The big challenge of Scientometrics 2.0 – exploring the broader impact of scientific research in public health

Wolfgang Glänzel[1,2], Pei-Shan Chi[3]

[1]*wolfgang.glanzel@kuleuven.be*
ECOOM and Dept. MSI, KU Leuven, Naamsestraat 61, 3000 Leuven, Belgium

[2]*glanzw@iif.hu*
Library of the Hungarian Academy of Sciences, Dept. Science Policy and Scientometrics, Budapest, Hungary

[3]*peishan.chi@kuleuven.be*
ECOOM, KU Leuven, Naamsestraat 61, 3000 Leuven, Belgium

Abstract

In the present study we discuss the challenge of "Scientometrics 2.0" as introduced by Priem & Hemminger (2010) in the light of possible applications to research evaluation. We use the Web of Science subject category *public, environmental & occupational health* to illustrate how indicators similar to those used in traditional scientometrics can be built, and we also discuss their opportunities and limitations. The discipline under study combines life sciences and social sciences in a unique manner and provides usable metrics reflecting both scholarly and wider impact. Nonetheless, metrics reflecting social media attention like tweets, retweets and Facebook likes, shares or comments are still subject to limitations in this research discipline as well. Furthermore, Usage metrics clearly point to the manipulation proneness of this measure. Although the counterparts of important bibliometric indicators proved to work for several altmetrics too, their interpretation and application to research assessment requires proper context analysis.

Background and introduction

Scientific communication was long dominated by scholarly communication in basic research in the sciences. Scholarly communication in fundamental research, in the so-called "hard sciences", took and partially still takes place in journal literature. A simple look into typical documents in these research areas reveals this kind of communication patterns: Most references point to other journal articles and scientific literature in periodicals thus covers most sources and targets of scholarly communication. The mission of the Scientometrics 1.0 version, which came up in the 1960s-1970s of the last century, was to model and measure documented scholarly communication in basic science and impact on scientific communities. The identification of both actors and users of scientific information was easy, as those could be found within the scientific communities. The even sometimes disputed use of citation measures for evaluative purposes was rather clearly defined as citations marked the information use in the process of knowledge production and dissemination with well-defined rules.

When scientometrics opened towards new data sources (including conference proceedings and books) and broadened towards the measurement of research performance in other fields than basic research, it became apparent that the above-mentioned framework proved too narrow for those fields. Researchers pointed to the fact that, e.g., in medical and applied sciences, a large share of information targets is found outside the research community and citations are therefore yet partial measures of impact and use of information. In the social sciences and humanities (SSH), a large share of both sources and targets are located outside the research community, thus citations based on periodicals can only be considered an insufficient measure of impact and use of information. Scientometricians attempted to catch up with this challenge and to keep pace with the new developments in research evaluation by broadening the scope and improving their methods.

Yet, the new millennium came up with new challenges to be met by scientometricians. These challenges partially result from the new demands through policy and society needs, new movements, like 'open science', are also caused by the new electronic communication forms becoming prevalent in scientific communication as well. These intra-scientific, societal, policy-driven and technical demands lead to the evolution of a new concept called "Scientometrics 2.0" (Priem & Hemminger, 2010). Priem and Hemminger considered *open science*, *social media metrics* and *alternative metrics* groundwork and components for this new concept. They also compiled a list of possible sources for its implementation. In particular, network-based approaches analysing relationship and interactions among different actors on social media, e.g. communities of attention, hashtag coupling analysis, and reader pattern analysis (Wouters et al., 2019), may contribute to the social capital-based system of scientific impact assessment (Hoffman et al., 2014).) Yet, as so often reality is quickly running past visions and nowadays a plethora of measures and metrics are in use, sometimes in a rather uncritical manner and even repeating or imitating typical errors of the early Scientometrics 1.0 (Gumpenberger et al., 2016). This study adds some novel aspects to the existing literature on altmetrics, by going beyond the traditional regression analyses and by attempting to systematically integrate metric profiles and mathematical models (the method of Characteristic Scores and Scales and the negative-binomial model previously used to model citation processes) in order to check their applicability in the context of the new metrics as well.

Previous results and research questions

Like in our previous studies special focus is laid on the comparison with traditional, mainly publication and citation-based indicators. In addition, we will show that several advanced methods and indicators developed for traditional 'productivity' and citation analysis are still fit for the new environment. For the present study we use publication, citation data and usage statistics from Clarivate Analytics' Web of Science Core Collection (WoS) in conjunction with altmetrics data from Plum Analytics.

In our previous studies, we analysed selected fields from the sciences and social sciences to uncover specific patterns of impact, information ageing. Thus, the results could readily be compared with those of traditional scientometric studies. We could also show that (full text) download processes generally mirror the characteristics of citation processes but not always to the same extent and mostly with a certain field specific "translation coefficient" (cf. Glänzel & Heeffer, 2014). This implies that one citation roughly corresponds to a certain number of downloads, which amounted to about 100 in our Elsevier sample of 80,000 journal documents put online in 2008 and followed up for downloads and citations with a five-year window. The citation process mirrors the increments of downloads, however with a certain 'phase shift' in accordance to our expectations. The correlation between the impact and the usage measure proved very strong, which partially confirmed results of earlier studies by others (e.g., Moed, 2005; Brody et al., 2006; Thelwall, 2012). Further studies by Chi and Glänzel and most recently, by Chi et al. (2019) could confirm and deepen these results. We also showed that traditional concepts and methods can be integrated into the new metrics. We defined a *Journals Usage Index* (Chi & Glänzel, 2018) in analogy to the Garfield Impact Factor as well as the idea of relative citation indicators, and the Characteristic Scales and Scores proved to work for new metrics as well.

In our previous studies we could already make some specific observations. The most important one concerns the difference between the patterns in basic research of science and in SSH. In terms of WoS usage statistics of journal articles, social sciences displayed disproportionately higher "usage" than citation impact (Chi & Glänzel, 2018, 2019). This did not strike us unexpectedly because citations to periodicals play a less pronounced part than in the sciences. All the more, we found it interesting that the usage of authored and edited books did not reflect the same patterns (Chi & Glänzel, 2019). Figure 1 gives the correlation between the mean usage rate (MUR) and the mean citation rate (MCR) of two document types of book publications, authored and edited books as reflected by the 2013 volume of Clarivate Analytics *Book Citation Index* (BKCI).

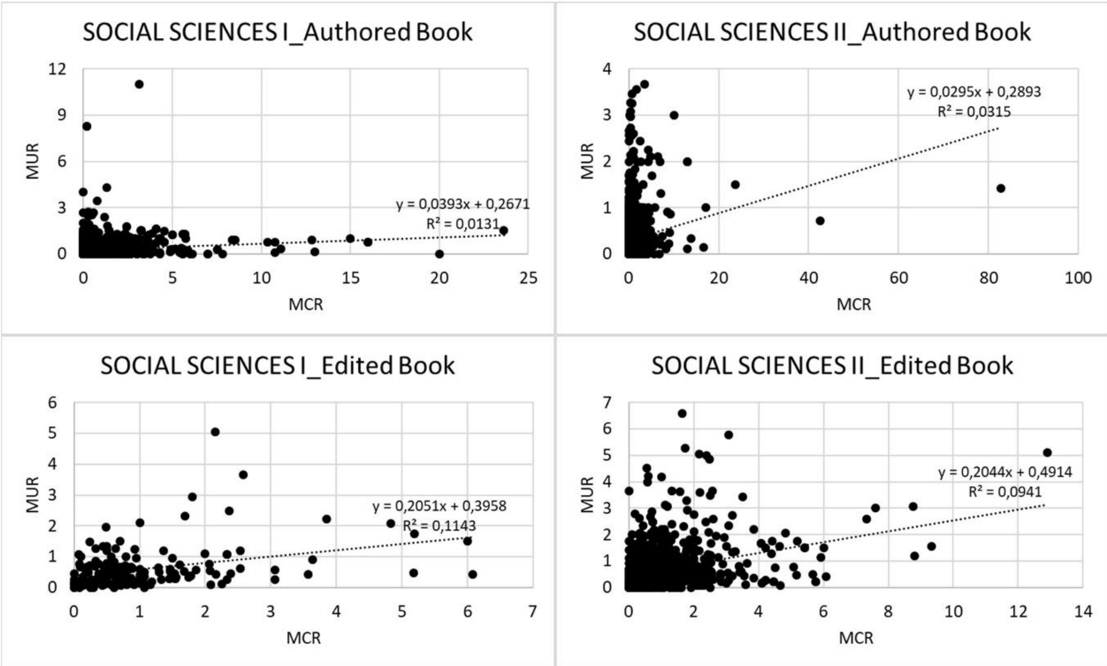

Figure 1. Scatter plots of MUR vs. MCR for two document types in two major fields according to Chi & Glänzel (2019)

In the present study, we will therefore further elaborate our methodology for the application and systematic analysis of an interdisciplinary field connecting both the life sciences and the social sciences. In particular, we have chosen the WoS subject category "*public, environmental & occupational health*". Furthermore, this discipline has already attracted our interest in terms of its growth and its emerging topics (Glänzel & Thijs, 2011, 2012).

In the light of previous results, we will attempt to answer mainly the following research questions.

1. In our previous studies we have found different extents of correlation between scholarly impact and Usage/Captures metrics with regard to disciplines and publication types. Will we find similar patterns as has been found in the (life) sciences?
2. Will the altmetric indicators distinctly exceed the scholarly impact with a factor for ‘‘translating’’ impact to Usage and Captures amounts?
3. Can we observe specific national patterns and can we find journals with significant deviation of their altmetrics from their traditional bibliometric characteristics?

In addition to the counts, shares and mean values, we will also apply the method of *Characteristic Scores and Scales* (CSS, see Glänzel & Schubert, 1988; Chi & Glänzel, 2018) to analyse distributional aspects of the metrics and to identify extreme values and outliers.

Data sources and data processing

For this study we have selected the WoS subject category *public, environmental & occupational health* ("public health" in short) in the Social Sciences Citation Index (SSCI) to emphasise the social sciences interaction in the selected field. All documents published in 2015 and indexed as articles and reviews in the SSCI have been downloaded from Clarivate Analytics' WoS database. All papers have been assigned to countries on the basis of the authors' corporate address and to the journals in which the papers have been published. Citations and usage have been counted till October 2018, that is, based on three years on an average. For matching the WoS data with altmetrics, the document DOIs and the PubMed IDs have been used. In total, 18,729 out of 18,824 articles and reviews, that is, 99.5% of all retrieved documents proved to have a valid DOI or PMID.

In a second step Usage, Captures, Social Media and Citation metrics have been downloaded from Plum Analytics (https://plumanalytics.com/) in October 2018 for each individual document using its DOI or PubMed ID. Possible errors in the identifiers have been corrected manually. Data downloaded from WoS and PlumX have been carefully cleaned and processed to bibliometric indicators.

Methodological considerations

From the conceptual-methodological viewpoint, five categories, namely, *Usage, Captures, Mentions, Social Media* and *Citations* are distinguished according to PlumX (PlumX, 2019). These five categories still form just a minor part of what can be covered by *broader* or *social impact* (cf. Lewison, 2004, 2008). "Usage" stands for the lowest level. This comprises, for instance, clicks, downloads or views and could rather be considered measures of the intention to use something than their actual usage (Gorraiz et al., 2014). "Captures" expresses somewhat more as it indicates repeated usage, for instance, as bookmarks, favourites, readers or watchers. The category "Citation" representing the highest level can be considered an extension of the concept of citations within the framework of traditional bibliometrics as this category reaches out beyond the framework of scholarly communication. In a previous study, we have not used the categories "Mentions" and "Social media", which partially require full-text and because of the document unavailability to a broader community and the "zero inflated" distributions resulting from it; therefore, we decided to omit these two indicators (cf. Chi, et al., 2019). However, the distinct perspective of "Social media" representing the social impact, which would be much more favourable in public health than in the subjects in the sciences, deserves a further investigation. We include this metrics in this study to broaden the discovery based on alternative metrics despite its very low percentage of data availability.

On the basis of the metrics selection we obtained a set of twelve metrics, Usage, Captures, Social Media supplemented by nine specific metrics: Scopus and CrossRef citations, EBSCO Abstract-views, EBSCO full text views, EBSCO link-outs, Tweets and Facebook from PlumX and Usage Count and Times Cited from the Web of Science (see Table 1). We applied our standard statistics, zero frequencies, mean values and Characteristic Scores and Scales to this set of metrics and we have broken down the data to the country level and to individual journals. In addition, we have conducted analysis to detect possible correlation between these metrics and to find the "translation" coefficient, provided the correlation is strong enough.

Table 1. Overview of the twelve metrics measured in this study

Overall categories	Specific metrics	Sources
Usage	EBSCO Abstract-views EBSCO full text views EBSCO link-outs	PlumX
	Usage Counts	WoS
Captures		PlumX
Social Media	Tweets Facebook	
Citations	Scopus citations CrossRef citations	
	Times Cited	WoS

Results

Correlation analysis

As with our previous studies, first we have applied a correlation analysis to study the relationship between the selected altmetric indicators and the traditional measures of scholarly impact. Instead of the Pearson correlation we have again applied Spearman correlation because of the skewed distributions underlying all metrics. The results of the analysis are presented in Table 2. As in the results by Chi et al. (2019), the three citation indicators correlate with each other strongly, especially between Scopus and WoS. Captures and WoS Usage have moderate correlations with citation indicators. PlumX Social Media has moderate to weak correlation with all the other metrics.

PlumX usage shows the most distinct patterns from other indicators. It only correlates strongly with PlumX captures indicator and has weak correlations with all the citation and Social Media indicators. The three EBSCO usage indicators correlate with PlumX captures and WoS Usage at moderate to strong level but have weaker correlation with citation and social media indicators. Among the three usage indicators, EBSCO full text views has the most distinct patterns from others. It is extremely weakly correlated with other indicators except for EBSCO abstract views. Its weak correlation with other metrics is even weaker than that of Facebook metrics with others which has, otherwise, also the highest share of zero-frequencies.

Another interesting finding is, that the social media metrics is stronger correlated with citations than Usage and Capture. This result is slightly different from the previous studies of the positive but weak correlation between citation and altmetric indicators (e.g., Costas et al., 2015; Zahedi et al., 2014) due the much higher degree of the correlation between the two metrics in the present study. The different subject coverage of each study and the higher percentage of altmetric data in this study may result in the different degree of positive correlation, although the finding that publications with more altmetrics tend to have more citations is confirmed in all of these studies. The message conveyed by Table 2 substantiates the different dimensions of usage, captures, social media and citation measurements and their relative degrees to the scholarly contribution.

Table 2. Spearman correlation between twelve metrics for the 18,729 documents in public health

	Usage	*Abstract Views*	*Full tekst Views*	*Link-outs*	*WoS Usage*	***Cap-tures***	*Scopus Cites*	*Cross Ref*	*WoS Cites*	***Social Media***	*Twitter*	*Face-book*
Usage	-	0.83	0.56	0.60	0.31	0.62	0.19	0.23	0.18	0.19	0.18	0.10
Abstract Views		-	0.56	0.75	0.38	0.74	0.28	0.33	0.26	0.24	0.26	0.10
Full tekst Views			-	0.21	0.13	0.40	0.08	0.11	0.06	0.11	0.11	0.05
Link-outs				-	0.44	0.68	0.31	0.36	0.29	0.24	0.25	0.13
WoS Usage					-	0.49	0.43	0.42	0.42	0.24	0.24	0.14
Captures						-	0.46	0.47	0.44	0.33	0.33	0.17
Scopus Cites							-	0.88	0.94	0.32	0.32	0.18
CrossRef								-	0.86	0.32	0.31	0.19
WoS Cites									-	0.31	0.31	0.17
Social Me-dia										-	0.89	0.61
Twitter											-	0.30
Facebook												-

CSS analysis

Table 3 presents the CSS-related indicators on the complete data set. We firstly explain the design of CSS method in short. Putting $b_0 := 0$ as the very first characteristic score for the random variable X represented by the sample, we then obtain the subsequent scores as $b_k := E(X|X \geq b_{k-1})$ for all non-negative integer values $k = 1, 2, \ldots$. These intervals $[b_{k-1}, b_k)$ between two adjoining scores define the performance classes from "poor" (Class 1) to "outstanding" (Class 4). The share of zero-frequency (f_0), the values of the three scores (b_1, b_2, b_3) and the scales of the four classes in public health are presented in Table 3.

Table 3. Bibliometric indicators for twelve metrics of the 18,729 documents in public health

	Usage	*Abstract Views*	*Full text Views*	*Link-outs*	*WoS Usage*	*Captures*	*Scopus Cites*	*CrossRef*	*WoS Cites*	*Social Media*	*Twitter*	*Facebook*
f_0	6.8%	7.3%	49.4%	10.1%	2.7%	1.4%	10.1%	19.1%	10.8%	45.3%	51.0%	78.7%
b_1	1001.94	601.66	108.51	64.87	11.98	80.74	7.07	4.67	6.31	10.00	3.80	6.20
b_2	3662.68	2322.30	533.34	200.60	23.98	172.27	16.48	10.63	14.56	63.92	16.30	82.23
b_3	7985.20	6112.24	1499.92	425.23	42.10	302.20	30.00	19.29	26.42	293.59	51.88	465.69
CSS1	78.3%	79.5%	81.7%	74.4%	65.0%	66.8%	69.9%	66.3%	69.0%	86.3%	79.5%	92.8%
CSS2	15.4%	15.4%	14.0%	18.4%	24.5%	23.2%	21.3%	23.5%	21.7%	11.6%	16.4%	6.2%
CSS3&4	6.3%	5.0%	4.3%	7.2%	10.6%	10.0%	8.9%	10.3%	9.3%	2.1%	4.1%	0.9%

The share of uncitedness is in line with the expectations. In Table 3, about one tenth of the documents remains uncited in WoS and Scopus in the period of roughly three years. Only the share of uncited documents according to CrossRef is distinctly larger. The zero-usage share on WoS platform is much lower than citations. The share of zero-captures is almost minute and even much lower than the shares of PlumX usage and social media and the other aggregated categories containing several individual metrics. The three EBSCO usage metrics keep the same zero-usage share as citations, except for the full text views, which may be the consequence of partial unavailability of the full texts of the underlying documents. Social media metrics, unsurprisingly, have the highest zero-frequencies.

The mean values reveal even more interesting patterns. Note that the mean value coincides with the first CSS score b_1. While all citation means are of the same order, Usage and Social Media reflect completely different patterns. The considerable difference between the mean value of PlumX Usage and those of other EBSCO usage metrics reflects another main component of usage metric apart from EBSCO: SciELO. Even though we do not report the SciELO metrics due to their very low percentages of data availability, which is probably caused by the regional coverage, the extremely high usage of some articles in SciELO lifted up the average of general usage metrics on PlumX. The same phenomenon could be observed in the case of Social Media metrics as well, however to a much lesser extent.

Figure 2 presents the proportion distributions of three CSS classes for nice indicators. Captures, citations and WoS usage have better preformance in terms of proportion of the fairly to highly cited/used/captured articles. By contrast, social media related metrics are highly skewed with the smallest proportion of highly followed articles. Their highest zero-frequencies may result in this skewed distribution; however, the high zero-usage of full text views seems not show the same effect to the scale distributions among classes.

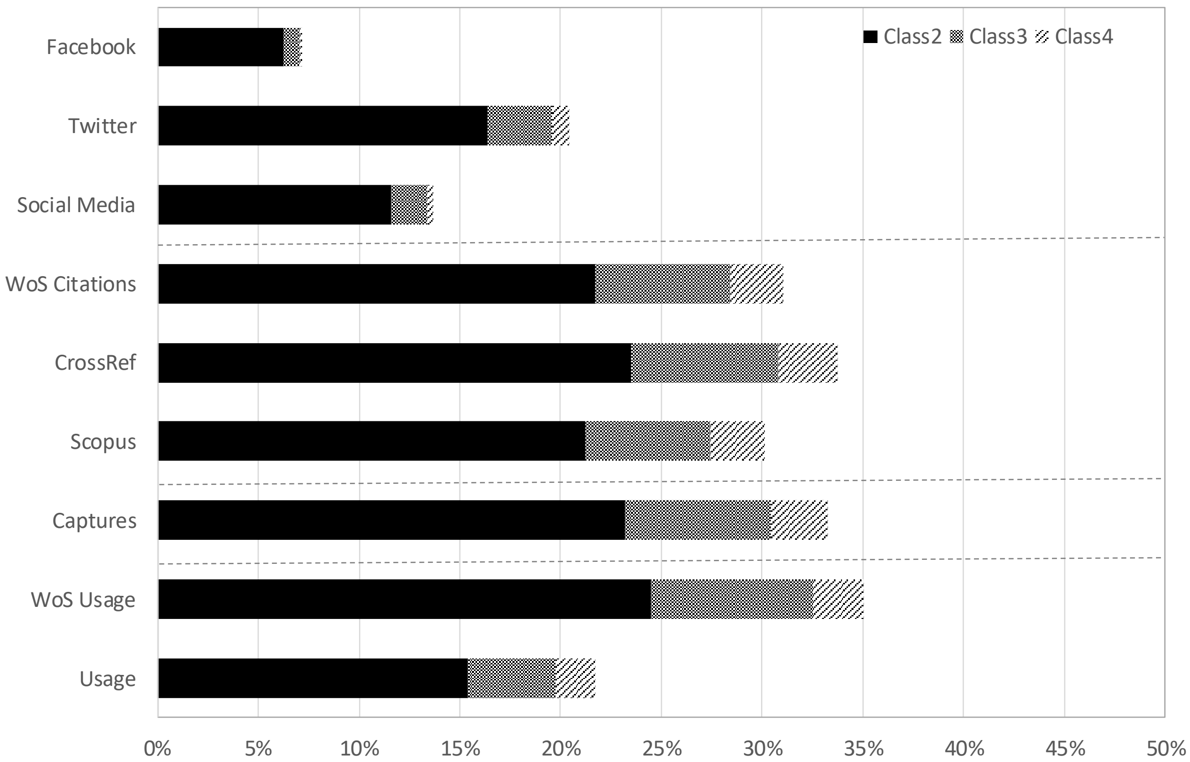

Figure 2. Distributions of CSS classes in public health

Distributions based on the metrics values

The citation distributions in Table 3 have their specifically skew shape in terms of the obsolete numbers of citation. By contrast, the value distribution of captures is very flat with typical probabilities around 1% each for $0 \leq k \leq 50$ captures. The other half is distributed less evenly between 51 and the maximum of 4416 captures. The two usage metrics show, however, large discrepancies. The Usage distribution is extremely flat with maximum usage frequency at $k = 0$ with more than 200 extreme values, each being larger than 10,000. By contrast, the WoS usage has a more "regular" flat-tailed distribution with the mode of relative frequency of 7% at $k = 6$. We have found only one single outlier ($k = 8791$) here.

The properties of these distributions are also reflected by the CSS scores ($b_1, \ldots, b_3$) in Table 3. The three EBSCO usage metrics follow by and large the patterns of Usage. The PlumX Usage and Captures scores are, in fact, of a one or two orders of magnitude higher than the citation scores and the WoS usage counts. The three citation distributions, indeed, substantiate strong relatedness. The negative binomial distribution-model described shape and characteristics of the empirical distributions quite well although the fit to the individual frequencies is, because of the long-stretched distribution, not perfect. This relatedness is also reflected by the parameters N and P of the negative binomial distribution (see Table 4) with

$$p_k = \binom{N+k-1}{k}\left(\frac{1}{P+1}\right)^N\left(\frac{P}{P+1}\right)^k, k = 0, 1, 2, \ldots .$$

N values greater than 1 indicate that the modus of the distribution might be at $k > 0$, which is the case for the WoS and Scopus citations as well as for the WoS Usage counts. The parameters of these distributions along with those of the CrossRef citations are in the same range. The large P value of the Captures distributions reflects its flatness and the extremely high P value of the PlumX Usage metrics substantiates that the negative binomial model here actually fails. Although this effect seems to be caused by the regional effect expressed by the SciELO data, the polarised distribution pattern also holds for the EBSCO Views, albeit to a much lesser extent.

Only Link-outs have an almost acceptable distribution, which is quite "close" to the case of Capture metrics. The (*N, P*) parameter-value pair of the Social Media metrics reflect interesting details. While the Twitter metrics still follow a regular, however very skewed distribution, the distribution of Facebook metrics can already be considered degenerate.
It is interesting to observe that the CSS classes nevertheless are in the same range for all indicators; only the PlumX Usage metrics shows a distinct deviation from the other ones. However, the range of all classes indicates that the CSS method is applicable in all cases. We will use this method later, in the context of journal and country statistics.

Table 4. Parameters of the negative binomial distributions fitted to the twelve metrics in public health

Metric	***N***	***P***
Usage	0.34	2978.55
EBSCO Abstract Views	0.35	1712.07
EBSCO Full text Views	0.10	1074.72
EBSCO Link-outs	0.46	139.72
Capture	0.96	84.03
Scopus	1.18	5.09
CrossRef	0.92	6.00
WoS Citations	1.22	5.18
WoS Usage	1.74	6.88
Social Media	0.20	49.46
Twitter	0.24	15.98
Facebook	0.05	125.40

Cases with extreme values in certain indicators

Among the 18,792 documents, some articles with *extreme values* in one indicator are not necessarily to have high values in other indicators. Table 5 lists the most extreme values of each indicators, and shows that those articles with high usage values come more often with high captures values while extremely highly cited articles keep their dominant positions among all the three citation sources. The most used articles on the WoS platform is a special case that only initiates high usage within the database but does not show high influence anywhere else (see Document #9). Pars pro toto, we will have a look at #1 with outstanding PlumX usage counts and otherwise low EBSCO usage and citation rates. #2 has very high PlumX Usage, Capture and Social Media metrics. Documents #6 – #8 have attracted above-average PlumX usage (CSS Class 3) and can be considered outstandingly cited (CSS Class 4).
The language of document #1 is Portuguese although there is an English version as well. This paper by Brazilian authors is entitled "The field of Collective Health in Brazil: definitions and debates on its constitution" and its topic and its strong regional/local focus might explain the enormous attraction in terms of usage, on the one hand, and the discrepancy between PlumX usage and EBSCO usage metrics and citation impact on the other hand. The high PlumX usage are contributed by the usage on SciELO platform. Document #2 entitled "Mental Illness, Mass Shootings, and the Politics of American Firearms" was cooperated by USA authors and drew a lot of attentions on social media and content provider platforms. This may be because of its topic coordinating public concerns and shows its societal influence resulting a vigorous discussion in society than in academia.
#6 is a review article on "The Prescription Opioid and Heroin Crisis: A Public Health Approach to an Epidemic of Addiction", which already presages the general interest and citation attractivity. The paper is published by five US institutions and, being a review it is also expected

to exhibit higher citation rates than research articles. The research paper entitled "Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research" (document #7) is the result of a collaboration of six US institutions and funded by the National Institute of Mental Health. Although it has a different document type, it shows very similar patterns in terms of Usage, Captures and citations as the previous one.

Finally, document #3 is another co-publication of a US institution with highest PlumX usage, which is mainly contributed by EBSCO usage metrics, we have found in our data set. Captures and Social Media are high but not extreme. By contrast, the citation rates it has received are rather moderate (cf. Table 3). The same applies to the WoS usage count. This document is a research article on "Social Science Collaboration with Environmental Health". The importance and strong social attraction of this topic speaks for itself. These five examples may just illustrate the effect of the thematic peculiarities of research in this specialty. In the following subsections we exclude WoS Usage count as we have already reviewed this metrics and its relationship with citation measures in previous studies (Chi & Glänzel, 2018, 2019; Chi et al., 2019).

Table 5. Extreme values of the twelve metrics of the 18,792 documents in public health

#	*DOI*	*Usage*	*A*	*F*	*L*	*WoS_U*	*Captures*	*Scopus*	*Cross Ref*	*WoS _C*	*Social Media*	*Twitter*	*Facebook*
1	10.1590/S0104-12902015S01018	46897	361	172	12	6	72	1	2	6	0	0	0
2	10.2105/AJPH.2014.302242	63131	43780	18295	263	75	4416	41	21	38	23586	4746	18840
3	10.1289/ehp.1409283	86660	43208	43328	32	12	269	14	12	11	83	5	78
4	10.2105/AJPH.2014.302393	44066	30639	13175	252	53	3597	27	12	18	2	2	0
5	10.1037/a0039045	14518	6762	173	7568	108	622	14	9	13	0	0	0
6	10.1146/annurev-publhealth-031914-122957	7694	4818	4	2528	101	1093	240	114	218	149	137	12
7	10.1007/s10488-013-0528-y	4467	2249	26	1644	81	1295	335	287	286	6	3	3
8	10.1093/ntr/ntu191	813	538	178	87	41	261	256	174	252	2	2	0
9	10.1111/jrh.12095	1217	1180	3	34	8791	55	7	4	9	0	0	0
10	10.1111/jrh.12078	904	851	2	51	332	143	2	1	2	0	0	0
11	10.1016/s2214-109x(15)70002-1	742	236	0	104	77	542	157	90	141	13106	881	12225

Legend: A: EBSCO Abstract views; F: EBSCO full-text views; L: EBSCO link-outs; WOS_U: WoS usage; WOS_C: WoS citations

Journal analysis

This subsection deals with a concise journal analysis. We have selected journals with at least 250 papers of document type *article* or *review* in 2015. Eleven journals met this criterion. In order to facilitate tabulating, we use the following official acronyms in the following: AAP (Accident Analysis and Prevention), AB (Aids and Behavior), AJPH (American Journal of Public Health), APJPH (Asia-Pacific Journal of Public Health), BMCPH (BMC Public Health), CSC (Ciencia & Saude Coletiva), HE (Health Expecta-tions), IJERPH (International Journal of Environmental Research and Public Health), MCHJ (Maternal and Child Health Journal), QLR (Quality of Life Research) and SSM stands for Social Science & Medicine. The list comprises journals ranging from the highest through the lowest impact quartile.

Table 6 shows both AJPH and CSC have very high average usage values; especially, every CSC paper has more than 1500 usage counts each. Unlike the tight relation between EBSCO abstract views and PlumX usage measure in AJPH, CSC has disproportionate EBSCO abstract views, full-text views and link-outs compared to PlumX usage counts. This may imply the high impact of SciELO usage counts in this Brazilian journal. The citation counts of CSC are generally the lowest among the eleven journals while AJPH has much higher citation counts than other

journals. The high SciELO usage counts seem not substantially affect international scholarly visibility at all.
The similarity of citation patterns, including CrossRef, is striking. *Social Science & Medicine* and AJPH reflect the most advantageous situation in terms of citations (about 20% highly and about 50% or less poorly cited papers, see Table 7). Nevertheless, the three EBSCO usage metrics completely contradict these patterns. While eighty percent of AJPH papers are "highly used" according to EBSCO abstract views, about eighty percent of *Accident Analysis and Prevention* are "poorly used" and only less than 2% of its papers can be considered "highly used".
Another striking observation is the converse distribution between PlumX Usage and the three EBSCO usage metrics over the lowest (Class 1) and highly used (Class 3&4) papers published in CSC. The CSC journal has zero percent of papers with PlumX Usage lower than the average, which means locating in Class 1, but owns almost zero percent of papers with EBSCO usage metrics in highly used group Class 3 &4. This indicates the high SciELO usage counts of papers in this Brazilian journal as discussed above based on the average values of various usage measures.

Table 6. Average values of the twelve metrics in public health

Journal	*Usage*	*Abstract Views*	*Fulltext Views*	*Links-outs*	*Capture*	*Scopus*	*CrossRef*	*WoS Cites*	*Social Media*
AAP	581.3	489.7	0.0	91.0	76.0	10.4	6.5	8.9	6.4
AB	203.9	158.9	13.1	29.9	77.4	9.6	9.3	8.9	8.7
AJPH	8092.4	7024.2	1033.2	28.5	255.2	11.5	7.3	10.4	58.2
APJPH	51.3	36.3	0.0	14.8	26.2	4.1	2.7	2.5	0.8
BMCPH	1215.3	728.7	450.1	29.5	141.9	7.7	4.7	7.0	7.7
CSC	5029.4	43.4	48.2	12.2	29.1	2.5	1.4	2.3	3.4
HE	521.5	330.0	110.9	78.7	90.5	8.1	5.1	6.2	6.3
IJERPH	91.5	46.2	22.5	19.1	42.6	7.5	7.0	6.7	12.6
MCHJ	953.5	706.4	168.8	76.3	131.0	6.5	5.8	5.6	14.1
QLR	701.5	586.0	87.5	26.8	62.2	7.8	7.1	7.0	5.1
SSM	817.9	624.2	0.1	184.6	119.8	11.2	7.2	10.0	21.2

The CSS classes for Capture reflect situations somewhere "in between" those of citation and usage. Social Media keeps its pattern of high shares of Class 1. Capture metrics of *Social Science & Medicine* are rather following its citation distribution, while those of IJERPH are closer to its EBSCO usage patterns. Since the number of papers in the selected journals is large enough (250 or more), the lack of clear correlation patterns can be considered significant without any specific test. Furthermore, these indefinite patterns intersect the individual journals as well. For instance, 32 out of the 512 AJPH papers (i.e., 6.25%) can be found in Capture CSS-Class 4 and in WoS Citation class 1. The same applies to 25 (i.e., 3.0%) papers published in *BMC Public Health*. Nevertheless, 19 (i.e., 3.7%) of the AJPH and 14 (1.7%) of the BMCPH papers are in the highest class according to both, Captures and WoS citations.
To conclude, Usage and Capture metrics supplement traditional journals impact measures but their interpretation is rather difficult while Social Media metrics keeps its own highly skewed pattern. Those cases, journals and individual publications, where metrics reflect contradicting situations, further, preferably context related analysis would be necessary for correct interpretation.

Table 7. Percentage of lowest and highest CSS-class documents of the eleven largest journals in public health (2015)

Journal	*N*	U_l in %	U_h in %	A_l in %	A_h in %	F_l in %	F_h in %	L_l in %	L_h in %	P_l in %	P_h in %	S_l in %	S_h in %	R_l in %	R_h in %	C_l in %	C_h in %	M_l in %	M_h in %
AAP	305	91.8	1.6	79.7	1.6	100.0	0.0	70.5	5.9	63.0	3.3	45.2	17.4	44.6	16.1	45.6	17.0	83.3	1.0
AB	251	100.0	0.0	100.0	0.0	99.6	0.0	91.6	0.0	64.1	4.4	52.6	15.1	33.9	26.3	52.6	17.1	77.3	3.2
AJPH	512	10.4	74.4	7.8	82.0	8.2	55.3	90.4	1.2	21.3	46.7	50.8	20.9	48.4	19.9	50.0	21.5	72.5	3.5
APJPH	355	100.0	0.0	100.0	0.0	100.0	0.0	96.3	0.0	97.2	0.0	84.5	1.1	81.4	2.0	90.7	1.1	98.9	0.0
BMCPH	830	63.1	4.8	62.2	4.5	18.6	22.8	88.9	1.6	38.4	25.1	65.5	9.3	65.3	8.8	62.8	10.0	81.1	1.9
CSC	357	0.0	54.1	99.7	0.0	89.4	0.3	96.6	0.8	96.1	0.3	94.4	0.8	95.2	0.6	93.6	0.8	95.2	1.7
HE	269	89.2	0.0	86.2	0.0	66.5	1.5	62.8	7.8	55.8	10.0	64.3	10.4	61.7	10.8	70.6	8.9	84.0	0.7
IJERPH	343	99.4	0.3	99.4	0.3	99.1	0.3	96.2	1.2	90.4	1.7	66.2	8.7	46.9	17.5	63.6	8.5	81.9	4.7
MCHJ	295	70.2	1.4	54.9	2.0	52.2	5.8	63.7	7.5	35.6	21.7	68.8	5.4	54.9	14.6	71.5	5.8	83.4	2.7
QLR	290	76.2	0.0	57.6	0.7	71.7	1.0	89.0	0.3	73.1	3.4	65.5	9.7	43.1	19.3	63.4	11.0	90.7	0.7
SSM	626	78.4	1.6	67.4	2.1	100.0	0.0	33.5	24.1	36.3	20.0	47.3	20.1	44.1	19.8	45.4	20.9	62.1	8.3

Legend: N: number of papers; U: PlumX Usage; A: EBSCO Abstract views; F: EBSCO Full text views; L: EBSCO Link-outs; P: PlumX Captures; S: Scopus citations; R: CrossRef citations; C: WoS citations; M: Plumx Social Media. Index *l* denotes the lowest CSS class 1, index *h* the highest two classes 3&4 combined.

Table 8. Metrics means and percentage of lowest and highest CSS-class documents of the twenty countries with largest publication output in public health (2015)

Country	*N*	U_{b_1}	U_l in %	U_h in %	A_{b_1}	A_l in %	A_h in %	P_{b_1}	P_l in %	P_h in %	S_{b_1}	S_l in %	S_h in %	C_{b_1}	C_l in %	C_h in %	M_{b_1}	M_l in %	M_h in %
USA	8524	1127.19	77.3	6.8	885.96	73.2	8.3	98.46	59.0	13.3	8.28	64.7	11.0	7.44	63.5	11.7	11.92	85.8	2.3
GBR	1989	761,79	82.5	3.5	546.36	78.4	4.3	91.96	59.4	12.2	9.75	58.7	15.2	8.63	56.4	15.5	15.04	73.4	3.9
AUS	1639	823.65	79.5	3.9	554.13	76.8	4.0	88.96	61.0	12.4	8.04	63.7	10.8	7.18	63.1	11.3	8.06	84.1	2.3
CAN	1399	805.16	81.0	3.8	594.31	77.9	4.6	87.29	60.3	11.2	7.96	65.5	11.0	7.11	64.2	11.4	9.19	82.4	2.6
BRA	991	3837.03	22.9	39.7	139.56	96.0	0.7	38.23	90.9	2.7	4.38	87.0	3.5	3.98	86.4	3.4	18.89	91.6	2.3
NLD	732	579.73	84.7	1.1	423.57	80.3	1.8	77.86	65.3	8.6	8.88	60.2	13.1	7.89	59.6	13.5	9.63	83.6	2.0
CHN	706	418.37	90.1	1.3	311.75	86.4	2.4	52.92	81.3	5.1	7.19	66.1	8.8	6.28	67.0	9.1	3.14	95.2	1.1
DEU	624	498.14	88.6	1.6	337.17	83.3	0.0	58.05	76.8	6.3	7.21	71.8	9.3	6.35	71.3	11.1	4.43	88.3	1.0
SWE	572	539.73	88.8	2.1	384.14	86.2	2.4	72.73	70.6	6.1	7.67	68.5	9.1	6.71	68.5	9.1	8.74	83.4	2.1
ESP	516	933.23	72.7	4.1	322.90	87.8	1.7	61.16	77.3	6.8	6.47	73.3	7.9	5.73	74.2	8.3	10.39	83.1	2.5
ZAF	391	558.94	84.9	1.5	423.66	81.6	3.1	97.39	56.5	11.0	8.74	61.9	9.5	7.91	61.9	12.3	6.58	81.1	1.3
ITA	343	338.10	92.7	0.9	258.43	89.5	1.2	52.84	79.3	4.7	8.41	61.8	14.9	7.35	63.6	15.2	11.18	86.3	3.5
NOR	322	658.33	85.1	2.8	451.78	83.5	2.8	79.12	69.3	11.5	8.79	64.3	12.1	7.77	63.7	12.7	12.07	81.1	4.0
DNK	315	401.35	91.4	0.6	291.20	87.0	0.6	69.16	70.8	7.0	7.65	65.7	10.5	6.88	62.5	10.8	9.62	82.5	2.9
FRA	296	537.41	88.5	1.7	324.30	84.8	1.0	63.93	72.6	6.8	9.91	58.1	18.9	8.87	57.1	19.6	12.02	82.8	3.4
CHE	296	621.68	86.8	3.0	440.95	83.4	3.7	92.57	62.5	12.5	11.99	52.4	18.2	10.90	50.3	20.3	12.62	79.1	3.4
JPN	287	368.19	92.0	1.4	257.97	91.6	1.0	46.42	86.1	2.4	5.74	76.7	5.9	5.01	76.3	7.7	7.21	88.2	3.8
IND	261	468.15	88.5	1.5	367.25	85.1	2.3	68.50	71.3	6.5	7.46	72.8	9.2	6.26	75.1	9.6	8.18	86.6	2.7
IRN	253	181.74	96.0	0.4	128.01	93.7	0.4	30.35	89.7	1.6	2.46	90.1	2.4	3.74	85.0	4.0	5.06	97.2	0.8
BEL	251	582.36	84.1	1.6	413.61	81.7	1.6	93.69	60.6	13.9	10.17	59.0	15.1	8.78	60.2	15.1	6.66	79.7	0.4
World	18729	1002.05	78.3	6.3	611.66	79.5	5.0	80.75	66.7	10.0	7.07	69.9	8.9	6.32	69.0	9.3	10.00	86.3	3.9

Legend: N: number of papers; U: PlumX Usage; A: EBSCO Abstract views; P: PlumX Captures; S: Scopus citations; C: WoS citations; M: PlumX Social Media. Index *l* denotes the lowest CSS class 1, index *h* the highest two classes 3&4 combined. Index b_1 denotes the mean value of the corresponding metrics.

Country analysis

The last subsection is devoted to a comparative analysis of countries. We have selected those countries that have (co-)authored at least 250 papers of document type *article* or *review* in 2015. Twenty countries met this condition. For data presentation, we will use their three-literal ISO codes (see the code list in Appendix). For the comparison we have used only PlumX Usage, Capture, Social Media metrics, EBSCO abstract views and Scopus and WoS citation rates. In addition to the CSS classes, we have added the mean values of the metrics, which actually coincide with the corresponding CSS b_1 scores.

The indicators are given in Table 8. The citation indicators by and large reflect a well-known situation. Research in the US and several countries in West- and North-Europe exhibit high citation impact. This is reflected by both Scopus and WoS citations and includes mean citation rates as well as citation distribution over CSS classes. In particular, Switzerland, France and the UK show the most favourable patterns in terms of citation impact. Brazil and Iran form the low-end of the selection. This is contrasted by the PlumX usage indicators.

Except for the US, which are slightly above the world standard (see Table 8), Brazil is the only country in the selection with usage-metric values that are considerably above the expectation. The mean value is almost four times the expectation and the share of highly used papers exceeds that of poorly used papers. However, Brazil has very low mean value of EBSCO abstract views among the overall usage value. The spearman correlation between PlumX usage and EBSCO abstract views for Brazil is not significant ($\rho = 0.007268$, $p > 0.05$). The extreme high PlumX usage value was confirmed by an additional check contributed from SciELO usage counts. The effect of SciELO usage is not that distinct in any other countries. The usage of the remaining 18 countries falls distinctly short of the expectations with relatively correlated abstract view usage. For example, Table 9 shows that the top 5 countries except for Brazil all have strong spearman correlations between PlumX usage and EBSCO abstract views. This trend reveals the dominant role of EBSCO abstract views in the PlumX usage indicators for the most countries except for Brazil.

Table 9. Spearman correlations between PlumX usage and EBSCO abstract views for the five countries with largest publication output in public health (2015)

Country	ρ
USA	0.9790391***
GBR	0.976407***
AUS	0.9609342***
CAN	0.7345188***
BRA	0.007268

$^{***}p < 0.001$

For Brazil, we find a clear contradiction between (PlumX) Usage and EBSCO usage/Citation metrics instead. Without further analysis of the background and motivation of usage and its user community, this metrics does not convey any clear message. Capture, by contrast, provides a more differentiated and less polarised picture. Deviations from the reference standard are less extreme and more in line with what one would expect from an impact measure. This measure seems indeed to provide added value to the scholarly impact. The effect of the outliers in altmetrics is once more expressively shown by the comparison of Social Media metrics of Brazil and the USA (cf. Table 8). Both countries have the same percentage high Social Media mentions (2.3%) while Brazil has a distinctly higher share of low mentions (91.6% – vs. 85.8% for the USA). Therefore, one would expect a mean value of this metrics of Brazil much lower than in the case of the USA. By contrast, the opposite case can be observed. With b_1=18.9 Brazil clearly “outperforms” the USA (b_1=11.9). However, this result is the effect of one single outlier: Brazil has one document mentioned, shared or commented more than 13,000 times on social media, mainly on Facebook pages (#11 in Table 5). Although the USA has also one extreme outlier (>23,000, #2 in Table 5), the effect of their document is absorbed by their large publication output, which is of one order of magnitude larger than that of Brazil. In the CSS model both documents are just one item in the highest class, where their actual numerical value does not have any further effect. This example may illustrate that mean-value based altmetric indicators should be used with the utmost caution, most notably in the case of smaller publication sets.

Discussion

The PlumX Usage metrics provide – at least in the subject under study – usable numerical information on abstract and full-text view as well as on EBSCO link outs. Other forms of usage were less or not significant. Above all, SciELO seems to be responsible for the outstanding usage counts of the Brazilian papers in public health. The biases, the extremely flat distribution and the experienced low robustness of this metrics make it less appropriate for application in research assessment.

The more robust WoS usage count lacks clear interpretation and requires access and use of the WoS database. This metrics might be an interesting companion to the WoS citation data as it leaves the scope of scholarly communication (cf. Chi & Glänzel, 2018; Chi et al., 2019). All citation metrics proved to correlate, most notably WoS and Scopus but these are restricted to scholarly communication.

Captures and social media may have the potential to provide additional information to citation impact. The two important components of Capture were Mendeley and Exports/Save counts. CiteUlike was, however, not significant. The usefulness of Mendeley readership as early impact indicator has recently been shown by Thelwall (2018) and was also confirmed in the present study, but he also pointed to limitations for their use in research evaluation (Thelwall, 2017a; 2017b).

In the present study we primary explored the possibility to compare the social media metrics to other altmetric and bibliometric indicators. Tweets is the most influential component of PlumX social media metrics, while Facebook is not that common to be used to disseminate research for most public health publications and Google + is never used. The distribution of social media metrics is very skewed with zero frequencies close to 50% in public health. This results in severe limitations for the general applicability of indicators based on these metrics.

Conclusions

The example of category *public, environmental & occupational health* has provided interesting and assumingly typical insight in the properties of altmetrics. Above all, these properties determine the opportunities and limitation for their possible application in an evaluative context. After our short digression to the world of altmetric indicators, as they represent the state-of-the-art, we can conclude that the indicators, in their present designed and availability, do not provide any comprehensive solution nor alternative to the well-elaborated and consistent system of scientometric tools, apart from those well-known conceptual and methodological limitations. Most strikingly, in this study, just like in our previous papers (e.g., Chi et al., 2019; Chi & Glänzel, 2019), we have found some lack of consistency in these measures. Adding, removing or just changing repositories or databases may result in dramatic changes and may turn local or regional effects into global phenomena. The database SciELO may just serve as an example for this effect. Just counting downloads, mentions, likes, tweeds and other social-media related measures without knowing the real purpose behind these actions certainly cannot provide unequivocally interpretable (quantitative) evidence.

Once again, we have to point to the insightful and profound article by Abraham Bookstein (1997) on the demons to measurement in social sciences. In his study, he characterised, in the context

of informetrics distributions, the three most essential ones as *randomness*, *fuzziness* and *ambiguity*. In the world of altmetrics, all three demons, randomness, fuzziness and most notably ambiguity, may become even more crucial than in traditional informetrics. In this context we would also like to refer to the arguments by Sugimoto (2016) and Glänzel & Chi (2016) in demand for more transparency and clarity in the data covered and the need for clear definition of actors on both sides. In particular, if one talks about impact – impact upon whom is meant and, furthermore, what are the potential biases in terms of actor and user profiles? Without clarification of such issues any attempt of standardization, normalization and benchmarking of metrics would remain unsuccessful.
We have, similarly to other recent studies in this paper too, analysed the correlation between altmetrics and traditional bibliometric indicators, but we did not aim at evaluating the usefulness of metrics on the basis of that correlation, nor at searching for causal relationship between the metrics under study. We expect the real (added) value of the new metrics in providing additional information that cannot directly conclude indicators of scholarly communication.
Summarising our results and observations, we can say that the example of public health has confirmed that the extension of scientometrics beyond the scope of scholarly communication remains a challenge. Significance and robustness of measures did not yet meet the standards of traditional bibliometric tools and the interpretability of altmetrics indicators requires even more context analysis than those of scholarly communication. At this moment, we find that the currently used altmetric metrics to measure the broader impact of research still fall short of the enormous expectations and the sometimes nonreflective enthusiasm in their use. Nonetheless, some of these new metrics may already provide useful information based on the feedback of broader, often heterogeneous groups of users that could be useful as supplement to traditional bibliometric indicators indeed.

Acknowledgement

This paper is an extended version of a previous work presented at the 17^{th} ISSI Conference in Rome, Italy (Glänzel & Chi, 2019).

References

Bookstein, A. (1997). Informetric Distributions. III. Ambiguity and Randomness. *JASIS*, 48(1), 2–10.
Brody, T., Harnad, S. & Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact, *JASIST*, 57(8), 1060–1072.
Chi, P.-S. & Glänzel, W. (2018). Comparison of citation and usage indicators in research assessment in scientific disciplines and journals. *Scientometrics*, 116(1), 537–554.
Chi, P.-S. & Glänzel, W. (2019), Citation and usage indicators for monographic literature in the Book Citation Index in the social sciences. *ISSI Newsletter*, 14(4), 80–86.
Chi, P.-S., Gorraiz, J. & Glänzel, W. (2019). Comparing capture, usage and citation indicators: an alt-metric analysis of journal papers in chemistry disciplines. *Scientometrics*, 120(3), 1461–1473.
Costas, R., Zahedi, Z. & Wouters, P. (2015). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *JASIST*, 66(10), 2003–2019.
Glänzel, W. & Chi, P.-S. (2016). Scientometrics 2.0 – and beyond? Background, promises, challenges and limitations. *ISSI Newsletter*, 12(3), 33–36.
Glänzel, W. & Chi, P.-S. (2019). Research beyond scholarly communication – The big challenge of scientometrics 2.0. In: G. Catalano, C. Daraio, M. Gregori, H. Moed, G. Ruocco (eds.), Proceedings of the ISSI Conference 2019, Rome, Italy, 424–436.
Glänzel, W. & Thijs, B. (2011). Using 'core documents' for the representation of clusters and topics. *Scientometrics*, 88(1), 297–309.
Glänzel, W. & Thijs, B. (2012). Using 'core documents' for detecting and labelling new emerging topics. *Scientometrics*, 91(2), 399–416.

Glänzel, W. & Heeffer, S. (2014). *Cross-national preferences and similarities in downloads and citations of scientific articles: A pilot study*. In: E. Noyons (Ed.), Context Counts: Pathways to Master Big and Little Data. Proceedings of the STI Conference 2014, Leiden University, 207–215.
Glänzel, W. & Schubert, A. (1988). Characteristic Scores and Scales in Assessing Citation Impact. *Journal of Information Science*, 14(2), 123–127.
Gorraiz, J., Gumpenberger, C. & Schloegl, C. (2014). Usage versus citation behaviours in four subject areas. *Scientometrics*, 101(2), 1077–1095.
Gumpenberger, Ch., Glänzel, W. & Gorraiz, J. (2016). The ecstasy and the agony of the altmetric score. *Scientometrics*, 108(2), 977–982.
Hoffmann, Ch. P., Lutz, Ch. & Meckel, M. (2014), Impact Factor 2.0: Applying Social Network Analysis to Scientific Impact Assessment. SSRN.
Lewison, G. (2004). *Citations to papers from other documents. Evaluation of the Practical Effects of Biomedical Research*. In: H.F. Moed et al. (eds.), Handbook of Quantitative Science and Technology Research, 457–472.
Lewison, G. (2008). The returns to society from medical research (in Spanish). *Medicina Clínica*, 131 Suppl. 5, 42–47.
Moed, H. F. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal, *JASIST*, 56(10), 1088–1097.
PlumX (2019), *PlumX Metrics*. Plum Analytics. Accessible at: https://plumanalytics.com/learn/about-metrics/.
Priem, J. & Hemminger, B. H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social Web. *First Monday,* DOI:10.5210/fm.v15i7.2874.
Sugimoto, C. (2016), *Unlocking social data for science indicator*s. (White paper), NSF Workshop on Bibliometric Indicators, Arlington.
Thelwall, M. (2012). Journal impact evaluation: a webometric perspective, *Scientometrics*, 92(2), 429–441.
Thelwall, M. (2017a). Are Mendeley reader counts high enough for research evaluations when articles are published? *Aslib Journal of Information Management*, 69(2), 174–183.
Thelwall, M. (2017b). Are Mendeley reader counts useful impact indicators in all fields? *Scientometrics*, 113(3), 1721–1731.
Thelwall, M. (2018). Early Mendeley readers correlate with later citation counts. *Scientometrics*, 115(3), 1231–1240.
Wouters, P., Zahedi, Z. & Costas, R. (2019). *Social media metrics for new research evaluation*. In: W. Glänzel, H. Moed, U. Schmoch, M. Thelwall (eds.) Handbook of science and technology indicators. Springer Berlin-Heidelberg. 687–714.
Zahedi, Z., Costas, R. & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics*, 101(2), 1491–1513.

Appendix

List of ISO 3166-1 alpha-3 country codes used in the study

ISO code	Country
USA	United States of America
GBR	United Kingdom
AUS	Australia
CAN	Canada
BRA	Brazil
NLD	Netherlands
CHN	China
DEU	Germany
SWE	Sweden

ESP	Spain
ZAF	South Africa
ITA	Italy
NOR	Norway
DNK	Denmark
FRA	France
CHE	Switzerland
JPN	Japan
IND	India
IRN	Iran
BEL	Belgium