Scientometrics (2019) 118:1011–1025
https://doi.org/10.1007/s11192-018-2993-2

Highly cited researchers: a moving target

Domingo Docampo[1] · **Lawrence Cram**[2,3]

Received: 26 June 2018 / Published online: 25 January 2019

Abstract
Highly cited researchers are a category of researchers defined by scientometric rules relating to counts of citations to their scholarly articles. The designation often refers to researchers identified according to scientometric rules specified by the Institute of Scientific Information (ISI) and its commercial affiliates; we denote these categories as HCR. The 2001 ISI rules (HRC.1) used membership thresholds derived from the total citation counts to an author's corpus in a specified research field and time window. The modified 2013 rules also include counts of individual highly cited publications (HCR.2), while the foreshadowed 2018 rules introduce the concept of cross-field influence (HCR.3). The HCR category is a popular, albeit flawed, indicator of outstanding individual researchers. HCR membership has been used as the basis for many studies of research excellence, including the use of an institution's HCR count as an indicator in the Academic Ranking of World Universities (ARWU). The paper traces the development of the HCR category and its use by ARWU, providing insights into the social construction of research indicators and their potential to change research practice.

Keywords Academic rankings · World class universities · Indicators · Highly cited researcher

Introduction

The genesis of the scientometric category *Highly Cited Researcher* can be traced to the founding by Eugene Garfield of the Institute for Scientific Information Inc. (ISI) and the construction of the Science Citation Index (SCI) in the 1950s (Garfield 1998, 1967). The earliest reference to the phrase "highly cited research" in the JSTOR database (Burns et al. 2009) is Useem (1976), who used the ISI Social Science Citation Index in 1973 to reveal a correlation between a researcher's citation rate and the level of their government funding

✉ Lawrence Cram
Lawrence.Cram@anu.edu.au

1 Atlantic Research Center for Information and Communication Technologies, Universidade de Vigo, Campus Universitario, 36310 Vigo, Spain

2 Research School of Physical Science and Engineering, Australian National University, Canberra, ACT 0200, Australia

3 Charles Darwin University, Darwin, NT 0811, Australia

of academic social research. Individual and/or group membership of a category of highly cited researchers has been used in several levels of research performance evaluation, from the appointment and advancement of individual researchers (e.g. Petersen et al. 2014) to inter- and intra-university resource allocation (e.g. Moed et al. 1985) and comparison of national systems (e.g. King 2004).

In a series of papers published through the 1950s Garfield and colleagues foreshadowed the value and wide applicability of citation indexing by contrast with the then-common practice of subject indexing (e.g. Adair 1955; Garfield 1955, 1957). Among the applications mentioned were searching for the provenance and criticism of ideas, measuring the "*impact factor*" (quotation marks in the original—Garfield 1955) of works, researchers and journals, and increasing the information that researchers have about the uptake of their work. Garfield (1970) summarizes the utility of citation analysis in supporting research including the capability to identify and track creative people and their networks.

As ISI developed the SCI and its related computer-based analysis tools, the foreshadowed opportunities were progressively implemented. Lists of researchers identified by high citation counts to their corpus of work became a feature of ISI's *Current Contents.* By 1968 Garfield (see Garfield and Welljams-Dorof 1992) could report optimistically on the prediction of Nobel Prize winners using citation analysis and other factors, strengthening the notion that outstanding research performance could be identified by citation analysis. *Current Contents* generated an abundance of lists of productive researchers across many time windows and within many disciplines (e.g. Garfield 1978).

An important ISI innovation was the *Citation Laureate,* a term we believe was first used in Thomson's *Science Watch* to refer to a group of 30 Japanese scientists who had published a number of high impact papers (Thomson-ISI 2000). Thomson announced similar groups in other countries (e.g. Australia—Anon 2004). The meaning of the Citation Laureate term has evolved and now it is used to refer to a small subset of researchers of "Nobel calibre" (Pendlebury 2014).

Building on the experience of Garfield and his colleagues in identifying exceptional researchers from citation resources, Thomson-ISI announced in 2001 the commercial category Highly Cited Researcher (hereinafter we use the abbreviation HCR to designate variants of this Thomson-ISI category). A press release (Thomson-ISI 2001) described the HCR list as a "new and invaluable resource" that would allow researchers and scientists to "identify key individuals, departments, and laboratories that have made fundamental contributions to the development of science and technology in recent decades" and referred to the web address *highlycited.com* which Thomson-ISI had trademarked (USPTO 2001).

The original HCR rules used the total citations acquired by a researcher's indexed corpus over a specified time window and in a specified research field. Denoted herein as HCR.1, the rules were used to update HCR membership lists over the years 2000–2008 and to maintain affiliation lists subsequently. In late 2012, Thomson Reuters revised the rules to include counts of highly cited publications as well as total citations in a specified field (HCR.2). Clarivate, the current owner of the HCR product, has foreshadowed a significant broadening of the HCR category in 2018 with the introduction of rules to count HCR's with substantial cross-field contributions (HCR.3). We compare the methods below.

ISI's HCR.1 resource listed names, affiliations, disciplines and countries for a defined class of elite researchers, providing convenient data that could be used to study research success within disciplinary, institutional and geopolitical contexts. For example, Batty (2003) showed that the affiliations reported in the December 2002 HCR.1 list revealed remarkable geographical and institutional concentrations of highly cited scientists. While acknowledging the developing and likely volatile nature of the database, Batty foreshadowed that such findings would have implications for education policy, research resource allocations and changes in comparative research quality between regions on decadal time scales.

The geographic properties of the listed affiliations of HCR members were used to discuss emerging threats to national security (Paarlberg 2004) and to explore differences among indicators of national scientific impact (King 2004), spatial scientometrics (Frenken et al. 2009) and comparative national research performance (Basu 2006; Bornmann et al. 2015). HCR lists have also been used as the basis for studies of professional and personal characteristics of this class of highly cited researchers, including talent management (Van den Brink et al. 2013), their use of social media (Más-Bleda and Aguillo 2013) and their academic standing (Bornmann et al. 2017). The lists have been used to investigate social correlates of success as a researcher in a discipline (Parker et al. 2010, 2013) and to suggest staged administrative actions that might improve institutional research performance (Amara et al. 2015).

It was fortuitous that ISI released the first list of individual highly cited researchers around the time that the architects of the Academic Ranking of World Universities (ARWU—also known as the Shanghai Ranking) were developing this university ranking instrument. Given the purposes and academic values that determined how ARWU indicators are constructed (Liu and Cheng 2005) it is unsurprising that the count of an institution's HCRs was selected as an indicator. As a consequence of this choice, volatility in the HCR list contributes to the annual revision of an institution's position in the ARWU rankings. Since the ARWU ranking is widely used and may be "the most credible" (Fernández-Cano et al. 2018 make this observation in a thorough albeit critical study of the ARWU) it is important to understand how volatility arises not only from changes in researchers' citation profiles but also from methodological shifts and changes in other characteristics of researchers such as their affiliations and the fields of journals they select for their publications.

The use of HCR membership lists to evaluate researchers or institutions is a blunt approach to bibliometric analysis, subject to serious pitfalls (e.g. Moed et al. 1985; van Raan 2005) that can be mitigated by nuanced measures (e.g. van Leeuwen et al. 2003; Bornmann et al. 2008; Hicks et al. 2015). Mitigation is not always possible, so that the utility of citation analysis has been robustly and enduringly critiqued from many perspectives. For example, Cronin (2005) reviews the "fervid debate" while De Bellis (2009, p. 335) concludes that bibliometricians are perhaps driven "*toward an exclusive choice: either you do believe … or you don't.*" Our article acknowledges this debate but does not aim to engage it. Rather, we aim to explore some implications of citation analysis emerging from the HCR category and its volatility.

This paper has several aims: (1) to present the rules that establish membership of various HCR lists, (2) to exhibit how volatility arises from these rules, (3) to explain the use of HCR lists by ARWU and (4) to examine institutional management of HCRs. Our conclusions follow.

Thomson-ISI highly cited researchers (HCR)

As the utility of citation databases became recognised in the 1950–1960s, public institutions such as international consortia of national libraries or research universities might have decided to generate and maintain comprehensive and reliable citation databases and query tools, representing an innovative development of traditional library catalogues. This would have been consistent with their public service roles in supporting discovery, access and evaluation of public scholarly work. However, these institutions did not take up this opportunity. Private interests then had an opportunity to deal commercially in the archival knowledge networks built mainly on publicly funded research.

Provision of access to citation indices has commercial value, and accordingly there are business corporations that maintain commercial citation databases and provide access to them in a business context. ISI and its stable of indexes and publications was itself established by Garfield and his business colleagues as a private-sector initiative and commercial venture (Garfield 1998). A majority shareholding of ISI was sold to JPT Publishing in 1988 which was then acquired by a subsidiary of the Thomson Corporation in 1992 as a step towards Thomson's venture into the business of specialized digital information services (Lane 1992). The Thomson Corporation and Reuters merged in 2008 forming a multi-national media and information conglomerate. Throughout these developments, ISI remained both a brand and an identifiable business activity within the broader and evolving Thomson business setting. The ISI citation indices became widely available through subscription to the Web of Science and bespoke services offered by the Scientific and Scholarly Research branch of the IP and Science division of Thomson Reuters. This division was sold to Clarivate Analytics (Yap 2016) which subsequently "re-established the prestigious Institute for Scientific Information (ISI)" (Anon 2018).

Throughout the 1990s, the potential utility of electronic access to digital records of research had become apparent to a widening community. In particular, citation analysis was becoming widely used and more controversial for evaluation and other studies of the research enterprise (MacRoberts and MacRoberts 1996; Moed et al. 1995; Seglan 1997). Through its HCR product ISI could foster positive relationships with a subset of outstanding researchers likely to be (a) influential in their home institutions and (b) favourably inclined to the use of ISI's data products. As reported by Thomson-ISI (2008), the HCR list could "identify and honour researchers" who had contributed to its index of "the world's finest scholarly literature" as well as furthering ISI's corporate mission by providing "some of the most important and rich sources of citation navigation in the Web of Knowledge™."

Initial HCR rules (HCR.1)

The rules for inclusion in the first HCR lists (HCR.1) derived from the total number of citations received by a researcher's indexed articles published and cited within a specified time interval (1981–1999 for the first HCR.1 list, 1983–2002 for the second, 1984–2003 for the third, and hence forth) and within 21 broad categories of research. A research category is induced onto an individual article through its appearance in a journal which is part of the "journal set" responsible for that field. A journal might be included in more than one research category. In detail, the HCR.1 procedures for forming and extending the list included the following steps (McVeigh 2004; Thomson-ISI 2008):

- In sequence, select every article (source) in the ISI index that had been published between 1981 and 1999 inclusive.
- Attach that article to a field or fields of research, through the assignment of its publishing journal (outlet) to the journal set associated with the field/s.
- Extract the author list for that article and attach the field/s to each author of that article.

- Identify every other article in the ISI index that cites the selected article. For each citation to that article, add one citation to the citation count for every author of that article.
- Once all articles and all citing articles in the index have been processed, sort the list of authors attached to each field from the highest to the lowest number of counted citations.
- Within a field, commence a detailed editorial analysis of author names using the first 250 as the threshold for highly cited authors.
- Consider potential variants of the names of a researcher, eliminate homographs, seek consistency of field/s and address/es for researchers, and for each (potential) highly cited researcher converge to the “hallmarks of a single body of work”.
- Once sufficient information is assembled and winnowed regarding the set of highly cited authors within a field, directly contact the authors asking for a curriculum vitae that includes standardized information and a list of publications.
- Cross-check the list of publications, especially for researchers with common and/or homographic names, to confirm the validity of the identification.
- Publish the initial profile of all selected Highly Cited Researchers. As a matter of policy, each HCR was given “editorial control” over the published record, allowing addition, updating, or suppression “of any aspect of their biographical, contact, or publication information at their discretion.” To do this, researchers were provided with an ID and a password.
- On an annual or biannual basis, repeat the process by advancing the time window, and identifying any new researcher who enters the threshold for the most highly cited researchers. Repeat the analysis steps for potential new HCRs and publish the expansion of the list.
- In 2008, conduct the final update to include new HCRs. Maintain a list of names, fields and affiliations for the set of HCRs created by this process between 2001 and 2008, and support processes for adjusting affiliations.

The complexity of this process is, of course, symptomatic of the widely-recognized uncertainties that arise in the application of data bases such as the ISI Citation indexes. Critics of applications of these databases sometimes appear to believe that the creators and users are ignorant of the problems.

An ideal HCR.1 record contained both biographical and publication information. Biographical data included education, faculty or professional posts, society memberships and/or offices, and current research interests. Publication information listed journal articles, books or book chapters, conference presentations, and Web sites or other Internet resources. The information was search-able by individual, category, country, or institutional affiliation. Although the rich HCR.1 data base is no longer directly accessible, at its zenith the HCR.1 data potentially included the following information about each researcher, generally obtained from the researcher or their agent: Author Information, ISI Information, Contact Information, Personal Information, Education, Appointments and Affiliations, Honours and Awards, Memberships, Research Interests, Research Funding and Grants, Publications, and Profile Details.

The HCR.1 data base was launched in March 2001 with entries in the fields of chemistry, engineering, neuroscience and physics. A total of 397 researchers were included. Over subsequent months and years whole new fields were added, and the membership of already-published fields was enlarged. A new field was added when approximately 100 replies had been received in response to the solicitation described above. Newly added researchers in a field would comprise (a) those that had provided the requisite personal data in response to the initial mail-out to the top 250 researchers, as well as (b) researchers whose citation count grew to the point where they became eligible. Thomson-ISI (2001) reported that the initial HCR count comprised "less than one half of one percent of the almost 5 million researchers in the ISI Citation Database 1981–1999." This estimate is conservative: if there were 250 HCRs in each of 21 categories, the initial number of HCR would have been around 5250 or approximately 0.1% of the total number of researchers.

For several years that HCR.1 database could be updated at the request of a listed person and in some circumstances a representative of the affiliation of a listed person. As discussed below, HCR.1 members sometimes initiated a change in their affiliation, including a shift to an institution where they had performed none of the research reported in their highly-cited publications.

The archived HCR.1 list (Clarivate Analytics 2018a—denoted as the 2001 HCR list) contains 7032 entries with approximately 6533 distinct names and 2167 distinct (in a strict lexical sense) institutional names. Only one affiliation element is recorded, although multiple institutions are sometimes encoded in this element. Note that some researchers' highly cited articles have been published in journals that are coded by ISI to more than one field, while others may publish in journals that belong to the journal set for different fields. Therefore, the total count of records exceeds the count of individual HCRs.

Revised HCR rules (HCR.2)

After a fallow period 2008–2012 in which HCR.1 affiliations were updated but no further researcher were added, in late 2012 Thomson Reuters rekindled the HCR product "to make the Highly Cited Researcher methodology consistent with Essential Science Indicators process" (Thomson Reuters 2014). The HCR.2 rules identified researchers whose work exceeded the threshold for total citations as well as publishing the greatest number of high "citation impact" publications in a set of 21 fields of research over a 10-year window. The measure "citation impact" identifies outputs that are highly cited relative to other papers in the journal or journal set in question, adjusting for differences between fields. The measure is reminiscent of the Crown Indicator (Waltman et al. 2011) and broadly consistent with the advice of Bornmann and Leydesdorff (2018).

Co-authors of a highly cited publication get the same credit regardless of their role in the paper. HCR.2 status accrues to highly cited individuals who have also co-authored more than a defined threshold of highly cited papers within their discipline within the time frame of analysis. The initially proposed new method had certain methodological difficulties and one of us (Docampo) has contributed to the critical assessment of these. After significant further development of the rules across 2013–2015 the HCR.2 rules are as follows (Clarivate Analytics 2018b):

- The Clarivate Analytics *Essential Science Indicators* (ESI) product and methodology underpins HCR.2. ESI draws on the subset *Core Collection* of some 11,800 journals indexed in Science Citation Index Expanded (SCI) and Social Science Citation Index (SSCI). The types of publication considered are reviews and articles.
- Each journal in the Core Collection is assigned to one and only one research field. A publication in a journal classified as Multidisciplinary is allocated to a unique field by inspection.
- A Highly Cited Paper (HCP) is defined to rank in the top 1% of eligible papers in the Core Collection, accounting for field and publication year.

- Each author of an HCP is included in the subsequent analysis. Disambiguation of authors' names is conducted without reference to the authors themselves, through algorithms and expert review of original publications and web sites.
- To be considered for HCR.2 status, an author's total citation count must exceed the threshold of the top 1% of the field. Authors in this set are then ranked according to their number of HCPs. A candidate HCR has a publication count exceeding a threshold which is set to relate the number of HCR to the total number of authors in the field (in practice, proportional to the square root of the total number of authors in that field). Other candidates are added if their publication count is one less than the threshold provided that the total citations to their HCPs rank in the top half of the citation count profile of the first selected set.
- Some names occur frequently yet refer to different authors so that iteration is required between disambiguation and threshold count. Some papers in some fields have a extraordinary number of authors. These publications are omitted if their institution/affiliation count exceeds 30. Retracted publications and authors found to have committed scientific misconduct are excluded.
- A researcher can have more than one entry in HCR.2 if their publications appear in journals allocated to different fields. Of the 31% of members of the 2016 HCR.2 list who can be uniquely identified via their ORCIDs, around 7% are listed with two fields.

Table 1 compares the profile across fields of the archived HCR.1 list ("2001") and the 2017 HCR.2 list. The total number of HCRs is exhibited, and the proportion for each field. The ratio of proportions is displayed in the final column. Notable differences are the reduction in 2017 total membership to approximately one-half that of 2001, and significant shifts in the proportions. In HCR.2 over 10% of all entries relate to clinical medicine, a proportion approaching three times that in the older list. On the other hand, the relative size of the small fields of Economics & Business, Mathematics, Microbiology and Space Science are approximately one half that in HCR.1.

HCR.2 methodology tends to be more responsive both to the emergence and to the ongoing success of preeminent researchers. Table 2 explores the cross-year continuity of membership of the HCR.2 category of the members of the 2017 list. Of HCRs identified by the "new" methodology, just under one-half of the 2017 members were also members in 2014, whilst of all members in 2014, just over one-half were in the 2017 list. The corresponding statistics for 2015 and 2016 are also exhibited. Note that each year's membership is constructed de novo so (for example) there are HCRs in the 2014 and 2017 lists who are not in the 2015 and/or 2016 lists. On the other hand, the HRC.1 rules left a researcher in the HCR.1 list after they met the entry criteria.

Approximately 6.1% of the 2017 HCR list are also members of the archived list of HCRs defined by the "old" methodology. Around 3.3% of the members of the archived list also appear in the 2017 HCR list.

It is premature to discuss in this paper the foreshadowed addition of "cross-field" desiderata to the HCR rules, beyond noting that Clarivate indicates that the additional number will be approximately 2000 added to the approximately 4000 members of HCR.2 (Analytics 2018). This is the third major variant of HCR rules since 2001, highlighting one of several ways that the membership changes from 1 year to the next:

1. The bibliometric rules for HCR membership change,

Table 1 Counts and proportions of HCRs in the HCR.1 archive and the 2017 HCR.2 list

Field	2001		2017		2017/2001
Agricultural Sciences	291	4.1%	156	4.4%	1.07
Biology and Biochemistry	298	4.2%	212	6.0%	1.41
Chemistry	277	3.9%	215	6.1%	1.54
Clinical Medicine	280	4.0%	389	11.0%	2.76
Computer Science	373	5.3%	152	4.3%	0.81
Economics and Business	346	4.9%	93	2.6%	0.53
Engineering	276	3.9%	176	5.0%	1.27
Environment/Ecology	365	5.2%	160	4.5%	0.87
Geosciences	350	5.0%	141	4.0%	0.80
Immunology	364	5.2%	138	3.9%	0.75
Materials Science	308	4.4%	150	4.2%	0.97
Mathematics	356	5.1%	96	2.7%	0.54
Microbiology	401	5.7%	106	3.0%	0.53
Molecular Biology and Genetics	345	4.9%	204	5.8%	1.18
Neuroscience and Behavior	349	5.0%	187	5.3%	1.06
Pharmacology and Toxicology	328	4.7%	134	3.8%	0.81
Physics	328	4.7%	194	5.5%	1.18
Plant and Animal Science	346	4.9%	207	5.9%	1.19
Psychiatry/Psychology	319	4.5%	138	3.9%	0.86
Social Sciences, general	369	5.2%	188	5.3%	1.01
Space Science	363	5.2%	102	2.9%	0.56
Total	7032	100%	3538	100%	

Table 2 Continuity of HCR 2017 membership from previous years. “2001” refers to Clarivate’s archived list of HCR according to the HCR.1 rules

	2001	2014	2015	2016	2017
Count in year who are in HCR 2017	216	1731	2220	2620	3535
Proportion of HCR 2017 in year	6.1%	49.0%	62.8%	74.1%	100%
Proportion in year who are HCR 2017	3.3%	53.9%	71.1%	80.3%	100%

2. An HCR’s data changes (name, affiliation/s, retraction),
3. An HCR’s affiliation’s data changes (merging, de-merging), and/or
4. A researcher begins to satisfy or fails to continue to satisfy the bibliometric rules.

Year-on-year research performance tracked through the HCR databases will reflect the full interplay of all of these changes. We now consider how this relates to ARWU rankings.

Use of the HCR databases in ARWU

A first-hand account of the Shanghai ranking indicators can be found in Liu and Cheng (2005) and a then-contemporary criticism in van Raan (2005). Initially, some authors questioned the reproducibility of the ARWU data (e.g. Florian 2007) but those concerns are now diminished (Docampo 2013; Docampo and Cram 2014) and the methodology is well characterised. Fernández-Cano et al. (2018) offer a recent integrative review of criticisms of the ARWU.

The HiCi indicator score for institution i having $N_{\mathrm{HCR}}(i)$ HCR researchers in a given year is calculated by the equation (Docampo 2013)

$$\mathrm{HiCi}(i) = 100\sqrt{N_{\mathrm{HCR}}(i)/N_{\mathrm{Harvard}}} \tag{1}$$

Harvard appears in this equation because ARWU normalises its individual indicators to the highest performer, and Harvard is the university with the largest number of affiliations in the HCR databases. The ARWU value of N_{Harvard} in each year must be found from the HiCi data itself. For the sake of specificity, we use ARWU 2017 as a reference, inferring $N_{\mathrm{Harvard}} = 84$ so that Eq. (1) can be written as

$$\mathrm{HiCi}(i) = 10.91\sqrt{N_{\mathrm{HCR}}(i)} \tag{2}$$

Accordingly, an institution with one HCR in ARWU 2017 will have a HiCi score of 10.9, two will have 15.4, and three will have 18.9. Each of these values is readily found in the ARWU 2017 list. The HiCi indicator receives a weight of 0.2 in the total score (apart from institutions whose Nature and Science publications are not counted and slightly re-adjusted to ensure that the total ARWU score for Harvard is 100). Equation (2) implies that the contribution of a single HCR to the total ARWU score is approximately (in 2017)

$$\Delta(\text{ARWU total score}) = 2.2(\sqrt{N+1} - \sqrt{N}) \tag{3}$$

for an institution with N HCRs.

While Eq. (3) is a straightforward non-linear relationship between ARWU total score and the HCR count, institutions' ARWU ranks are also shaped by the values of other indicators. The dynamics of this link are illustrated in Fig. 1 which exhibits the rise in ranking that would occur if one hypothetical HCR were added to the actual count of each institution listed in ARWU 2017 while no other changes were made.

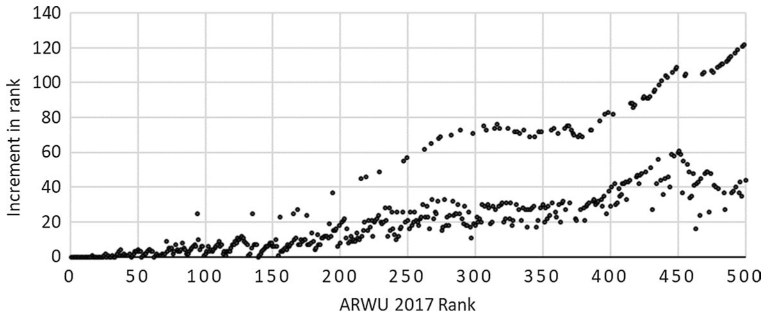

Fig. 1 Potential rise in rank due to the addition of one (fictitious) HCR to each institution ranked in ARWU 2017. The upper irregular curve represents the hypothetical rise in rank for those institutions that have a zero value for the 2017 HiCi indicator

To illustrate how the dynamics of Fig. 1 have played out over the time of the ARWU ranking, it is useful to introduce the concept of "volatility" $V_I(t)$ of AWRU measure $I(t)$ in year t where

$$V_I(t) = 1/n \sum_{\substack{\text{Ranked}\\\text{institutions}}} |I(t) - I(t-1)| \tag{4}$$

The average of the summed absolute change in the indicator is taken over the n institutions that are ranked in both years t and $t-1$. Figure 2 exhibits the volatility in HiCi and ARWU Rank over the years 2005/4–2018/17. The volatility of the HiCi indicator falls from 1.7 to 0.9 over 2005/4 and 2006/5 as the initial HCR list is stabilised with the addition of more fields and researchers. HiCi volatility continues to decline to less than 0.2 in 2013/12 as the HCR.1 list is "frozen", before rising as Thomson-ISI introduces the HCR.2. The HiCi volatility cycles as ARWU responds to the rule changes by averaging across HCR.1 and HCR.2 lists, before moving to the HCR.2 rules which yield more volatility owing to the fact that under HCR.2 a researcher may exit as well as enter the list (Table 2). Higher volatility in HiCi is linked to higher volatility in ranking. Note that our study refers to the rank based on individual total scores. In practice, ARWU publishes banded rankings to address the volatility of ranks derived from combined indicators that have small variability between institutions and over time in one institution.

While Fig. 2 emphasises the link between elevated HiCi volatility and elevated ranking volatility, volatility in ARWU ranking is also driven by other factors. This is illustrated in a scatter plot (Fig. 3) of the year-on-year change $I(t) - I(t-1)$ for indicators "rank" and "HiCi" for all institutions ranked in year t and $t-1$. A dozen or so points featured in the top right quadrant of this Figure represent institutions and years in which a HiCi increase has driven a higher ranking. However, the amorphous scatter of the plot indicates that ranking changes are also driven by changes other than HiCi.

The ARWU changes illustrated in Figs. 1, 2 and 3 are principally due to changes in HCR rules. As we noted above, ARWU changes can also arise from changes initiated by researchers (e.g. affiliation) and institutions (e.g. merging). For Harvard, by definition, the HiCi indicator provides 20% of its total score. We can say that a university with proportionately low HiCi (say, less than 10%) establishes its rank on indicators other than HiCi, while one with proportionately high HiCi (say, greater than 30%) relies on the HiCi indicator.

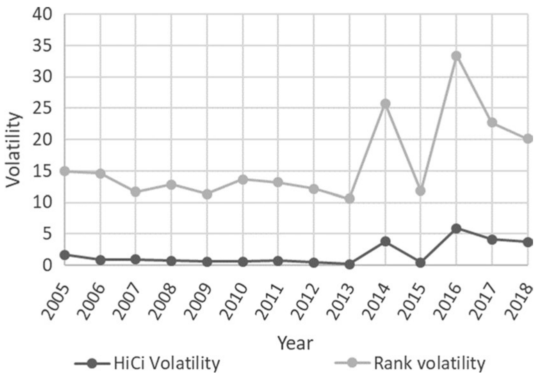

Fig. 2 Volatility (Eq. 4) of the ARWU HiCi indicator and annual ranking over the years 2005/4 to 2018/17. The higher values in 2014 and from 2016 onwards reflect the influence of the rule change from HCR.1 to HCR.2

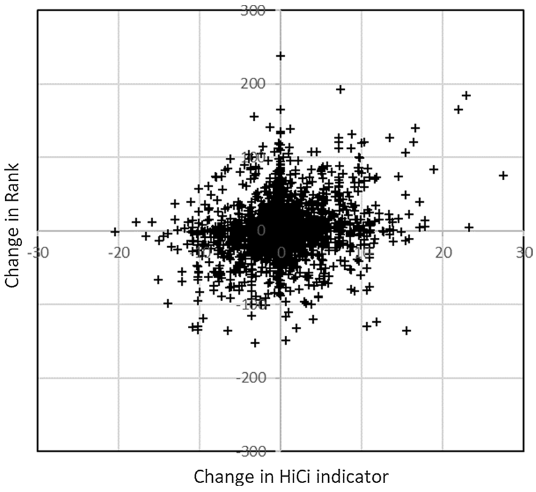

Fig. 3 Scatter plot of the annual change in HiCi indicator value versus the corresponding change in ranking for all ranked institutions in the years 2005–2018

For ARWU 2017, a total of 89 universities have a HiCi indicator of 0, and another two have less than 10% of their total score provided by HiCi. Notably, out of these 91 universities, 76 are in non-English-speaking countries. Marked differences arise between nations' citation rates owing to national research practice and article indexing decisions (Van Leeuwen et al. 2001). It is important to acknowledge the achievements of institutions that overcome English-language biases against citation to enter the ARWU. There are 35 universities for which HiCi represents more than 30% of their total score and for 3 of these it is more than 40% (two are in Saudi Arabia and one in ROC). For around half of these universities, their relatively high HiCi score relies on affiliation with only five or fewer HCR.

Bhattacharjee (2011) reported that two Saudi universities were acquiring the affiliations of overseas scientists with the apparent intention of increasing visibility and ranking. ARWU (2012) announced that if an HCR's affiliation changed in HCR-T, ARWU would reassign a weighted score depending on the declared percentage of affiliation. This open clarification by ARWU that its methodology recognizes the mobility of HCR counts between institutions and may have prompted some institutions to recruit the affiliations of HCRs. This may have been intensified by the use of counts of highly cited publications in the US News and World Report Best Global Universities (BGU) ranking (Morse and Krivian 2017).

HCR are talented and there are many drivers for a university to attract them beyond the pursuit of higher ranking. Reviews of recent affiliations of HCRs using the Web of Science reveal that some appear to have relocated, perhaps with a research group or to a leadership position. Some have long-standing, fertile inter-institutional collaborative arrangements. Some have recently commenced joint publication with researchers at the new affiliation, while others do not publish research at or with their current affiliation. Some aspects of the recruitment of HCRs are a legitimate element of academic capitalism, but others have led to the warning "such practices could detract from the genuine efforts that […] universities are making to transform themselves into world-class research centres" (Bhattacharjee 2011).

Conclusion

Citation counts arguably constitute the main way to *measure* the impact and relevance of academic research, standing alongside the *judgements* involved in peer review. Counts of citations to an individual's corpus or to sets of an individual's papers lie behind the creation of the category of highly cited researchers.

The ISI Highly Cited Researcher category HCR.1 emerged at the turn of the century following developments in citation counting spanning a period of approximately 40 years. The original focus was ostensibly to celebrate individual researchers' achievements. Around 2012–2014 the methodology for identifying HCRs shifted from simple counts of citations in a specified time interval towards counting both total citations and the number of highly cited publications normalised by disciplinary publication practice. The revised HCR.2 list is more responsive to preeminent researchers emerging earlier in their career (Clarivate Analytics 2018b). A further significant change introducing cross-field achievements has been foreshadowed (Clarivate Analytics 2018c; HCR.3). Changes in the rules for allocating a researcher to the HCR category have a significant influence on the membership of the list: the extent of consequential influences on institutional rankings depends on the way that rule changes re-distribute the population of HCR researchers among their affiliated institutions.

The HCR category took on significant additional meaning when the Shanghai Ranking decided to use the total number of HCRs at an institution as the raw data for one of its indicators of global university ranking. Commentary on ARWU rankings often relates to reported year-to-year changes, and it is sometimes overlooked that these may be due simply to changes in HCR rules.

Changes in affiliations in the HCR lists can also be made by researchers themselves. A small number of institutions have raised their standing in the Shanghai Ranking by encouraging such changes. Arguably this subverts the original intent of ARWU's use of the HiCi indicator and "such practices could detract from the genuine efforts that […] universities are making to transform themselves into world-class research centres" (Bhattacharjee 2011).

Despite these concerns, use of HCR lists as a basis for recruitment can be a legitimate element of sound academic policy. HCRs are by and large talented and there are many reasons for a university to attract them beyond the pursuit of higher ranking. Reviews of recent realigned affiliations of HCRs using the Web of Science reveal frequent relocation with a research group or to a leadership position. Some HCRs have initiated joint publication with researchers at the new affiliation while others have long-standing, fertile inter-institutional collaborative arrangements. Academic stakeholders respond in many ways to the glimpse of individual excellence revealed through inclusion in yearly HCR lists. The HCR lists provide fertile data for further studies of the broader implications for nations and for society at large.

Acknowledgements The work of D. Docampo was supported by the European Regional Development Fund (ERDF) and the Galician Regional Government under agreement for funding the Atlantic Research Center for Information and Communication Technologies (AtlantTIC), as well as by the Government of Spain through the Salvador de Madariaga Program.

Data Additional information relating to the work reported here is deposited for open access in Technical Reports and other data sets at https://www.researchgate.net/profile/Domingo_Docampo/publications.

References

Adair, W. C. (1955). Citation indexes for scientific literature? *American Documentation, 6*(1), 31–32.

Amara, N., Landry, R., & Halilem, N. (2015). What can university administrators do to increase the publication and citation scores of their faculty members? *Scientometrics, 103*(2), 489–530. https://doi.org/10.1007/s11192-015-1537-2.

Analytics, C. (2018). *New this year: Cross-field category*. *Clarivate Analytics*. https://feedback.hcr.clarivate.com/new-this-year/. Accessed November 23, 2018.

Anon. (2004). Citation Laureate Awarded to psychiatrist. *Australiasian Psychiatry, 12*(2), 204.

Anon. (2018). *Back to the future: Institute for Scientific Information re-established within Clarivate Analytics*. https://clarivate.com/blog/news/back-future-institute-scientific-information-re-established-within-clarivate-analytics/. Accessed March 28, 2018.

ARWU. (2012). *Methodology for the computation of the HiCi indicator*. http://www.shanghairanking.com/ARWU-Methodology-2012.html. Accessed October 19, 2015.

Basu, A. (2006). Using ISI's' Highly Cited Researchers' to obtain a country level indicator of citation excellence. *Scientometrics, 68*(3), 361–375.

Batty, M. (2003). The geography of scientific citation. *Environment and Planning A, 35*(5), 761–765.

Bhattacharjee, Y. (2011). Saudi universities offer cash in exchange for academic prestige. *Science, 34*(6061), 1344–1345.

Bornmann, L., Bauer, J., & Schlagberger, E. M. (2017). Characteristics of highly cited researchers 2015 in Germany. *Scientometrics, 111*(1), 543–545. https://doi.org/10.1007/s11192-017-2248-7.

Bornmann, L., & Leydesdorff, L. (2018). Count highly-cited papers instead of papers with h citations: Use normalized citation counts and compare "like with like"! *Scientometrics, 115*(2), 1119–1123. https://doi.org/10.1007/s11192-018-2682-1.

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*(1), 93–102.

Bornmann, L., Wagner, C., & Leydesdorff, L. (2015). BRICS countries and scientific excellence: A bibliometric analysis of most frequently cited papers. *Journal of the Association for Information Science and Technology, 66*(7), 1507–1513.

Burns, J., Brenner, A., Kiser, K., Krot, M., Llewellyn, C., & Snyder, R. (2009). JSTOR-data for research. In *International conference on theory and practice of digital libraries* (pp. 416–419). Springer.

Clarivate Analytics. (2018a). *Highly cited researchers: Archived lists*. https://clarivate.com/hcr/researchers-list/archived-lists/. Accessed June 12, 2018.

Clarivate Analytics. (2018b). *Highly cited researchers: Metholdology*. https://clarivate.com/hcr/methodology/. Accessed June 12, 2018.

Clarivate Analytics. (2018c). *New this year: Cross-field category*. https://feedback.hcr.clarivate.com/new-this-year/. Accessed 23 Nov 2018.

Cronin, B. (2005). A hundred million acts of whimsy? *Current Science - Bangalore, 89*(9), 1505.

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

Docampo, D. (2013). Reproducibility of the Shanghai academic ranking of world universities results. *Scientometrics, 94*(2), 567–587.

Docampo, D., & Cram, L. (2014). On the internal dynamics of the Shanghai ranking. *Scientometrics, 98*(2), 1347–1366. https://doi.org/10.1007/s11192-013-1143-0.

Fernández-Cano, A., Curiel-Marin, E., Torralbo-Rodríguez, M., & Vallejo-Ruiz, M. (2018). Questioning the Shanghai Ranking methodology as a tool for the evaluation of universities: An integrative review. *Scientometrics, 116*(3), 2069–2083. https://doi.org/10.1007/s11192-018-2814-7.

Florian, R. V. (2007). Irreproducibility of the results of the Shangai academic ranking of world universities. *Scientometrics, 72*(1), 25–32.

Frenken, K., Hardeman, S., & Hoekman, J. (2009). Spatial scientometrics: Towards a cumulative research program. *Journal of Informetrics, 3*(3), 222–232.

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science, 122,* 108–111.

Garfield, E. (1957). Breaking the subject index barrier-a citation index for chemical patents. *Journal of the Patent Office Society, 39,* 583.

Garfield, E. (1967). Current contents ninth anniversary. *Essays of an Information Scientist, 1*(12), 12–15.

Garfield, E. (1970). Citation indexing for studying science. *Nature, 227,* 669–671.

Garfield, E. (1978). The 100 most-cited SSCI authors, 1969–1977. *Essays of an Information Scientist, 3,* 633–639.

Garfield, E. (1998). *On the origins of current contents and ISI*. http://www.garfield.library.upenn.edu/papers/origins_cc_isi.html. Accessed January 13, 2016.

Garfield, E., & Welljams-Dorof, A. (1992). Of Nobel class: A citation perspective on high impact research authors. *Theoretical Medicine, 13,* 117–135.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature, 520*(7548), 429–431. https://doi.org/10.1038/520429a.

King, D. A. (2004). The scientific impact of nations. *Nature, 430*(6997), 311–316.

Lane, P. (1992). Thomson corporation acquires majority interest in ISI. *Information Today, 9*(5), 1–2.

Liu, N. C., & Cheng, Y. (2005). Academic ranking of world universities: Methodologies and problems. *Higher Education in Europe, 30*(2), 127–136.

MacRoberts, M. H., & MacRoberts, B. R. (1996). Problems of citation analysis. *Scientometrics, 36,* 435–444.

Más-Bleda, A., & Aguillo, I. F. (2013). Can a personal website be useful as an information source to assess individual scientists? The case of European highly cited researchers. *Scientometrics, 96*(1), 51–67. https://doi.org/10.1007/s11192-013-0952-5.

McVeigh, M. E. (2004). *How does ISI identify highly cited researchers*. Retrieved from https://web.archive.org/web/20040213102413/http://highlycited.com/isi_copy/comm_news08.htm. Accessed 1 Oct 2013.

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & Van Raan, A. F. J. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy, 14*(3), 131–149.

Moed, H. F., De Bruin, R. E., & Van Leeuwen, T. N. (1995). New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. *Scientometrics, 33*(3), 381–422.

Morse, R., & Krivian, A. (2017). *How U.S. News calculated the best global universities rankings*. https://www.usnews.com/education/best-global-universities/articles/methodology. Accessed June 17, 2018 .

Paarlberg, R. L. (2004). Knowledge as power: Science, military dominance, and U.S. security. *International Security, 29*(1), 122–151.

Parker, J., Allesina, S., & Lortie, C. (2013). Characterizing a scientific elite (B): Publication and citation patterns of the most highly cited scientists in environmental science and ecology. *Scientometrics, 94*(2), 469–480. https://doi.org/10.1007/s11192-012-0859-6.

Parker, J., Lortie, C., & Allesina, S. (2010). Characterizing a scientific elite: The social characteristics of the most highly cited scientists in environmental science and ecology. *Scientometrics, 85*(1), 129–143. https://doi.org/10.1007/s11192-010-0234-4.

Pendlebury, D. A. (2014). *Methodology: Thomson reuters citation laureates*. http://sciencewatch.com/sites/sw/files/images/basic/methodology_citation_laureate_2014.pdf. Accessed June 23, 2015.

Petersen, A. M., Fortunato, S., Pan, R. K., Kaski, K., Penner, O., Rungi, A., et al. (2014). Reputation and impact in academic careers. *Proceedings of the National Academy of Sciences, 111*(43), 15316–15321.

Seglan, P. O. (1997). Education and debate. *British Medical Journal, 314,* 498–502.

Thomson-ISI. (2000). *Japan's citation laureates, 1981–98*. Retreived from http://web.archive.org/web/20081014043920/http://archive.sciencewatch.com/nov-dec2000/sw_nov-dec2000_page1.htm. Accessed 2 March 2015.

Thomson-ISI. (2001). *ISI Launches ISIHighlyCited.com*. http://www.infotoday.com/IT/apr01/news7.htm. Accessed June 25, 2015 .

Thomson-ISI (2008). *How do we identify highly cited researchers?* http://archive.li/XJtCF. Accessed November 19, 2018.

Thomson Reuters. (2014). *Methodology for identifying highly-cited researchers*. http://www.highlycited.com/methodology/. Accessed October 19, 2015.

Useem, M. (1976). State production of social knowledge: Patterns in government financing of academic social research. *American Sociological Review, 41*(4), 613–629.

USPTO. (2001). highlycited.com. United States Patent and Trademark Office. http://tmsearch.uspto.gov/bin/showfield?f=doc&state=4810:do7p7v.2.1. Accessed November 19, 2018.

Van den Brink, M., Fruytier, B., & Thunnissen, M. (2013). Talent management in academia: Performance systems and HRM policies. *Human Resource Management Journal, 23*(2), 180–195.

Van Leeuwen, T. N., Moed, H. F., Tijssen, R. J. W., Visser, M. S., & Van Raan, A. F. J. (2001). Language biases in the coverage of the Science Citation Index and its consequencesfor international comparisons of national research performance. *Scientometrics, 51*(1), 335–346.

van Leeuwen, T. N., Visser, M. S., Moed, H. F., Nederhof, T. N., & van Raan, A. F. J. (2003). The Holy Grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics, 57*(2), 257–280.

van Raan, A. F. J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics, 62*(1), 133–143.

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics, 5*(1), 37–47.

Yap, S. (2016). *Onex, Baring PE Asia complete $3.55b acquisition of Thomson Reuters unit*. https://www.dealstreetasia.com/stories/54724-54724/. Accessed March 28, 2018.