researchers: The Code of Conduct for the Recruitment of Researchers.* Brussels: European Commission:Directorate-General for Research.

RAISE. (2013). *Recognizing the acheivements of Women in Science, technology, eningeering, maths and medicine*, at http://www.raiseproject.org/index.php [accessed 10 June 2013]
Salisbury, L (2009) Web of Science and Scopus: A Comparative review of Content and Searching Capabilities. *The Charleston Advisor* (July), pp.5-18.

Slife, J. (2008) Self-evaluation: A disconnect in our values. *Air and Space Power Journal*, Winter, pp.52-59

Stumpp, T., Muck, P.M., Hülsheger, U.R., Judge, T.A., & Maier, G.W. (2010) Core self-evaluations in Germany: Validation of a German Measure and its relationships with career success. *Applied Psychology: An International Review*, 59 (4), pp.674-700.

Tafarodi, R.W. & Swann, W.B. Jr. (1996). Individualism-collectivism and global self-esteem: Evidence for a cultural trade-off. *Journal of Cross-Cultural Psychology*, 27, pp.651-672

Takata, T. (2003) Self-enhancement and self-criticism in Japanese culture. An experimental analysis. *Journal of Cross-Cultural Psychology,* 34 (5), pp.542-551.

Toncich, D. (2006). University Research - Evaluation, Bibliometrics and Conduct Issues. In D. Toncich, *Key Factors in Postgraduate Research - A Guide for Students* (pp. 281-306). Australia: Chrystobel Engineering.

Cheng, C.H.K., & Watkins, D. (2000). Age and gender invariance of self-concept factor structure: An investigation of a newly developed Chinese self-concept instrument. *International Journal of Psychology,* 35 (5), pp.186-193.

Tyser, M.P., McCrea, S.M., & Knüpfer, K. (2012). Pursuing perfection or pursuing protection? Self-evaluation motives moderate the behavioral consequences of counterfactual thoughts. *European Journal of Social Psychology*, 42, pp.372-382.

Vallacher, R.R., Nowak, A., Froehlich, M., & Rockloff, M. (2002) The dynamics of self-evaluation. *Personality and Psychology Review*, 6 (4), pp.370-379.

Wildgaard, L., Schneider, J., & Larsen, B. (2013). Bibliometric Self-Evaluation: A review of the characteristics of 114 indicators of individual performance. *Submitted.*

Yin, P., & Fan, X. (2003) Assessing the factor structure invariance of self-concept measurement across ethnic and gender groups: Findings from a national sample. *Educational and Psychological Measurement,* 63 (2), pp.296-318.

Zitt, M. B. (2008). Challenges for scientometric indicators: data demining, knowledge-flow measurements and diversity issues. *Ethics in Science and Environmental Politics, 8*, pp. 49-60.

ACUMEN
academic careers understood through measurement and norms

FP7 Grant Agreement	266632
Deliverable No and Title	D5.8 Part 5 – Consequences of Indicators: using indicators on data from Google Scholar
Dissemination Level	PU (public)
Work Package	WP5 Bibliometric Indicators
Version	1.0
Release Date	28-02-2014
Author(s)	Lorna Wildgaard Birger Larsen Jesper W Schneider
Project Website	http://research-acumen.eu/

European Commission
7th Framework Programme
SP4 - Capacities
Science in Society 2010
Grant Agreement: 266632

Main conclusions of analysis of usefulness of Publish or Perish statistics on Google Scholar data

Abstract
We investigate if Publish or Perish ready-to-use bibliometric indicators can be used by individual scholars to enrich their curriculum vitae. Selected indicators were tested in four different fields and across 5 different academic seniorities. The results show performance in bibliometric evaluation is highly individual and using indicators as "benchmarks" unwise. Further the simple calculation of cites per publication per years-since-first-publication is a more informative indicator than the ready-to-use ones and can also be used to estimate if it is at all worth the scholar's time to apply indicators to their CV.

Keywords: bibliometrics, research evaluation, ready-to-use indicators, micro-level, individual, impact, curriculum vitae

Introduction
As bibliometric techniques have become more available and easier to apply at the micro-level they have become increasingly used as both self-evaluation and third party evaluations (Wouters et al 2013). This increased use presents challenges for the correct application of bibliometric indicators on a small amount of data, the correct interpretation of these statistics and, if any, the conclusions that can be drawn. These challenges are discussed in many bibliometric studies (Glänzel & Wouters 2013, Bach, 2011, Costas et al 2011, Costas et al 2009, Sandström 2009), but at the current time it is still unclear which indicators are appropriate for which scholars and in which fields. This study examines this gap in knowledge and attempts to recommend useful indicators. We use ready-to-use indicators available to the scholar though Publish or Perish, and investigate if scholars can potentially use these indicators to enrich the information on their CVs.

Purpose
The purpose of this study is to investigate if ready-to-use bibliometric indicators are useful in enriching the CV of an individual scholar by the scholar. Aspects to be considered in the analyses of the indicators are:

1. If the indicators in this study more appropriate in some disciplines than others.
2. If the indicators in this study are more appropriate for some seniorities than others.
3. If the indicators in this study are gender appropriate
4. If indicator produces useful information that scholars can use to enrich their CV. 5. If the indicator produces information that is redundant if used in combination with other indicators.
6. If the indicator would have a positive or negative effect on the profile of the scholar.

Method

Dataset

The dataset consists of a sample of 750 researchers: 584 men and 165 (22%) women, Table 1.

Table 1. Distribution of seniorities and gender across the disciplines in the sample

	nPhD	nPost Doc	nAssis Prof	nAssoc Prof	nProf	Total
Astronomy	15	48	26	67	37	193
Gender M/F	*12:3*	*37:11*	*20:6*	*58:9*	*35:2*	*162:31*
Environment	3	17	39	85	51	195
Gender M/F	*3:0*	*11:6*	*30:9*	*72:13*	*44:7*	*160:35*
Philosophy	9	22	45	75	78	229
Gender M/F	*6:3*	*20:2*	*37:8*	*57:18*	*63:15*	*183:46*
Public Health	9	14	31	50	29	133
Gender M/F	*2:7*	*7:7*	*18:13*	*34:16*	*19:10*	*79:53*
Total	36	101	140	277	195	750
Discipline M/F	*23:13*	*75:26*	*105:36*	*221:56*	*161:34*	*585:165*

Indicator identification

The ready-to-use indicators tested in this study are the cumulative indicators of individual performance from Publish or Perish[1]. They are: Total number of papers *(P)*, years since first publication *(PY)*, total number of citations *(C)*, cites per paper *(CPP)*, average number of citations per paper normalized for years since first publication *(CPAY)*, h-index *(h)*, g-index *(g)*, e-index *(e)* and age-weighted index[2] *(AW)*. With this information the scholar can easily calculate the m-quotient *(m)* and the mg-quotient[3] *(mg)*. These indicators are often defined as indicators of "quality" and do not adjust for the amount of authors-per-paper or add age-weighting parameters to each cited article. They were chosen based on selections criteria presented in our previous review (D5.8 Part 1) of 114 bibliometric indicators used in individual evaluation.

Data Collection

Publication data and ready-to-use bibliometric indicators were sourced for European scholars in the field of Astronomy, Environmental studies, Philosophy and Public Health. Scholars in these fields were sampled from a questionnaire study of scholarly web-presence undertaken by the University of Wolverhampton in December 2011[4]. Of the 2154 scholars who responded, 793 provided a link to an online CV and/or publication list. We collected publication, citation data and indicators in Google Scholar via Publish or Perish[5] from June 13th to July 20th 2013, figure 1. Publications were verified using the publication list the scholar provided a link to.

[1] http://www.harzing.com/pophelp/metrics.htm

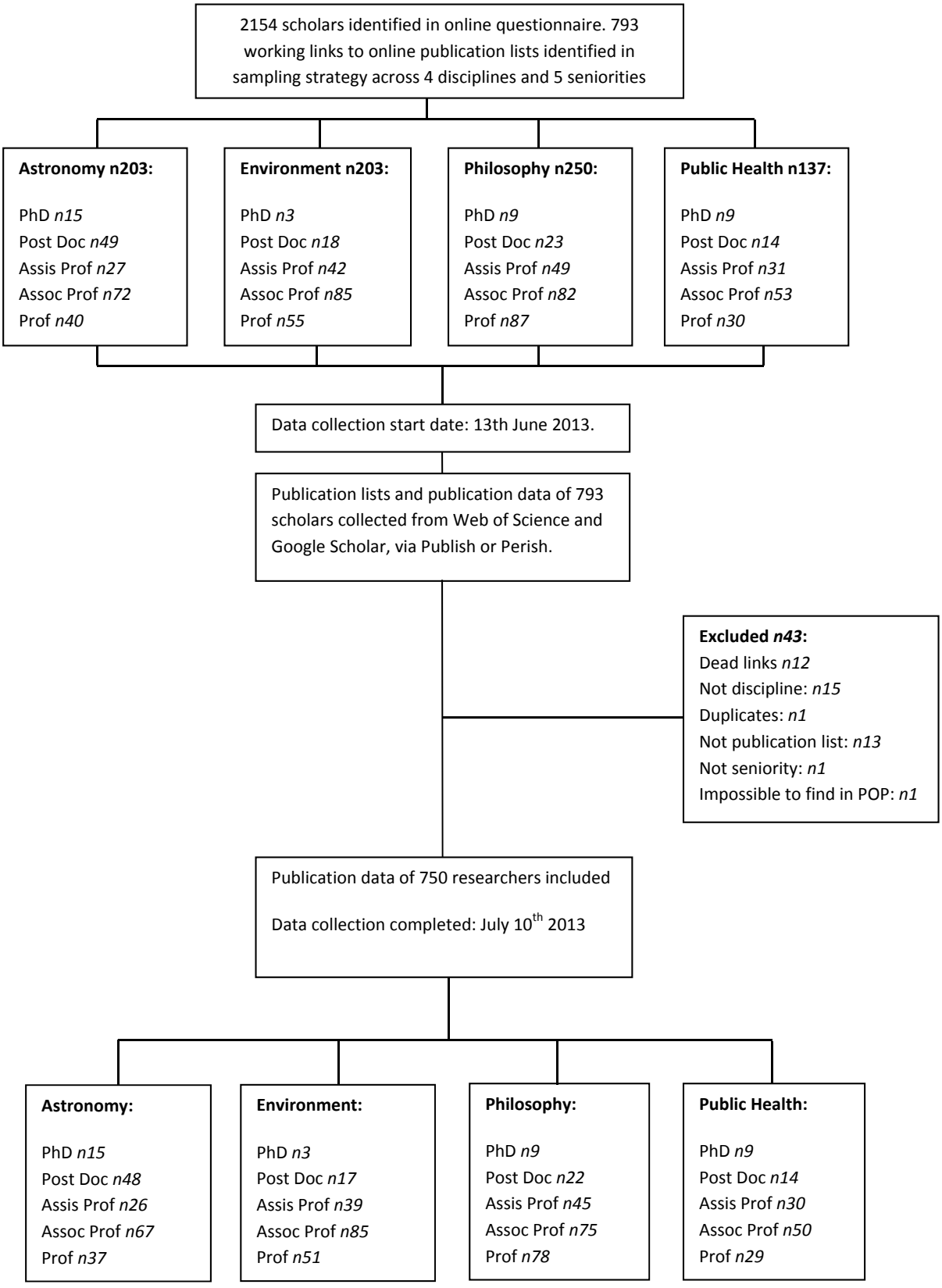
[2] AW index: AW is the square root of the number of citations to a given body of work divided by the total number of papers, it approximates the h-index if the average citation rate remains more or less constant over the years.

[3] Mg-index: mg is the m-quotient, h adjusted for the number of years since first publication, calculated with g-index instead of the h-index.

[4] http://cybermetrics.wlv.ac.uk/survey-acumen.html

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

Fig. 1. Flowchart of data-collection



Main results and discussion

Women make up 22% of the overall sample reflecting the European ratio of men to women in science, 3:1[6]. In the junior categories, PhD students, post docs and assistant professors, the ratio men to women is 2:1, while in the senior categories, associate professor and professor, the ratio is 4:1. This reflects the 2012 SHE figures of gender in research, confirming that our sample patterns the share of women employed in academia across Europe where gender imbalance increases with seniority[7].

However, the size and content of the seniority categories were not homogenous. The spread of publication and citation data within categories and across fields was highly skewed and it was difficult to estimate effects of indicators and detect homogeneity, which is important if we wish to establish the performance benchmarks. We used quartiles to illustrate the spread of the data and the median or second quartile as the best estimate of average performance within group. In all seniorities there were outliers that pulled the average performance up or down. Therefore the relative interquartile range (RIQR) was calculated. Even when outliers were removed, the variation in the number of publications a scholar produces, within each seniority, in Astronomy, Environmental Studies and Philosophy was still very large, but in Public Health there was less variation.
To understand if we need to recommend gender specific indicators, we studied the career trajectory of scholars in our sample. Our hypothesis was a longer publication history in the junior seniorities could be an indirect sign of possible female discrimination or other disruption in career promotion. *PY* was calculated and analyzed in panel box plots by gender and seniority to identify differences in length of publication history between male and female scientists. According to our data, advancement from PhD to associate professor for both genders was based on a 9 to 11 yearlong publication history. Professors had *PY* 3 to 6 years longer than associate professors in Astronomy and Public Health and additional 9 to 11 years in Philosophy and Environmental Studies. Women do not appear to need a higher number of publication years to advance. We compared the performance of female scholars to male scholars within seniority using the other indicators in this study. The performance of each indicator was highly individual and no gender-specific patterns were identified.

We took Astronomy as a case study. Scholars were ranked per seniority in descending order for each indicator, *P, PY, C, CPAY, h, g, e, AW, m, mg*. Each ranking was copied to a table depicting the performance of all scholars, within seniority, across all indicators. The tables were divided into lower and upper quartiles. Each scholar's placement in the rankings of each indicator was mapped manually and categorized as high (3rd quartile), middle (second quartile) or low (1st quartile). This resulted in the identification of two groups of indicators. The first group showed predictive relations: *h, g, e, AW, m, mg* where a high, middle or low score on one predicts a high, middle or low score on another. The *e, AW, m* supplemented *h* while *mg* supplemented *g.* The top 25%, middle 50% or bottom 25% scholars remained the same but ranked in a different order.
The second indicator group was "unpredictive indicators": *PY, P, C, CPP, CPAY*. For example, a low *P* doesn't result in a high *C* likewise a high *PY* doesn't predict a high *P*. The threshold where the ratio *C* to *P* results in a high *CPP* was also highly individual. No

[6] Directorate-General for Research and Innovation, Unit B6 (2012) SHE Figures 2012: Gender in Research and Innovation. European Commision: Brussells.
Retrieved from:
http://ec.europa.eu/research/science-society/document_library/pdf_06/she-figures-2012_en.pdf
[7] SHE figures 2012.

individual or seniority patterns were found across this sub-group of indicators, and ranking resulted in different scholars appearing in the top, middle or bottom quartiles. No difference was observed between *CPAY* and *m*, resulting in redundant information.

We suspected a ratio relationship between *PY, P and C* that controls level of performance across ALL indicators. The ratio "years since first publication to amount of publications" was calculated for each scholar, then the ratio "years since first publication to total citations". This is the math behind the CPAY indicator, but the ratio is more informative than the single number CPAY produces, eg. Scholar A averages 2 papers per year which over his career and receives 28 citations per year=1 (year): 2(papers):28 (citations) = 1:2:28 (CPAY=28). By comparing the scholar's rank to their ratio we found the indicators favour scholars with the ratio short "career:many papers:high citation count" over scholars with different "career:paper:citation" ratios. To investigate if it is the amount of citations per paper per year that dictate how useful the indicators will be to the scholar, we divided the amount of citations per year by the amount of publications per year for all the scholars identified in the top, middle and low quartile, eg. Scholar A ratio score 1:2:28, citation score per publication per year = 28/2=14. We compared this ratio score to their rank position and found the ratios within seniorities fit for the whole group, which in our dataset is a proxy for the disciplinary level, Table 2.

Table 2. Citations per publication per year across disciplines and seniorities

		PHD	Post Doc	Assis. Prof	Assoc. Prof	Prof
Top 25%	Astronomy	-	≥ 18	≥ 19	≥ 27	≥ 28
	Environment	-	≥ 7.3	≥ 14	≥ 16.3	≥ 19.1
	Philosophy	-	≥ 4	≥ 4.1	≥ 6.8	≥ 10.4
	Public Health	-	≥ 24.4	≥ 38	≥ 18.3	≥ 23.2
Middle 50%	Astronomy	-	≥ 3 cites ≤ 8	≥ 7 cites ≤18	≥ 10 cites ≤15	≥15 cites ≤ 27
	Environment	-	≥ 3 cites ≤ 4	≥ 4 cites ≤9.6	≥4.1 cites ≤13,1	≥5.4 cites ≤17.6
	Philosophy	-	≥ 1 cites ≤3.6	≥ 1.4 cites ≤3.7	≥ 1.7 cites ≤4.8	≥2.6 cites ≤9.5
	Public Health	-	≥5.5 cites ≤ 13	≥ 2.4 cites ≤28.9	≥ 7.9 cites ≤17.1	≥19.2 cites ≤21.8
Bottom 25%	Astronomy	≤ 2	≤ 3	≤ 8	≤ 7	≤ 9
	Environment	-	≥ 0.6 cites ≤ 1	≤ 2	≤3.8	≤ 5
	Philosophy	-	≤0.99	≤ 0.7	≤1.2	≤ 2.2
	Public Health	≤ 1	≤2.3	≤ 2.4	≤ 5	≤6.4

How to read the table

The dataset was divided into disciplines and seniorities with disciplines. The performance of each scholar, within seniority, was ranked from highest to lowest scores using the indicators CPP, h, g, AW, e, m, mg. Each scholar was then mapped across the indicators to find out if

they ranked in the top, middle or bottom quartile of their seniority. Scholars that placed across all indicators in the top 25%, middle 50% or bottom 25% were used as the expected performance scores and are represented in the above table. Each cell shows the expected amount of citations per publication per year a scholar has to accrue to be ranked in the top, middle or bottom of their discipline if they choose to use the most common readily available bibliometric indicators h, g, e, AW, m quotient, mg-quotient.

The table clearly illustrates the different citation cultures in the disciplines and hence how unwise performance comparisons across disciplines are. Also, the expected performance of scholars according to their seniority is very different. Public Health, Astronomy and Environmental Science have a strong citation culture whereas Philosophy appears more selective. It is not surprising that 3 of the disciplines exhibit similar behavior, as in our dataset, they have a strong culture of multiple authorships and tradition for publishing articles in journals whereas the philosophers seem to prefer sole authorship and other publication forms. PHD students do not appear to have the accumulated enough citation and publication data or years of experience to use classic bibliometric indicators.

Conclusions

The publication and citation data was highly skewed, and using simple average based indicators, such as *CPP*, as an indicator of performance or disciplinary benchmark misrepresents the individual. The heterogeneity of the data made comparisons to peers and disciplinary benchmarks uninformative about the performance of the individual scholar. Gender specific indicators were not necessary. The variance in the amount of publications between scholars differs from discipline to discipline, but there are clear differences in the quantities of publications a discipline produces as a whole. Public Health shows potential for the development of useful expected performance benchmarks, as within seniority variation was low. The *h,g,e, AW, m* or *mg* indices supplemented each other and useful combinations need further investigation. Further these indices showed a predictive relationship, raising the question if it is at all informative to calculate more than one of these indicators. There was information redundancy between *CPAY* and *m*. Normalizing publications and citations to the length of a scholar's career within the seniorities predicted if it is worth the scholar's time to use the indicators. Scholars whose ratio scores place them in the low 25% of their seniority should not expect to perform well and the information these statistics provide will not positively enrich their CV. The top 25% can expect the indicators to add value to their publication lists.

Recommendations for use of indicators based on data from Google Scholar

1. The h, g, e, AW indicators show a predictive relationship, ie if you score high on one, you will score high on the others; low on one, you'll score low on the others. This correlation also applies to the m and mg quotient, cites is not as stable.
2. There is no seniority or disciplinary trend between the amount of years active as a scholar, number of papers and number of citations. This is highly individual. The predictive indicators favour scholars with the ratio short "career:many papers:high citation count" over scholars with different "career:paper:citation"
3. CPP is more informative than raw citation counts or the tested indicators as it allows for field specific citing behavior.
4. We suggest that ACUMEN could present these ratios as a baseline for performance, and as such this ratio can be used to inform the scholar if bibliometrics indicators recommended in the portfolio are useful for them. If the indicators are not deemed useful or scholars simply do not wish to use them, ACUMEN needs to recommend

alternative ways of contextualizing the scholar's published work. Our suggestion is to provide standard formulations to help the scholar construct their narrative, such as encouraging the scholar to present their total publications, years of experience and citations as well as the citations and publications adjusted for years of activity. For example, the scholar fills in the year, number of papers, number of citations and ratio in the following:

"I have been publishing since the year 2000 and have in that time published 24 papers that have received in total 342 citations. This averages out at roughly 2 published papers per year over my career, which each have accrued on average 14,2 citations. On a yearly basis my articles each attract on average 7 citations According to the ACUMEN table of field citing behaviours, this places me on the border between middle and top performing scholars in (insert field) Astronomy according to my current seniority of a Post Doc."

Limitations

The data in this table is based on Google Scholar and needs to be repeated with data from WOS to understand if the results are database dependent or can be generalized. Clearly scholars have to use the same database to collect their citations as the database used to construct the disciplinary/seniority benchmarks. This could be a challenge for both WOS, as disciplines and nationalities are not equally represented, and for Google Scholar which is not always accepted as a reliable source of citation and publication activity by scientist and evaluators.

References

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

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., & 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.

Glänzel, W., & Wouters, P. (2013) The dos and don'ts of individual-level bibliometrics. Paper presented at the 14th ISSI Conference, Vienna, 15-18 July 2013. Accessed September 2013 via url: http://t.co/gYnVxcpvwk

Sandström, E., & Sandstrøm, 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.

Wouters, P., Glänzel, W., Gläser, J., Rafols, I. (2013) Individual-level evaluative bibliometrics-the politics of use and abuse. Brief report at the STI 2013 plenary on the methodological aspects of individual-level bibliometrics. Berlin September 2013.

ACUMEN
academic careers understood through measurement and norms

FP7 Grant Agreement	266632
Deliverable No and Title	D5.8 Part 6 - Cluster analysis of bibliometric indicators of individual scientific performance
Dissemination level	Public (PU)
Work Package	WP5-Bibliometric Indicators
Version	1.0
Release Date	28. February 2014
	Lorna Wildgaard Birger Larsen Jesper W Schneider
Project Website	http://research-acumen.eu/

European Commission

7th Framework Programme

SP4 - Capacities

Science in Society 2010

Grant Agreement: 266632

Cluster analysis of bibliometric indicators of individual scientific performance
ACUMEN Deliverable 5.8 Part 6

Lorna Wildgaard[a] Jesper W Schneider[b] Birger Larsen[c]
[a] Royal School of Library and Information Science, Birketinget 6, 2300 Copenhagen, Denmark

[b] Institut for Statskundskab - Dansk Center for Forskningsanalyse, Bartholins Allé 7, 8000 Aarhus C, Denmark
[c] Department of Communication and Psychology, Aalborg University Copenhagen, A. C. Meyers Vænge 15, 2450 Copenhagen SV, Denmark

Contents

1 Introduction 3
2 Data 3
3 Methods 5
3.1 Data analysis 5
3.2 Limitations of the analyses 6
4 Results 10
4.1 Association between seniority and bibliometric indicators 10
4.2 Identifying central indicators across disciplines 12
4.3 Identifying central indicators for each discipline 13
4.3.1 Astronomy 13
4.3.2 Environmental Science 15
4.3.3 Philosophy 18
4.3.4 Public Health 19
4.3.5 Discussion and recommendations 22
5 Conclusion and recommendations 23
6 Acknowledgements 24
7 References 24
Appendix 1: Effect of excluding proceedings papers 26
Appendix 2: Calculation of indicators 34
Appendix 3: Correlation matrix *Astronomy* 36
Appendix 4: Correlation matrix *Environmental Science* 37
Appendix 5: Correlation matrix *Philosophy* 38
Appendix 6: Correlation matrix *Public Health* 39

1 Introduction

As discussed in Wildgaard, Schneider and Larsen (2014) bibliometricians are cautious of evaluation at the level of individuals, as the context and variables affecting the results of analyses are many, and often unsatisfactorily explored. Hence, the debate on the shortcomings of performance indicators generated by bibliometric methods at the micro-level continues (Bach, 2011; Bornmann & Werner, 2012; Burnhill & Tubby Hille, 1994; Sandström & Sandström, 2009; Wagner et al., 2011). Despite of the concerns from the bibliometric community, evaluation of the individual through bibliometric indices is already being performed as a form of 'pseudo peer review' in selection of candidates for tenure, in background checks of potential employees' publication- and citation impact, and in appraisal of funding applications. As part of developing the ACUMEN portfolio we therefore carried of an extensive review of 114 bibliometric indicators in WP5 Deliverable 5.8 Part 1 to identify 1) which indices are useful in individual self-evaluation to document activities listed on the CV and contextualize publication performance, 2) identify which scientific activities it is possible to measure and with which indices, 3) analyse the applicability of these indices by discussing the strengths and weakness of each one, and 4) identify if there is a need for any additional novel indicators to measures the performance of individuals.

The analysis showed that there is no immediate need to develop new bibliometric indicators. There is a wealth of indicators to choose from, some used in practice and some theoretical only. There is therefore a need to understand the usefulness of existing indicators and which ones represent independent research activities of authors. In this paper, we investigate how 1) traditional and novel indicators complement each other, 2) if there is a redundancy among indicators, i.e. two or more indicators measure the same thing, and 3) which indicators are the "best" choice in regards to our four predefined disciplines. The main parameter we judge the usefulness of indicators on is their simplicity, as investigated in Wildgaard, Schneider and Larsen (2014) and their sensitivity to publishing and citation traditions within disciplines.

2 Data

The analysis in this paper is based on citation and publication data of European researchers. The data is drawn from the shared ACUMEN data set of 2,554 researchers in four scientific disciplines who responded to an online survey of web-presence conducted by WP2. In the analysis in the present paper the researchers to have 1) an active curriculum vitae on the web, and 2) a publication list on the web. A subset of 741 researchers from the shared ACUMEN data set fulfilled both conditions[1].

In the survey the respondents reported their academic discipline and seniority, and these are used to group the 741 researchers analysed in this paper. We extracted their publications from the CVs and searched the Thomsen Reuters Web of Science (WoS) to identify these publications. We identified 34,660 citable papers indexed in WoS, written by 741 European researchers in the disciplines of *Astronomy*, *Environmental Science*, *Philosophy* and *Public Health*. Additional publication and citation information on articles and reviews in this data set was kindly provided for the purposes of this study by the Centre for Science and Technology Studies (CWTS) at Leiden University, the Netherlands from their custom version of the WoS. This custom database contains records from the Science Citation Index Expanded, Social Sciences Citation Index and Arts & Humanities Citation Index portions of WoS, and has been specially prepared for bibliometric analysis. The data delivered

[1] Please refer to the following WP5 deliverables D5.8 Part 1 "Literature Review" and D5.8 Part 2 "selection of samples".

by CWTS thus contains a wide range of bibliometric indicators for each paper including field normalised indicators using CWTS standard procedures. As the CWTS data does not contain data from the Conference Proceedings Citation Indexes we do not have additional data on 3,693 citable papers and these are excluded from the present analysis. Our final data set thus consists of 30,967 publications with additional citation information.

Table 1. Sample of 741 researchers, distribution of publications and citations across disciplines and seniorities.

	Publications				Citations		
Discipline	**Sample**	**Range**	**Median (CI)**	**Mean (CI)**	**Range**	**Median**	**Mean (CI)**
***Astrology*, 192 researchers**							
PhD	15	2-36	7(5.0;14.2)	10.8(5.6;15.9)	8-529	150(27.9;209.7)	149.4 (64;234.7)
Post Doc	48	3-103	19.5(14;26.5)	26 (19.9;32.1)	3-3177	201.5(140.4;479.4)	561.1(339,7;782.4)
Assis Prof	26	10-142	39.5(30;65.9)	51 (37.3;64.8)	69-4009	702 (432.2;1327.5)	1118,6 (675;1562.1)
Assoc Prof	66	7-292	61.5(48.5;75.4)	77.7(63.2;92.2)	19-9083	1214(783.6;1622.8)	1981.1(1477.8;2484.4)
Professor	37	34-327	90(75.2;109.6)	121.3(92.8;149.8)	177-16481	1889(1292.9;3245.3)	3579.1(2170.9;4988.2)
***Environmental Science*, 195 researchers**							
PhD,	3	3-5	4	4	16-60	34	36
Post Doc	17	2-59	9(6;12.9)	12.8(5.6;20)	10-642	41(25;56)	91.7(11.1;172.2)
Assis Prof	39	2-46	18(13.9;20)	19(15.6;22.5)	0-573	148(90.6;167.6)	185.4(133.7;237.1)
Assoc Prof	85	1-103	29(25;41)	36.8(31.7;42)	2-2519	326(232.9;459.4)	520.1(404.4;635.7)
Professor	51	1-425	51.5(39.3;64.2)	59.7(46.8;72.5)	6-14141	435(324.5;722.6)	998.1(614.7;1381.5)
***Philosophy*, 222 researchers**							
PhD	8	1-5	1(1;4.1)	2(0.6;3.3)	1-33	0.5(0;13.5)	6.2(-3.2;15.7)
Post Doc	22	1-31	4(3;8)	7(3.8;10.1)	0-235	8(1-10)	21.4(-1.9;44.7)
Assis Prof	44	1-106	6.5(4;8.9)	10.8(5.7;15.9)	0-1829	6.5(3;20)	74.3(-11.5;160.2)
Assoc Prof	73	1-45	7(6;9)	10(7.8;12.1)	0-565	8(5;13)	50.7(22.7;78.7)
Professor	75	1-140	18(13.5;23.4)	28.1(21;35.2)	0-3495	29(20.5;65.6)	157(52.1;262)
***Public Health*, 132 researchers**							
PhD	9	4-27	8(7.1;17.8)	12.2(6.6;17.8)	7-253	60(34.5;146.7)	82.2(23.5;140.8)
Post Doc	14	1-23	11(8.8;14.4)	12(8.6;15.3)	0-353	80.5(21.5;203.9)	113.6(49.4;177.6)
Assis Prof	30	3-288	22(13.1;29.6)	36.2(15.6;56.7)	10-3796	167(107.8;350.8)	417.4(131.4;703.3)
Assoc Prof	50	4-221	43(30.6;56.3)	54.6(41.6;67.7)	4-3649	518(312.6;701.7)	778.5(539.4;1017.5)
Professor	29	5-661	76(53.6;107.6)	110.2(62.7;157.7)	13-13520	954(554,2;2394.7)	2104(1065.3;3142.6)

Table 1 provides an overview of the data set used in this study showing publication and citation data distributions across the four disciplines and the academic seniorities of the 741 researchers in the sample. The four disciplines are very broad and comparison of scientists within each discipline and across sub disciplines is not recommended in practice as publication and citation behaviour differ greatly. However in this quantitative study, trends of indicator performance on a disciplinary level are identifiable. Preliminary data exploration shows that *Astronomy* has a strong preference for multi-authorship and article publication; *Environmental Science* publishes a great amount of conference papers and are only partially represented in Web of Science; *Philosophy* is a dialogue-based discipline, preferring single authorship and publishing in blogs, books and in national languages whereas *Public Health* has a strong tradition of publishing articles in international journals indexed in the citation databases, but also publishes a fair amount of articles in local journals in national languages as issues often concern local health issues and regulations. Only *Public Health* researchers exhibit regular publication trends that can be captured by average measures at the seniority level; the other three disciplines suggest highly individual production rates where averages rates do not match well with seniority level.

3 Methods

As reported in Wildgaard, Schneider and Larsen (2014), the usability of indices is a major consideration therefore the complexity of each indicator was assessed. The indices were graded on a 5 point numerical scale to assess 1) the availability of citation data and, 2) the intricacy of the mathematical model required to compile the indicator. This assessment might result in a reduction of the granularity and sophistication of the indices we identify as useful, and might even encourage the use of rougher measures over more accurate ones. The indices have to measure what they purport to measure, however, usability is lost if correct measurement requires data that is not readily available to the researcher, difficult mathematical calculations, and intricate interpretations of complicated data output. We assume the user of the indicators has a complete publication list and would only need to find citations and calculate the indicator. Only indicators that we scored ≤3 (on a scale where 5 was highest complexity / data collection required) were considered for the analysis. Simplicity is an important criterion for researcher-level indicators because it is more often than not librarians, information specialists, administrators or even researcher's themselves that use them to compare and discriminate between scholars in an evaluation. This results in 37 potentially useful indicators at the individual level that are analysed in this paper. These indicators are supplemented by 17 field level performance indicators supplied by CWTS. For an overview see Table 4 where the indicators are briefly presented along with information of the data they have been derived from and the various factors that are applied in their calculation. For details on their calculation please refer to Appendix 2 as well as Wildgaard, Schneider and Larsen (2014).

The set of selected indicators is intended to capture the major output and effects of a researcher's published work that can be captured using publication and citation data. Figure 1 provides a systematic overview of the indicators and the relations between them. Indicators in *blue* pertain to publication output, and counts publications in various ways. Indicators in *green* measure the effect of output and are based on raw citation count such as **C** or fractionalised citation counts, as well as average citations of the entire portfolio, for example **CPP**. Indicators in *red* measure impact over time, e.g. with citations adjusted for length of academic career such as **AW**, and are often adjusted to field norms such as **IQP**. Indicators in *purple* measure citations to core or selected publications, e.g. **H**. All these indicators are simple to calculate but in prioritizing simplicity our method may resulted in choosing coarse measures of performance. Therefore, we compare these relatively simple indicators to the more sophisticated indicators of expected performance that are CWTS field standards, indicated in *yellow* such as **pp top prop**, **mnjs**, etc.

3.1 Data analysis

The primary purpose of this report is to analyse and compare different bibliometric indicators using the citation and publication records of individual scientists. We wish to investigate if the simple or sophisticated indicators discriminate just as well between the scientists of different academic seniorities and disciplines. From this point of view, the best choice of indicators will be dependent discipline, academic seniority and complexity. We will address the recommendation of indicators using standard statistical methods.

For each discipline we also computed a correlation matrix for the indicators using Kendall's tau rank correlation coefficient, which is a standard correlation measure for non-parametric data. Kendall's tau is a non-parametric test that measures the correlation of the ranks of the samples instead of the actual values. This means it bases the correlation on the extent pairs of variables agree, and is effective for smaller sample sizes and is insensitive to errors. Perfect agreement tau=1, independence tau=0 and

increasing values between -1 and 1=increasing agreement between the variables. We used IBM SPSS version 19 for the statistics.

3.2 Limitations of the analyses

The exclusion of the 3,693 records that were mainly in conference proceedings had a great effect on the *Astronomy* sample; see Table 2 and Table 3. Some researchers lost up to 80% of their publications. Appendix 1 presents a detailed overview. Basic citation data on these publications can be identified in WoS and it will be possible to calculate a selection of the indicators in Table 2 for these publications. This is, however, beyond the scope of this paper and we leave this for future work.

Our experience with the missing data, illustrates how important it is in a bibliometric evaluation to report the version of the citation index the data is collected from, e.g. version of WoS. In our case, the publication and citation analysis in the present study is limited to articles and reviews and is based on information indexed in the version of WoS data that we use. Such information must be reported in an evaluation report to enable third parties to understand what is included and is not included in the evaluation.

Table 2. Effect of removing papers on a disciplinary level.

	***N* with publication and citation information**	***N* without publication and citation information**	**Total**	**%**
Astronomy	12,359	2,467	14,826	16,6
Environment	7,820	863	8,683	9,9
Philosophy	3,494	264	3,758	7
Public Health	7,294	99	7,393	1,3
total	**30,967**	**3,693**	**34,660**	

Table 3. Percent missing publications by level of seniority.

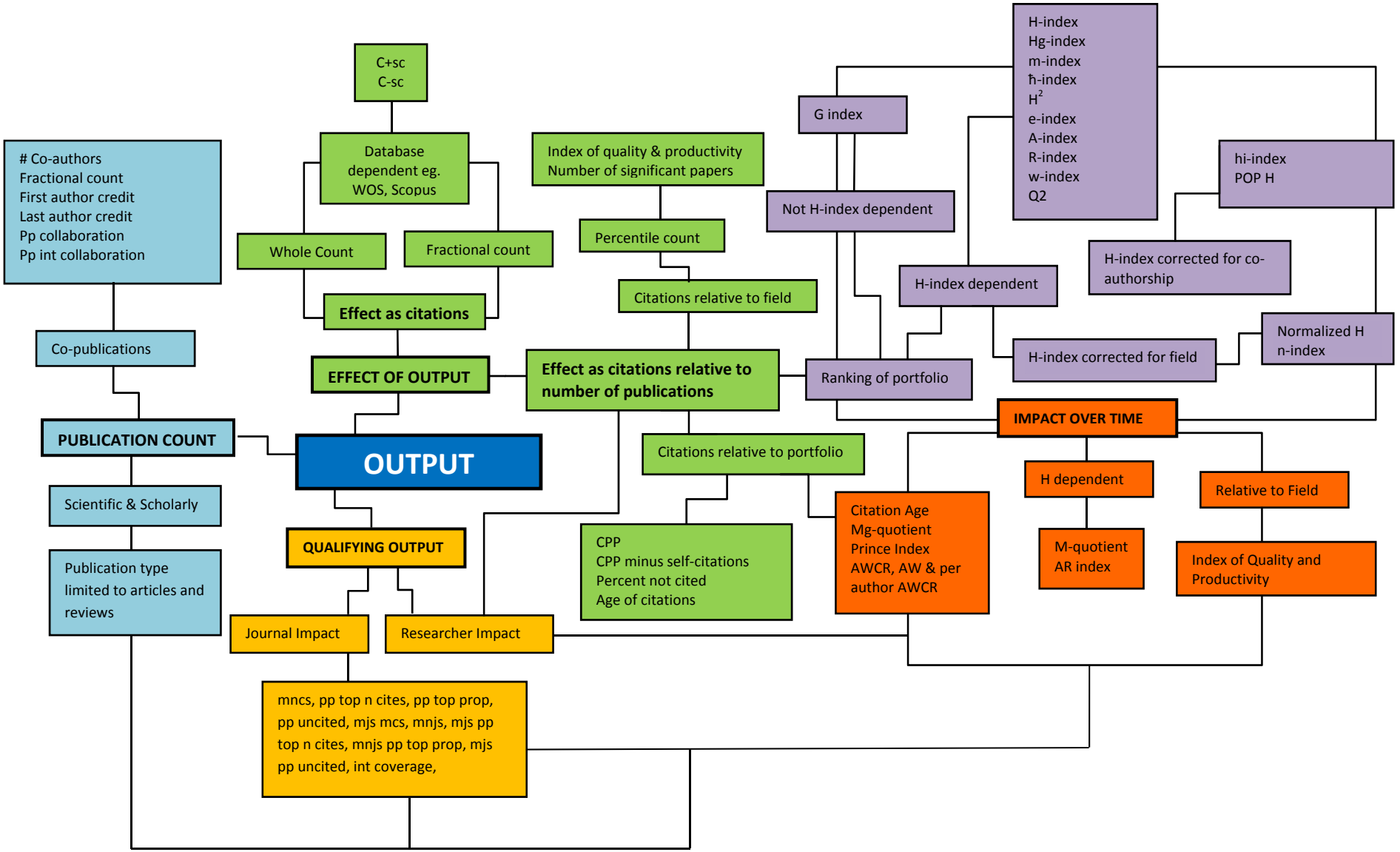
	PhD	**Post Doc**	**Assistant Prof.**	**Associate Prof.**	**Professor**
Astronomy	12,4	13	13,9	16,6	18,4
Environmental	7,6	20,1	7	6,9	12,2
Philosophy	0	6,6	7,3	3,3	8,2
Public Health	0	0	0,9	0,9	1,9

Table 4. Indicators of individual impact as well as discipline benchmarks analysed in this study.

ID	Type	Abbr.	Indicator	Intention
Productivity metrics				
1	Publication	**P**	Publication count	Total count of production used in formal communication. Limited in our dataset to ISI processed publications
2	Publication	**Fp**	Fractionalized publication count	Each of the authors receive a score equal to 1/n to give less weight to collaborative works
3	Publication	**App**	Average papers per author	Indicates average amount of collaboration per paper
4	Publication/time	**Pyrs**	Years since first publication	Length of publication career from 1st article in dataset to 2013
Impact metrics				
5	Citation	**C**	Citation count	Use of all publications
6	Citation	**C-sc**	Citation count minus self-citations.	Use of publications, minus self-use.
7	Citation	**Sig**	Highest cited paper	Most significant paper
8	Citation	**minC**	Minimum citations	Minimum number of citations
9	Citation	**%sc**	Percent self-citations	Disambiguate self-citations from external citations
10	Citation/author	**Fc**	Fractional citation count	Remove dependence of co-authorship, all authors receive equal share of citations.
11	Citation/time	**C<5**	Citations less than 5 years old	Age of citations
Hybrid metrics				
12	Citation/publication/field	**IQP**	Index of Quality & Productivity	Number of citations a scholar's work would receive if it is of average quality in the field
13	Citation/publication/field	**Tc>a**	(part of IQP)	Actual times scholar's core papers are cited more than average quality of field
14	Citation/publication/field	**H norm**	Normalized h	Normalizes h-index (to compare scientists across fields).
15	Citation/publication	**Cage**	Age of citation	If citations are due to recent or past articles
16	Citation/publication	**%PNC**	Percent not cited	If citations are due to a few or many articles
17	Citation/publication	**CPP**	Citations per paper	Average citations per paper
18	Citation/publication	**h**	h index	Cumulative achievement
19	Citation/publication	**g**	g index	Distinction between and order of scientists
20	Citation/publication	**m**	m index	Median citations to publications included in h to reduce impact of highly cited papers
21	Citation/publication	**e**	e index	Supplements h, by calculating impact of articles with excess h citations
22	Citation/publication	**w**	wu index	Impact of researcher's most excellent papers
23	Citation/publication	**hg**	Hg index	Balanced view of production by keeping advantages of h and g, and minimizing their disadvantages
24	Citation/publication	**H^2**	Kosmulski index	Weights most productive papers
25	Citation/publication	**A**	A index	Magnitude of researcher's citations to publications
26	Citation/publication	**R**	R index	Improvement of A-index
27	Citation/publication	**AR**	AR-index	Citation intensity and age of articles in the h core
28	Citation/publication	**ħ**	Miller's h	Overall structure of citations to papers
29	Citation/publication	**Q^2**	Quantitative & Quality index	Relates the number of papers and their impact
30	Citation/publication/author	**hi**	individual h	Number of papers with at least h citations if researcher had worked alone

ID	Citation/publication/author	POP h	Harzing's publish or perish h index	Accounts for co-authorship effects
31	Citation/publication/author/time	AWCR	age weighted citation rate	Number of citations to all publications adjusted for age of each paper
32	Citation/publication/author/time	AW	Age weighted h	Square root of AWCR to avoid punishing researcher's with few very highly cited papers. Approximates h index
33	Citation/publication/author/time	AWCRpa	Per-author AWCR	Number of citations to all publications adjusted for age of each paper and number of authors
34	Citation/publication /time	M quotient	m-quotient	Age weighted h. H divided by years since first publication
35	Citation/publication/time	Mg	Mg-quotient	Age weighted g. G divided by years since first publication
36	Citation/publication/time	PI	Price Index	Percentage references to documents not older than 5 years at the time of publication of the citing sources
37	Citation/publication/field	IQP	Index of Quality & Productivity	Number of citations a scholar's work would receive if it is of average quality in the field
Journal-field benchmarks, calculated by CWTS				
38	Crown Indicator	mcs	mcs	Mean citation score
39	Crown Indicator	mncs	mncs	Mean normalized citation score.
40	Crown Indicator	pp top n cites	pp top n cites	Proportion of top papers
41	Crown Indicator	pp top prop	pp top prop	Proportion in top 10% of world
42	Crown Indicator	pp uncited	pp uncited	Proportion uncited
43	Crown Indicator	mjs mcs	mjs mcs	Crown-type indicator
44	Crown Indicator	mnjs	mnjs	Mean normalized journal score
45	Crown Indicator	mjs pp top n cits	mjs pp top n cits	Crown-type indicator
46	Crown Indicator	mnjs pp top prop	mnjs pp top prop	Crown-type indicator
47	Crown Indicator	mjs pp uncited	mjs pp uncited	Crown type indicator
48	Crown Indicator	prop self cits	prop self cits	Proportion self-citations
49	Crown Indicator	int coverage	int coverage	Internal coverage.
50	Crown Indicator	pp collaboration	pp collaboration	collaboration
51	Crown Indicator	pp int collab	pp int collab	International collaboration
52	Crown Indicator	n self cites	n self cites	Number of self-citations

Figure 1. Relationship between the analysed indicators and the publication activities they purport to measure.



4 Results

4.1 Association between seniority and bibliometric indicators

The assumption behind this analysis is that knowing the seniority of the researcher will improve the prediction of the performance of the indicator.

We used *gamma* as the symmetric measure of association and cross-tabulated seniority and the bibliometric indicators, discipline by discipline. The value of gamma tends to be large due to how it is calculated, so Kendall's tau-c (for non-square tables – like a 2 x 3 table) are often preferred. Gamma is a Proportional Reduction of Error, which is interpreted as the improvement in predicting the dependent variable that can be attributed to knowing a case's value on the independent variable. Because gamma is a proportional reduction in error we can suggest that the following indicators are potential useful predictors of discipline specific seniority performance, Table 5. For simplicity we report only the indicators that are improved by ≥10%.

Astronomy
Knowing the seniority of the researcher will improve the prediction of the performance of **minimum number of citations** (51%), **Price Index** (20%), **minimum mjs mcs** (23%), **average mjs** (12%) and **normalized h** (16%).

Environmental Science
Knowing the seniority of the researcher will improve the prediction of the performance of **minimum citations** (25%), **Years since first publication** (24%), **Citations** (11%), **Publications** (16%), **Fractionalized papers** (18%), **number not cited papers** (17%), **Citation age** (18%), **Most significant paper** (10%), **Cites minus self-citations** (12%), **Fractional citations** (14%), **sum pp top n cites** (12%), **sum pp top prop** (16%), **h index** (14%), **g** (10%), **h2** (11%) and **POP h** (13%).

Philosophy
Knowing the seniority of the researcher will improve the prediction of the performance of **Years since first publication** (18%) and **Wu** (16%).

Public Health
Knowing the seniority of the researcher will improve the prediction of the performance of **AWCR_pp** (13%), **minimum citations** (36%), **minimum mjs mcs** (13%), and **times cited more frequently than the average paper in the discipline** (12%).

Across all disciplines
Knowing the seniority of the researcher will improve the prediction of the performance of **number not cited** (19%) and **percent not cited** (49%). All other indicators displayed minimum or no association.

Table 5. Analysis of prediction power of bibliometric indicators when knowing the seniority of a researcher. Proportional Reduction of Error gamma values of 10% or more are interpreted as indicating an association.

Discipline	No association	Minimal association ≤10%	Moderate association 11~50%	Strong association ≥51%

Astronomy	**App, Pyrs, cpp, c, p, fp, nnc, %nc, %sc, cage, AWCR_c, AW, AWCR_au, Sig, ħ, C-sc, Fc, sum pp top n cit, sum pp top prop, average mjs mcs, max mjs mcs, IQP, mg, e q2, h2, AR, POPh, productivity adjusted papers, h, mquot, m, A, R, g, hg, WU, cites <5 yrs**	**AWCR_pp, times cited more frequently than average papers**	**PI, min mjs mcs, average mjs, h norm**	**min n cites (51%)**
Environmental Science	**App, %sc, %nc, AWCR-pp, PI, min mjsmcs, times cited more frequently than average papers, mquot, hnorm, wu, mg, AR**	**Cpp, sc, AWCR_c, AWCR_au, AW, max cites, average mjs mcs, max mjs mcs, IQP, m, A, R, e, q2, h2, cites <5yrs**	**Pyrs, C, P, fp, nnc, cage, sig, ħ, min cites, max cites, c-sc, fc, sum pp top n cites, sum pp top prop, Nproductivity adjusted papers, h, g, hg, poph**	**-**
Philosophy	**%sc, %nc, AWCR_pp, AWCR_au, min cites, PI, min mjs mcs, gennemsnit mnjs, times cited more frequently than average papers, mquot, hnorm, mg**	**App, cpp, c, sc, p, fp, nnc, cage, AWCR_c, AW, sig, ħ, C-sc, fc, sum pp top n cites, sum pp top prop, average mjs mcs, max mjs mcs, IQP, h, m, A, R, g, hg, wu, e, q2, h2, AR, hpop, cites <5yrs**	**Pyrs, nproductivity adjusted papers,**	**-**
Public Health	**Pyrs, P, Fp, nnc, %nc, cage, AWCR_au, max cites, sig, Fc, PI, productivity adjusted papers, h, Q2, poph**	**App, cpp, c, sc, %sc, AWCR_c, AW, cites <5yrs, AR, ħ, c-sc, sum pp top n cites, sum pp top prop, average mjs mcs, min mjs mcs, max mjs mcs, average mnjs, IQP, mquot, hnorm, m, A, R, g, hg, mg, e, h2**	**AWCR_pp, min cites, times cited more frequently than average paper,**	**-**

Generally the prediction of the performance of h-type indicators to seniority was minimal or no association. This makes sense, as these indicators are dependent on citations and publications also being predictors of performance on a seniority level, which is only the case in Environmental Science. That is why we can only indicate a trend towards h-type indicators being a performance predictor on seniority level in the discipline of *Environmental Science*, and that said the improvement is only between 9-14%. Across *Astronomy*, *Environmental Science* and *Public Health* there appears to be a trend towards a minimum citation limit within seniority, as **minimum citations** is a moderate to strong indicator of performance, 25-51%. This echoes our findings in the Google Scholar data (D5.8 Part 5) where we concluded that minimum citations per paper (**minCPP**) can be used as expected seniority performance benchmarks. Whereas in Google Scholar **minCPP** was a strong indicator, on this WoS data **minimum total citations** is a better associative indicator, thus illustrating that indicators do not only perform differently between disciplines but also between citation indexes or versions of the same citation index used to collect the data.

4.2 Identifying central indicators across disciplines

In this analysis we are inspired by Franceschet (2009) and analyse which indicators display high correlations to other indicators. The purpose is on one hand to identify indicators that are highly correlated to other indicators, and on the other to identify indicators that practically measure the same inherent properties. If indicators can be grouped by such an analysis into clusters of highly similar indicators, then the simpler alternatives from each cluster can be recommended over more complex ones – thus making it more feasible for individuals to calculate them.

We first attempt to identify central indicators for each discipline and then compare across disciplines. To answer this question we constructed correlation matrixes of the sample for each discipline. The Kendall correlation matrices are shown in Appendix 3-6.

Table 6 uses data from the correlation matrices to highlight isolated indicators, meaning that they do not have any strong links, defined as over 0.7, to any other indicator in the correlation. In the third column of the table the most central indicators are highlighted, that is the indicators with the highest number of links over 0.7 to other indicators in the matrix (indicated in column 4).

Table 6. Isolated and highly correlated indicators across disciplines.

Discipline	Isolated Indicators	Central Indicators	Number of links to other indicators
Astronomy	**App, sum sc, AWCR_pp, fp, %nc, average mjs mcs, min mjs mcs, maxs mjs mcs, average mnjs, h norm, wu**	**Hg** **IQP, AR**	25 24
Environmental Science	**Pyrs, App, %sc, Fp, nnc, %nc, Cage, AWCR_pp, PI, average mnjs, min mjs mcs, maxs mjs mcs, nproductivity adjusted papers, wu, AR**	**H, h2** **popH, Q2, e, IQP**	26 25
Philosophy	**App, %sc, nnc, &nc, PI, sum pp top prop, average mjs mcs, max mjs mcs, average mnjs, nproductivity adjusted papers, hnorm, Wu**	**IQP** **AR, h2, Q2, e, g, h**	28 27
Public Health	**Pyrs, app, %sc, nnc, %nc, cage, AWCR_pp, minC, PI, min mjs mcs, average mnjs, nproductivity adjusted papers, hnorm, Wu**	**g** **Hg, ħ, h2**	23 22

The central indicators all hybrid indicators, that is, indicators that in their calculations adjust in some form citations to number of publications. To investigate the role of the identified central indicators, we ranked researchers within disciplines and mapped how their position in the ranks changes when using the central indicators as the control. We identified the top 10%, top 25%, middle 50% and bottom 25% in each set. In *Astronomy* we used the **hg** index as the ranking factor, in *Environmental Science* the **h** index, in *Philosophy* the **IQP** index and in *Public Health* we used the **g** index. Across all disciplines we observed the same trend. If a researcher is placed in the top 10% of the sample by the central indicator, the researcher is placed in the top 10% using the other indicators that the central indicator has strong links to. Likewise, for researchers in the top 25%, middle 50% and bottom 25%. For example a researcher in **Public Health** scores in the middle 50% on the **g** index, will be placed in

the middle 50% on the other 23 indicators the **g** index has strong links to. The **g** index has strong links to **C**, **sc**, **P**, **AWCR**, **AWCR_au**, **AW**, **max cites**, **Sig**, **Fc**, **sum top pp prop**, **sum pp top prop**, **IQP**, **h** , **m**, **A**, **R**, **hg**, **e**, **h**, **Q2**, **h2**, **AR** and **POPh**. This group represents indicators of production, crown type indicators, hybrid indicators and raw publication and citation counts. Further we noticed that the isolated indicators produce a very random rank, placing a researcher sometimes in the top 10% and sometimes in the bottom 25%. This observation needs to be supported by further statistical analyses, where we investigate the overlap between the central indicators and the indicators they link to, to understand which aspects of the effect of a researchers' production they capture.

4.3 Identifying central indicators for each discipline

Here we attempt to apply clustering techniques to recommend single indicators that represent independent aspects of research performance. To continue the analysis of central indicators and how they cluster other indicators around them we now consider the output of the correlation analysis using the ALSCAL procedure in SPSS. The clustering is shown as two-dimensional models of Euclidean distance (i.e. maps), which illustrate the association between indicators by measuring the distance between them as points on a two-dimensional plane with coordinates (x,y) and (a,b). To get an idea of how well the clustering model fits the data, we report the S-stress as a measure of fit ranging from 1 (worst possible fit) to 0 (perfect fit) and R-square to illustrate how much of the variance in the model is explained by the two dimensions. In general, in the results presented below the fit is low and the stress high indicating that the maps do not capture the complexity of higher dimensions that well when transformed into 2 dimensions. For this reason we choose to supplement the maps with a hierarchical clustering algorithm that starts the clustering with the pair of indicators that have smallest squared Euclidean distance between them. The output is a dendogram – i.e. a tree diagram that illustrates the arrangement of clusters. The branch-like nature of the dendogram allows you to trace backward or forward to any individual case or cluster at any level. In addition it gives an idea of how great the distance is between cases or groups that are clustered in a particular step, using a 0-25 scale along the top of the chart. While it is difficult to interpret distance in the early clustering phases (the extreme left of the chart), as you move to the right relative distance become more apparent. The bigger the distances before two clusters are joined, the bigger the differences in these clusters. To find membership of a particular cluster trace backwards down the branches to the name.

4.3.1 Astronomy

The central indicator for astronomy is the hg index, marked with an arrow. S-stress=0,375 and R^2 =0,253, only 25% variance is explained by the model. This is a very coarse grouping of indicators.

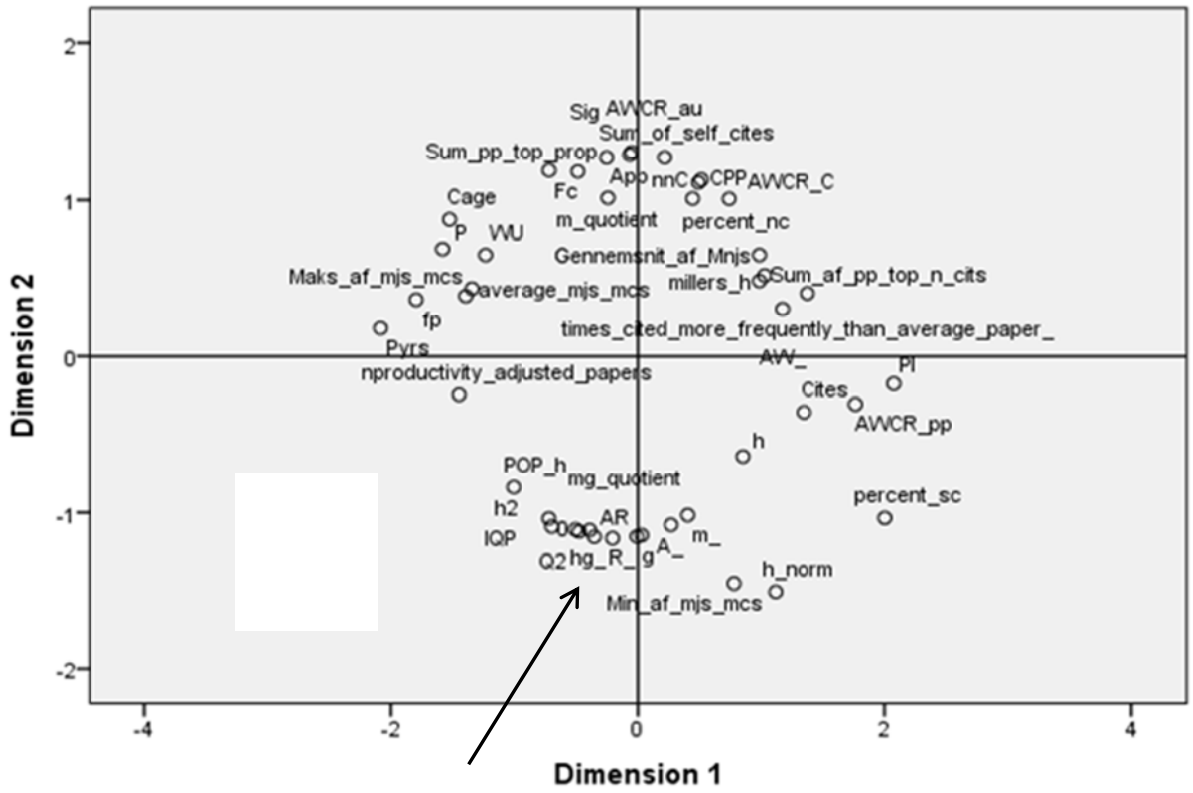


Figure 2. Multidimensional Scaling map of the studied bibliometric for *Astronomy*.

The indicators are roughly grouped into 3 correlation clusters, the most intense cluster is the hybrid indicators that group around the **hg** index. The second cluster is heavily dominated by publication based indicators, which gather in an arch at the top of the figure from **number of productivity adjusted papers** through to **AW** index. The third is a cluster of isolated indicators **%sc**, **PI**, **AWCR_pp**, **hnorm** and **min mjs mcs**. **Citations** (**C**) and **h** index appear to fall outside the clusters.

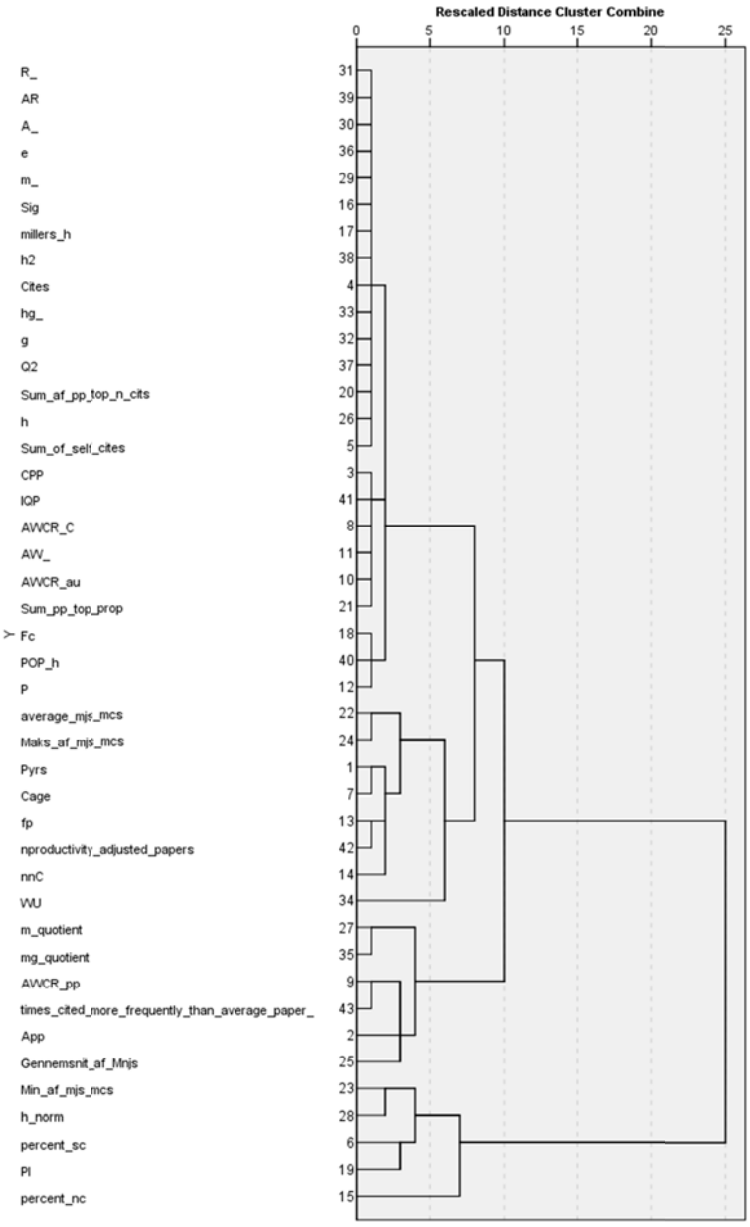


Figure 3. Hierarchical clustering dendrogam of the studied bibliometric for *Astronomy*.

Our observations about isolated indicators are confirmed. These indicators potentially measure researcher impact not covered by the other indicators. The resulting partition contains 4 clusters. One main cluster of hybrid indicators (**R** through **Sc**), and three smaller clusters that illustrate less intense relationships between the indicators. These clusters have expected field performance indicators (crown indicators) mixed in with them: paper-based metrics (**CPP** to **Sum pp top prop**), production adjusted for age or discipline (**average mjs mcs** through **Wu**) and finally a mix of time dependent metrics and researcher-adjusted metrics.

4.3.2 Environmental Science

The model explains 24% of the variance (R2), S-stress=0.378. The central indicators **h** and **h2** are marked with arrows and fall within the same cluster. Four clear clusters are visible with **percent sc** falling outside of these. These four identifiable groups are hybrid indicators, cite-based indicators, indicators of production and crown type indicators (expected field performance).

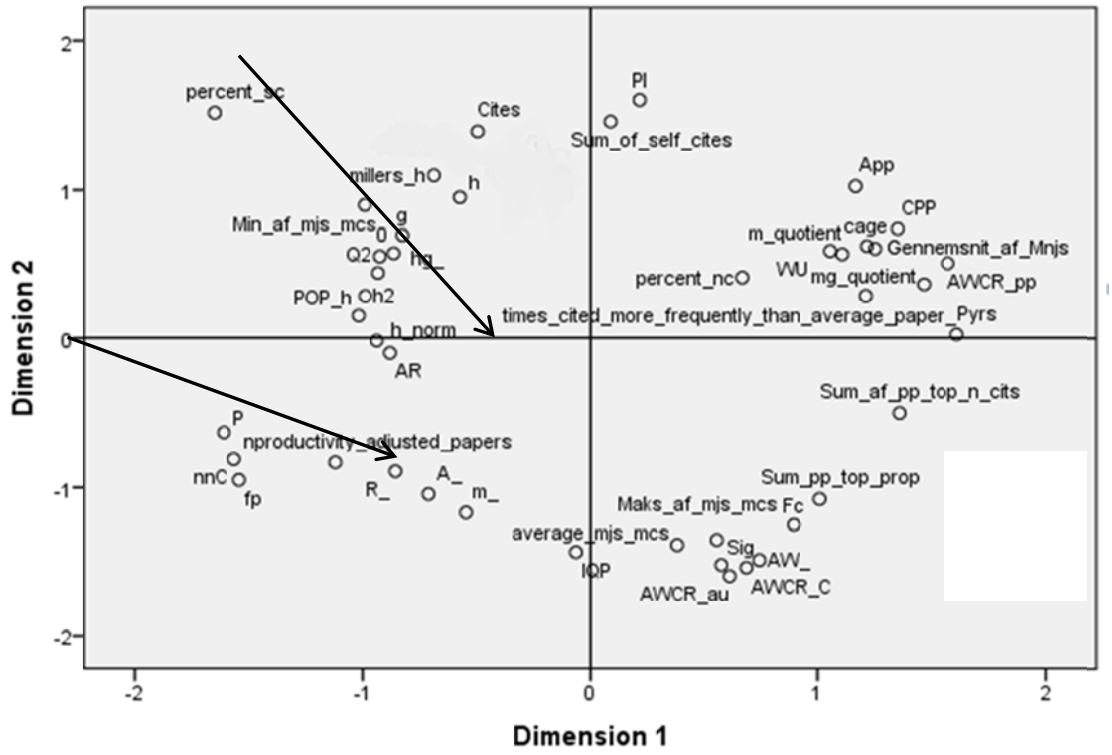


Figure 4. Multidimensional Scaling map of the studied bibliometric for *Environmental Science*.

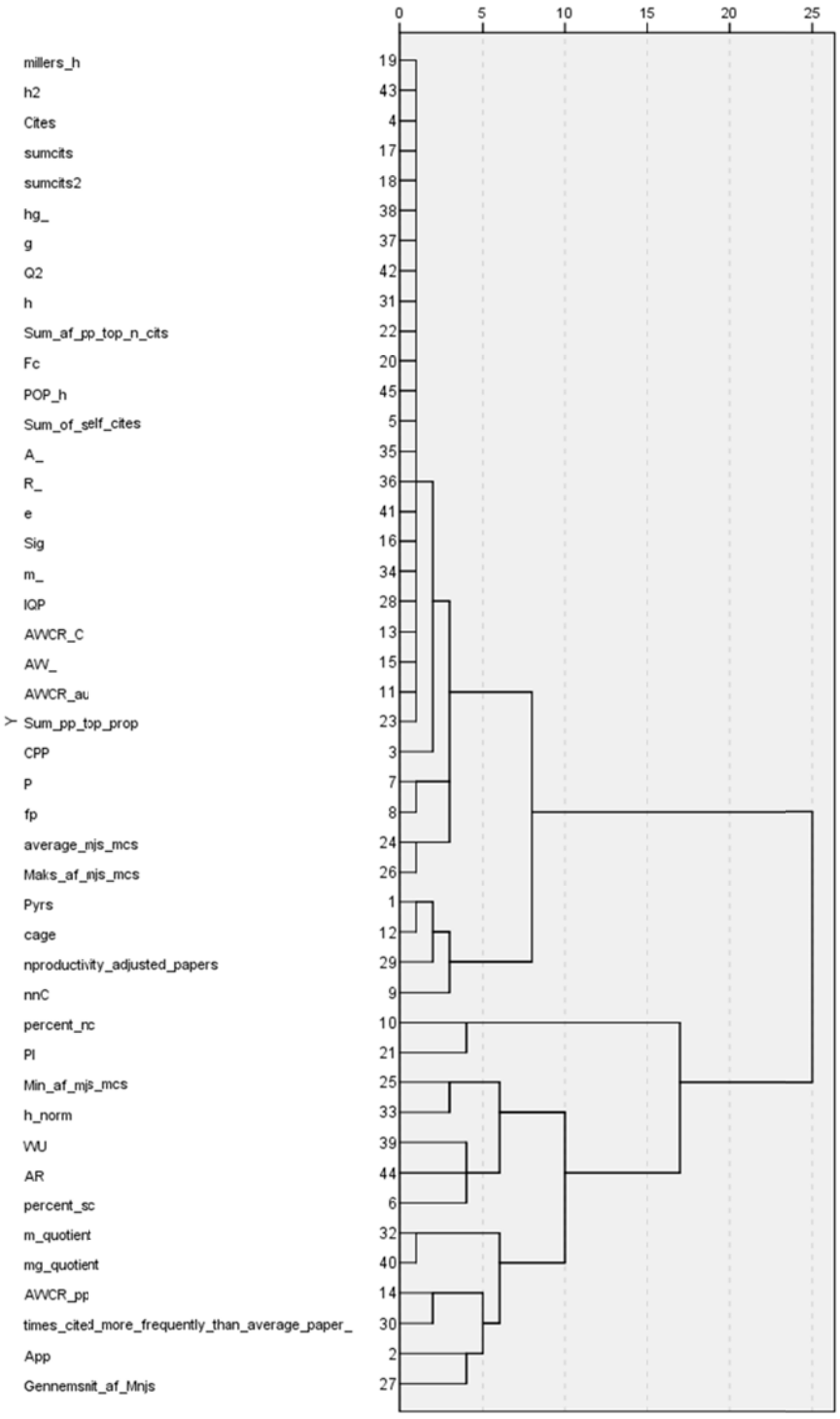


Figure 5. Hierarchical clustering dendrogam of the studied bibliometric for *Environmental Science*.

The distance between the clusters is easier to read in the dendogram. The hybrid h indicators (**millers_h** through **sum pp top prop**) form a tight group, while the remaining indicators form 6 smaller and more loosely related groups. The paper-based indicators **p** and **fp** form one group, indicators of production another group (**Pyrs, Cage, nproductivity adjusted papers** and **nnc**), the isolated indicators (**%nc** and **PI**); a seemingly random cluster of indicators (**min mjs mcs** to **%sc**), the crown indicators **average mjs mcs, max mjs mcs**; and finally indicators that account for age or time (**m-quotient** through **average of mnjs**).

4.3.3 Philosophy

The model is a better fit, R^2 explaining 47% of the variance. S-stress=0.38. The central indicator **IQP** is marked with an arrow. Three clusters are presented. Hybrid indicators group at the top of the figure (**A** through **mg**-quotient), a group of paper-based indicators in the top left **(times cited more frequently than average paper** to **P**) and a large mixed group of the remaining indicators that includes our central indicator. The **Percent not cited** indicator falls outside any grouping.

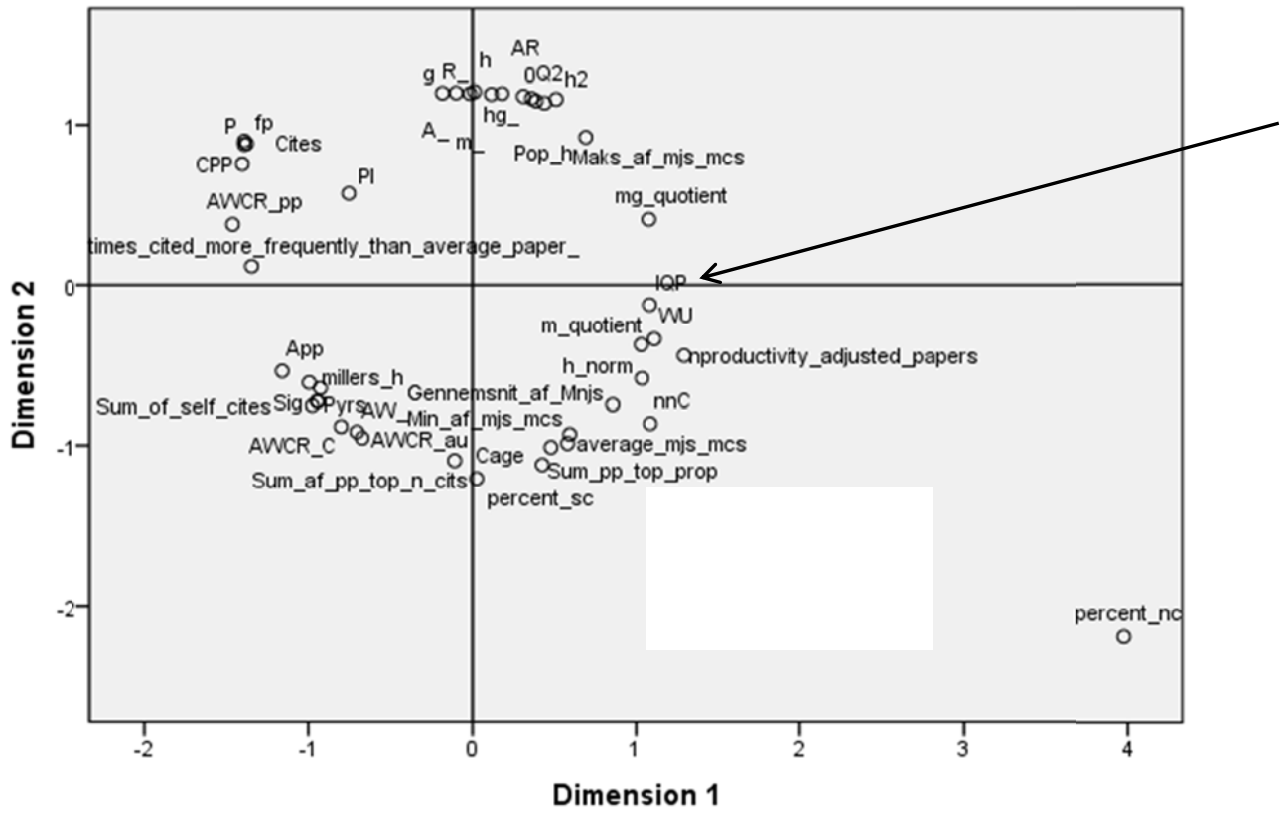


Figure 6. Multidimensional Scaling map of the studied bibliometric for *Philosophy*.

The dendrogam illustrates the distance of the groups of indicators from each other. The hybrid and crown-type indicators are closely related and group strongly with a second cluster of production indicators (**p** through **average of mnjs**). More distant relations with the cluster of ratio based indicators are illustrated, **AWCR_pp** through **h_norm**, and with the fourth group that consists of a mix of time, citation and paper adjusted indicators. **Percent not cited** and **PI** (price index) are only related to the other indicators on a very distant level.

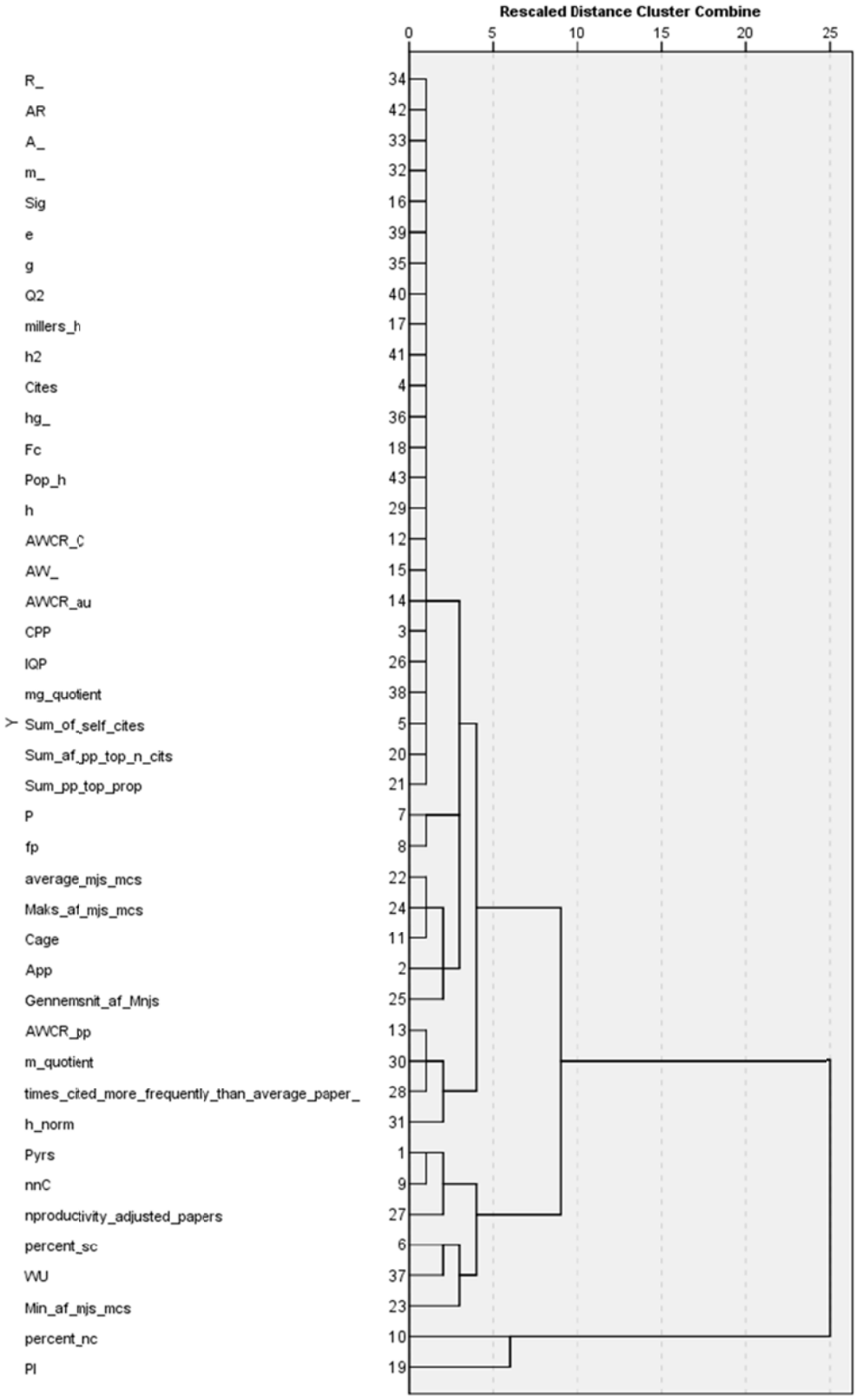


Figure 7. Hierarchical clustering dendrogam of the studied bibliometric for *Philosophy*.

4.3.4 Public Health

38% of the variance is explained by the model (R^2), S-stress=0.499. The central indicator **g** is marked with an arrow. It is very difficult to deduce independent clusters in the distance model, below. We suggest two clusters. The small cluster in the bottom right of the frame, from **AWCR_pp** to **min mjs**

mcs, and the large cluster of remaining indicators that spread across the centre of the diagram. **Publication years** (Pyrs) is the clear outlier.

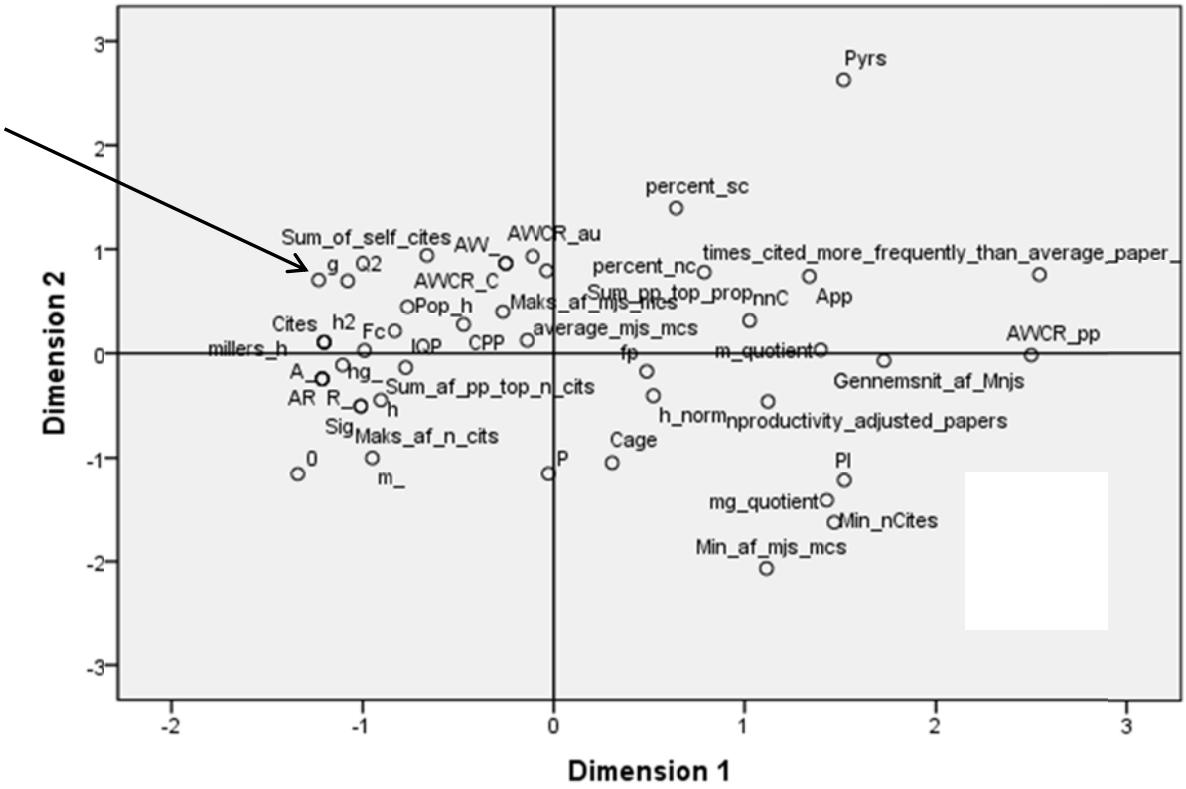


Figure 8. Multidimensional Scaling map of the studied bibliometric for *Public Health*.

The dendrogam is more informative. Hybrid indicators and indicators adjusted for author contribution form one large cluster, and are closely related to two crown indicators (**average mjs mcs** and **maks mjs mcs**). Paper-based metrics form their own cluster (**Pyrs** through **productivity adjusted papers**). The last three clusters are distantly related to the aforementioned clusters and the indicators within these three only loosely related to each other. Hence they present groupings of miscellaneous indicators. Again the **%not cited**, **% self-citations** and **Price Index** (**PI**) are only very distantly related to the other indicators.

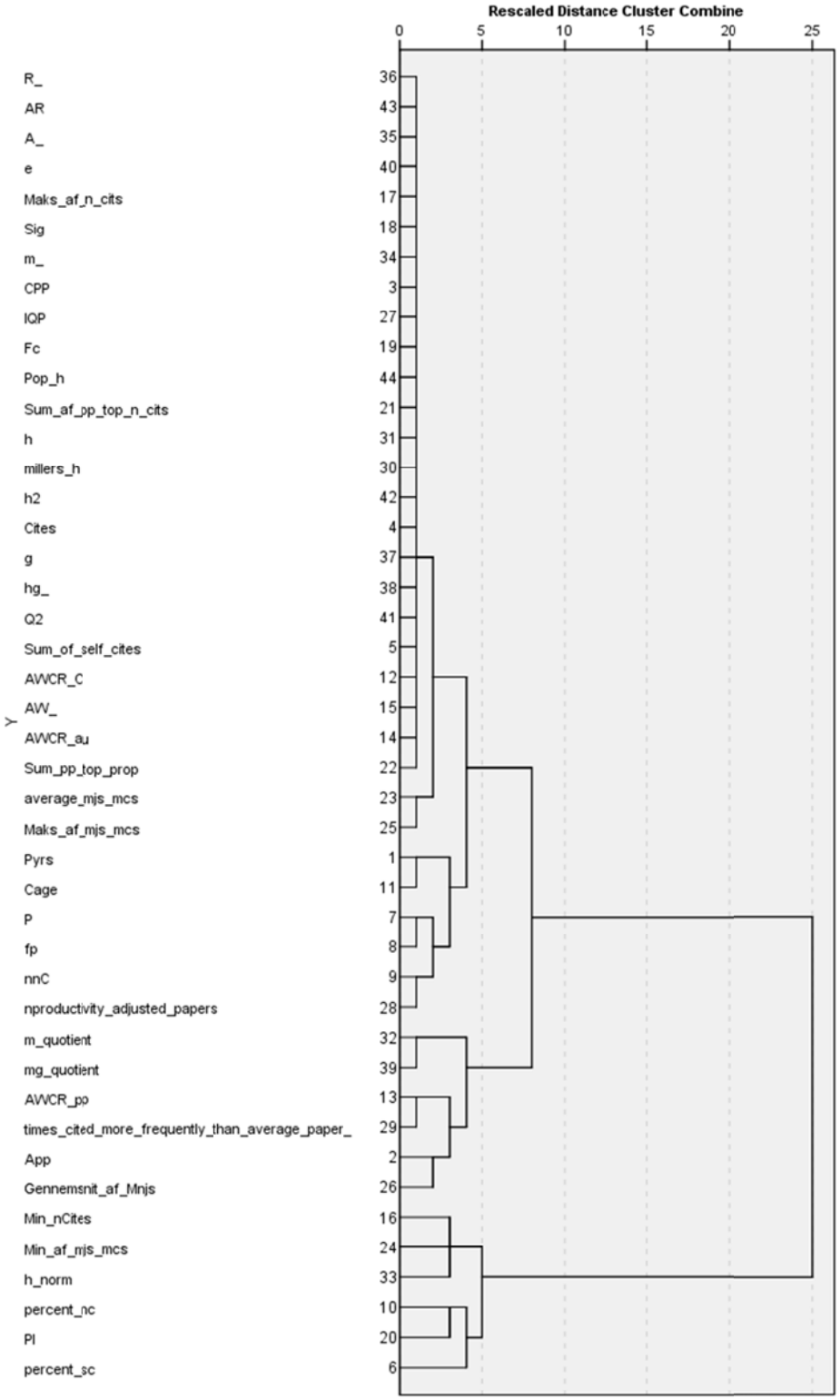


Figure 9. Hierarchical clustering dendrogam of the studied bibliometric for *Public Health*.

4.3.5 Discussion and recommendations

We posed the question if using clustering structures is a good method to recommend single indicators that represent independent aspects of research performance. The hierarchical clustering illustrates that choosing one central indicator will not measure all aspects of the effects of a researchers publication. At an overall level, the indicators group together in indicators of production, citations, production & citations, production adjusted for time, production adjusted for discipline and miscellaneous isolated indicators that measure the more subjective aspects of a researcher's publishing portfolio. We note that the clustering of indicators is different from discipline to discipline, and no unified picture emerges across the disciplines. However, in each of the disciplines our analysis has identified central indicators and isolated indicators.

Isolated indicators are interesting because they measure aspects of the effect of publications not captured by other indicators. The **Price Index** for instance, identifies the currency of citations to papers: Is a citation count due to recent papers or papers published many years ago? A moderate association was found between knowing the seniority of the researcher and predicting the researcher's performance using isolated indicators.

Identifying central indicators illustrates the different roles of citations in the four disciplines and the power a single indicator has in researcher rankings. Interestingly for Philosophy it is an indicator that adjusts for disciplinary expected average citations and publishing age of the researcher, the **IQP** indicator. The other three disciplines that have a strong tradition for publishing and citations display the same preference for hybrid indicators. In *Astronomy* the **Hg** index is central. **Hg** is more granular than **h** and **g** indices, minimizes the effect of very highly cited papers to calculate a fairer version of the **h** index. This makes sense, as it is a disciplinary trait in our *Astronomy* set, that researchers commonly have one or two multi-authored papers that are very highly cited. In *Public Health* the **g** index is the central indicator, and as such is sensitive to highly cited papers – a criticism of the **h** index that ignores high performing papers. Further it is usual to find different scientists with same **h** but different number of publications and cites. The **g** index presents a granular solution good for a discipline that has a strong tradition of publishing and citing. *Environmental Science* groups also around the **h** and **h2** index, which can be used together as **h** suffers from the flaw of ignoring highly cited papers and the aforementioned flaw on granularity.

If we were to recommend a performance indicator for each discipline, for each type of indicator of activity, we would need to investigate the role of the indicators within their cluster: what they measure, if they overlap, how complicated they are and which are redundant.

Table 7. Calculation of the central indicators.

Discipline	Indicator	Calculation	Type
Astronomy	**Hg**	The square root of (**h** multiplied by g).	Citation/publication
Environmental Science	**H or H2**	Publications are ranked in descending order after number of citations. Where number of citations and rank is the same, this is the **h** index Cube root of total citations	Citation/publication
Philosophy	**IQP**	a) **A**= (**mnjs** x **Pyrs** x **p**+1)/2. (number of citations if author was of average quality for field) b) **A**/number of papers (estimated performance per paper) c) define actual number of citations d) **IQP**=actual citations/b+number of papers e) calculate field impact per paper x number of papers **IQP**= expected average performance of scholar in the field, amount of papers that are cited more frequently than average and how much more than average they are cited (**Tc>a**)	Citation/publications adjusted to field and age
Public Health	**g**	Publications are ranked in descending order after number of citations. The cumulative sum of citations is calculated, and where the square root of the cumulative sum is equal to the rank this is **g**-index	Citation/publication

5 Conclusion and recommendations

The clustering identified central and isolated indicators for each discipline. To investigate the role of the identified central indicators, we ranked authors within disciplines and mapped how their position in the ranks change when using the central indicators as the control. We identified the top 10%, top 25%, middle 50% and bottom 25% researchers in each set and found that certain indicators appear to control rank position These central indicators differed from discipline to discipline. In Astronomy the hg index was the central indicator, in Environmental Science the h index, in Philosophy the IQP index and in Public Health the g index. Across all disciplines we observed the same trend. If a researcher is placed in the top 10% of the sample ranking by the central indicator, the researcher is placed in the top 10% using the other

indicators the central indicator has strong links to. The same holds for authors in the top 25%, middle 50% and bottom 25%. We also noticed that isolated indicators, PI, %nc, %sc have no strong links to other indicators and produce a very random rank positions. However, they do indicate activities that are not covered by the other indicators.

These observations need to be explored and deepened in further statistical analyses that investigate the overlap between the central indicators and the indicators they link to as well as the aspects of the effect of an authors' production they capture. Using a hierarchical clustering model that illustrated how closely related the indicators are to each other, we discovered that indicators group together in descriptors of production, citations, production & citations, production adjusted for time, production adjusted for field and miscellaneous measures that describe the more subjective aspects of a researcher's publishing portfolio. The clustering of indicators is different from discipline to discipline, as is the strength of their relation. If we were to recommend a performance indicator for each field, for each type of indicator of activity, we would need to investigate the role of the indicators within their cluster: what they measure, if they overlap, how complicated they are and which of them are redundant. The m-quotient displayed stability within disciplines and comparability across databases, please see the continuation of this study in the supplementary material.

6 Limitations

The bibliometric indicators tested in our study discriminate between high and low performing researchers, but proved ineffective in discriminating between mediocre researchers in the middle quartiles.

The values of citation analysis in junior researchers is questioned as papers accumulate citations over many years after publication, and junior researchers do not in this respect have time on their side in bibliometric evaluation. Time is a factor that must be adjusted for when comparing researcher impact.

The number of publications and citations required to make meaningful researcher assessments of junior scholars, scholars who publish in national languages and scholars who publish in other formats than articles in journals indexed in citation databases.. Other indicators of a researcher's scientific activities, not limited to publications in journals, must be considered such as altmetrics, network analysis and surveys. Our object has been to find that indicator most useful in five academic seniorities within four broad disciplines.

7 Acknowledgements

The authors wish to the *Centre for Science and Technology Studies* (CWTS) in Leiden, the Netherlands for generously providing citation data for the purposes of this study. In particular we wish to thank Paul Wouters, Clara Calero-Medina, Erik van Wijk and Rodrigo Costas for their help in extracting the data.

8 References

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

Bornmann, L., & Werner, M. (2012). How good is research really? *EMBO Reports*, 14, 226-230.
Burnhill, P., & Tubby Hille, M. (1994). On measuring the relation between social science research activity and research publication. *Research Evaluation,* 4(3), 130-152.

Lehamnn, S., Jackson, A.D., Lautrup,B.E. (2007) A quantitative analysis of indicators of scientific performance. *Scientometrics*, 76(2), pp.369-390

Franceschet, M (2009) A cluster analysis of scholar and journal bibliometric indicators. *Journal of the American Society for Information Science and Technology* 60(10). Pp.1950-1964

Sandström, E., & Sandstrøm, 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.

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.

Wildgaard, L. Schneider, J.W & Larsen, B (2014) *Bibliometric Self-Evaluation: A review of the characteristics of 108 indicators of individual performance*. Manuscript submitted for publication and under revision.

WP5 Deliverable D5.8 Part 1 ACUMEN WP5 (2013) *Literature Review (Grant Agreement no.266632)* 7th Framework Programme, SP4- Capacities. Science in Society 2010

WP5 Deliverable D5.8 Part 3 ACUMEN WP5 (2014) *Selection of two samples (Grant Agreement no.266632)* 7th Framework Programme, SP4- Capacities. Science in Society 2010

WP5 Deliverable D5.8 Part 4 ACUMEN WP5 (2014) *Consequences of indicators: effects on the users (Grant Agreement no.266632)* 7th Framework Programme, SP4- Capacities. Science in Society 2010

WP5 Deliverable 5.8 Part 5 ACUMEN WP5 (2014) *Consequences of indicators: Using indicators on data from Google Scholar (Grant Agreement no.266632)* 7th Framework Programme, SP4- Capacities. Science in Society 2010

Appendix 1: Effect of excluding proceedings papers

Researcher	Proceedings	All publications	% Proc.	Discipline	Seniority
1	4	40	10,0	astro	phd
3	1	16	6,3	astro	phd
4	2	27	7,4	astro	phd
5	3	5	60,0	astro	phd
8	4	11	36,4	astro	phd
9	5	8	62,5	astro	phd
14	4	22	18,2	astro	phd
17	1	15	6,7	astro	Post doc
22	3	16	18,8	astro	Post doc
23	6	59	10,2	astro	Post doc
24	4	13	30,8	astro	Post doc
27	4	13	30,8	astro	Post doc
28	4	44	9,1	astro	Post doc
30	2	11	18,2	astro	Post doc
31	50	100	50,0	astro	Post doc
32	4	19	21,1	astro	Post doc
33	2	48	4,2	astro	Post doc
36	5	108	4,6	astro	Post doc
37	12	29	41,4	astro	Post doc
38	1	4	25,0	astro	Post doc
39	7	21	33,3	astro	Post doc
40	3	23	13,0	astro	Post doc
41	4	40	10,0	astro	Post doc
43	3	17	17,6	astro	Post doc
45	1	29	3,4	astro	Post doc
47	8	33	24,2	astro	Post doc
48	2	32	6,3	astro	Post doc
49	1	70	1,4	astro	Post doc
50	1	27	3,7	astro	Post doc
51	3	35	8,6	astro	Post doc
52	5	41	12,2	astro	Post doc
53	6	49	12,2	astro	Post doc
54	5	87	5,7	astro	Post doc
56	1	59	1,7	astro	Post doc
57	4	25	16,0	astro	Post doc
58	1	17	5,9	astro	Post doc
59	4	12	33,3	astro	Post doc
60	19	30	63,3	astro	Post doc
61	4	14	28,6	astro	Post doc
62	3	22	13,6	astro	Post doc
63	1	12	8,3	astro	Post doc
64	3	16	18,8	astro	Post doc
65	11	105	10,5	astro	Assis Prof
68	8	43	18,6	astro	Assis Prof
69	20	121	16,5	astro	Assis Prof
70	19	45	42,2	astro	Assis Prof

71	3	13	23,1	astro	Assis Prof
72	10	37	27,0	astro	Assis Prof
73	15	157	9,6	astro	Assis Prof
74	6	41	14,6	astro	Assis Prof
75	2	32	6,3	astro	Assis Prof
76	13	61	21,3	astro	Assis Prof
78	2	48	4,2	astro	Assis Prof
79	5	76	6,6	astro	Assis Prof
81	6	56	10,7	astro	Assis Prof
83	2	34	5,9	astro	Assis Prof
84	4	71	5,6	astro	Assis Prof
85	7	74	9,5	astro	Assis Prof
86	25	58	43,1	astro	Assis Prof
87	21	65	32,3	astro	Assis Prof
88	2	25	8,0	astro	Assis Prof
89	1	17	5,9	astro	Assis Prof
90	10	40	25,0	astro	Assis Prof
91	23	151	15,2	astro	Assis Prof
92	4	68	5,9	astro	Assoc
93	4	52	7,7	astro	Assoc
96	4	47	8,5	astro	Assoc
97	8	28	28,6	astro	Assoc
98	14	153	9,2	astro	Assoc
99	9	83	10,8	astro	Assoc
100	24	84	28,6	astro	Assoc
101	10	62	16,1	astro	Assoc
102	31	154	20,1	astro	Assoc
103	11	28	39,3	astro	Assoc
104	3	27	11,1	astro	Assoc
105	24	124	19,4	astro	Assoc
106	1	8	12,5	astro	Assoc
107	23	315	7,3	astro	Assoc
108	62	149	41,6	astro	Assoc
109	3	38	7,9	astro	Assoc
110	16	91	17,6	astro	Assoc
111	15	104	14,4	astro	Assoc
112	3	17	17,6	astro	Assoc
113	4	30	13,3	astro	Assoc
114	7	57	12,3	astro	Assoc
115	24	105	22,9	astro	Assoc
116	5	60	8,3	astro	Assoc
117	9	78	11,5	astro	Assoc
118	34	163	20,9	astro	Assoc
119	25	94	26,6	astro	Assoc
120	16	59	27,1	astro	Assoc
122	24	197	12,2	astro	Assoc
123	14	169	8,3	astro	Assoc
124	1	50	2,0	astro	Assoc
125	19	47	40,4	astro	Assoc
126	1	87	1,1	astro	Assoc
127	29	131	22,1	astro	Assoc
128	13	76	17,1	astro	Assoc
129	23	68	33,8	astro	Assoc

130	18	66	27,3	astro	Assoc
132	8	46	17,4	astro	Assoc
134	30	211	14,2	astro	Assoc
135	28	149	18,8	astro	Assoc
136	12	48	25,0	astro	Assoc
137	15	87	17,2	astro	Assoc
138	4	56	7,1	astro	Assoc
139	7	16	43,8	astro	Assoc
140	11	104	10,6	astro	Assoc
143	7	44	15,9	astro	Assoc
144	8	133	6,0	astro	Assoc
146	16	58	27,6	astro	Assoc
147	15	194	7,7	astro	Assoc
148	11	62	17,7	astro	Assoc
149	3	32	9,4	astro	Assoc
150	5	28	17,9	astro	Assoc
151	13	35	37,1	astro	Assoc
152	1	79	1,3	astro	Assoc
153	21	97	21,6	astro	Assoc
154	12	121	9,9	astro	Assoc
155	73	279	26,2	astro	Assoc
156	40	51	78,4	astro	Assoc
157	27	93	29,0	astro	Assoc
158	9	132	6,8	astro	Assoc
159	75	334	22,5	astro	Assoc
160	8	40	20,0	astro	Assoc
161	4	149	2,7	astro	Assoc
162	20	178	11,2	astro	Assoc
163	7	40	17,5	astro	Assoc
164	14	334	4,2	astro	Prof
165	16	64	25,0	astro	Prof
166	14	75	18,7	astro	Prof
168	17	113	15,0	astro	Prof
169	3	50	6,0	astro	Prof
170	4	82	4,9	astro	Prof
171	12	116	10,3	astro	Prof
172	22	58	37,9	astro	Prof
173	100	271	36,9	astro	Prof
174	33	252	13,1	astro	Prof
175	8	121	6,6	astro	Prof
176	3	78	3,8	astro	Prof
177	8	118	6,8	astro	Prof
178	26	110	23,6	astro	Prof
179	3	56	5,4	astro	Prof
180	5	60	8,3	astro	Prof
181	32	137	23,4	astro	Prof
182	100	427	23,4	astro	Prof
183	93	372	25,0	astro	Prof
184	15	88	17,0	astro	Prof
185	58	239	24,3	astro	Prof
186	9	43	20,9	astro	Prof
187	9	105	8,6	astro	Prof
189	7	88	8,0	astro	Prof

190	39	120	32,5	astro	Prof
191	36	140	25,7	astro	Prof
192	45	166	27,1	astro	Prof
193	16	227	7,0	astro	Prof
194	5	95	5,3	astro	Prof
195	6	94	6,4	astro	Prof
196	93	353	26,3	astro	Prof
197	7	143	4,9	astro	Prof
198	51	123	41,5	astro	Prof
199	8	60	13,3	astro	Prof
200	97	422	23,0	astro	Prof
203	3	73	4,1	astro	Prof
206	1	6	16,7	enviro	Phd
212	1	9	11,1	enviro	Post doc
214	10	18	55,6	enviro	Post doc
216	5	38	13,2	enviro	Post doc
220	12	25	48,0	enviro	Post doc
221	13	72	18,1	enviro	Post doc
222	4	13	30,8	enviro	Post doc
223	10	25	40,0	enviro	Post doc
228	1	21	4,8	enviro	Assis Prof
234	1	20	5,0	enviro	Assis Prof
237	2	17	11,8	enviro	Assis Prof
238	1	46	2,2	enviro	Assis Prof
240	1	10	10,0	enviro	Assis Prof
241	1	10	10,0	enviro	Assis Prof
244	8	27	29,6	enviro	Assis Prof
245	1	13	7,7	enviro	Assis Prof
247	5	40	12,5	enviro	Assis Prof
249	1	12	8,3	enviro	Assis Prof
250	6	45	13,3	enviro	Assis Prof
253	5	28	17,9	enviro	Assis Prof
255	5	51	9,8	enviro	Assis Prof
258	2	15	13,3	enviro	Assis Prof
259	5	34	14,7	enviro	Assis Prof
260	9	23	39,1	enviro	Assis Prof
261	1	9	11,1	enviro	Assis Prof
264	2	40	5,0	enviro	Assis Prof
266	1	17	5,9	enviro	Assis Prof
268	3	76	3,9	enviro	Assoc
269	2	48	4,2	enviro	Assoc
270	5	67	7,5	enviro	Assoc
271	3	55	5,5	enviro	Assoc
272	1	21	4,8	enviro	Assoc
274	1	19	5,3	enviro	Assoc
275	1	7	14,3	enviro	Assoc
278	2	43	4,7	enviro	Assoc
280	1	49	2,0	enviro	Assoc
282	1	53	1,9	enviro	Assoc
284	7	32	21,9	enviro	Assoc
285	3	51	5,9	enviro	Assoc
286	11	102	10,8	enviro	Assoc
287	4	16	25,0	enviro	Assoc

288	1	4	25,0	enviro	Assoc
290	1	10	10,0	enviro	Assoc
291	1	30	3,3	enviro	Assoc
292	4	20	20,0	enviro	Assoc
294	1	15	6,7	enviro	Assoc
297	3	27	11,1	enviro	Assoc
298	1	11	9,1	enviro	Assoc
299	8	33	24,2	enviro	Assoc
300	2	10	20,0	enviro	Assoc
301	1	37	2,7	enviro	Assoc
302	1	44	2,3	enviro	Assoc
303	8	41	19,5	enviro	Assoc
304	5	50	10,0	enviro	Assoc
306	3	51	5,9	enviro	Assoc
307	7	65	10,8	enviro	Assoc
308	12	30	40,0	enviro	Assoc
309	2	25	8,0	enviro	Assoc
311	5	54	9,3	enviro	Assoc
312	8	66	12,1	enviro	Assoc
313	1	50	2,0	enviro	Assoc
314	1	25	4,0	enviro	Assoc
315	1	28	3,6	enviro	Assoc
316	8	49	16,3	enviro	Assoc
317	1	6	16,7	enviro	Assoc
318	11	38	28,9	enviro	Assoc
319	2	14	14,3	enviro	Assoc
322	3	32	9,4	enviro	Assoc
323	3	27	11,1	enviro	Assoc
325	4	74	5,4	enviro	Assoc
328	5	39	12,8	enviro	Assoc
329	5	69	7,2	enviro	Assoc
330	1	61	1,6	enviro	Assoc
331	2	22	9,1	enviro	Assoc
332	1	26	3,8	enviro	Assoc
333	15	29	51,7	enviro	Assoc
334	3	28	10,7	enviro	Assoc
335	1	6	16,7	enviro	Assoc
338	4	50	8,0	enviro	Assoc
340	3	57	5,3	enviro	Assoc
341	5	15	33,3	enviro	Assoc
343	13	28	46,4	enviro	Assoc
344	4	23	17,4	enviro	Assoc
345	11	113	9,7	enviro	Assoc
347	1	41	2,4	enviro	Assoc
348	1	27	3,7	enviro	Assoc
350	1	65	1,5	enviro	Assoc
351	3	20	15,0	enviro	Assoc
352	12	90	13,3	enviro	Prof
353	6	71	8,5	enviro	Prof
354	1	2	50,0	enviro	Prof
355	2	53	3,8	enviro	Prof
356	1	151	0,7	enviro	Prof
357	53	233	22,7	enviro	Prof

358	26	154	16,9	enviro	Prof
359	40	102	39,2	enviro	Prof
361	3	12	25,0	enviro	Prof
362	2	27	7,4	enviro	Prof
363	25	113	22,1	enviro	Prof
364	1	14	7,1	enviro	Prof
365	3	127	2,4	enviro	Prof
366	8	72	11,1	enviro	Prof
367	1	7	14,3	enviro	Prof
368	10	73	13,7	enviro	Prof
369	20	106	18,9	enviro	Prof
371	2	44	4,5	enviro	Prof
373	2	101	2,0	enviro	Prof
374	9	88	10,2	enviro	Prof
375	18	106	17,0	enviro	Prof
376	5	78	6,4	enviro	Prof
377	3	67	4,5	enviro	Prof
378	7	77	9,1	enviro	Prof
379	1	90	1,1	enviro	Prof
380	16	48	33,3	enviro	Prof
381	16	76	21,1	enviro	Prof
383	9	157	5,7	enviro	Prof
384	1	33	3,0	enviro	Prof
387	1	18	5,6	enviro	Prof
388	5	78	6,4	enviro	Prof
389	11	76	14,5	enviro	Prof
390	14	135	10,4	enviro	Prof
391	4	21	19,0	enviro	Prof
392	1	16	6,3	enviro	Prof
393	3	23	13,0	enviro	Prof
394	24	112	21,4	enviro	Prof
395	3	10	30,0	enviro	Prof
397	25	192	13,0	enviro	Prof
398	3	72	4,2	enviro	Prof
399	29	454	6,4	enviro	Prof
400	4	55	7,3	enviro	Prof
401	7	23	30,4	enviro	Prof
402	67	166	40,4	enviro	Prof
404	10	103	9,7	enviro	Prof
406	2	61	3,3	enviro	Prof
424	1	3	33,3	Phil	Post doc
427	1	4	25,0	Phil	Post doc
432	1	18	5,6	Phil	Post doc
434	8	14	57,1	Phil	Post doc
439	1	11	9,1	Phil	Assis Prof
446	5	13	38,5	Phil	Assis Prof
450	5	6	83,3	Phil	Assis Prof
451	1	9	11,1	Phil	Assis Prof
454	1	5	20,0	Phil	Assis Prof
455	2	10	20,0	Phil	Assis Prof
459	2	20	10,0	Phil	Assis Prof
460	5	111	4,5	Phil	Assis Prof
462	1	10	10,0	Phil	Assis Prof

463	2	5	40,0	Phil	Assis Prof
465	5	6	83,3	Phil	Assis Prof
469	1	21	4,8	Phil	Assis Prof
479	1	4	25,0	Phil	Assis Prof
482	1	39	2,6	Phil	Assis Prof
484	1	4	25,0	Phil	Assis Prof
486	2	8	25,0	Phil	Assis Prof
487	2	7	28,6	Phil	Assis Prof
495	1	4	25,0	Phil	assoc
500	1	33	3,0	Phil	assoc
502	1	9	11,1	Phil	assoc
520	1	7	14,3	Phil	assoc
523	2	37	5,4	Phil	assoc
525	1	12	8,3	Phil	assoc
528	1	14	7,1	Phil	assoc
544	6	28	21,4	Phil	assoc
548	1	29	3,4	Phil	assoc
554	1	9	11,1	Phil	assoc
555	4	12	33,3	Phil	assoc
560	3	42	7,1	Phil	assoc
562	2	6	33,3	Phil	assoc
570	1	3	33,3	Phil	Prof
571	1	48	2,1	Phil	Prof
574	3	59	5,1	Phil	Prof
575	7	16	43,8	Phil	Prof
576	1	16	6,3	Phil	Prof
577	1	36	2,8	Phil	Prof
580	6	125	4,8	Phil	Prof
581	4	32	12,5	Phil	Prof
582	3	24	12,5	Phil	Prof
585	1	4	25,0	Phil	Prof
586	22	79	27,8	Phil	Prof
588	1	10	10,0	Phil	Prof
590	1	54	1,9	Phil	Prof
591	5	28	17,9	Phil	Prof
592	8	86	9,3	Phil	Prof
600	2	44	4,5	Phil	Prof
601	4	14	28,6	Phil	Prof
602	4	30	13,3	Phil	Prof
604	2	18	11,1	Phil	Prof
606	6	25	24,0	Phil	Prof
612	2	22	9,1	Phil	Prof
615	8	121	6,6	Phil	Prof
617	4	33	12,1	Phil	Prof
620	2	29	6,9	Phil	Prof
626	7	25	28,0	Phil	Prof
629	1	34	2,9	Phil	Prof
630	1	120	0,8	Phil	Prof
631	1	65	1,5	Phil	Prof
633	2	39	5,1	Phil	Prof
636	1	10	10,0	Phil	Prof
639	1	22	4,5	Phil	Prof
642	6	23	26,1	Phil	Prof

645	1	16	6,3	Phil	Prof
648	2	28	7,1	Phil	Prof
649	1	34	2,9	Phil	Prof
650	3	64	4,7	Phil	Prof
653	3	14	21,4	Phil	Prof
654	58	173	33,5	Phil	Prof
655	2	55	3,6	Phil	Prof
656	1	14	7,1	Phil	Prof
693	1	156	0,6	Pub Health	Assis Prof
705	1	15	6,7	Pub Health	Assis Prof
706	2	290	0,7	Pub Health	Assis Prof
708	3	16	18,8	Pub Health	Assis Prof
709	3	20	15,0	Pub Health	Assis Prof
714	1	60	1,7	Pub Health	Assoc
723	7	36	19,4	Pub Health	Assoc
724	2	36	5,6	Pub Health	Assoc
738	2	8	25,0	Pub Health	Assoc
746	2	49	4,1	Pub Health	Assoc
747	1	23	4,3	Pub Health	Assoc
748	1	147	0,7	Pub Health	Assoc
752	2	117	1,7	Pub Health	Assoc
756	3	13	23,1	Pub Health	Assoc
758	3	106	2,8	Pub Health	Assoc
760	3	77	3,9	Pub Health	Assoc
764	1	63	1,6	Pub Health	Prof
765	3	39	7,7	Pub Health	Prof
766	8	669	1,2	Pub Health	Prof
769	29	119	24,4	Pub Health	Prof
771	3	224	1,3	Pub Health	Prof
776	3	187	1,6	Pub Health	Prof
778	1	40	2,5	Pub Health	Prof
781	5	118	4,2	Pub Health	Prof
784	2	235	0,9	Pub Health	Prof
787	1	47	2,1	Pub Health	Prof
789	1	23	4,3	Pub Health	Prof
791	2	9	22,2	Pub Health	Prof
792	1	66	1,5	Pub Health	Prof
793	2	128	1,6	Pub Health	Prof

Appendix 2: Calculation of indicators

ID	Type	Indicator	Indicator	Calculation
Productivity metrics				
1	Publication	**P**	Publication count	Sum of total publications
2	Publication	**Fp**	Fractionalized publication count	Each publication divided by number of authors, limited to max. 10 authors
3	Publication	**App**	Average papers per author	Average number of author per paper over all publications
4	Publication/time	**Pyrs**	Years since first publication	Length of publication career from 1st article in dataset to 2013
Impact metrics				
5	Citation	**C**	Citation count	Sum of total citations
6	Citations	**minC**	Minimum number of citations	Smallest number of citations to a paper over all papers
7	Citation	**C-sc**	Citation count minus self-citations.	Total citations minus self citations. Self citations calculated by CWTS.
8	Citation	**Sig**	Highest cited paper	Highest cited paper
9	Citation	**%sc**	Percent self-citations	Number of self citations calculated by CWTS, as a percent of total citations
10	Citation/author	**Fc**	Fractional citation count	Citations divided by authors. Limited to max. 10 authors
11	Citation/time	**C<5**	Citations less than 5 years old	Number of citations less than 5 years old, from the publication of the paper. Publication year is Zero
Hybrid metrics				
12	Citation/publication/field	**IQP**	Index of Quality & Productivity	a) A= (mnjs x Pyrs x p+1)/2. (number of citations if author was of average quality for field) b) A/number of papers (estimated performance per paper) c) define actual number of citations d) IQP=actual citations/b+number of papers e) calculate field impact per paper x number of papers IQP= expected average performance of scholar in the field, amount of papers that are cited more frequently than average and how much more than average they are cited (Tc>a)
13	Citation/publication/field	**Tc>a**	(part of IQP)	As above-
14	Citation/publication/field	**H norm**	Normalized h	Define how many articles are included in the h-index and subtract these from total number of publications
15	Citation/publication	**Cage**	Age of citation	Average age of citations to all publications
16	Citation/publication	**%PNC**	Percent not cited	Total not cited papers divided by all papers, multiplied by 100.
17	Citation/publication	**CPP**	Citations per paper	Citations/papers
18	Citation/publication	**h**	h index	Publications are ranked in descending order after number of citations. Where number of citations and rank is the same, this is the h index
19	Citation/publication	**g**	g index	Publications are ranked in descending order after number of citations. The cumulative sum of citations is calculated, and where the square root of the cumulative sum is equal to the rank this is g-index
20	Citation/publication	**m**	m index	Median citations to publications included in h
21	Citation/publication	**e**	e index	Define total citations to articles in h-index. Subtract h^2 from total citations, giving e^2. Square root of e^2 is e.

ID	Type	Indicator	Indicator	Calculation
23	Citation/publication	**hg**	Hg index	The square root of the sum of h multiplied by g.
24	Citation/publication	**H^2**	Kosmulski index	Cube root of total citations
25	Citation/publication	**A**	A index	Average number of citations to articles in the h-index
26	Citation/publication	**R**	R index	Square root of the A-index
27	Citation/publication	**AR**	AR-index	Square root of average number of citations to articles in h-index
28	Citation/publication	**ħ**	Miller's h	Square root of half the number of total citations to all publications
29	Citation/publication	**Q^2**	Quantitative & Quality index	Square root of (Geometric mean of h multiplied by median number of citations to papers in h index)
30	Citation/publication/author	**hi**	individual h	H index divided median number of researcher in papers included in h
31	Citation/publication/author	**POP h**	Harzing's publish or perish h index	Divide the number of total citations by number of authors for each paper. Calculate h using this normalized citation count
32	Citation/publication/author/time	**AWCR**	age weighted citation rate	(Citations/Pyrs)/Papers
33	Citation/publication/author/time	**AW**	Age weighted h	Square root of AWCR
34	Citation/publication/author/time	**AWCRpa**	Per-author AWCR	(citations/Pyrs)/average number of authors per paper
35	Citation/publication /time	**M quotient**	m-quotient	H divided by years since first publication
36	Citation/publication/time	**Mg**	Mg-quotient	G divided by years since first publication
37	Citation/publication/time	**PI**	Price Index	Citations<5 yrs old/total number citations. Publication year is Zero
Journal-field benchmarks, calculated by CWTS				
38		**mcs**	Mean citation score	Mean citation score of articles in publishing journal
39		**mncs**	Mean normalized citation score.	Relates article to world average in regards to document type, publication year and field. 0.9 means cited 10% below average, 1.2% cited 20% above.
40		**pp top n cites**	Proportion of top papers	Proportion papers that receive more than 10 citations. 1 is that the paper has more than 10 citations and 0 that is has less
41		**pp top prop**	Proportion in top 10% of world	Proportion of papers in the top 10% of the world. 100% means that the article belongs to this set of papers, 0 means not.
42		**pp uncited**	Proportion uncited	Proportion uncited
43		**mjs mcs**	Average number of citations for the journal	This is the MCS (mean citation score) of the publishing journal, ie the average number of citations of the journal
44		**mnjs**	Mean normalized journal score	
45		**mjs pp top n cits**		Proportion of papers from the publishing journal that have more than 10 citations
46		**mnjs pp top prop**		Proportion of papers of the publishing journal that are on the pp top prop of the world.
47		**mjs pp uncited**		Proportion of papers of publishing journal that are not cited
48		**prop self cits**	Proportion self-citations	Proportion of self citations to external citations
49		**int coverage**	Internal coverage.	The proportion of the cited references of the paper covered by WOS
50		**pp collaboration**	collaboration	Percentage inter-institutional collaboration
51		**pp int collab**	International/internal	Percentage
52		**n self cites**	Number of self-citations	Number of self-citations (author level)

Appendix 3: Correlation matrix *Astronomy*

Indicator	Pyrs	App	CPP	Cites	Sum_of_self_cites	percent_sc	Cage	AWCR_C	AWCR_pp	AWCR_au	AW_	P	fp	nnC	percent_nc	Sig	millers_h	Fc	PI	Sum_af_pp_top_n_cits	Sum_pp_top_prop	average_mjs_mcs	Min_af_mjs_mcs	Maks_af_mjs_mcs	Gennemsnit_af_Mnjs	h	m_quotient	h_norm	m_	A_	R_	g	hg_	WU	mg_quotient	e	Q2	h2	AR	POP_h	IQP	nproductivity_adjusted_papers	times_cited_more_frequently_than_average_paper_
Pyrs	1,00	-0,15	0,19	0,40	0,34	-0,21	0,75	0,15	-0,34	0,23	0,15	0,47	0,58	0,40	0,05	0,33	0,40	0,50	-0,62	0,44	0,23	0,43	-0,22	0,44	-0,10	0,42	-0,15	-0,32	0,34	0,32	0,32	0,39	0,40	0,22	-0,16	0,35	0,39	0,40	0,32	0,50	0,20	0,59	-0,37
App	-0,15	1,00	0,25	0,23	0,32	0,33	-0,20	0,35	0,32	0,08	0,35	0,17	-0,15	0,06	-0,12	0,22	0,23	-0,01	0,24	0,17	0,25	0,06	-0,06	0,20	0,16	0,22	0,41	-0,04	0,24	0,25	0,25	0,24	0,23	0,03	0,41	0,24	0,23	0,23	0,25	-0,06	0,27	0,02	0,31
CPP	0,19	0,25	1,00	0,62	0,54	-0,19	0,23	0,64	0,48	0,60	0,64	0,37	0,23	0,06	-0,39	0,67	0,62	0,51	-0,14	0,59	0,64	0,45	0,07	0,41	0,36	0,58	0,46	0,07	0,68	0,75	0,75	0,67	0,62	0,06	0,53	0,71	0,64	0,62	0,75	0,53	0,90	0,18	0,39
Cites	0,40	0,23	0,62	1,00	0,83	-0,11	0,33	0,75	0,20	0,72	0,75	0,76	0,53	0,38	-0,21	0,75	1,00	0,74	-0,28	0,87	0,73	0,45	-0,19	0,54	0,22	0,91	0,44	-0,26	0,81	0,81	0,81	0,96	0,95	0,13	0,44	0,86	0,90	1,00	0,81	0,70	0,70	0,51	0,14
Sum_of_self_cites	0,34	0,32	0,54	0,83	1,00	0,06	0,25	0,74	0,19	0,63	0,74	0,76	0,48	0,41	-0,19	0,67	0,83	0,62	-0,20	0,75	0,65	0,37	-0,23	0,52	0,17	0,83	0,47	-0,29	0,72	0,72	0,72	0,82	0,83	0,12	0,46	0,76	0,80	0,83	0,72	0,57	0,62	0,49	0,15
percent_sc	-0,21	0,33	-0,19	-0,11	0,06	1,00	-0,28	-0,04	-0,04	-0,18	-0,04	-0,03	-0,18	0,04	0,13	-0,14	-0,11	-0,25	0,29	-0,18	-0,15	-0,26	-0,13	-0,08	-0,16	-0,10	0,05	-0,08	-0,12	-0,14	-0,14	-0,13	-0,11	0,00	0,02	-0,14	-0,12	-0,11	-0,14	-0,31	-0,15	-0,07	0,02
Cage	0,75	-0,20	0,23	0,33	0,25	-0,28	1,00	0,11	-0,24	0,22	0,11	0,34	0,48	0,26	-0,05	0,28	0,33	0,45	-0,66	0,38	0,17	0,50	-0,12	0,39	-0,04	0,34	-0,16	-0,17	0,31	0,28	0,28	0,32	0,33	0,22	-0,16	0,30	0,33	0,33	0,28	0,47	0,21	0,44	-0,28
AWCR_C	0,15	0,35	0,64	0,75	0,74	-0,04	0,11	1,00	0,41	0,73	1,00	0,59	0,36	0,26	-0,27	0,69	0,75	0,56	-0,08	0,69	0,76	0,32	-0,12	0,42	0,30	0,73	0,68	-0,16	0,72	0,75	0,75	0,77	0,75	0,07	0,69	0,77	0,74	0,75	0,75	0,54	0,74	0,33	0,35
AWCR_pp	-0,34	0,32	0,48	0,20	0,19	-0,04	-0,24	0,41	1,00	0,31	0,41	0,00	-0,17	-0,24	-0,38	0,28	0,20	0,07	0,26	0,16	0,33	0,08	0,24	0,04	0,38	0,17	0,57	0,31	0,27	0,31	0,31	0,24	0,20	-0,08	0,66	0,28	0,22	0,20	0,31	0,09	0,46	-0,23	0,79
AWCR_au	0,23	0,08	0,60	0,72	0,63	-0,18	0,22	0,73	0,31	1,00	0,73	0,58	0,51	0,28	-0,25	0,65	0,72	0,71	-0,19	0,69	0,70	0,35	-0,11	0,38	0,26	0,69	0,52	-0,16	0,67	0,70	0,70	0,73	0,71	0,07	0,53	0,71	0,70	0,72	0,70	0,70	0,66	0,36	0,25
AW_	0,15	0,35	0,64	0,75	0,74	-0,04	0,11	1,00	0,41	0,73	1,00	0,59	0,36	0,26	-0,27	0,69	0,75	0,56	-0,08	0,69	0,76	0,32	-0,12	0,42	0,30	0,73	0,68	-0,16	0,72	0,75	0,75	0,77	0,75	0,07	0,69	0,77	0,74	0,75	0,75	0,54	0,74	0,33	0,35
P	0,47	0,17	0,37	0,76	0,76	-0,03	0,34	0,59	0,00	0,58	0,59	1,00	0,67	0,56	-0,08	0,57	0,76	0,69	-0,30	0,75	0,58	0,37	-0,33	0,52	0,09	0,78	0,33	-0,48	0,62	0,59	0,59	0,72	0,75	0,16	0,30	0,64	0,71	0,76	0,59	0,62	0,46	0,69	-0,04
fp	0,58	-0,15	0,23	0,53	0,48	-0,18	0,48	0,36	-0,17	0,51	0,36	0,67	1,00	0,51	-0,01	0,42	0,53	0,71	-0,45	0,57	0,40	0,34	-0,29	0,40	0,00	0,55	0,12	-0,46	0,44	0,43	0,43	0,51	0,53	0,16	0,09	0,46	0,51	0,53	0,43	0,67	0,29	0,66	-0,20
nnC	0,40	0,06	0,06	0,38	0,41	0,04	0,26	0,26	-0,24	0,28	0,26	0,56	0,51	1,00	0,39	0,28	0,38	0,39	-0,29	0,38	0,26	0,19	-0,45	0,34	-0,09	0,41	0,09	-0,56	0,30	0,27	0,27	0,36	0,38	0,10	0,05	0,30	0,36	0,38	0,27	0,35	0,13	0,57	-0,24
percent_nc	0,05	-0,12	-0,39	-0,21	-0,19	0,13	-0,05	-0,27	-0,38	-0,25	-0,27	-0,08	-0,01	0,39	1,00	-0,23	-0,21	-0,17	-0,05	-0,21	-0,27	-0,16	-0,25	-0,10	-0,25	-0,19	-0,25	-0,19	-0,24	-0,27	-0,27	-0,23	-0,21	-0,01	-0,29	-0,25	-0,22	-0,21	-0,27	-0,18	-0,36	0,05	-0,34
Sig	0,33	0,22	0,67	0,75	0,67	-0,14	0,28	0,69	0,28	0,65	0,69	0,57	0,42	0,28	-0,23	1,00	0,75	0,62	-0,25	0,67	0,66	0,42	-0,10	0,49	0,23	0,67	0,40	-0,17	0,69	0,84	0,84	0,77	0,72	0,14	0,46	0,83	0,70	0,75	0,84	0,59	0,73	0,36	0,22
millers_h	0,40	0,23	0,62	1,00	0,83	-0,11	0,33	0,75	0,20	0,72	0,75	0,76	0,53	0,38	-0,21	0,75	1,00	0,74	-0,28	0,87	0,73	0,45	-0,19	0,54	0,22	0,91	0,44	-0,26	0,81	0,81	0,81	0,96	0,95	0,13	0,44	0,86	0,90	1,00	0,81	0,70	0,70	0,51	0,14
Fc	0,50	-0,01	0,51	0,74	0,62	-0,25	0,45	0,56	0,07	0,71	0,56	0,69	0,71	0,39	-0,17	0,62	0,74	1,00	-0,40	0,78	0,61	0,46	-0,18	0,47	0,16	0,75	0,29	-0,27	0,67	0,66	0,66	0,74	0,74	0,13	0,29	0,69	0,73	0,74	0,66	0,88	0,55	0,56	0,02
PI	-0,62	0,24	-0,14	-0,28	-0,20	0,29	-0,66	-0,08	0,26	-0,19	-0,08	-0,30	-0,45	-0,29	-0,05	-0,25	-0,28	-0,40	1,00	-0,31	-0,14	-0,36	0,16	-0,30	0,07	-0,28	0,17	0,20	-0,23	-0,23	-0,23	-0,27	-0,27	-0,16	0,16	-0,25	-0,26	-0,28	-0,23	-0,42	-0,14	-0,39	0,29
Sum_af_pp_top_n_cits	0,44	0,17	0,59	0,87	0,75	-0,18	0,38	0,69	0,16	0,69	0,69	0,75	0,57	0,38	-0,21	0,67	0,87	0,78	-0,31	1,00	0,70	0,50	-0,18	0,55	0,23	0,89	0,40	-0,25	0,78	0,74	0,74	0,86	0,88	0,15	0,39	0,79	0,87	0,87	0,74	0,75	0,65	0,55	0,09
Sum_pp_top_prop	0,23	0,25	0,64	0,73	0,65	-0,15	0,17	0,76	0,33	0,70	0,76	0,58	0,40	0,26	-0,27	0,66	0,73	0,61	-0,14	0,70	1,00	0,36	-0,10	0,41	0,30	0,70	0,54	-0,15	0,71	0,74	0,74	0,75	0,72	0,07	0,56	0,75	0,72	0,73	0,74	0,59	0,70	0,37	0,26
average_mjs_mcs	0,43	0,06	0,45	0,45	0,37	-0,26	0,50	0,32	0,08	0,35	0,32	0,37	0,34	0,19	-0,16	0,42	0,45	0,46	-0,36	0,50	0,36	1,00	0,00	0,61	0,39	0,46	0,12	-0,02	0,47	0,45	0,45	0,47	0,46	0,17	0,14	0,46	0,47	0,45	0,45	0,47	0,41	0,42	-0,06
Min_af_mjs_mcs	-0,22	-0,06	0,07	-0,19	-0,23	-0,13	-0,12	-0,12	0,24	-0,11	-0,12	-0,33	-0,29	-0,45	-0,25	-0,10	-0,19	-0,18	0,16	-0,18	-0,10	0,00	1,00	-0,13	0,25	-0,21	-0,04	0,40	-0,13	-0,10	-0,10	-0,16	-0,19	-0,05	0,00	-0,12	-0,17	-0,19	-0,10	-0,16	0,00	-0,30	0,18
Maks_af_mjs_mcs	0,44	0,20	0,41	0,54	0,52	-0,08	0,39	0,42	0,04	0,38	0,42	0,52	0,40	0,34	-0,10	0,49	0,54	0,47	-0,30	0,55	0,41	0,61	-0,13	1,00	0,30	0,54	0,20	-0,21	0,52	0,51	0,51	0,54	0,54	0,17	0,21	0,53	0,55	0,54	0,51	0,44	0,42	0,54	-0,07
Gennemsnit_af_Mnjs	-0,10	0,16	0,36	0,22	0,17	-0,16	-0,04	0,30	0,38	0,26	0,30	0,09	0,00	-0,09	-0,25	0,23	0,22	0,16	0,07	0,23	0,30	0,39	0,25	0,30	1,00	0,22	0,32	0,18	0,26	0,27	0,27	0,24	0,23	0,00	0,34	0,26	0,24	0,22	0,27	0,16	0,31	0,11	0,17
h	0,42	0,22	0,58	0,91	0,83	-0,10	0,34	0,73	0,17	0,69	0,73	0,78	0,55	0,41	-0,19	0,67	0,91	0,75	-0,28	0,89	0,70	0,46	-0,21	0,54	0,22	1,00	0,45	-0,26	0,78	0,74	0,74	0,89	0,95	0,13	0,43	0,79	0,90	0,91	0,74	0,71	0,66	0,55	0,11
m_quotient	-0,15	0,41	0,46	0,44	0,47	0,05	-0,16	0,68	0,57	0,52	0,68	0,33	0,12	0,09	-0,25	0,40	0,44	0,29	0,17	0,40	0,54	0,12	-0,04	0,20	0,32	0,45	1,00	-0,01	0,45	0,45	0,45	0,46	0,45	-0,03	0,84	0,46	0,46	0,44	0,45	0,28	0,52	0,09	0,52
h_norm	-0,32	-0,04	0,07	-0,26	-0,29	-0,08	-0,17	-0,16	0,31	-0,16	-0,16	-0,48	-0,46	-0,56	-0,19	-0,17	-0,26	-0,27	0,20	-0,25	-0,15	-0,02	0,40	-0,21	0,18	-0,26	-0,01	1,00	-0,15	-0,15	-0,15	-0,22	-0,23	-0,14	0,01	-0,17	-0,20	-0,26	-0,15	-0,20	0,00	-0,53	0,30
m_	0,34	0,24	0,68	0,81	0,72	-0,12	0,31	0,72	0,27	0,67	0,72	0,62	0,44	0,30	-0,24	0,69	0,81	0,67	-0,23	0,78	0,71	0,47	-0,13	0,52	0,26	0,78	0,45	-0,15	1,00	0,82	0,82	0,84	0,81	0,10	0,47	0,84	0,88	0,81	0,82	0,66	0,74	0,41	0,21
A_	0,32	0,25	0,75	0,81	0,72	-0,14	0,28	0,75	0,31	0,70	0,75	0,59	0,43	0,27	-0,27	0,84	0,81	0,66	-0,23	0,74	0,74	0,45	-0,10	0,51	0,27	0,74	0,45	-0,15	0,82	1,00	1,00	0,86	0,79	0,11	0,51	0,94	0,80	0,81	1,00	0,64	0,81	0,37	0,25
R_	0,32	0,25	0,75	0,81	0,72	-0,14	0,28	0,75	0,31	0,70	0,75	0,59	0,43	0,27	-0,27	0,84	0,81	0,66	-0,23	0,74	0,74	0,45	-0,10	0,51	0,27	0,74	0,45	-0,15	0,82	1,00	1,00	0,86	0,79	0,11	0,51	0,94	0,80	0,81	1,00	0,64	0,81	0,37	0,25
g	0,39	0,24	0,67	0,96	0,82	-0,13	0,32	0,77	0,24	0,73	0,77	0,72	0,51	0,36	-0,23	0,77	0,96	0,74	-0,27	0,86	0,75	0,47	-0,16	0,54	0,24	0,89	0,46	-0,22	0,84	0,86	0,86	1,00	0,94	0,13	0,47	0,91	0,92	0,96	0,86	0,71	0,74	0,48	0,17
hg_	0,40	0,23	0,62	0,95	0,83	-0,11	0,33	0,75	0,20	0,71	0,75	0,75	0,53	0,38	-0,21	0,72	0,95	0,74	-0,27	0,88	0,72	0,46	-0,19	0,54	0,23	0,95	0,45	-0,23	0,81	0,79	0,79	0,94	1,00	0,13	0,44	0,84	0,92	0,95	0,79	0,71	0,70	0,51	0,14
WU	0,22	0,03	0,06	0,13	0,12	0,00	0,22	0,07	-0,08	0,07	0,07	0,16	0,16	0,10	-0,01	0,14	0,13	0,13	-0,16	0,15	0,07	0,17	-0,05	0,17	0,00	0,13	-0,03	-0,14	0,10	0,11	0,11	0,13	0,13	1,00	-0,03	0,12	0,12	0,13	0,11	0,11	0,06	0,17	-0,09
mg_quotient	-0,16	0,41	0,53	0,44	0,46	0,02	-0,16	0,69	0,66	0,53	0,69	0,30	0,09	0,05	-0,29	0,46	0,44	0,29	0,16	0,39	0,56	0,14	0,00	0,21	0,34	0,43	0,84	0,01	0,47	0,51	0,51	0,47	0,44	-0,03	1,00	0,49	0,45	0,44	0,51	0,28	0,58	0,05	0,59
e	0,35	0,24	0,71	0,86	0,76	-0,14	0,30	0,77	0,28	0,71	0,77	0,64	0,46	0,30	-0,25	0,83	0,86	0,69	-0,25	0,79	0,75	0,46	-0,12	0,53	0,26	0,79	0,46	-0,17	0,84	0,94	0,94	0,91	0,84	0,12	0,49	1,00	0,85	0,86	0,94	0,67	0,78	0,42	0,21
Q2	0,39	0,23	0,64	0,90	0,80	-0,12	0,33	0,74	0,22	0,70	0,74	0,71	0,51	0,36	-0,22	0,70	0,90	0,73	-0,26	0,87	0,72	0,47	-0,17	0,55	0,24	0,90	0,46	-0,20	0,88	0,80	0,80	0,92	0,92	0,12	0,45	0,85	1,00	0,90	0,80	0,70	0,72	0,48	0,16
h2	0,40	0,23	0,62	1,00	0,83	-0,11	0,33	0,75	0,20	0,72	0,75	0,76	0,53	0,38	-0,21	0,75	1,00	0,74	-0,28	0,87	0,73	0,45	-0,19	0,54	0,22	0,91	0,44	-0,26	0,81	0,81	0,81	0,96	0,95	0,13	0,44	0,86	0,90	1,00	0,81	0,70	0,70	0,51	0,14
AR	0,32	0,25	0,75	0,81	0,72	-0,14	0,28	0,75	0,31	0,70	0,75	0,59	0,43	0,27	-0,27	0,84	0,81	0,66	-0,23	0,74	0,74	0,45	-0,10	0,51	0,27	0,74	0,45	-0,15	0,82	1,00	1,00	0,86	0,79	0,11	0,51	0,94	0,80	0,81	1,00	0,64	0,81	0,37	0,25
POP_h	0,50	-0,06	0,53	0,70	0,57	-0,31	0,47	0,54	0,09	0,70	0,54	0,62	0,67	0,35	-0,18	0,59	0,70	0,88	-0,42	0,75	0,59	0,47	-0,16	0,44	0,16	0,71	0,28	-0,20	0,66	0,64	0,64	0,71	0,71	0,11	0,28	0,67	0,70	0,70	0,64	1,00	0,56	0,50	0,04
IQP	0,20	0,27	0,90	0,70	0,62	-0,15	0,21	0,74	0,46	0,66	0,74	0,46	0,29	0,13	-0,36	0,73	0,70	0,55	-0,14	0,65	0,70	0,41	0,00	0,42	0,31	0,66	0,52	0,00	0,74	0,81	0,81	0,74	0,70	0,06	0,58	0,78	0,72	0,70	0,81	0,56	1,00	0,23	0,39
nproductivity_adjusted_papers	0,59	0,02	0,18	0,51	0,49	-0,07	0,44	0,33	-0,23	0,36	0,33	0,69	0,66	0,57	0,05	0,36	0,51	0,56	-0,39	0,55	0,37	0,42	-0,30	0,54	0,11	0,55	0,09	-0,53	0,41	0,37	0,37	0,48	0,51	0,17	0,05	0,42	0,48	0,51	0,37	0,50	0,23	1,00	-0,33
times_cited_more_frequently_	-0,37	0,31	0,39	0,14	0,15	0,02	-0,28	0,35	0,79	0,25	0,35	-0,04	-0,20	-0,24	-0,34	0,22	0,14	0,02	0,29	0,09	0,26	-0,06	0,18	-0,07	0,17	0,11	0,52	0,30	0,21	0,25	0,25	0,17	0,14	-0,09	0,59	0,21	0,16	0,14	0,25	0,04	0,39	-0,33	1,00

Appendix 4: Correlation matrix *Environmental Science*

Indicator	Pyrs	App	CPP	Cites	Sum_of_self_cites	percent_sc	P	fp	nnC	percent_nc	AWCR_au	cage	AWCR_C	AWCR_pp	AW_	Sig	sumcits	sumcits2	millers_h	Fc	PI	Sum_af_pp_top_n_cits	Sum_pp_top_prop	average_mjs_mcs	Min_af_mjs_mcs	Maks_af_mjs_mcs	Gennemsnit_af_Mnjs	IQP	nproductivity_adjusted_papers	times_cited_more_frequently_than_average_paper_	h	m_quotient	h_norm	m_	A_	R_	g	hg_	WU	mg_quotient	e	Q2	h2	AR	POP_h
Pyrs	1,00	0,01	0,32	0,50	0,47	-0,05	0,55	0,56	0,36	-0,20	0,27	0,68	0,28	-0,26	0,28	0,42	0,50	0,50	0,50	0,51	-0,57	0,48	0,36	0,39	-0,11	0,46	0,06	0,33	0,59	-0,34	0,50	-0,08	-0,24	0,39	0,42	0,42	0,48	0,49	0,15	-0,07	0,43	0,46	0,50	-0,08	0,48
App	0,01	1,00	0,15	0,16	0,17	0,08	0,12	-0,08	0,05	-0,06	0,01	-0,02	0,19	0,14	0,19	0,18	0,16	0,16	0,16	0,02	-0,02	0,16	0,13	0,11	0,02	0,15	0,07	0,19	0,04	0,13	0,16	0,18	0,03	0,18	0,18	0,18	0,16	0,16	0,00	0,18	0,17	0,17	0,16	0,02	0,04
CPP	0,32	0,15	1,00	0,66	0,54	-0,21	0,42	0,37	0,12	-0,42	0,62	0,40	0,67	0,43	0,67	0,75	0,66	0,66	0,66	0,61	-0,37	0,68	0,60	0,59	0,12	0,51	0,32	0,86	0,22	0,28	0,63	0,39	0,14	0,73	0,78	0,78	0,71	0,66	0,04	0,50	0,76	0,69	0,66	0,02	0,67
Cites	0,50	0,16	0,66	1,00	0,81	-0,05	0,77	0,66	0,40	-0,34	0,73	0,47	0,79	0,17	0,79	0,76	1,00	1,00	1,00	0,85	-0,44	0,86	0,71	0,53	-0,06	0,64	0,24	0,75	0,51	0,06	0,91	0,41	-0,14	0,76	0,79	0,79	0,95	0,95	0,07	0,43	0,83	0,90	1,00	-0,04	0,84
Sum_of_self_cites	0,47	0,17	0,54	0,81	1,00	0,15	0,78	0,66	0,45	-0,30	0,66	0,42	0,72	0,12	0,72	0,64	0,81	0,81	0,81	0,74	-0,39	0,72	0,63	0,44	-0,11	0,58	0,20	0,64	0,52	0,02	0,83	0,41	-0,19	0,65	0,65	0,65	0,78	0,81	0,09	0,40	0,69	0,77	0,81	-0,03	0,74
percent_sc	-0,05	0,08	-0,21	-0,05	0,15	1,00	0,06	0,04	0,13	0,11	-0,06	-0,12	-0,03	-0,16	-0,03	-0,15	-0,05	-0,05	-0,05	-0,06	0,10	-0,13	-0,10	-0,19	-0,15	-0,07	-0,10	-0,14	0,08	-0,11	-0,01	0,05	-0,14	-0,15	-0,16	-0,16	-0,08	-0,05	0,08	-0,05	-0,13	-0,07	-0,05	0,06	-0,09
P	0,55	0,12	0,42	0,77	0,78	0,06	1,00	0,78	0,58	-0,23	0,59	0,45	0,63	-0,03	0,63	0,57	0,77	0,77	0,77	0,73	-0,41	0,70	0,60	0,40	-0,16	0,58	0,15	0,53	0,66	-0,11	0,79	0,32	-0,36	0,58	0,58	0,58	0,72	0,75	0,08	0,30	0,62	0,71	0,77	-0,08	0,70
fp	0,56	-0,08	0,37	0,66	0,66	0,04	0,78	1,00	0,53	-0,20	0,58	0,46	0,54	-0,07	0,54	0,50	0,66	0,66	0,66	0,74	-0,39	0,61	0,55	0,35	-0,17	0,51	0,10	0,46	0,64	-0,15	0,68	0,24	-0,35	0,50	0,51	0,51	0,63	0,65	0,06	0,22	0,55	0,61	0,66	-0,10	0,69
nnC	0,36	0,05	0,12	0,40	0,45	0,13	0,58	0,53	1,00	0,23	0,32	0,18	0,34	-0,18	0,34	0,29	0,40	0,40	0,40	0,39	-0,23	0,37	0,31	0,16	-0,28	0,36	0,00	0,22	0,51	-0,21	0,42	0,15	-0,47	0,31	0,30	0,30	0,38	0,40	0,11	0,13	0,33	0,38	0,40	-0,13	0,38
percent_nc	-0,20	-0,06	-0,42	-0,34	-0,30	0,11	-0,23	-0,20	0,23	1,00	-0,31	-0,35	-0,32	-0,26	-0,32	-0,32	-0,34	-0,34	-0,34	-0,33	0,20	-0,33	-0,33	-0,34	-0,18	-0,24	-0,24	-0,38	-0,12	-0,17	-0,35	-0,22	-0,08	-0,30	-0,31	-0,31	-0,34	-0,34	0,04	-0,22	-0,32	-0,33	-0,34	-0,02	-0,34
AWCR_au	0,27	0,01	0,62	0,73	0,66	-0,06	0,59	0,58	0,32	-0,31	1,00	0,30	0,82	0,34	0,82	0,66	0,73	0,73	0,73	0,75	-0,28	0,70	0,68	0,44	-0,03	0,50	0,27	0,72	0,34	0,22	0,71	0,57	-0,07	0,67	0,68	0,68	0,74	0,73	0,01	0,60	0,71	0,72	0,73	-0,02	0,79
cage	0,68	-0,02	0,40	0,47	0,42	-0,12	0,45	0,46	0,18	-0,35	0,30	1,00	0,29	-0,10	0,29	0,43	0,47	0,47	0,47	0,49	-0,63	0,47	0,32	0,48	-0,01	0,47	0,08	0,38	0,42	-0,17	0,48	-0,02	-0,09	0,39	0,41	0,41	0,46	0,47	0,06	0,00	0,42	0,44	0,47	-0,06	0,47
AWCR_C	0,28	0,19	0,67	0,79	0,72	-0,03	0,63	0,54	0,34	-0,32	0,82	0,29	1,00	0,35	1,00	0,72	0,79	0,79	0,79	0,71	-0,29	0,75	0,71	0,46	-0,04	0,54	0,28	0,79	0,36	0,23	0,77	0,61	-0,06	0,73	0,74	0,74	0,80	0,79	0,02	0,65	0,77	0,79	0,79	-0,02	0,75
AWCR_pp	-0,26	0,14	0,43	0,17	0,12	-0,16	-0,03	-0,07	-0,18	-0,26	0,34	-0,10	0,35	1,00	0,35	0,27	0,17	0,17	0,17	0,14	0,10	0,20	0,25	0,20	0,23	0,09	0,31	0,39	-0,24	0,69	0,15	0,55	0,37	0,29	0,29	0,29	0,21	0,18	-0,10	0,65	0,27	0,21	0,17	0,06	0,20
AW_	0,28	0,19	0,67	0,79	0,72	-0,03	0,63	0,54	0,34	-0,32	0,82	0,29	1,00	0,35	1,00	0,72	0,79	0,79	0,79	0,71	-0,29	0,75	0,71	0,46	-0,04	0,54	0,28	0,79	0,36	0,23	0,77	0,61	-0,06	0,73	0,74	0,74	0,80	0,79	0,02	0,65	0,77	0,79	0,79	-0,02	0,75
Sig	0,42	0,18	0,75	0,76	0,64	-0,15	0,57	0,50	0,29	-0,32	0,66	0,43	0,72	0,27	0,72	1,00	0,76	0,76	0,76	0,70	-0,43	0,73	0,66	0,53	0,00	0,58	0,24	0,80	0,36	0,16	0,70	0,37	-0,02	0,74	0,87	0,87	0,81	0,75	0,07	0,46	0,87	0,75	0,76	-0,02	0,72
sumcits	0,50	0,16	0,66	1,00	0,81	-0,05	0,77	0,66	0,40	-0,34	0,73	0,47	0,79	0,17	0,79	0,76	1,00	1,00	1,00	0,85	-0,44	0,86	0,71	0,53	-0,06	0,64	0,24	0,75	0,51	0,06	0,91	0,41	-0,14	0,76	0,79	0,79	0,95	0,95	0,07	0,43	0,83	0,90	1,00	-0,04	0,84
sumcits2	0,50	0,16	0,66	1,00	0,81	-0,05	0,77	0,66	0,40	-0,34	0,73	0,47	0,79	0,17	0,79	0,76	1,00	1,00	1,00	0,85	-0,44	0,86	0,71	0,53	-0,06	0,64	0,24	0,75	0,51	0,06	0,91	0,41	-0,14	0,76	0,79	0,79	0,95	0,95	0,07	0,43	0,83	0,90	1,00	-0,04	0,84
millers_h	0,50	0,16	0,66	1,00	0,81	-0,05	0,77	0,66	0,40	-0,34	0,73	0,47	0,79	0,17	0,79	0,76	1,00	1,00	1,00	0,85	-0,44	0,86	0,71	0,53	-0,06	0,64	0,24	0,75	0,51	0,06	0,91	0,41	-0,14	0,76	0,79	0,79	0,95	0,95	0,07	0,43	0,83	0,90	1,00	-0,04	0,84
Fc	0,51	0,02	0,61	0,85	0,74	-0,06	0,73	0,74	0,39	-0,33	0,75	0,49	0,71	0,14	0,71	0,70	0,85	0,85	0,85	1,00	-0,46	0,79	0,68	0,51	-0,06	0,61	0,22	0,69	0,51	0,03	0,83	0,37	-0,15	0,71	0,73	0,73	0,84	0,84	0,06	0,39	0,76	0,82	0,85	-0,06	0,89
PI	-0,57	-0,02	-0,37	-0,44	-0,39	0,10	-0,41	-0,39	-0,23	0,20	-0,28	-0,63	-0,29	0,10	-0,29	-0,43	-0,44	-0,44	-0,44	-0,46	1,00	-0,44	-0,32	-0,42	0,05	-0,44	-0,08	-0,36	-0,39	0,17	-0,44	0,00	0,12	-0,41	-0,42	-0,42	-0,44	-0,44	-0,12	-0,03	-0,43	-0,43	-0,44	0,05	-0,44
Sum_af_pp_top_n_cits	0,48	0,16	0,68	0,86	0,72	-0,13	0,70	0,61	0,37	-0,33	0,70	0,47	0,75	0,20	0,75	0,73	0,86	0,86	0,86	0,79	-0,44	1,00	0,70	0,57	-0,02	0,63	0,27	0,75	0,49	0,07	0,86	0,42	-0,09	0,80	0,78	0,78	0,87	0,87	0,08	0,44	0,82	0,87	0,86	-0,04	0,83
Sum_pp_top_prop	0,36	0,13	0,60	0,71	0,63	-0,10	0,60	0,55	0,31	-0,33	0,68	0,32	0,71	0,25	0,71	0,66	0,71	0,71	0,71	0,68	-0,32	0,70	1,00	0,38	-0,04	0,46	0,24	0,67	0,39	0,14	0,70	0,44	-0,10	0,67	0,69	0,69	0,73	0,71	0,06	0,48	0,70	0,71	0,71	-0,03	0,72
average_mjs_mcs	0,39	0,11	0,59	0,53	0,44	-0,19	0,40	0,35	0,16	-0,34	0,44	0,48	0,46	0,20	0,46	0,53	0,53	0,53	0,53	0,51	-0,42	0,57	0,38	1,00	0,19	0,66	0,48	0,53	0,38	-0,01	0,53	0,23	0,08	0,55	0,55	0,55	0,56	0,54	0,06	0,27	0,55	0,55	0,53	-0,04	0,55
Min_af_mjs_mcs	-0,11	0,02	0,12	-0,06	-0,11	-0,15	-0,16	-0,17	-0,28	-0,18	-0,03	-0,01	-0,04	0,23	-0,04	0,00	-0,06	-0,06	-0,06	-0,06	0,05	-0,02	-0,04	0,19	1,00	-0,02	0,27	0,04	-0,12	0,12	-0,07	0,01	0,28	0,03	0,01	0,01	-0,04	-0,05	0,01	0,04	-0,01	-0,03	-0,06	0,09	-0,04
Maks_af_mjs_mcs	0,46	0,15	0,51	0,64	0,58	-0,07	0,58	0,51	0,36	-0,24	0,50	0,47	0,54	0,09	0,54	0,58	0,64	0,64	0,64	0,61	-0,44	0,63	0,46	0,66	-0,02	1,00	0,34	0,53	0,51	-0,07	0,63	0,27	-0,09	0,58	0,58	0,58	0,64	0,63	0,06	0,29	0,61	0,63	0,64	-0,04	0,62
Gennemsnit_af_Mnjs	0,06	0,07	0,32	0,24	0,20	-0,10	0,15	0,10	0,00	-0,24	0,27	0,08	0,28	0,31	0,28	0,24	0,24	0,24	0,24	0,22	-0,08	0,27	0,24	0,48	0,27	0,34	1,00	0,26	0,25	0,00	0,24	0,28	0,14	0,28	0,26	0,26	0,25	0,25	0,04	0,30	0,27	0,26	0,24	0,04	0,25
IQP	0,33	0,19	0,86	0,75	0,64	-0,14	0,53	0,46	0,22	-0,38	0,72	0,38	0,79	0,39	0,79	0,80	0,75	0,75	0,75	0,69	-0,36	0,75	0,67	0,53	0,04	0,53	0,26	1,00	0,28	0,28	0,72	0,49	0,07	0,78	0,83	0,83	0,81	0,76	0,03	0,57	0,83	0,78	0,75	0,00	0,75
nproductivity_adjusted_papers	0,59	0,04	0,22	0,51	0,52	0,08	0,66	0,64	0,51	-0,12	0,34	0,42	0,36	-0,24	0,36	0,36	0,51	0,51	0,51	0,51	-0,39	0,49	0,39	0,38	-0,12	0,51	0,25	0,28	1,00	-0,42	0,54	0,08	-0,41	0,37	0,36	0,36	0,48	0,50	0,11	0,04	0,40	0,47	0,51	-0,09	0,48
times_cited_more_frequently_than_average_paper_	-0,34	0,13	0,28	0,06	0,02	-0,11	-0,11	-0,15	-0,21	-0,17	0,22	-0,17	0,23	0,69	0,23	0,16	0,06	0,06	0,06	0,03	0,17	0,07	0,14	-0,01	0,12	-0,07	0,00	0,28	-0,42	1,00	0,04	0,43	0,32	0,16	0,17	0,17	0,09	0,07	-0,10	0,50	0,15	0,09	0,06	0,07	0,08
h	0,50	0,16	0,63	0,91	0,83	-0,01	0,79	0,68	0,42	-0,35	0,71	0,48	0,77	0,15	0,77	0,70	0,91	0,91	0,91	0,83	-0,44	0,86	0,70	0,53	-0,07	0,63	0,24	0,72	0,54	0,04	1,00	0,45	-0,13	0,73	0,72	0,72	0,89	0,95	0,07	0,43	0,77	0,89	0,91	-0,04	0,85
m_quotient	-0,08	0,18	0,39	0,41	0,41	0,05	0,32	0,24	0,15	-0,22	0,57	-0,02	0,61	0,55	0,61	0,37	0,41	0,41	0,41	0,37	0,00	0,42	0,44	0,23	0,01	0,27	0,28	0,49	0,08	0,43	0,45	1,00	0,10	0,41	0,39	0,39	0,43	0,44	-0,09	0,80	0,41	0,44	0,41	0,04	0,43
h_norm	-0,24	0,03	0,14	-0,14	-0,19	-0,14	-0,36	-0,35	-0,47	-0,08	-0,07	-0,09	-0,06	0,37	-0,06	-0,02	-0,14	-0,14	-0,14	-0,15	0,12	-0,09	-0,10	0,08	0,28	-0,09	0,14	0,07	-0,41	0,32	-0,13	0,10	1,00	-0,03	-0,03	-0,03	-0,09	-0,11	-0,01	0,13	-0,04	-0,09	-0,14	0,14	-0,10
m_	0,39	0,18	0,73	0,76	0,65	-0,15	0,58	0,50	0,31	-0,30	0,67	0,39	0,73	0,29	0,73	0,74	0,76	0,76	0,76	0,71	-0,41	0,80	0,67	0,55	0,03	0,58	0,28	0,78	0,37	0,16	0,73	0,41	-0,03	1,00	0,84	0,84	0,81	0,77	0,06	0,48	0,83	0,85	0,76	-0,04	0,77
A_	0,42	0,18	0,78	0,79	0,65	-0,16	0,58	0,51	0,30	-0,31	0,68	0,41	0,74	0,29	0,74	0,87	0,79	0,79	0,79	0,73	-0,42	0,78	0,69	0,55	0,01	0,58	0,26	0,83	0,36	0,17	0,72	0,39	-0,03	0,84	1,00	1,00	0,85	0,78	0,07	0,48	0,94	0,81	0,79	-0,02	0,76
R_	0,42	0,18	0,78	0,79	0,65	-0,16	0,58	0,51	0,30	-0,31	0,68	0,41	0,74	0,29	0,74	0,87	0,79	0,79	0,79	0,73	-0,42	0,78	0,69	0,55	0,01	0,58	0,26	0,83	0,36	0,17	0,72	0,39	-0,03	0,84	1,00	1,00	0,85	0,78	0,07	0,48	0,94	0,81	0,79	-0,02	0,76
g	0,48	0,16	0,71	0,95	0,78	-0,08	0,72	0,63	0,38	-0,34	0,74	0,46	0,80	0,21	0,80	0,81	0,95	0,95	0,95	0,84	-0,44	0,87	0,73	0,56	-0,04	0,64	0,25	0,81	0,48	0,09	0,89	0,43	-0,09	0,81	0,85	0,85	1,00	0,95	0,07	0,46	0,89	0,93	0,95	-0,04	0,86
hg_	0,49	0,16	0,66	0,95	0,81	-0,05	0,75	0,65	0,40	-0,34	0,73	0,47	0,79	0,18	0,79	0,75	0,95	0,95	0,95	0,84	-0,44	0,87	0,71	0,54	-0,05	0,63	0,25	0,76	0,50	0,07	0,95	0,44	-0,11	0,77	0,78	0,78	0,95	1,00	0,07	0,45	0,83	0,92	0,95	-0,04	0,85
WU	0,15	0,00	0,04	0,07	0,09	0,08	0,08	0,06	0,11	0,04	0,01	0,06	0,02	-0,10	0,02	0,07	0,07	0,07	0,07	0,06	-0,12	0,08	0,06	0,06	0,01	0,06	0,04	0,03	0,11	-0,10	0,07	-0,09	-0,01	0,06	0,07	0,07	0,07	0,07	1,00	-0,07	0,07	0,06	0,07	0,13	0,05
mg_quotient	-0,07	0,18	0,50	0,43	0,40	-0,05	0,30	0,22	0,13	-0,22	0,60	0,00	0,65	0,65	0,65	0,46	0,43	0,43	0,43	0,39	-0,03	0,44	0,48	0,27	0,04	0,29	0,30	0,57	0,04	0,50	0,43	0,80	0,13	0,48	0,48	0,48	0,46	0,45	-0,07	1,00	0,48	0,47	0,43	0,03	0,45
e	0,43	0,17	0,76	0,83	0,69	-0,13	0,62	0,55	0,33	-0,32	0,71	0,42	0,77	0,27	0,77	0,87	0,83	0,83	0,83	0,76	-0,43	0,82	0,70	0,55	-0,01	0,61	0,27	0,83	0,40	0,15	0,77	0,41	-0,04	0,83	0,94	0,94	0,89	0,83	0,07	0,48	1,00	0,85	0,83	-0,03	0,80
Q2	0,46	0,17	0,69	0,90	0,77	-0,07	0,71	0,61	0,38	-0,33	0,72	0,44	0,79	0,21	0,79	0,75	0,90	0,90	0,90	0,82	-0,43	0,87	0,71	0,55	-0,03	0,63	0,26	0,78	0,47	0,09	0,89	0,44	-0,09	0,85	0,81	0,81	0,93	0,92	0,06	0,47	0,85	1,00	0,90	-0,04	0,85
h2	0,50	0,16	0,66	1,00	0,81	-0,05	0,77	0,66	0,40	-0,34	0,73	0,47	0,79	0,17	0,79	0,76	1,00	1,00	1,00	0,85	-0,44	0,86	0,71	0,53	-0,06	0,64	0,24	0,75	0,51	0,06	0,91	0,41	-0,14	0,76	0,79	0,79	0,95	0,95	0,07	0,43	0,83	0,90	1,00	-0,04	0,84
AR	-0,08	0,02	0,02	-0,04	-0,03	0,06	-0,08	-0,10	-0,13	-0,02	-0,02	-0,06	-0,02	0,06	-0,02	-0,02	-0,04	-0,04	-0,04	-0,06	0,05	-0,04	-0,03	-0,04	0,09	-0,04	0,04	0,00	-0,09	0,07	-0,04	0,04	0,14	-0,04	-0,02	-0,02	-0,04	-0,04	0,13	0,03	-0,03	-0,04	-0,04	1,00	-0,04
POP_h	0,48	0,04	0,67	0,84	0,74	-0,09	0,70	0,69	0,38	-0,34	0,79	0,47	0,75	0,20	0,75	0,72	0,84	0,84	0,84	0,89	-0,44	0,83	0,72	0,55	-0,04	0,62	0,25	0,75	0,48	0,08	0,85	0,43	-0,10	0,77	0,76	0,76	0,86	0,85	0,05	0,45	0,80	0,85	0,84	-0,04	1,00

Appendix 5: Correlation matrix *Philosophy*

Indicator	Pyrs	App	CPP	Cites	Sum_of_self_cites	percent_sc	P	fp	nnC	percent_nc	Cage	AWCR_C	AWCR_pp	AWCR_au	AW_	Sig	millers_h	Fc	PI	Sum_af_pp_top_n_cits	Sum_pp_top_prop	average_mjs_mcs	Min_af_mjs_mcs	Maks_af_mjs_mcs	Gennemsnit_af_Mnjs	IQP	nproductivity_adjusted_papers	times_cited_more_frequently_than_average_paper_	h	m_quotient	h_norm	m_	A_	R_	g	hg_	WU	mg_quotient	e	Q2	h2	AR	Pop_h
Pyrs	1,00	0,15	0,21	0,39	0,34	0,00	0,52	0,53	0,49	-0,07	0,40	0,18	-0,08	0,19	0,18	0,34	0,39	0,41	-0,45	0,30	0,23	0,33	-0,03	0,39	0,08	0,20	0,45	-0,08	0,40	-0,10	-0,03	0,34	0,34	0,34	0,38	0,38	0,22	0,06	0,33	0,38	0,39	0,34	0,40
App	0,15	1,00	0,45	0,44	0,47	0,19	0,32	0,23	0,14	-0,33	0,26	0,46	0,39	0,37	0,46	0,45	0,44	0,38	-0,04	0,41	0,28	0,39	0,11	0,41	0,16	0,47	-0,02	0,36	0,44	0,36	0,21	0,42	0,44	0,44	0,46	0,45	0,03	0,45	0,46	0,44	0,44	0,44	0,41
CPP	0,21	0,45	1,00	0,74	0,60	0,01	0,39	0,34	0,11	-0,64	0,56	0,78	0,73	0,75	0,78	0,79	0,74	0,70	-0,11	0,61	0,52	0,55	0,18	0,56	0,36	0,88	-0,07	0,61	0,70	0,60	0,44	0,76	0,77	0,77	0,76	0,72	0,02	0,73	0,76	0,76	0,74	0,77	0,71
Cites	0,39	0,44	0,74	1,00	0,77	0,12	0,67	0,61	0,38	-0,47	0,53	0,81	0,51	0,80	0,81	0,88	1,00	0,94	-0,25	0,69	0,62	0,55	0,06	0,64	0,35	0,77	0,22	0,42	0,88	0,51	0,27	0,82	0,85	0,85	0,94	0,93	0,08	0,67	0,85	0,94	1,00	0,85	0,86
Sum_of_self_cites	0,34	0,47	0,60	0,77	1,00	0,39	0,66	0,59	0,39	-0,43	0,43	0,71	0,43	0,69	0,71	0,70	0,77	0,75	-0,14	0,63	0,56	0,44	-0,04	0,54	0,28	0,66	0,20	0,37	0,80	0,51	0,23	0,65	0,67	0,67	0,77	0,78	0,08	0,61	0,69	0,74	0,77	0,67	0,76
percent_sc	0,00	0,19	0,01	0,12	0,39	1,00	0,17	0,14	0,12	-0,01	-0,09	0,13	0,03	0,11	0,13	0,06	0,12	0,11	0,13	0,02	0,04	-0,08	-0,25	0,01	-0,06	0,08	0,03	0,04	0,17	0,14	-0,03	0,03	0,03	0,03	0,10	0,13	0,01	0,14	0,04	0,09	0,12	0,03	0,11
P	0,52	0,32	0,39	0,67	0,66	0,17	1,00	0,93	0,71	-0,21	0,34	0,53	0,19	0,53	0,53	0,58	0,67	0,68	-0,29	0,51	0,51	0,38	-0,08	0,49	0,22	0,46	0,49	0,14	0,69	0,30	-0,02	0,54	0,55	0,55	0,64	0,66	0,15	0,41	0,56	0,62	0,67	0,55	0,67
fp	0,53	0,23	0,34	0,61	0,59	0,14	0,93	1,00	0,74	-0,16	0,32	0,47	0,14	0,49	0,47	0,53	0,61	0,64	-0,30	0,45	0,49	0,33	-0,11	0,44	0,21	0,40	0,51	0,09	0,63	0,25	-0,06	0,49	0,50	0,50	0,58	0,60	0,16	0,36	0,51	0,57	0,61	0,50	0,62
nnC	0,49	0,14	0,11	0,38	0,39	0,12	0,71	0,74	1,00	0,13	0,09	0,25	-0,06	0,27	0,25	0,33	0,38	0,40	-0,30	0,30	0,33	0,18	-0,21	0,27	0,09	0,18	0,53	-0,08	0,38	0,07	-0,27	0,31	0,31	0,31	0,37	0,37	0,15	0,18	0,32	0,35	0,38	0,31	0,39
percent_nc	-0,07	-0,33	-0,64	-0,47	-0,43	-0,01	-0,21	-0,16	0,13	1,00	-0,59	-0,53	-0,65	-0,50	-0,53	-0,46	-0,47	-0,45	-0,07	-0,34	-0,32	-0,37	-0,25	-0,37	-0,31	-0,59	0,19	-0,57	-0,51	-0,57	-0,52	-0,42	-0,43	-0,43	-0,46	-0,46	0,00	-0,50	-0,43	-0,45	-0,47	-0,43	-0,45
Cage	0,40	0,26	0,56	0,53	0,43	-0,09	0,34	0,32	0,09	-0,59	1,00	0,43	0,38	0,44	0,43	0,51	0,53	0,54	-0,25	0,40	0,33	0,44	0,17	0,47	0,25	0,51	0,05	0,35	0,55	0,30	0,37	0,48	0,49	0,49	0,52	0,51	0,11	0,34	0,49	0,51	0,53	0,49	0,52
AWCR_C	0,18	0,46	0,78	0,81	0,71	0,13	0,53	0,47	0,25	-0,53	0,43	1,00	0,67	0,92	1,00	0,81	0,81	0,77	-0,11	0,64	0,62	0,48	0,08	0,55	0,37	0,85	0,05	0,55	0,78	0,69	0,33	0,77	0,79	0,79	0,81	0,80	0,04	0,83	0,77	0,81	0,81	0,79	0,77
AWCR_pp	-0,08	0,39	0,73	0,51	0,43	0,03	0,19	0,14	-0,06	-0,65	0,38	0,67	1,00	0,65	0,67	0,57	0,51	0,48	0,11	0,44	0,40	0,38	0,20	0,37	0,33	0,70	-0,29	0,75	0,50	0,74	0,53	0,55	0,55	0,55	0,52	0,50	-0,04	0,74	0,53	0,52	0,51	0,55	0,50
AWCR_au	0,19	0,37	0,75	0,80	0,69	0,11	0,53	0,49	0,27	-0,50	0,44	0,92	0,65	1,00	0,92	0,79	0,80	0,79	-0,12	0,63	0,63	0,46	0,07	0,52	0,38	0,80	0,08	0,52	0,76	0,67	0,31	0,75	0,77	0,77	0,79	0,78	0,04	0,79	0,75	0,80	0,80	0,77	0,78
AW_	0,18	0,46	0,78	0,81	0,71	0,13	0,53	0,47	0,25	-0,53	0,43	1,00	0,67	0,92	1,00	0,81	0,81	0,77	-0,11	0,64	0,62	0,48	0,08	0,55	0,37	0,85	0,05	0,55	0,78	0,69	0,33	0,77	0,79	0,79	0,81	0,80	0,04	0,83	0,77	0,81	0,81	0,79	0,77
Sig	0,34	0,45	0,79	0,88	0,70	0,06	0,58	0,53	0,33	-0,46	0,51	0,81	0,57	0,79	0,81	1,00	0,88	0,84	-0,23	0,71	0,59	0,57	0,08	0,63	0,35	0,81	0,13	0,48	0,78	0,52	0,29	0,91	0,95	0,95	0,90	0,84	0,08	0,72	0,93	0,90	0,88	0,95	0,82
millers_h	0,39	0,44	0,74	1,00	0,77	0,12	0,67	0,61	0,38	-0,47	0,53	0,81	0,51	0,80	0,81	0,88	1,00	0,94	-0,25	0,69	0,62	0,55	0,06	0,64	0,35	0,77	0,22	0,42	0,88	0,51	0,27	0,82	0,85	0,85	0,94	0,93	0,08	0,67	0,85	0,94	1,00	0,85	0,86
Fc	0,41	0,38	0,70	0,94	0,75	0,11	0,68	0,64	0,40	-0,45	0,54	0,77	0,48	0,79	0,77	0,84	0,94	1,00	-0,27	0,67	0,63	0,53	0,05	0,61	0,35	0,73	0,25	0,39	0,87	0,49	0,24	0,79	0,81	0,81	0,89	0,89	0,09	0,64	0,81	0,90	0,94	0,81	0,87
PI	-0,45	-0,04	-0,11	-0,25	-0,14	0,13	-0,29	-0,30	-0,30	-0,07	-0,25	-0,11	0,11	-0,12	-0,11	-0,23	-0,25	-0,27	1,00	-0,26	-0,15	-0,22	0,09	-0,25	0,03	-0,11	-0,25	0,13	-0,22	0,16	0,22	-0,24	-0,24	-0,24	-0,26	-0,24	-0,16	-0,01	-0,25	-0,24	-0,25	-0,24	-0,27
Sum_af_pp_top_n_cits	0,30	0,41	0,61	0,69	0,63	0,02	0,51	0,45	0,30	-0,34	0,40	0,64	0,44	0,63	0,64	0,71	0,69	0,67	-0,26	1,00	0,57	0,51	0,04	0,54	0,29	0,62	0,13	0,36	0,66	0,41	0,17	0,68	0,71	0,71	0,73	0,69	0,04	0,55	0,74	0,71	0,69	0,71	0,70
Sum_pp_top_prop	0,23	0,28	0,52	0,62	0,56	0,04	0,51	0,49	0,33	-0,32	0,33	0,62	0,40	0,63	0,62	0,59	0,62	0,63	-0,15	0,57	1,00	0,32	0,00	0,39	0,40	0,54	0,25	0,27	0,63	0,46	0,14	0,55	0,57	0,57	0,63	0,64	0,06	0,55	0,59	0,61	0,62	0,57	0,66
average_mjs_mcs	0,33	0,39	0,55	0,55	0,44	-0,08	0,38	0,33	0,18	-0,37	0,44	0,48	0,38	0,46	0,48	0,57	0,55	0,53	-0,22	0,51	0,32	1,00	0,29	0,82	0,42	0,50	0,20	0,26	0,54	0,31	0,27	0,55	0,56	0,56	0,57	0,55	0,09	0,42	0,57	0,56	0,55	0,56	0,54
Min_af_mjs_mcs	-0,03	0,11	0,18	0,06	-0,04	-0,25	-0,08	-0,11	-0,21	-0,25	0,17	0,08	0,20	0,07	0,08	0,08	0,06	0,05	0,09	0,04	0,00	0,29	1,00	0,16	0,27	0,11	-0,04	0,12	0,06	0,12	0,27	0,09	0,08	0,08	0,06	0,06	-0,04	0,09	0,09	0,06	0,06	0,08	0,05
Maks_af_mjs_mcs	0,39	0,41	0,56	0,64	0,54	0,01	0,49	0,44	0,27	-0,37	0,47	0,55	0,37	0,52	0,55	0,63	0,64	0,61	-0,25	0,54	0,39	0,82	0,16	1,00	0,39	0,54	0,24	0,27	0,62	0,34	0,23	0,61	0,61	0,61	0,64	0,63	0,12	0,47	0,62	0,64	0,64	0,61	0,62
Gennemsnit_af_Mnjs	0,08	0,16	0,36	0,35	0,28	-0,06	0,22	0,21	0,09	-0,31	0,25	0,37	0,33	0,38	0,37	0,35	0,35	0,35	0,03	0,29	0,40	0,42	0,27	0,39	1,00	0,29	0,34	0,09	0,35	0,34	0,25	0,32	0,32	0,32	0,35	0,34	0,00	0,36	0,33	0,33	0,35	0,32	0,34
IQP	0,20	0,47	0,88	0,77	0,66	0,08	0,46	0,40	0,18	-0,59	0,51	0,85	0,70	0,80	0,85	0,81	0,77	0,73	-0,11	0,62	0,54	0,50	0,11	0,54	0,29	1,00	-0,08	0,65	0,74	0,64	0,40	0,77	0,79	0,79	0,79	0,76	0,04	0,77	0,77	0,79	0,77	0,79	0,75
nproductivity_adjusted_papers	0,45	-0,02	-0,07	0,22	0,20	0,03	0,49	0,51	0,53	0,19	0,05	0,05	-0,29	0,08	0,05	0,13	0,22	0,25	-0,25	0,13	0,25	0,20	-0,04	0,24	0,34	-0,08	1,00	-0,53	0,25	-0,14	-0,37	0,11	0,12	0,12	0,19	0,21	0,12	-0,05	0,13	0,18	0,22	0,12	0,22
times_cited_more_frequently_tl	-0,08	0,36	0,61	0,42	0,37	0,04	0,14	0,09	-0,08	-0,57	0,35	0,55	0,75	0,52	0,55	0,48	0,42	0,39	0,13	0,36	0,27	0,26	0,12	0,27	0,09	0,65	-0,53	1,00	0,42	0,63	0,53	0,47	0,47	0,47	0,43	0,41	-0,02	0,60	0,45	0,44	0,42	0,47	0,42
h	0,40	0,44	0,70	0,88	0,80	0,17	0,69	0,63	0,38	-0,51	0,55	0,78	0,50	0,76	0,78	0,78	0,88	0,87	-0,22	0,66	0,63	0,54	0,06	0,62	0,35	0,74	0,25	0,42	1,00	0,59	0,33	0,72	0,74	0,74	0,86	0,91	0,08	0,64	0,74	0,86	0,88	0,74	0,86
m_quotient	-0,10	0,36	0,60	0,51	0,51	0,14	0,30	0,25	0,07	-0,57	0,30	0,69	0,74	0,67	0,69	0,52	0,51	0,49	0,16	0,41	0,46	0,31	0,12	0,34	0,34	0,64	-0,14	0,63	0,59	1,00	0,47	0,47	0,48	0,48	0,51	0,52	-0,04	0,72	0,48	0,51	0,51	0,48	0,52
h_norm	-0,03	0,21	0,44	0,27	0,23	-0,03	-0,02	-0,06	-0,27	-0,52	0,37	0,33	0,53	0,31	0,33	0,29	0,27	0,24	0,22	0,17	0,14	0,27	0,27	0,23	0,25	0,40	-0,37	0,53	0,33	0,47	1,00	0,25	0,26	0,26	0,27	0,27	-0,03	0,36	0,25	0,27	0,27	0,26	0,26
m_	0,34	0,42	0,76	0,82	0,65	0,03	0,54	0,49	0,31	-0,42	0,48	0,77	0,55	0,75	0,77	0,91	0,82	0,79	-0,24	0,68	0,55	0,55	0,09	0,61	0,32	0,77	0,11	0,47	0,72	0,47	0,25	1,00	0,95	0,95	0,85	0,79	0,07	0,69	0,89	0,88	0,82	0,95	0,79
A_	0,34	0,44	0,77	0,85	0,67	0,03	0,55	0,50	0,31	-0,43	0,49	0,79	0,55	0,77	0,79	0,95	0,85	0,81	-0,24	0,71	0,57	0,56	0,08	0,61	0,32	0,79	0,12	0,47	0,74	0,48	0,26	0,95	1,00	1,00	0,88	0,81	0,09	0,70	0,93	0,89	0,85	1,00	0,81
R_	0,34	0,44	0,77	0,85	0,67	0,03	0,55	0,50	0,31	-0,43	0,49	0,79	0,55	0,77	0,79	0,95	0,85	0,81	-0,24	0,71	0,57	0,56	0,08	0,61	0,32	0,79	0,12	0,47	0,74	0,48	0,26	0,95	1,00	1,00	0,88	0,81	0,09	0,70	0,93	0,89	0,85	1,00	0,81
g	0,38	0,46	0,76	0,94	0,77	0,10	0,64	0,58	0,37	-0,46	0,52	0,81	0,52	0,79	0,81	0,90	0,94	0,89	-0,26	0,73	0,63	0,57	0,06	0,64	0,35	0,79	0,19	0,43	0,86	0,51	0,27	0,85	0,88	0,88	1,00	0,96	0,09	0,73	0,91	0,94	0,94	0,88	0,90
hg_	0,38	0,45	0,72	0,93	0,78	0,13	0,66	0,60	0,37	-0,46	0,51	0,80	0,50	0,78	0,80	0,84	0,93	0,89	-0,24	0,69	0,64	0,55	0,06	0,63	0,34	0,76	0,21	0,41	0,91	0,52	0,27	0,79	0,81	0,81	0,96	1,00	0,09	0,70	0,85	0,92	0,93	0,81	0,90
WU	0,22	0,03	0,02	0,08	0,08	0,01	0,15	0,16	0,15	0,00	0,11	0,04	-0,04	0,04	0,04	0,08	0,08	0,09	-0,16	0,04	0,06	0,09	-0,04	0,12	0,00	0,04	0,12	-0,02	0,08	-0,04	-0,03	0,07	0,09	0,09	0,09	0,09	1,00	0,03	0,07	0,08	0,08	0,09	0,08
mg_quotient	0,06	0,45	0,73	0,67	0,61	0,14	0,41	0,36	0,18	-0,50	0,34	0,83	0,74	0,79	0,83	0,72	0,67	0,64	-0,01	0,55	0,55	0,42	0,09	0,47	0,36	0,77	-0,05	0,60	0,64	0,72	0,36	0,69	0,70	0,70	0,73	0,70	0,03	1,00	0,71	0,69	0,67	0,70	0,69
e	0,33	0,46	0,76	0,85	0,69	0,04	0,56	0,51	0,32	-0,43	0,49	0,77	0,53	0,75	0,77	0,93	0,85	0,81	-0,25	0,74	0,59	0,57	0,09	0,62	0,33	0,77	0,13	0,45	0,74	0,48	0,25	0,89	0,93	0,93	0,91	0,85	0,07	0,71	1,00	0,88	0,85	0,93	0,84
Q2	0,38	0,44	0,76	0,94	0,74	0,09	0,62	0,57	0,35	-0,45	0,51	0,81	0,52	0,80	0,81	0,90	0,94	0,90	-0,24	0,71	0,61	0,56	0,06	0,64	0,33	0,79	0,18	0,44	0,86	0,51	0,27	0,88	0,89	0,89	0,94	0,92	0,08	0,69	0,88	1,00	0,94	0,89	0,87
h2	0,39	0,44	0,74	1,00	0,77	0,12	0,67	0,61	0,38	-0,47	0,53	0,81	0,51	0,80	0,81	0,88	1,00	0,94	-0,25	0,69	0,62	0,55	0,06	0,64	0,35	0,77	0,22	0,42	0,88	0,51	0,27	0,82	0,85	0,85	0,94	0,93	0,08	0,67	0,85	0,94	1,00	0,85	0,86
AR	0,34	0,44	0,77	0,85	0,67	0,03	0,55	0,50	0,31	-0,43	0,49	0,79	0,55	0,77	0,79	0,95	0,85	0,81	-0,24	0,71	0,57	0,56	0,08	0,61	0,32	0,79	0,12	0,47	0,74	0,48	0,26	0,95	1,00	1,00	0,88	0,81	0,09	0,70	0,93	0,89	0,85	1,00	0,81
Pop_h	0,40	0,41	0,71	0,86	0,76	0,11	0,67	0,62	0,39	-0,45	0,52	0,77	0,50	0,78	0,77	0,82	0,86	0,87	-0,27	0,70	0,66	0,54	0,05	0,62	0,34	0,75	0,22	0,42	0,86	0,52	0,26	0,79	0,81	0,81	0,90	0,90	0,08	0,69	0,84	0,87	0,86	0,81	1,00

Appendix 6: Correlation matrix *Public Health*

Indicator	Pyrs	App	CPP	Cites	Sum_of_self_cites	percent_sc	P	fp	nnC	percent_nc	Cage	AWCR_C	AWCR_pp	AWCR_au	AW_	Min_nCites	Maks_af_n_cits	Sig	Fc	PI	Sum_af_pp_top_n_cits	Sum_pp_top_prop	average_mjs_mcs	Min_af_mjs_mcs	Maks_af_mjs_mcs	Gennemsnit_af_Mnjs	IQP	nproductivity_adjusted_papers	times_cited_more_frequently_than_average_paper_	millers_h	h	m_quotient	h_norm	m_	A_	R_	g	hg_	mg_quotient	e	Q2	h2	AR	Pop_h
Pyrs	1,00	0,10	0,39	0,50	0,45	-0,18	0,48	0,47	0,34	-0,22	0,65	0,26	-0,20	0,26	0,26	-0,04	0,43	0,43	0,51	-0,60	0,51	0,28	0,49	-0,05	0,48	-0,01	0,38	0,48	-0,22	0,50	0,50	-0,06	-0,17	0,44	0,43	0,43	0,50	0,50	-0,06	0,46	0,49	0,50	0,43	0,51
App	0,10	1,00	0,26	0,28	0,29	0,01	0,25	0,02	0,26	-0,03	0,06	0,30	0,17	0,11	0,30	-0,06	0,30	0,30	0,13	-0,02	0,27	0,28	0,23	0,11	0,25	0,18	0,28	0,20	0,16	0,28	0,27	0,26	-0,11	0,27	0,31	0,31	0,29	0,28	0,28	0,29	0,28	0,28	0,31	0,11
CPP	0,39	0,26	1,00	0,64	0,55	-0,24	0,40	0,34	0,21	-0,32	0,48	0,60	0,43	0,55	0,60	0,00	0,70	0,70	0,61	-0,37	0,63	0,56	0,65	0,12	0,59	0,33	0,89	0,25	0,33	0,64	0,59	0,32	0,09	0,74	0,77	0,77	0,68	0,63	0,43	0,75	0,68	0,64	0,77	0,63
Cites	0,50	0,28	0,64	1,00	0,83	-0,14	0,76	0,64	0,51	-0,28	0,50	0,77	0,21	0,71	0,77	-0,08	0,77	0,77	0,83	-0,40	0,89	0,70	0,58	-0,03	0,66	0,22	0,73	0,55	0,14	1,00	0,89	0,43	-0,24	0,76	0,79	0,79	0,97	0,94	0,45	0,86	0,90	1,00	0,79	0,80
Sum_of_self_cites	0,45	0,29	0,55	0,83	1,00	0,04	0,77	0,64	0,51	-0,30	0,45	0,74	0,18	0,68	0,74	-0,08	0,65	0,65	0,75	-0,33	0,79	0,64	0,49	-0,03	0,57	0,20	0,64	0,56	0,12	0,83	0,85	0,48	-0,26	0,67	0,67	0,67	0,82	0,84	0,46	0,73	0,79	0,83	0,67	0,74
percent_sc	-0,18	0,01	-0,24	-0,14	0,04	1,00	-0,03	-0,04	-0,05	-0,06	-0,17	-0,09	-0,11	-0,10	-0,09	-0,01	-0,23	-0,23	-0,14	0,25	-0,15	-0,15	-0,26	-0,03	-0,22	-0,11	-0,21	-0,02	-0,07	-0,14	-0,07	0,07	-0,06	-0,21	-0,24	-0,24	-0,16	-0,11	-0,04	-0,20	-0,14	-0,14	-0,24	-0,14
P	0,48	0,25	0,40	0,76	0,77	-0,03	1,00	0,76	0,68	-0,20	0,39	0,65	0,02	0,61	0,65	-0,12	0,58	0,58	0,71	-0,33	0,75	0,60	0,41	-0,13	0,54	0,11	0,49	0,71	-0,02	0,76	0,81	0,42	-0,45	0,57	0,57	0,57	0,74	0,77	0,38	0,63	0,70	0,76	0,57	0,69
fp	0,47	0,02	0,34	0,64	0,64	-0,04	0,76	1,00	0,57	-0,22	0,41	0,52	-0,04	0,60	0,52	-0,09	0,49	0,49	0,72	-0,34	0,63	0,51	0,35	-0,18	0,47	0,04	0,42	0,64	-0,08	0,64	0,68	0,33	-0,42	0,50	0,48	0,48	0,62	0,65	0,28	0,53	0,60	0,64	0,48	0,70
nnC	0,34	0,26	0,21	0,51	0,51	-0,05	0,68	0,57	1,00	0,15	0,14	0,45	-0,08	0,41	0,45	-0,18	0,44	0,44	0,46	-0,18	0,50	0,44	0,28	-0,18	0,40	0,06	0,29	0,62	-0,10	0,51	0,54	0,32	-0,54	0,39	0,41	0,41	0,50	0,52	0,28	0,45	0,48	0,51	0,41	0,45
percent_nc	-0,22	-0,03	-0,32	-0,28	-0,30	-0,06	-0,20	-0,22	0,15	1,00	-0,46	-0,26	-0,18	-0,27	-0,26	-0,17	-0,20	-0,20	-0,31	0,20	-0,29	-0,21	-0,20	-0,06	-0,18	-0,06	-0,32	-0,09	-0,17	-0,28	-0,30	-0,18	-0,14	-0,24	-0,23	-0,23	-0,28	-0,29	-0,17	-0,26	-0,28	-0,28	-0,23	-0,32
Cage	0,65	0,06	0,48	0,50	0,45	-0,17	0,39	0,41	0,14	-0,46	1,00	0,33	0,00	0,33	0,33	0,06	0,42	0,42	0,53	-0,61	0,50	0,29	0,53	0,04	0,45	0,05	0,46	0,31	-0,01	0,50	0,49	0,03	0,02	0,46	0,44	0,44	0,50	0,49	0,05	0,47	0,51	0,50	0,44	0,55
AWCR_C	0,26	0,30	0,60	0,77	0,74	-0,09	0,65	0,52	0,45	-0,26	0,33	1,00	0,38	0,81	1,00	-0,09	0,71	0,71	0,70	-0,25	0,73	0,75	0,47	-0,01	0,55	0,30	0,71	0,43	0,29	0,77	0,74	0,63	-0,22	0,69	0,73	0,73	0,78	0,76	0,68	0,76	0,75	0,77	0,73	0,69
AWCR_pp	-0,20	0,17	0,43	0,21	0,18	-0,11	0,02	-0,04	-0,08	-0,18	0,00	0,38	1,00	0,35	0,38	0,06	0,29	0,29	0,18	0,06	0,19	0,33	0,20	0,18	0,16	0,38	0,42	-0,16	0,74	0,21	0,16	0,45	0,28	0,31	0,33	0,33	0,23	0,20	0,59	0,29	0,23	0,21	0,33	0,20
AWCR_au	0,26	0,11	0,55	0,71	0,68	-0,10	0,61	0,60	0,41	-0,27	0,33	0,81	0,35	1,00	0,81	-0,07	0,64	0,64	0,75	-0,25	0,68	0,69	0,42	-0,05	0,51	0,28	0,64	0,41	0,26	0,71	0,69	0,59	-0,20	0,63	0,65	0,65	0,71	0,70	0,62	0,69	0,69	0,71	0,65	0,75
AW_	0,26	0,30	0,60	0,77	0,74	-0,09	0,65	0,52	0,45	-0,26	0,33	1,00	0,38	0,81	1,00	-0,09	0,71	0,71	0,70	-0,25	0,73	0,75	0,47	-0,01	0,55	0,30	0,71	0,43	0,29	0,77	0,74	0,63	-0,22	0,69	0,73	0,73	0,78	0,76	0,68	0,76	0,75	0,77	0,73	0,69
Min_nCites	-0,04	-0,06	0,00	-0,08	-0,08	-0,01	-0,12	-0,09	-0,18	-0,17	0,06	-0,09	0,06	-0,07	-0,09	1,00	-0,08	-0,08	-0,05	-0,01	-0,07	-0,08	-0,02	0,01	-0,07	0,01	-0,04	-0,12	0,06	-0,08	-0,08	-0,06	0,16	-0,06	-0,07	-0,07	-0,08	-0,08	-0,05	-0,07	-0,07	-0,08	-0,07	-0,03
Maks_af_n_cits	0,43	0,30	0,70	0,77	0,65	-0,23	0,58	0,49	0,44	-0,20	0,42	0,71	0,29	0,64	0,71	-0,08	1,00	1,00	0,68	-0,38	0,70	0,63	0,59	0,00	0,64	0,25	0,76	0,40	0,23	0,77	0,68	0,38	-0,16	0,71	0,87	0,87	0,80	0,73	0,48	0,86	0,73	0,77	0,87	0,66
Sig	0,43	0,30	0,70	0,77	0,65	-0,23	0,58	0,49	0,44	-0,20	0,42	0,71	0,29	0,64	0,71	-0,08	1,00	1,00	0,68	-0,38	0,70	0,63	0,59	0,00	0,64	0,25	0,76	0,40	0,23	0,77	0,68	0,38	-0,16	0,71	0,87	0,87	0,80	0,73	0,48	0,86	0,73	0,77	0,87	0,66
Fc	0,51	0,13	0,61	0,83	0,75	-0,14	0,71	0,72	0,46	-0,31	0,53	0,70	0,18	0,75	0,70	-0,05	0,68	0,68	1,00	-0,42	0,81	0,66	0,54	-0,06	0,60	0,20	0,68	0,53	0,10	0,83	0,81	0,39	-0,21	0,72	0,71	0,71	0,82	0,82	0,41	0,76	0,81	0,83	0,71	0,88
PI	-0,60	-0,02	-0,37	-0,40	-0,33	0,25	-0,33	-0,34	-0,18	0,20	-0,61	-0,25	0,06	-0,25	-0,25	-0,01	-0,38	-0,38	-0,42	1,00	-0,41	-0,25	-0,49	0,01	-0,42	-0,01	-0,36	-0,30	0,08	-0,40	-0,37	0,04	0,09	-0,37	-0,38	-0,38	-0,41	-0,39	0,00	-0,40	-0,40	-0,40	-0,38	-0,43
Sum_af_pp_top_n_cits	0,51	0,27	0,63	0,89	0,79	-0,15	0,75	0,63	0,50	-0,29	0,50	0,73	0,19	0,68	0,73	-0,07	0,70	0,70	0,81	-0,41	1,00	0,70	0,58	-0,03	0,64	0,23	0,71	0,57	0,10	0,89	0,90	0,43	-0,24	0,77	0,74	0,74	0,89	0,91	0,43	0,80	0,89	0,89	0,74	0,82
Sum_pp_top_prop	0,28	0,28	0,56	0,70	0,64	-0,15	0,60	0,51	0,44	-0,21	0,29	0,75	0,33	0,69	0,75	-0,08	0,63	0,63	0,66	-0,25	0,70	1,00	0,43	-0,04	0,51	0,32	0,64	0,44	0,22	0,70	0,68	0,54	-0,20	0,66	0,66	0,66	0,71	0,70	0,57	0,68	0,70	0,70	0,66	0,66
average_mjs_mcs	0,49	0,23	0,65	0,58	0,49	-0,26	0,41	0,35	0,28	-0,20	0,53	0,47	0,20	0,42	0,47	-0,02	0,59	0,59	0,54	-0,49	0,58	0,43	1,00	0,14	0,73	0,34	0,62	0,38	0,07	0,58	0,55	0,19	0,00	0,59	0,63	0,63	0,61	0,58	0,25	0,63	0,59	0,58	0,63	0,56
Min_af_mjs_mcs	-0,05	0,11	0,12	-0,03	-0,03	-0,03	-0,13	-0,18	-0,18	-0,06	0,04	-0,01	0,18	-0,05	-0,01	0,01	0,00	0,00	-0,06	0,01	-0,03	-0,04	0,14	1,00	0,00	0,22	0,07	-0,11	0,11	-0,03	-0,05	0,00	0,29	0,00	0,03	0,03	-0,01	-0,03	0,03	0,01	-0,02	-0,03	0,03	-0,05
Maks_af_mjs_mcs	0,48	0,25	0,59	0,66	0,57	-0,22	0,54	0,47	0,40	-0,18	0,45	0,55	0,16	0,51	0,55	-0,07	0,64	0,64	0,60	-0,42	0,64	0,51	0,73	0,00	1,00	0,28	0,61	0,48	0,06	0,66	0,63	0,27	-0,15	0,60	0,65	0,65	0,67	0,65	0,32	0,66	0,64	0,66	0,65	0,61
Gennemsnit_af_Mnjs	-0,01	0,18	0,33	0,22	0,20	-0,11	0,11	0,04	0,06	-0,06	0,05	0,30	0,38	0,28	0,30	0,01	0,25	0,25	0,20	-0,01	0,23	0,32	0,34	0,22	0,28	1,00	0,31	0,16	0,12	0,22	0,21	0,29	0,16	0,27	0,28	0,28	0,24	0,22	0,33	0,27	0,24	0,22	0,28	0,23
IQP	0,38	0,28	0,89	0,73	0,64	-0,21	0,49	0,42	0,29	-0,32	0,46	0,71	0,42	0,64	0,71	-0,04	0,76	0,76	0,68	-0,36	0,71	0,64	0,62	0,07	0,61	0,31	1,00	0,31	0,34	0,73	0,67	0,41	0,00	0,79	0,84	0,84	0,77	0,72	0,50	0,82	0,76	0,73	0,84	0,69
nproductivity_adjusted_papers	0,48	0,20	0,25	0,55	0,56	-0,02	0,71	0,64	0,62	-0,09	0,31	0,43	-0,16	0,41	0,43	-0,12	0,40	0,40	0,53	-0,30	0,57	0,44	0,38	-0,11	0,48	0,16	0,31	1,00	-0,28	0,55	0,61	0,25	-0,50	0,39	0,39	0,39	0,53	0,57	0,18	0,44	0,51	0,55	0,39	0,52
times_cited_more_frequently_t	-0,22	0,16	0,33	0,14	0,12	-0,07	-0,02	-0,08	-0,10	-0,17	-0,01	0,29	0,74	0,26	0,29	0,06	0,23	0,23	0,10	0,08	0,10	0,22	0,07	0,11	0,06	0,12	0,34	-0,28	1,00	0,14	0,09	0,37	0,25	0,24	0,25	0,25	0,16	0,12	0,50	0,22	0,16	0,14	0,25	0,12
millers_h	0,50	0,28	0,64	1,00	0,83	-0,14	0,76	0,64	0,51	-0,28	0,50	0,77	0,21	0,71	0,77	-0,08	0,77	0,77	0,83	-0,40	0,89	0,70	0,58	-0,03	0,66	0,22	0,73	0,55	0,14	1,00	0,89	0,43	-0,24	0,76	0,79	0,79	0,97	0,94	0,45	0,86	0,90	1,00	0,79	0,80
h	0,50	0,27	0,59	0,89	0,85	-0,07	0,81	0,68	0,54	-0,30	0,49	0,74	0,16	0,69	0,74	-0,08	0,68	0,68	0,81	-0,37	0,90	0,68	0,55	-0,05	0,63	0,21	0,67	0,61	0,09	0,89	1,00	0,47	-0,25	0,72	0,70	0,70	0,88	0,94	0,44	0,77	0,88	0,89	0,70	0,82
m_quotient	-0,06	0,26	0,32	0,43	0,48	0,07	0,42	0,33	0,32	-0,18	0,03	0,63	0,45	0,59	0,63	-0,06	0,38	0,38	0,39	0,04	0,43	0,54	0,19	0,00	0,27	0,29	0,41	0,25	0,37	0,43	0,47	1,00	-0,15	0,38	0,39	0,39	0,44	0,45	0,80	0,42	0,43	0,43	0,39	0,41
h_norm	-0,17	-0,11	0,09	-0,24	-0,26	-0,06	-0,45	-0,42	-0,54	-0,14	0,02	-0,22	0,28	-0,20	-0,22	0,16	-0,16	-0,16	-0,21	0,09	-0,24	-0,20	0,00	0,29	-0,15	0,16	0,00	-0,50	0,25	-0,24	-0,25	-0,15	1,00	-0,10	-0,13	-0,13	-0,21	-0,22	-0,10	-0,15	-0,18	-0,24	-0,13	-0,17
m_	0,44	0,27	0,74	0,76	0,67	-0,21	0,57	0,50	0,39	-0,24	0,46	0,69	0,31	0,63	0,69	-0,06	0,71	0,71	0,72	-0,37	0,77	0,66	0,59	0,00	0,60	0,27	0,79	0,39	0,24	0,76	0,72	0,38	-0,10	1,00	0,82	0,82	0,79	0,76	0,46	0,81	0,85	0,76	0,82	0,74
A_	0,43	0,31	0,77	0,79	0,67	-0,24	0,57	0,48	0,41	-0,23	0,44	0,73	0,33	0,65	0,73	-0,07	0,87	0,87	0,71	-0,38	0,74	0,66	0,63	0,03	0,65	0,28	0,84	0,39	0,25	0,79	0,70	0,39	-0,13	0,82	1,00	1,00	0,83	0,76	0,49	0,93	0,79	0,79	1,00	0,70
R_	0,43	0,31	0,77	0,79	0,67	-0,24	0,57	0,48	0,41	-0,23	0,44	0,73	0,33	0,65	0,73	-0,07	0,87	0,87	0,71	-0,38	0,74	0,66	0,63	0,03	0,65	0,28	0,84	0,39	0,25	0,79	0,70	0,39	-0,13	0,82	1,00	1,00	0,83	0,76	0,49	0,93	0,79	0,79	1,00	0,70
g	0,50	0,29	0,68	0,97	0,82	-0,16	0,74	0,62	0,50	-0,28	0,50	0,78	0,23	0,71	0,78	-0,08	0,80	0,80	0,82	-0,41	0,89	0,71	0,61	-0,01	0,67	0,24	0,77	0,53	0,16	0,97	0,88	0,44	-0,21	0,79	0,83	0,83	1,00	0,94	0,47	0,90	0,92	0,97	0,83	0,81
hg_	0,50	0,28	0,63	0,94	0,84	-0,11	0,77	0,65	0,52	-0,29	0,49	0,76	0,20	0,70	0,76	-0,08	0,73	0,73	0,82	-0,39	0,91	0,70	0,58	-0,03	0,65	0,22	0,72	0,57	0,12	0,94	0,94	0,45	-0,22	0,76	0,76	0,76	0,94	1,00	0,45	0,83	0,91	0,94	0,76	0,82
mg_quotient	-0,06	0,28	0,43	0,45	0,46	-0,04	0,38	0,28	0,28	-0,17	0,05	0,68	0,59	0,62	0,68	-0,05	0,48	0,48	0,41	0,00	0,43	0,57	0,25	0,03	0,32	0,33	0,50	0,18	0,50	0,45	0,44	0,80	-0,10	0,46	0,49	0,49	0,47	0,45	1,00	0,48	0,45	0,45	0,49	0,42
e	0,46	0,29	0,75	0,86	0,73	-0,20	0,63	0,53	0,45	-0,26	0,47	0,76	0,29	0,69	0,76	-0,07	0,86	0,86	0,76	-0,40	0,80	0,68	0,63	0,01	0,66	0,27	0,82	0,44	0,22	0,86	0,77	0,42	-0,15	0,81	0,93	0,93	0,90	0,83	0,48	1,00	0,85	0,86	0,93	0,75
Q2	0,49	0,28	0,68	0,90	0,79	-0,14	0,70	0,60	0,48	-0,28	0,51	0,75	0,23	0,69	0,75	-0,07	0,73	0,73	0,81	-0,40	0,89	0,70	0,59	-0,02	0,64	0,24	0,76	0,51	0,16	0,90	0,88	0,43	-0,18	0,85	0,79	0,79	0,92	0,91	0,45	0,85	1,00	0,90	0,79	0,82
h2	0,50	0,28	0,64	1,00	0,83	-0,14	0,76	0,64	0,51	-0,28	0,50	0,77	0,21	0,71	0,77	-0,08	0,77	0,77	0,83	-0,40	0,89	0,70	0,58	-0,03	0,66	0,22	0,73	0,55	0,14	1,00	0,89	0,43	-0,24	0,76	0,79	0,79	0,97	0,94	0,45	0,86	0,90	1,00	0,79	0,80
AR	0,43	0,31	0,77	0,79	0,67	-0,24	0,57	0,48	0,41	-0,23	0,44	0,73	0,33	0,65	0,73	-0,07	0,87	0,87	0,71	-0,38	0,74	0,66	0,63	0,03	0,65	0,28	0,84	0,39	0,25	0,79	0,70	0,39	-0,13	0,82	1,00	1,00	0,83	0,76	0,49	0,93	0,79	0,79	1,00	0,70
Pop_h	0,51	0,11	0,63	0,80	0,74	-0,14	0,69	0,70	0,45	-0,32	0,55	0,69	0,20	0,75	0,69	-0,03	0,66	0,66	0,88	-0,43	0,82	0,66	0,56	-0,05	0,61	0,23	0,69	0,52	0,12	0,80	0,82	0,41	-0,17	0,74	0,70	0,70	0,81	0,82	0,42	0,75	0,82	0,80	0,70	1,00

ACUMEN
academic careers understood through measurement and norms

FP7 Grant Agreement	266632
Deliverable No and Title	D5.8 Part 7 - Comparison of indicators in Google Scholar and Web of Science
Dissemination level	Public (PU)
Work Package	WP5-Bibliometric Indicators
Version	1.0
Release Date	24. April 2014
Author(s)	Lorna Wildgaard Birger Larsen Jesper W Schneider
Project Website	http://research-acumen.eu/

European Commission

7th Framework Programme

SP4 - Capacities

Science in Society 2010

Grant Agreement: 266632

EUROPEAN COMMISSION

SEVENTH FRAMEWORK PROGRAMME

Comparision of indicators in Google Scholar and Web of Science

Lorna Wildgaard[a] Jesper W Schneider[b] Birger Larsen[c]

[a] Royal School of Library and Information Science, Birketinget 6, 2300 Copenhagen, Denmark

[b] Institut for Statskundskab - Dansk Center for Forskningsanalyse, Bartholins Allé 7, 8000 Aarhus C, Denmark

[c] Department of Communication and Psychology, Aalborg University Copenhagen, A. C. Meyers Vænge 15, 2450 Copenhagen SV, Denmark

Contents

Introduction 4
Coverage 4
Effect of database on author-level indicators 5
Age and seniority 6
Gender 8
Nationality 8
Citations per paper 10
Conclusions and recommendations 12

Introduction

We collected publication and citation data in two databases to investigate the extent performance of author-level indicators are effected by choice of database, the stability of indicators across databases and ultimately to illustrate how differences in the computed indicators change our perception of individual researchers. In this report we begin by comparing database coverage, coverage at seniority and gender-level and then the performance of four basic indicators computed in both databases. In the main deliverables D5.8 Part 5 and D5.8 Part 6, we investigate the performance of our previously identified indicators of author-level impact in Google Scholar and in Web of Science. Understanding the effect of the database used to source the data and the demographics of the researchers in our sample, will enable us to put the results of our cluster analysis in perspective and direct future studies.

Coverage

Out of the ACUMEN shared data set of 2154 researchers, 750 were identified as unique scholars having a working link to their curriculum vitae including/and a publication list. Publication and citation data was retrieved from Web of Science (Wos) and from Google Scholar (GS). A direct comparison between the two databases showed that WoS has about the same coverage for researchers as Google Scholar, Table 1.

Table 1. Overall coverage of Scholars in WoS and GS

Researchers with CV and publication list	Researchers covered in Web of Science	Researchers covered in Google Scholar
750	741	748
Difference to CV	9	2
Coverage	98%	99%

The researchers listed in total 62046 publications on their CVs and publication lists. Overall GS retrieved 41613 unique records more than WoS. Wos covered 50% of the records reported on CVs and publication lists, while GS covered 116%, Table 2. In both databases records that could be claimed by the searched researcher but not written on the CV or publication list were included. This is because CVs and publication lists sometimes only report selected papers or are not completely up-to-date.

Table 2. Overall coverage of publications in WoS and GS

Number of publications on CV	Number of records in WOS	Number of records in Google Scholar
62046	30967	72580
Difference to CV	31079	+10534
coverage	50%	116%

Researcher coverage differs only slightly from discipline to discipline in the two databases, Table 3. However the depth of coverage in the databases differs greatly between WoS and GS, which is of great importance for individual assessment. Further disciplinary coverage within WoS varies as well,

Table 4. In Wos Astronomy has a 58% coverage, while GS found more papers resulting in 132% coverage. Environmental Science has 46% coverage in WoS and 104% in GS, Philosophy 23% in WoS and 97% in GS and Public Health 80% in WoS and 136% in GS.

Table 3. Coverage of researchers in WoS and GS

Discipline	Researchers with CV & Publication list	Number in Wos	Difference	Coverage	Number in Google Scholar	Difference	Coverage
Astronomy	203	192	11	94%	193	10	95%
Environmental Science	203	195	8	96%	195	8	96%
Philosophy	250	222	28	88%	229	21	91%
Public Health	137	132	5	96%	132	5	96%

Table 4. Disciplinary coverage in Wos and GS

Discipline	Number of publications on CV	Number in WoS	Difference CV	Coverage	Number in Google Scholar	Difference CV	Coverage
Astronomy	21169	12359	8810	58%	28127	+6958	132%
Environmental Science	16720	7820	8900	46%	17453	+733	104%
Philosophy	15090	3494	11596	23%	14708	382	97%
Public Health	9067	7294	1773	80%	12387	+3320	136%

Effect of database on author-level indicators

Raw citation count alone is not an indicator of impact; citation counts need to be benchmarked or normalized to similar research. Citation patterns differ greatly between sub-disciplines and the types of publications a researcher publishes. Also citations accumulate over time, so the year of publication must be taken into account. Four common indicators computed in Web of Science and Google Scholar were compared, Table 5.

Table 5. Average difference between indicators computed in Google Scholar and Web of Science

Discipline	Difference in mean academic age GS : WoS	Difference in mean CPP GS:WoS	Difference in mean H-index GS:WoS	Difference in mean m-quotient GS:WoS	Difference in mean g-index GS:WoS
Astronomy	+3 years	-4.5 CPP	+3.6h	0	+8.7g
Environmental Science	+4 years	-0.3 CPP	+2.7h	+0.7	+5.3g
Philosophy	+6 years	+2.9 CPP	+4.6h	+0.17	+9.3g
Public Health	+3 years	+1.4 CPP	+3.5h	+0.1	+7.8g

Across all disciplines the academic age of researchers are on average 4 years older in Google Scholar than Web of Science. Academic age is the number of years since the first publication for the researcher recorded in the database. This information is used to adjust many indicators to the length of a researcher's career to enable comparability. The average number of citations per paper is however only 0.7 citations between the two databases and the m-quotient is similar as well, with only a difference of 0.2; the h-index is on average 3.7 h higher in Google Scholar than Web of Science and likewise the g-index is also higher by 8.1. However, the performance of indicators of individual impact should not be compared across disciplines. Within disciplinary analysis reveals larger differences that favour Google Scholar as it produces the higher numbers, however data collection proved more reliable in Web of Science and as such we assume the reliability of the indicators to represent the actual publications and reception of the individual scholar is more accurate in WoS, Table 5. Interestingly the m-quotient is very similar on average per researcher in both databases. The m-quotient makes the h-index comparable, as it divides h by the number of years since the researcher's first publication recorded in the database thus enabling the comparison of researchers with different length of career.

Age and seniority

Early career researchers are defined as PhD and Post Docs, middle career are Assistant professors and senior researchers are associate professors. In this report we call professors "established researchers". As expected early career researchers are not as highly cited as researchers who have had a longer career. This is not an indication of quality, but simply that during their short career the work of these early career researchers has not had enough time to accumulate citations. Comparing their citations to field norm is uninformative. However, comparing their citations per paper to the expected number of citations of the articles in journals they publish in (CWTS indicator *average mjs mcs*) can be an indication of impact. In the WoS data set 396 researchers performed under the *average mjs mcs* (Sample A) and 345 researchers performed better than *average mjs mcs*, (Sample B). Normally field benchmarks are computed using the average number of citations per paper for a WoS subject category which may or may not represent the sub-specialty of the researcher. However, as *average mjs mcs* is calculated with a two year citation window, the junior researcher needs to have been published for two years to allow fair comparison, Table 6. This indicator is only comparable as an expected performance benchmark to the number of citations received to articles and reviews retrieved from WoS. The Table shows that publications written by senior and established staff are only performing marginally better than junior or middle career researchers. Seniority is not a classification of academic age, a Post Doc can for example have 6 or 15 yearlong publishing history. Apart from age, gender and nationality can have an effect on researchers' career paths and research output.

Table 6: Summary of actual citations to expected seniority performance (WoS)

Seniority	Average mjs mcs	Number of researchers	Number of researchers performing better	% achieving ≥expected
Astronomy				
PHD	7,583046907	15	9	60%
Post Doc	12,4729792	48	21	43%
Assis Prof	12,54805936	26	11	42%
Assoc Prof	16,36060726	66	29	43%
Full Professor	18,64497503	37	17	45%
Environmental Science				
PHD	11,54813557	3	0	0
Post Doc	4,932046506	17	8	47%
Assis Prof	8,275902941	39	14	35%
Assoc Prof	10,08383101	85	37	43%
Full Professor	12,4342212	51	25	49%
Philosophy				
PHD	1,237678971	8	2	25%
Post Doc	2,110023794	22	6	27%
Assis Prof	4,261891167	44	8	18%
Assoc Prof	3,826703308	73	18	24%
Full Professor	5,019210551	75	22	29%
Public Health				
PHD	6,30695831	9	4	44%
Post Doc	8,843720756	14	6	42%
Assis Prof	9,154821404	30	14	46%
Assoc Prof	12,69529504	50	26	52%
Full Professor	14,6056222	29	15	51%

Table 7: Overall performance of researchers compared to disciplinary benchmark (WoS)

Discipline	Number of researchers	Number in WoS	% researchers performing better than expected citation score
Astronomy	192	12359	45%
Environmental Science	195	7820	43%
Philosophy	222	3494	25%
Public Health	132	7294	49%

Gender

In the WoS data set there are 580 male researchers and 161 female researchers. Overall 44% of the female researchers perform better than expected, while 47% of the male researchers perform better than expected. Performance on a disciplinary level is shown in Table 8.

Table 8. Gender performance better than expected on a disciplinary level (WoS)

	Number of researchers	Number of publications	% of researchers performing better than expected	Citations per paper Sample A	Citations per paper Sample B
Astronomy					
Male	162	11163	59%	14.1	29.8
Female	30	1196	80%	15.7	29.5
Environmental Science					
Male	160	6874	46%	11.1	16.6
Female	35	946	60%	7.5	20.8
Philosophy					
Male	179	2889	32%	3.2	8.2
Female	43	605	20%	2.9	14.3
Public Health					
Male	79	4458	55%	13.1	19.4
Female	53	2836	32%	14.7	17.0

The average academic age in Sample A and Sample B are the same, 14 years. However Sample B, the high performing group, have on a greater amount of citations to a smaller amount of papers than Sample A, resulting in a higher rate of Citations Per Paper. Even though they produce fewer papers the female researchers' publications are achieving on average a higher impact than their male counterparts in all disciplines except Public Health.

Nationality

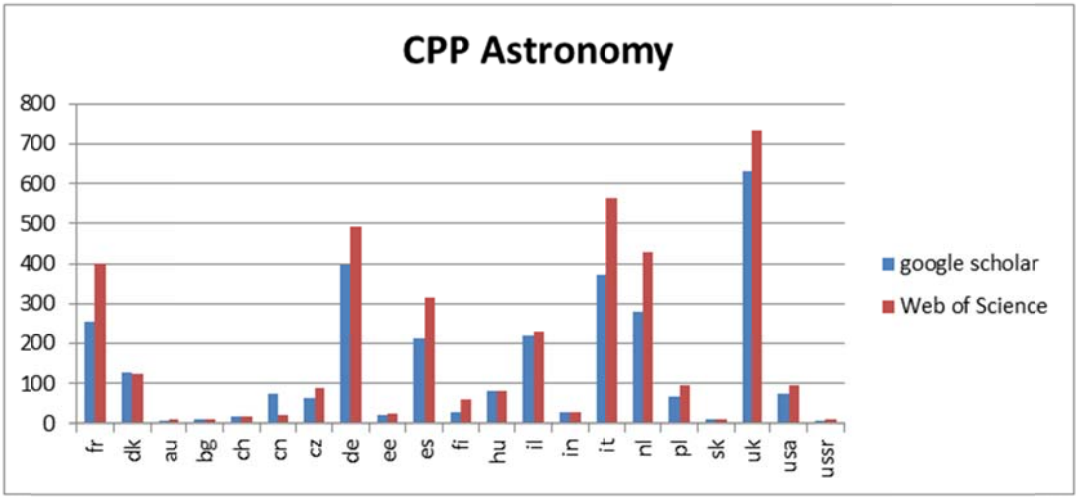
Nationality can also have an effect on researcher output and reception of their work. The researchers in our sample of researchers that are covered in GS and WoS are primarily western European, Table 9.

Table 9. Nationality of researchers

Nationality	nResearchers	% sample A	% sample B	Nationality	nResearchers	% sample A	% sample B
British	105	74	26	Finnish	14	85	15
Italian	78	78	12	Estonian	8	100	0
German	54	64	36	American	5	20	80
Spanish	46	80	20	Slovakian	4	100	0
Dutch	42	73	27	Bulgarian	2	100	0
French	33	54	46	Indian	2	100	0
Danish	27	92	8	Australian	1	0	100
Chzec	24	87	13	Chinese	1	0	100
Israelian	24	87	13	Greek	1	100	0
Polish	21	85	15	Russian	1	100	0
Hungarian	18	100	0	Swiss	1	0	100

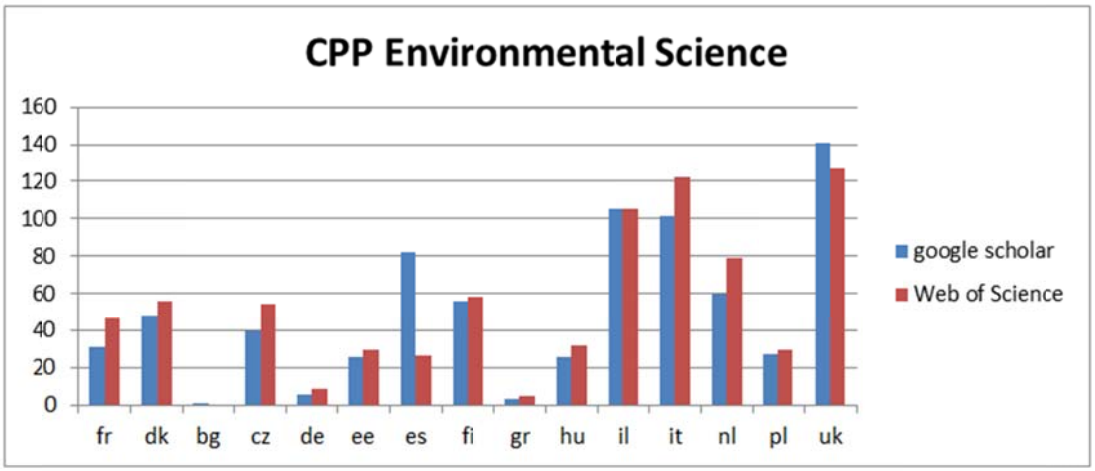
There is no clear grouping of nationalities in Sample A and Sample B. However, there is definite advantage for scholars of certain nationalities and disciplines to find citations in Google Scholar rather than WoS, Tables 10, 11 , 12, 13.

Table 10. Citations per paper in Astronomy



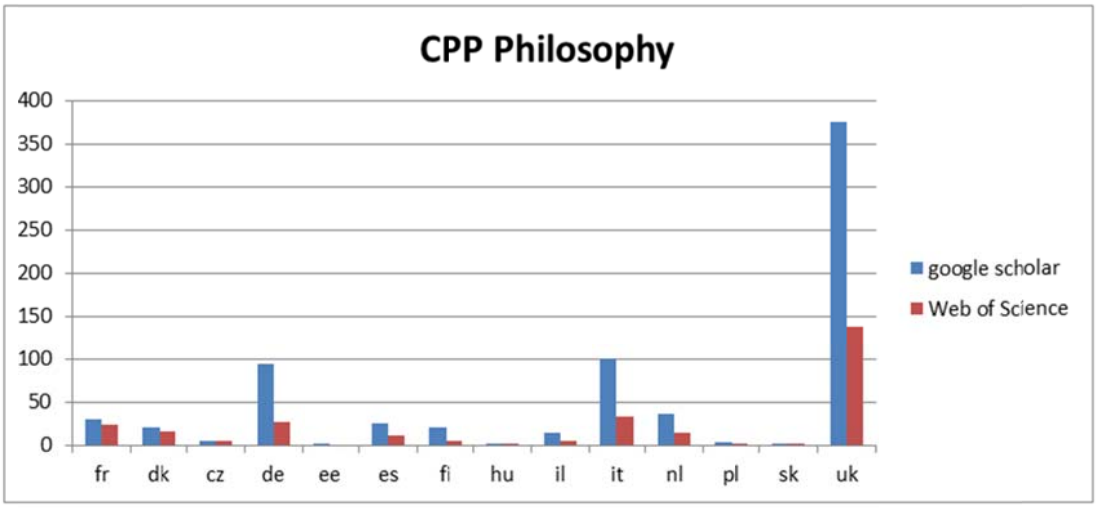
Conference papers are an important publication type for Astronomers, and as we experienced in our data-collection these were not available in our version of Web of Science and seriously reduced the amount of publications and citations per researcher. However, Web of Science still results in higher CPP for all researchers than Google Scholar.

Table 11. Citations per paper in Environmental Science



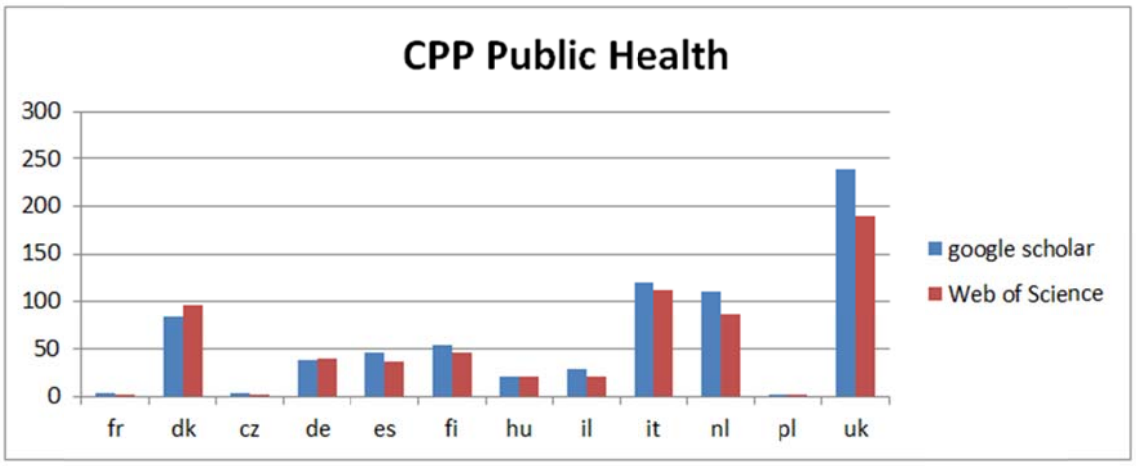
CPP is slightly improved in Web of Science across all nationalities apart from a noticeable improvement in Google Scholar for Spanish researchers.

Table 12. Citations per paper in Philosophy



Google Scholar clearly out performs Web of Science in indicating CPP for researchers in Philosophy, whereas for Public Health the resulting CPP is only slightly higher.

Table 13. Citations per paper in Public Health

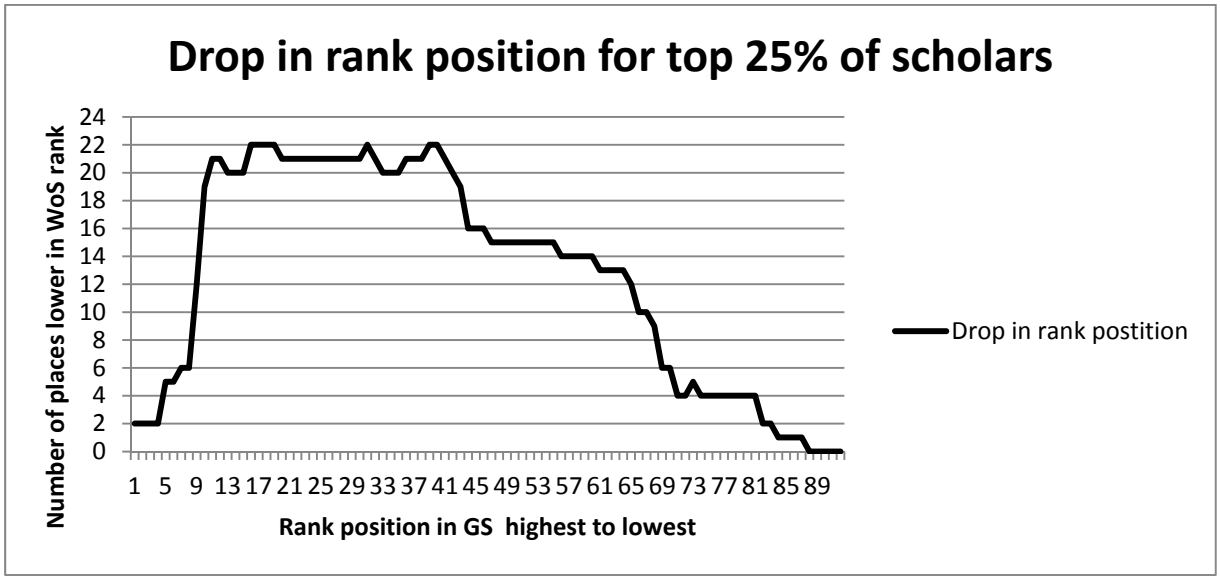


Citations per paper

In the previous section we exemplified database performance to nationalities using citaqtions per paper (CPP). CPP is considered a robust indicator of performance. But we wish to investigate if this indicator is database dependent or if it is database independent for the top performing researchers. It was possible to compute bibliometric indicators for 512 researchers in both WoS and GS. The number of CPP a researcher received in the Google Scholar data was compared to the Web of Science data. Even though there is a positive correlation between CPP in WoS and GS, r=0754, n=512, p=0.00, there is no correlation between the resulting ranks of the scholars. All scholars were ranked from highest to lowest CPP and there was no correlation between their rank position in Google Scholar and in Web of Science. The set was divided into quartiles to identify if the CPP was

stabile as a rank across databases for the top 25% of CPP scholars, r=0.051, n=128 and p=0.566. By manually investing the change of rank position in this top set, we found that 72% of the scholars appear in the top set in both databases, however the remaining 28% of scholars are entirely different from Google Scholar to Web of Science. On average the rank of the researcher in Google scholar was 12 places higher than the ranking of the same researchers in Web of Science, figure 1.

Figure 1: Number of places a scholar drops when ranked using CPP in Google Scholar compared to Web of Science



In the WoS data set 396 researchers performed under the *average mjs mcs* (Sample A) and 345 researchers performed better than *average mjs mcs*, (Sample B). Continuing the investigation of the stability of CPP, we investigated if researchers' whose publications out-perform the expected benchmarks, were well represented in the top 25% CPP. Eighty-one out of the 128 highest ranking CPP researchers in WoS, 63%, were from Sample B, while 65 researchers from Sample B where ranked top 25% CPP in Google Scholar, making up 50% of this sample.

Conclusions and recommendations

Our main finding is that indicators are highly dependent on the database used to compute them and the resulting impact-rankings of researchers are different. As such it is of utmost importance that the database used to collect the publication and citation data is reported alongside the indicators. Researchers who compute their indicators using Web of Science data should not be compared with researchers who compute indicators in Google Scholar. Further, our own data collection showed that different versions of the same database can also produce different results.

- Even though Google Scholar provided more publications and citations on an individual level, the work needed to clean the data to ensure researchers are only attributed with works that they authored is time consuming and sometimes impossible due to name ambiguities.

- The data retrieved from Web of Science was reliable, but limited in its coverage of the individual, which was detrimental to the outcome of the computed indicators in some disciplines and for some nationalities.

- Disciplinary and national coverage of a database should be established before author-level indicators are computed, as coverage can limit fair indications of the impact of work. Based on our study, we would recommend Philosophers use Google Scholar, well aware that this recommendation incurs increased work in cleaning and importing the publication and citation data.

- Raw citation count alone is not an indicator of impact; citation counts need to be benchmarked or normalized to similar research. Generally indicators computed using Google Scholar data are higher than indicators computed using Web of Science data.

- The m-quotient provides an indication of impact adjusted to the academic age of the researcher, and proved comparable across Google Scholar and Web of Science.

- A benchmark of expected citations for the researcher's speciality was calculated using only the Web of Science data. This was used to compare the impact of the individuals' publications. The results showed that even though female researchers produce fewer papers, they have a higher impact on average in their specialty than male researchers.