Scientometrics (2021) 126:4941–4974
https://doi.org/10.1007/s11192-021-03944-9

The citation impact of articles from which authors gained monetary rewards based on journal metrics

Gabriel-Alexandru Vîiu[1] · **Mihai Păunescu**[1]

Received: 17 September 2020 / Accepted: 4 March 2021 / Published online: 27 April 2021

Abstract
Monetary rewards granted on a per-publication basis to individual authors are an important policy instrument to stimulate scientific research. An inconsistent feature of many article reward schemes is that they use *journal-level citation metrics*. In this paper we assess the actual *article-level citation impact* of about 10,000 articles whose authors received financial rewards within the Romanian Program for Rewarding Research Results (PR3), an exemplary money-per-publication program that uses journal metrics to allocate rewards. We present PR3, offer a comprehensive empirical analysis of its results and a scientometric critique of its methodology. We first use a reference dataset of 1.9 million articles to compare the impact of each rewarded article from five consecutive PR3 editions to the impact of all the other articles published in the same journal and year. To determine the wider global impact of PR3 papers we then further benchmark their citation performance against the worldwide field baselines and percentile rank classes from the Clarivate Analytics Essential Science Indicators. We find that *within their journals* PR3 articles span the full range of citation impact almost uniformly. In the larger context of *global broad fields of science* almost two thirds of the rewarded papers are below the world average in their field and more than a third lie below the world median. Although desired by policymakers to exemplify excellence many PR3 articles are characterized by a rather commonplace individual citation performance and have not achieved the impact presumed and rewarded after publication based on journal metrics. Furthermore, identical rewards have been offered to articles with markedly different impact. Direct monetary incentives for articles may support productivity but they cannot guarantee impact.

Keywords Research articles · Citation impact · Journal metrics · Monetary rewards · Web of Science

✉ Gabriel-Alexandru Vîiu
gabriel.viiu@yahoo.com

[1] National University of Political Studies and Public Administration, Bucharest, Romania

Introduction

Performance-based research funding systems implemented in many countries since the 1980s with a view towards increasing the quantity and quality of scientific research have been a recurring topic of interest in the scholarly literature on higher education institutions and science policy (see for example Geuna & Martin, 2003; Hammarfelt & de Rijcke, 2015; Hicks, 2012; Sandström and Van den Besselaar, 2018). Such systems typically consist in comprehensive national evaluations whose outcomes guide the differential allocation of research funding in a way that *institutions* (or collective entities such as departments) with a proven track record for excellence receive more funding to further their work than similar entities with weaker performance. The United Kingdom's peer-review-based Research Excellence Framework is probably the best known and it epitomizes the so-called "hard" version of managerialism (Trow, 1994) which consists in the core idea of linking funding to performance, as measured against certain government-mandated criteria. A bibliometric-based alternative approach is embodied in the Norwegian model (Schneider, 2009; Sivertsen, 2016) that relies on a point system within which publications are weighed by type and by the prestige of their publication channel. Though explicitly designed for aggregated use at the national level the model's constituent elements were reported (Aagaard, 2015) to have found local use at the department level in universities, for example in staff monitoring and in recruitment decisions. Furthermore, since 2016 the model has been adapted in the form of an Output-Based Research Support Scheme at University College Dublin where monetary incentives for research are given to individual scholars based on their publications and on PhD supervision (Cleere & Ma, 2018).

Within the broader literature on performance-based research funding over the past few years a growing number of studies have drawn attention to the specific issue of direct monetary rewards allocated in some countries to *individual researchers*, rather than to collective units of assessment, with the specific aim of fostering an increase in the number of publications in highly visible and impactful international journals. The extent of per-publication rewards in China has received increased attention in studies which suggest that the practice of giving additional money to scholars for top publications is widespread in this country, while nonetheless varying in practical application from university to university (Quan et al., 2017; Shu et al., 2020). Abritis and McCook (2017) list no fewer than 13 countries where per-publication awards are reported to range from 754 US dollars in South Africa to as much as 165,000 US dollars in China. Direct per-publication rewards granted to individuals are distinct from other incentives to increase research publications, such as the fixed salary bonuses that have been implemented in Spain's "sexenio" (Jiménez-Contreras et al., 2003; Osuna et al., 2011), in Mexico's National System of Researchers (Neff, 2018; Sandoval-Romero & Larivière, 2020) or in Paraguay's National Programme of Research Support (Aboal & Tacsir, 2017).

In this paper we contribute to the ongoing debate on financial incentives awarded directly to individual scholars on a per-publication basis by presenting and critically evaluating the Romanian Program for Rewarding Research Results—abbreviated in this paper as *PR3*—which is a national money-per-publication policy that has been in operation since 2007 with the aim of improving the quality, impact and visibility of Romanian research. PR3 essentially consists in offering Romanian authors one time monetary rewards for each paper they publish in journals indexed in the Clarivate Analytics Web of Science, still often referred to colloquially as "ISI papers". Although in continuous operation for over a decade and well-known to Romanian academics and researchers this program has not been

the subject of any systematic work aiming to document its development, nor has it been the object of any kind of appraisal seeking to determine whether or not it has achieved its stated goals of promoting internationally visible and impactful research. We believe PR3 is an interesting case study for scholars and policymakers concerned with science policy and we have three specific aims for the current paper:

(1) to provide a factual account of the context and technical underpinnings of the program;
(2) to determine the actual, empirical citation impact of PR3 papers;
(3) to offer an appraisal of the PR3 methodology from a scientometric perspective.

The three aims are naturally interlinked and we believe the empirical one is an especially important topic because of the journal-centered (in particular journal impact factor-centered) approach adopted as a basis for allocating rewards: PR3 is an exemplary money-per-publication program that essentially uses *journal metrics* to evaluate individual articles. However, this top-down mechanism—which is common in the other national contexts mentioned above—is at odds with the technical features of citation distributions, especially their well-documented skewness (Albarrán et al., 2011; Ruiz-Castillo & Costas, 2018; Seglen, 1992) which prohibits generalizing from journal to paper impact. Therefore, as we explain in greater detail below, in this paper we are concerned strictly with determining the *article-level citation impact of individual PR3 papers*. Journal-level metrics play no part in our analyses and we emphasize that such collective indicators (especially the journal impact factor) should not be confounded with article-level citation impact.

There is a mixed tableau of negative and positive effects of the per-publication monetary rewards within which we may broadly contextualize the present work. In a study of paper submissions to *Science* Franzoni et al. (2011) found that cash bonuses awarded on an individual per-publication basis (introduced in Turkey, China and South Korea) only augment the number of papers being submitted (i.e. productivity), not their actual acceptance rate, and that it is primarily career incentives that correlate with actual publications. That cash rewards significantly increase aggregate output is also attested to in the case of South Africa where "publications listed in the Scopus database each year more than doubled in the decade after the payout programme began" (Hedding, 2019, p. 267). However, based on the comprehensive experience of the Turkish cash bonus system Tonta (2018) makes the point that such mechanisms do not have significant effects in increasing researchers' productivity. Demir (2018) on the other hand arguest that monetary rewards do have the effect of increasing output, but cautions that they also inadvertently stimulate publications in potentially fake journals and attendance to questionable conferences. Somewhat closer to our own intent—although only in very broad terms and adopting an altogether different methodological apparatus—Tonta and Akbulut (2020) use a sample of papers published between 2006 and 2015 to investigate the differences between the citation impact of papers included in the Turkish reward program and the impact of papers that were not rewarded. Their analyses do not include field-level distinctions (only a rough division between science and social science) but their results concerning citations to individual papers indicate only a moderate advantage for papers that were rewarded compared to papers that were not, a trend that is more pronounced for science papers.

In contrast to the rather negative or neutral effects noted above, favorable effects of cash bonuses have also been reported in narrower institutional settings. For example, analyses of the support policies in a South Korean university (Bak & Kim, 2019; Kim & Bak, 2016) indicate that financial bonuses can be effective tools in academia, that

the performance of professors is proportional to the incentives, and that such incentives can motivate even unproductive researchers to improve the quality and quantity of papers in their portfolios. A general positive reception of monetary incentives among scholars is also reported in a qualitative investigation of the effects of the Output-Based Research Support Scheme at University College Dublin by Ma (2019) who nonetheless also mentions important negative outcomes: a more competitive, individualistic approach to research, the emergence of gaming strategies, a shift in the focus of staff from teaching to research activities and a very uneven distribution of benefits, typically skewed toward senior research staff. The point that monetary rewards lead to a highly stratified increase in productivity that mainly comes from senior professors has also been made in the Chinese context (Heywood et al., 2011). Such findings ultimately echo problems of goal displacement and task reduction (see for example Müller & de Rijcke, 2017) generally associated with the use of performance indicators in evaluation.

At this point we note that the literature specifically addressing the effects of national or institutional policies involving per-publication monetary rewards given to scholars individually is still underdeveloped. To date it has largely been overshadowed by studies dealing with performance-based research funding systems (likely due to the greater prevalence and public character of such systems relative to direct monetary rewards). However, even previous studies on individual monetary rewards mostly address their effects from the perspective of scholars and institutions, focusing for example on changes in publication strategies, mentalities and attitudes towards research, or on changes in aggregate publication output. *Article-level analysis of the citation impact of rewarded papers* is still an area of scientometric research marked by deep knowledge gaps. The present work aims to contribute in charting this as yet insufficiently explored territory by discussing a representative national reward program addressing individual scholars. Importantly, we explicitly focus on the core objects of the program, the actual articles and their impact, years after the moment of reward, when impact can actually be measured and contextualized individually rather than be presumed based on past (collective) journal metrics. Future studies might of course address the distinct topic of effects on scholars and institutions.

Our paper is structured as follows: "The output-oriented nature of research evaluation in Eastern Europe" section contextualizes the Romanian PR3 in the broader landscape of research evaluation policies implemented in other East European countries, highlighting its specific status. "The Romanian Program for Rewarding Research Results" section addresses the first of our three aims by presenting PR3 and briefly describing its integration with other national policies. Two subsequent sections—"Toward an empirical assessment of the citation impact of PR3 papers" and "Data and methods for the comprehensive evaluation of PR3 articles"—begin to address our second aim and lay the foundations for the presentation of our empirical investigation, encompassing about 10,000 articles whose authors received monetary rewards in the 2011–2015 editions of PR3. Our research results are then reported in three distinct sections—"The citations of PR3 articles relative to those of all articles from the publishing journals", "The citation impact of PR3 articles within global scientific fields" and "A dual view of citation impact within journals and broad fields". The presentation of our results is followed by the section "Discussion: a methodological critique of PR3 from a scientometric perspective", wherein the third aim of our paper is addressed. A final section presents our "Summary and concluding remarks".

The output-oriented nature of research evaluation in Eastern Europe

The global output of scientific publications over the past two decades is dominated by Europe, which in 2016 accounted for 27.1% of outputs indexed in the Web of Science (WoS), and by the United States (with 19.5%), while Chinese contributions have surged to 16.7% (from 2.7% in 2000) as a result of significant recent investment in science (European Commission 2018). Despite its overall leadership (also visible in the proportion of highly cited papers) European scientific research has been marked by significant disparities when considering constituent countries. East European (EE) countries in particular have been shown to display lower levels of scientific output and impact, features which may be explained by their weaker economies which sustain significantly lower research and development expenditure compared to older European Union members (Must, 2006; Vinkler, 2008).

Owing, among others, to historical resource deficits, the development of science in EE countries over the past decades has also been marked by divergent patterns regarding international cooperation. For example, whereas Poland and Hungary had already benefited from collaboration with more developed countries since the early 1980s, for Romania and others this did not occur until the 1990s (Braun & Glänzel, 1996), i.e. following the collapse of the communist regime. In the next two decades, in a broader context in which international collaboration between EE countries and Western nations increased substantially, Romania also benefited from improved co-authorship, albeit quite unequally since collaboration was stronger for sciences than for social sciences and humanities (Teodorescu & Andrei, 2011). Overall, it took well over a decade for Romanian international collaboration and publication output to noticeably increase relative to pre-1990 levels, marking delayed effects of political and societal changes on the scientific system (Kozak et al., 2014).

In an effort to improve their scientific standing EE countries have resorted to a variety of policies which have sought in one form or another to increase their publication counts, with international publication being stressed against outputs in national outlets which are usually deemed to be less prestigious. The Czech Republic implemented what has been argued to be a uniquely radical performance-based research funding system focused, since 2009, exclusively on research outputs which were converted into concrete funding in accordance with a preset point system that strongly favored articles in journals indexed by WoS, especially those in *Nature* or *Science* (Good et al., 2015). Though it considered organizations as units of analysis the Czech evaluation is also reported (Pisár et al., 2019) to have trickled down to the level of individuals as academic salaries in universities changed based on the points collected for publications, further incentivizing publication quantity in lieu of quality. However, comparing the Czech Republic to a handful of other European countries with regard to WoS-indexed articles Vanecek (2014, p. 670) noted that "the growth of the publication impact has lagged behind the growth of the output". In its distinct effort of enhancing scientific research Slovakia has also implemented a performance-based system emphasizing publications in Scopus and WoS since 2011 and more recent methodological constraints entail the explicit use of journal impact factors to distinguish between six categories that correspond to different levels of funding (Pisár & Šipikal, 2017). A comprehensive evaluation aimed at categorizing all research units in four performance classes based mainly on bibliometric indicators is also in operation in Poland where the highest number of points-per-evaluation item can be obtained for articles from journals indexed in the Journal Citation Reports (JCR) having an impact factor (Korytkowski & Kulczycki,

2019; Kulczycki, 2017). Noting that most EE states have focused their national scientific evaluation policies on international, commercial bibliometric data from either Scopus or WoS, with preference in evaluation typically being given to the latter, Pajić (2015) also reports that in Serbian evaluations a paper from a JCR journal can be worth as much as eight papers from a national journal.

Without endeavoring to construct an exhaustive list of European practices it is obvious from the preceding examples that international publications, especially WoS-indexed articles, play a critical role in the evaluation of research in EE countries and that a linkage between *institutional* funding and such publications is common, although methodologically diverse. Romanian research evaluation practice is also heavily reliant on WoS and JCR, as we explain in more detail below, but the peculiar feature that seems to set Romania apart is that it appears to be the only EE country where a distinct practice of rewarding *individual authors* for their WoS papers has been implemented *on a national scale for more than a decade*. Since no systematic work has been carried out either to document or to assess this distinguishing characteristic of Romanian science policy from a technical, scientometric perspective, we undertake this process in the subsequent sections.

The Romanian program for rewarding research results

Romania is a remarkable epitome for appealing to international databases for research evaluation as it essentially built most of its assessment policies primarily around WoS and JCR standards on which it continues to rely heavily. It should in fact first be emphasized that the PR3 we discuss below is part of a much broader set of national policies and practices regarding the evaluation of research. These policies converge toward the idea that international publications, especially in WoS-indexed journals, are a highly desirable academic product. First, in a shift from "in-breeding localism and academic gerontocracy" to a meritocratic mode of career advancement (Vlăsceanu & Hâncean, 2015, p. 188), WoS publications (and in some cases JCR metrics such as the journal impact factor or article influence score) have played a prominent role in the national standards for associate and full professors since 2005 when the Ministry of Education issued orders explicitly identifying international visibility and impact with ISI papers or papers from other field-relevant international databases. Second, in a broader effort of relating quality with funding in higher education (Miroiu & Vlăsceanu, 2012), the introduction of so-called quality indicators that determined a part of each institution's annual revenues addressed, among others, the relevance and visibility of research and favored WoS papers against those from other databases and against national outputs. Third, in what to date remains a singular and still contested national exercise carried out in 2011 to classify universities and rank their study programs, the largely research-driven process (Vîiu et al., 2016) also emphasized WoS publications and article influence scores. Finally, WoS publications have also played an important role in the project-based competitive allocation of national research grants where project principal investigators are typically required to have a minimum number of publications and/or publication-based points for which article influence scores are taken into consideration.

Within the broader policy context sketched above the idea of providing direct monetary rewards for research outputs produced by Romanian scholars took shape under the 2007–2013 National Plan for Research, Development and Innovation. Within the Human Resources component outlined by the National Plan a specific program was set up to reward "excellent research results", namely the Program for Rewarding Research

Results—abbreviated as *PR3* in this paper. Unlike other policies designed to support the national research system, PR3 was (and continues to remain) the only program within which direct monetary rewards are given to individual Romanian scholars for their research outputs. The rewards are given for each possible output that is submitted and eligible and no formal limit exists on the number of submissions that an author can make to PR3. The use of the reward funds is at the full discretion of the recipient, exempt from any reporting or audit, unlike some other support schemes wherein monetary incentives can be allocated to individuals strictly for use in their research activities and subject to subsequent scrutiny.[1] The officially declared objective of PR3 (constantly repeated in all yearly information packages) has always been that of "increasing the quality, impact and international visibility of Romanian research". The program is managed by the Executive Agency for Higher Education, Research, Development and Innovation Funding (UEFISCDI) under the direction of the Ministry of Education and Research.

PR3 is voluntary and WoS-centric. On one hand, the only eligible research outputs that may be submitted to PR3 to claim a reward are articles and reviews published in journals that are indexed in the main citation indexes currently managed by Clarivate Analytics: Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI) and the Arts and Humanities Citation Index (AHCI). On the other hand, PR3 is fundamentally driven by the Clarivate Analytics journal-level metrics presented in the annual JCR. Specifically, the journal impact factor (JIF) or/and the article influence score (AIS) are used to rank journals—within their specific WoS subject category—in two main classes: a so-called "red" class (with higher monetary rewards) and a so-called "yellow" class (with smaller rewards). The red class is made up of the top 25% journals within each WoS subject category, as ranked by the JIF/AIS, and the yellow class is made up of the top 50% of journals (excluding the ones already in the top 25%). Readers familiar with JCR will recognize that the two regions correspond to the 1st (Q1) and 2nd (Q2) JIF/AIS quartiles that can be selected in the JCR online platform within any of approximately 230 WoS subject categories. For articles published in journals attributed to multiple subject categories the most favorable ranking is taken into consideration so if an article is marked as yellow in one category but red in another, the reward that authors can claim for it is that corresponding to the red class. Throughout the 14 consecutive years in which it has been implemented either the JIF, the AIS or both have been used to determine the two main ranking classes. Since 2013 (with the exception of 2016) all PR3 editions to date have allowed *both* JIF and AIS rankings, again with the provision that the more favorable of the two may be used when submitting a published paper for a reward in PR3.

In addition to the red and yellow classes, which make up the bulk of research publications submitted to PR3, the program delineates papers published in the journals *Nature* or *Science*—which are considered the most prestigious and are attributed the highest possible rewards—and AHCI papers. AHCI papers were initially (in 2009) separated in two classes based on the length of the period during which the publishing journal had been indexed but since 2011 only papers from journals indexed in AHCI for at least five years are accepted. Since 2016 an additional reward class has been established for papers that are published in the top ranked journal in a WoS category. These are valued above the red class but greatly below a *Nature* or *Science* paper. Papers from some *Nature* group journals (e.g. *Nature*

[1] Note also that the per-publication nature of PR3 precludes the necessity of translating the publications to any intermediate system of points that would then determine rewards.

Communications) have been attributed to the same reward class as those from *Nature* or *Science*.

Table 1 offers a summary of the evolution and main components of PR3 since its inception in 2007. Throughout the past 14 years the authors of 38,425 publications have been rewarded. Overall, there seems to be a visible difference in the total number of articles rewarded considering the eligibility criterion used: in the three years in which only the AIS was used there are fewer rewarded articles than in the others.[2] Note also that for 2007 and 2008 no specific indicator was used to distinguish the articles and rewards: outputs were accepted if they were simply published in any WoS indexed journal. Only one criterion was used to differentiate rewards: the number of Romanian authors listed for each article. In fact, *the only firm provision that has not changed under any PR3 edition is that the reward corresponding to an article is divided, equally, among all the Romanian authors of the article*, in line with the PR3 objective of increasing the international visibility and impact of *Romanian* research. Being a Romanian author is technically assimilated with the declared affiliation to a Romanian research entity in the WoS record of the indexed paper. The very lax initial provision regarding eligible documents was changed from 2010 onwards, with only articles, reviews and proceedings papers being considered valid submissions to PR3. Furthermore, from 2011 proceedings papers were only allowed for social sciences and from 2015 they have been removed entirely, with only articles and reviews now being accepted.[3]

The size of the monetary rewards within PR3 has oscillated considerably for the papers in journals indexed in AHCI, which have also been the smallest, whereas those for papers in the red and yellow region, as well as those from *Nature* or *Science* have been largely stable, with two distinct periods discernible with regard to the monetary values of the articles: 2009–2015 and 2016—present. Due to currency fluctuations as well as to increases in the reward amounts (since 2016) a single-author paper in a red group journal could gain a researcher an amount between about $998 in 2015 and $1522 in 2018, while a paper in a yellow group journal (setting aside 2016 which was an atypical year when the yellow class rewards were halved compared to all the other years) could gain the author somewhere between $472 in 2019 and $656 in 2009. A paper in *Nature* or *Science* could be worth as "little" as $4993 in 2015 and as much as $10,148 in 2018. For completeness it should be mentioned that all the values presented are subject to the (generally volatile) national fiscal policies but to our knowledge only the standard 10% income tax has been applied to the monetary rewards in recent years. Nonetheless, it should be stated explicitly that the rewards are not negligible in a still developing post-communist country in which the net average salary has ranged between about $427/month in 2007 (the year of the 1st PR3) and about $670/month in 2018.[4]

[2] We considered whether this might also be due to financial constraints in the respective years but cross-referencing budget specifications from the information packages with figures on the total amounts disbursed for the articles indicates that in fact the initially allocated budget was *not* exhausted in any of the three years. For 2011 and 2012 the exclusion of proceedings papers from eligible documents (except for social sciences which are, however, poorly represented) might also explain the decrease.

[3] As is known to WoS users some documents can have multiple attribution. Those listed as "article; proceedings paper", "article; book chapter", etc. are not eligible submissions in PR3.

[4] These figures are based on calculations considering the annual exchange rate from the National Bank of Romania (listed in Table 1) and net income data from the Romanian National Institute of Statistics (https://insse.ro/cms/ro/content/c%C3%A2%C8%99tiguri-salariale-din-1938-serie-anual%C4%83-0).

Table 1 Synopsis of PR3 throughout 2007–2020

PR3 edition	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Papers rewarded	1586	3109	2883	2343	1522	1500	2577	2676	3072	2422	3101	3119	3467	5048
Eligible documents	Any journal item		A, R, PP		A, R, PP (only for social sciences)				A, R					
Criteria	–		JIF	JIF	AIS	AIS	JIF & AIS			AIS	JIF & AIS			
Reward classes (monetary rewards in RON)	Single author (3500) Multiple authors (2000/N) (*N* = number of Romanian authors)		Red (4000) Yellow (2000) Blue (500) AHCI 1 (2000) AHCI 2 (500) *Nature/Science* (20,000)	Red (4000) Yellow (2000) AHCI (1500) *Nature/Science* (20,000)	Red (4000) Yellow (2000) AHCI (400) *Nature/Science* (20,000)	Red (4000) Yellow (2000) AHCI (2000) *Nature/Science* (20,000)				Red (6000) Yellow (1000) AHCI (1000) *Nature/Science* (40,000) 1st in WoS category (10,000)	Red (6000) Yellow (2000) AHCI (2000) *Nature/Science* (40,000) 1st in WoS category (10,000)			
RON/EUR	3.34	3.68	4.24	4.21	4.24	4.46	4.42	4.44	4.45	4.49	4.57	4.65	4.75	4.84
RON/USD	2.44	2.52	3.05	3.18	3.05	3.47	3.33	3.35	4.01	4.06	4.05	3.94	4.24	4.24

The table mentions the total papers rewarded, eligible document types (*A* article, *R* review, *PP* proceedings paper), criteria used (*JIF* Journal impact factor, *AIS* Article influence score), reward categories (*AHCI* Arts & Humanities Citation Index) based on the activity reports of UEFISCDI (These are available at https://uefiscdi.gov.ro/rapoarte-de-activitate for the 2016–2020 period and http://old.uefiscdi.ro/Public/cat/820/Rapoarte-activitate-UEFISCDI.html for the previous years) and on yearly program information packages (These information packages are available at http://old.uefiscdi.ro/articole/1722/Articole.html (for 2007–2015) and https://uefiscdi.gov.ro/premierea-rezultatelor-cercetarii-articole (for 2016–2020)]. Annual average exchange rates between the Romanian RON and the Euro as well as the US dollar are provided for contextual reference based on data from the National Bank of Romania (Retrieved from https://www.bnro.ro/StatisticsReportHTML.aspx?icid=800&table=705&column=); the currency values have been rounded to two digits

Without delving into unessential aspects, two additional notes are necessary: within each PR3 edition since 2011 both articles published in the actual PR3 year and articles published in the previous year—which had not been rewarded, subject to the PR3 regulations specific to that preceding year—could be submitted to the program. Finally, it must also be stressed that rewards are granted following the chronological order in which the articles are submitted to PR3 (i.e. a first come, first served basis) and *strictly upon completion of administrative checks* rather than based on any substantive evaluation, either in the form of peer review or grounded in any article-level metrics. All submissions are simply checked to verify their existence in WoS and the reward class based on the journal metrics, the (eligible) document type recorded in the database, the explicit affiliation of the author(s) to a Romanian organization and to confirm the number of Romanian authors reported during the submission to PR3 in order to divide the reward.

Toward an empirical assessment of the citation impact of PR3 papers

Based on the above factual description it is readily seen that in its entirety PR3 rests on the implicit assumption that prospective evaluation based on journal-level metrics is adequate in determining what constitutes "quality", "impactful", "internationally visible" research. No consideration is given to retrospective, article-level assessment, based for example on actual citation counts. For comparison, Quan et al. (2017) report that in China some universities reward scholars based on actual paper citations, either by setting a minimum threshold themselves or by using the WoS Essential Science Indicators citation thresholds for hot (i.e. top 0.1%) or highly cited papers (top 1%). In PR3 however, being published in a top journal, regardless of the further history of the actual paper, is considered sufficient to warrant a monetary reward. Though we agree that being published is a mark of quality to the extent it shows successful navigation of the rigorous editorial and referee process one expects to find in a top journal, we also believe that publication is only a necessary condition for achieving visibility and impact, not a guarantee for either, especially for the latter, no matter the publishing journal.

There are strong scientometric caveats against the practice of evaluating articles via journal metrics, as conducted in PR3. We believe these caveats[5] create a legitimate need for an ex post empirical assessment of the articles whose authors benefited from monetary rewards across the years. The question that remains is how to conduct such an appraisal and we believe that a reasonable answer must fundamentally take stock of the citations accrued by the individual articles themselves. We remark that this approach is in fact in the very spirit of the underlying PR3 methodology which operates with the implicit acceptance of the critical relevance of citations in research evaluation (since the JIF and AIS heavily emphasized in PR3 are nothing more than citation-based indicators). There is, of course, ample support for our citation-based assessment in the scientometric literature. For example, writing with regard to the (then) Research Assessment Exercise in the United Kingdom Moed (2007, p. 581) highlighted the need to use "indicators measuring actual citation impact", further emphasizing that "journal-impact factors should not play an important role, and probably no role at all." More recently the evaluation of papers in the

[5] Knowledgeable readers of the journal are undoubtedly familiar with at least some and we therefore relegate their exposition to the part of the paper where we discuss our results.

context of the full citation distribution of all the other articles from the same journal has been advocated by Larivière et al. (2016) and the case has also explicitly been made that "if one wants to measure the quality or visibility[6] of a particular item, one must look at the citations actually received in the years following its publication" (Gingras, 2016, p. 48). In what follows we subscribe to the view that the retrospective, citation-based evaluation of papers is a reasonable way to gauge their actual, achieved impact. The general question that informs our empirical investigation is therefore concerned with determining the actual citation impact of the papers that were rewarded in PR3.

For a comprehensive evaluation we consider two distinct reference sets for each PR3 paper: *journals* and *broad fields of science*. We are motivated by the underlying idea of comparing focal PR3 articles with homogeneous reference sets of papers. Our approach implicitly mitigates the problem of disciplinary differences that typically place limitations on the comparative analysis of citation impact.

Journals remain the first legitimate reference standard when assessing an individual paper because—with the exception of multidisciplinary journals, including *Nature* and *Science*—they tend to contain coherent sets of publications (Schubert & Braun, 1996). Because the PR3 evaluation is also conducted having journals in the foreground we find it reasonable, at a *first analytical level*, to assess the ex post citation impact of the rewarded articles in the specific context of their publishing journals, i.e. by comparing the citations of a PR3 paper to the citations received by all the other papers published in the same year and journal. This allows a fine-grained analysis as it compares PR3 papers to their nearest peers.[7]

In spite of its obvious advantages an assessment which takes as reference only the publishing journal does not account for the global international impact of a paper, which is ultimately what PR3 aims to support. For this reason, at a *second analytical level*, we also study the citation performance of each PR3 paper relative to the worldwide citation performance of all the comparable items in the same broad scientific field and publication year. To this end we consider the position of each PR3 paper within its relevant Essential Science Indicators (ESI) category. There are 22 distinct ESI categories, a paper is uniquely assigned to only one of them and Clarivate Analytics regularly computes citation baselines (i.e. the "Citation Rate" table) which represent "the average performance of a global set of publications with the same subject area, document type, and year" (Clarivate Analytics, 2018, p. 11). These baselines offer the possibility of determining whether a specific paper has had a citation impact above or below the worldwide average in a specific field and are better suited for measuring the actual impact of individual papers than JIFs (Bornmann & Pudovkin, 2017). In addition to the citation rate baselines—and considering the same worldwide sets of WoS-indexed publications—distinct ESI tables also present selected *percentile baselines* which can provide a more nuanced appraisal as they allow the inclusion of a paper in one of six (explicit) rank classes: top 50%, top 20%, 10%, 1%, 0.1% and top 0.01% within each distinct category. Having the citation count of a given paper, its ESI category and the citation thresholds for the six classes above for that category and publication year, one can determine to what class a focal paper of interest belongs to (Bornmann et al., 2013). This allows a more nuanced appraisal compared to a benchmark against worldwide

[6] Note that together with impact these are exactly the aspects PR3 states that it is rewarding.

[7] It also has the benefit of avoiding the problems associated with the classification of journals in the WoS categories, especially the issue of multiple assignment of papers to more than one category, with which we engage in the discussion section.

averages. We note that in the calculation of ESI citation baselines only articles and reviews are taken into consideration.

Data and methods for the comprehensive evaluation of PR3 articles

Our investigation starts from the approximately 11,000 articles rewarded in PR3 in the 2011–2015 editions. We opted to limit our analysis to this five-year period owing to several considerations. First, since our main effort is that of determining the actual impact of the articles via citation counts we need to allow a reasonable window of several years within which citations can accumulate; this naturally excludes the analysis of articles rewarded in the more recent PR3 editions. For the specific window we have selected each article has at least four years to be cited (in the case of articles published in 2015) and as much as eight years (in the case of articles published in 2011). We emphasize that our main concern is that of comparing each of these articles with publications from the same year (and journal/ESI category) and the difference in citation window is therefore not fundamentally relevant. A second reason for opting for the 2011–2015 PR3 articles is that the earlier editions (especially the first two) were more permissive regarding eligible document types that could be submitted, a feature which subsided since 2011 when the eligibility criteria and general PR3 framework increasingly stabilized. Finally, 2011 was the first year in which the submission and evaluation processes were conducted online, with each article submission receiving a unique identification code.

We only analyze *articles* rewarded in the 2011–2015 PR3 editions and explicitly published in one of these years. We exclude other document types,[8] items published in 2010 that were rewarded in the 2011 PR3 edition[9] and publications indexed only in AHCI. Our main dataset comprises the remaining 10,281 papers and contains the essential identification information published by UEFISCDI for the articles that were admitted to PR3 in the 2011–2015 editions[10]: the unique ID internally allocated to each article upon submission, the article title, the name of the publishing journal, the total number of Romanian authors as well as the sum of money requested for payment. Based on this last item of information we derived the reward category of each article. For each of the articles in our main dataset we retrieved the information contained in the Web of Science (Core Collection) "full record", most notably the total citation counts.

With a view towards the fine-grained, relative assessment of the impact and visibility of the articles accepted in the 2011–2015 PR3 editions (i.e. for our first analytical level) we constructed, separately, a second dataset with the 1,871,212 articles published by all

[8] These mostly include proceedings papers and reviews but also two *retracted* publications, both included in the red reward class, published in 2014 (one in the *International Journal of Obesity*, the other in *Diabetes*) and rewarded in the 2015 PR3. The PR3 information packages do not mention the unlikely eventuality of article retraction and any steps that might be taken in such cases.

[9] Similarly, our investigation does not include articles published in 2015 but rewarded in the 2016 PR3.

[10] The information was published (and is still published for recent PR3 editions) in multiple *pdf* files made public as submissions are processed. In total, 36 such files contain the information for the 2011–2015 competitions considered in our investigation. They are available (under the heading for each year) at the following link: http://old.uefiscdi.ro/articole/1722/Articole.html. There is a notable inconsistency in the way information was published in the lists from one year to another and sometimes even between the lists belonging to the same year. We opted to keep only the minimal set of variables that were reported consistently across the entire five-year window and only for the articles that were accepted for a reward.

the journals relevant to PR3, i.e. every journal in which at least one PR3 article was published in any of the five relevant years (2011–2015). Our 10,281 focal articles were published in 2098 unique journals and there are 4323 unique relevant journal-year combinations as some journals contain more PR3 articles, sometimes published in the same year. With regard to the main variables of interest (publishing journals, article titles and citation counts) the second dataset with about 1.9 million records is an effective superset[11] of the initial 10,281 articles that allows us to compare the citation performance of each rewarded article to the performance of all the similar articles, published by the same journal and in the same year. When constructing the secondary dataset we also limited its extent to items indexed as articles and we explicitly excluded retracted publications. The final citation data for our two datasets were retrieved from WoS on the 20th and 21st of August 2020 and were then further augmented with the most up-to-date ESI field baselines for citation rates and percentile rank classes.[12]

For each PR3 article, considering its own citations and the citations achieved by all its peer publications from the same journal and year, we determine whether it is above or below the mean, whether it is above or below the median and whether or not it belongs to the top 10% most cited papers, a threshold typically associated with high impact. Although these benchmarks are very informative they have the disadvantage of ultimately sacrificing important information since they convert numerical data to a categorical, dichotomous assessment. We therefore also appeal to a more information-inclusive indicator—the CP-EX—very recently proposed in the ongoing discussion of percentiles as state-of-the-art instruments for cross-time and cross-field comparisons of bibliometric data. Starting from the citation count of a selected, focal paper, "CP-EX reveals exactly the percentage of papers with lower citation impact" (Bornmann & Williams, 2020, p. 1464). For example, in a reference set with 100 papers, supposing each to have distinct citation counts and sorting them from low to high, the paper on the 10th rank position would have a CP-EX of 9—the *cumulative percent* of papers with strictly fewer citations—, the one on the 50th rank a CP-EX of 49 and the one on the 90th rank a value of 89.[13] The CP-EX can take values between 0 and 100 (though technically never exactly 100) as it determines the cumulative frequencies, in percentages, of the papers that have fewer citations than a chosen focal paper. In our study the focal papers are, of course, PR3 papers.

Though CP-EX has been proposed for combinations of subject categories and publication years it is sufficiently flexible to be adapted in our investigation to combinations of journals and publication years. One complication we must address in this adaptation is that

[11] Note that this superset does not include all WoS-indexed publications from 2011–2015 since many of the thousands of indexed journals did not publish any PR3 article or were not eligible for submission in PR3.

[12] Our assignment of PR3 papers to ESI categories was based on the mapping of their publishing journals to the broad categories most recently updated in the June 2020 ESI master journal list. A limitation we must acknowledge is that papers assigned to the *multidisciplinary* category in our datasets are not necessarily assigned to this particular category in the ESI where information on the cited references is used to attribute the papers to the other categories. Of the 10,281 PR3 papers 229 have been attributed to the multidisciplinary category.

[13] Bornmann and Williams also propose the alternative CP-IN indicator which has a different interpretation: it represents the cumulative percentage of papers having a citation impact *lower than or equivalent to* a focal paper. For the three examples above the CP-IN values for the three focal papers would be exactly 10, 50 and 90. A disadvantage of CP-IN is that in reference sets with uncited items the papers with no citations would appear to have some sort of citation impact as a consequence of the *or equivalent to* clause. In CP-EX the percentile corresponding to the first empirical citation count in the reference distribution is 0 and this value can be expected to be overrepresented compared to the others.

when considering journals as reference sets it is likely to find journal-year combinations with a very small number of papers. For example, no fewer than 1495 articles from the 10,281 in our primary dataset were published in journals that, in the respective years, had less than 100 publications. Since we are reluctant to remove such a large share of publications but also recognize that undesirable numeric artefacts may emerge from very small sets we opt for a compromise: in our journal-level analysis we further restrict our attention to articles published in journals that had at least 50 publications in the year in which a PR3 paper was present. Since only 469 papers fail to meet this criterion we are left with a final subset of 9812 PR3 papers to analyze and a slightly diminished corresponding journal papers superset of 1,858,944 articles.

At the broader level of ESI categories (i.e. at our second level of analysis) we retain all the 10,281 papers since the problem of very small publishing journals is no longer relevant in a global field comparison. At this second analytical level we compute the ratio of each PR3 papers' citations to the worldwide citation rate in the corresponding category and year. This reveals whether, independent from the "local" performance in the context of its publishing journal, a PR3 paper performed better or worse than the world average.[14] Since high citation skewness and its effects on global averages are a cause for concern in this context we also determine the selected ESI percentile rank class to which each PR3 paper belongs to. This permits a more nuanced appraisal and is arguably more relevant for policy discussions.

The citations of PR3 articles relative to those of all articles from the publishing journals

Looking to our empirical data and considering the set or 9812 PR3 papers described above we find that since their publication they have accumulated a total of about 206,000 citations, while the approximately 1.9 million papers in the journals superset have amassed about 47.5 million citations. Only 35.5% (3480) of PR3 papers turn out to have a citation impact greater than the mean for their corresponding journal and year, 50% (4901) have an impact above the median and 10.6% (1035) belong in the top 10% of the most cited papers.

To make our journal-level analysis tractable and also offer a more meaningful context Table 2 presents the citations-based comparison of PR3 papers with all the similar papers in their relevant journals and years by further grouping the articles in the 22 broad fields of science (the ESI categories) which we later also use for the global analysis. In addition to the distribution of PR3 papers across the ESI categories the table also offers the distribution of the articles in the journal papers superset as well as total and per item citation counts. The last three columns show the share of PR3 papers within each field above the mean, above the median and within the top 10% in their distinct reference journals and publication years. We may note that three of the 22 broad fields—*chemistry*, *materials science* and *physics*—account for slightly more than half of PR3 papers and that by adding those from *engineering*, *mathematics* and *clinical medicine* more than ¾ of the 2011–2015 rewarded articles are accounted for. In contrast, in each of six other broad fields (for instance *immunology* and *microbiology*) we find fewer than 100 PR3 papers. It is not

[14] The averages for broad fields—and the percentile rank class thresholds—are calculated from all articles and reviews indexed in WoS and therefore reflect the complete citation performance in a broad field. Our journal superset data capture only a part of this performance.

Table 2 Distribution of PR3 papers and journals superset papers across broad fields of science (ESI categories), total citations and citations per item, percent share of PR3 papers within each field above the mean, above the median and within the top 10% most cited in the reference journal and year (rows are sorted by frequency of PR3 articles)

ESI category	Articles in superset	Citations to articles in superset	Citations per article in superset	PR3 articles	Citations to PR3 articles	Citations per PR3 article	% PR3 papers above journal mean	% PR3 papers above journal median	% PR3 papers in journal top 10%
Chemistry	438,908	10,522,124	23.97	2221	39,979	18.00	31.07	44.84	7.79
Materials science	205,935	5,705,947	27.71	1441	24,029	16.68	33.03	47.47	8.26
Physics	296,304	6,145,703	20.74	1371	23,119	16.86	31.80	49.09	9.04
Engineering	165,229	4,199,508	25.42	1016	24,581	24.19	36.52	53.35	11.32
Mathematics	64,533	545,301	8.45	996	9301	9.34	35.84	51.00	11.65
Clinical medicine	93,908	3,467,755	36.93	436	25,836	59.26	48.85	58.72	19.27
Geosciences	55,240	1,465,054	26.52	325	9375	28.85	43.38	58.15	11.69
Multidisciplinary	165,697	5,618,574	33.91	229	7177	31.34	37.12	51.09	12.66
Environment/ ecology	54,096	1,567,789	28.98	223	7470	33.50	44.39	55.61	19.73
Biology & biochemistry	66,135	1,979,444	29.93	214	5118	23.92	38.32	49.53	13.55
Social sciences, general	19,086	394,606	20.68	213	2176	10.22	29.11	44.13	7.98
Agricultural sciences	41,411	951,281	22.97	196	4433	22.62	44.90	57.65	12.24
Computer science	27,282	570,022	20.89	177	4080	23.05	40.11	55.93	12.43
Plant & animal science	28,976	570,437	19.69	169	3233	19.13	40.24	52.07	14.20
Pharmacology & toxicology	26,778	586,616	21.91	140	3454	24.67	47.14	57.86	12.86
Molecular biology & genetics	21,408	874,316	40.84	87	4429	50.91	47.13	59.77	14.94
Psychiatry/psychology	11,897	262,312	22.05	74	2715	36.69	45.95	64.86	25.68

Table 2 (continued)

ESI category	Articles in superset	Citations to articles in superset	Citations per article in superset	PR3 articles	Citations to PR3 articles	Citations per PR3 article	% PR3 papers above journal mean	% PR3 papers above journal median	% PR3 papers in journal top 10%
Space science	38,338	1,141,255	29.77	74	1265	17.09	12.16	25.68	4.05
Economics & business	1879	35,538	18.91	72	342	4.75	30.56	33.33	5.56
Neuroscience & behavior	14,589	446,189	30.58	50	2065	41.30	44.00	64.00	18.00
Microbiology	12,837	323,629	25.21	48	1315	27.40	52.08	66.67	12.50
Immunology	8478	187,425	22.11	40	1172	29.30	55.00	60.00	12.50
Column totals	1,858,944	47,560,825		9812	206,664				

surprising to find many papers in chemistry and physics since post-communist Romanian science inherited a disciplinary structure largely focused precisely on these two disciplines, similar to other Central and East European countries (Kozlowski et al., 1999). This inherited disciplinary focus also helps to explain the high prevalence of papers from *materials science*, *engineering* and *mathematics*.

Moving to the actual citation performance within journal-year reference sets, we note that within 20 of the 22 broad scientific fields the majority of PR3 papers (i.e. more than 50%) do not meet the average of their peer items; in particular, for the five categories that pool the most numerous PR3 papers (*chemistry*, *materials science*, *physics*, *engineering*, *mathematics*) only about a third of the rewarded papers exceed the average citation performance of the articles published in the same journal and year. In the case of *chemistry*, for example, of 2221 rewarded papers only 31% had an above-average performance compared to their journal peers. The highest share of PR3 papers with above-average citation counts (about 49%) is found in the case of *clinical medicine*. The outlook on citation impact improves if one considers the median thresholds which are less affected by the skewness of the distributions of journal citations[15]: in most of the 22 fields—though not for the dense *chemistry*, *materials science* and *physics*—more than half of PR3 papers achieve an above-median impact. With regard to the papers whose citation counts place them in the top 10% in their journal, noticeably higher shares than 10% are more commonly found in fields where PR3 papers are fewer in number. For example, a quarter of the 74 papers in the *psychiatry/psychology* category and a fifth of the 223 in *environment/ecology* meet the more demanding thresholds required to place them in the top 10% most cited articles in their journals. However, in the three best represented broad categories at most 9% of the papers are found in the top 10% articles.

To offer a comprehensive picture of the within-journal performance of all PR3 papers we present the values of the CP-EX indicator for the 9812 articles. Recall we use the indicator to show the cumulative percent of papers with strictly fewer citations than a focal PR3 paper of interest in a relevant journal-year combination. As CP-EX is a time and field (in our case journal) normalized indicator it permits comparison of all the different papers in our dataset. Figure 1 reveals an almost uniform distribution of the rewarded papers across the range of CP-EX values, save for an over-representation of articles in the 0 percentile and an under-representation of those in the 100th. On the whole, it seems that if one were to select a PR3 paper at random the empirical probability of it being, for example, better than 90% of its journal peers with regard to citation impact is essentially the same as the probability of it being better than only 10% of its journal peers.

As discussed in the previous paragraphs, PR3 papers are not evenly distributed across fields of science and we may expect variation of CP-EX values for each field. We show in Fig. 2 that there is indeed some noticeable between field variation especially for fields with fewer papers compared to those with more PR3 papers: for larger fields (*physics*, *mathematics*, *materials science*, *engineering* but less for *chemistry*) the distribution of CP-EX values is typically centered close to 50 (i.e. the median CP-EX value is about 50) and symmetric. For fields with fewer PR3 articles, although the papers also tend to span the entire range of citation impact the distribution is shifted away from the center, either towards

[15] A routine statistical analysis for the citation counts of all articles in all the 3971 journal-year combinations from the superset corresponding to the 9812 PR3 papers shows skewness coefficients above 1 for all but 28 cases, above 2 for all but 784 and above 3 for no fewer than 1978.

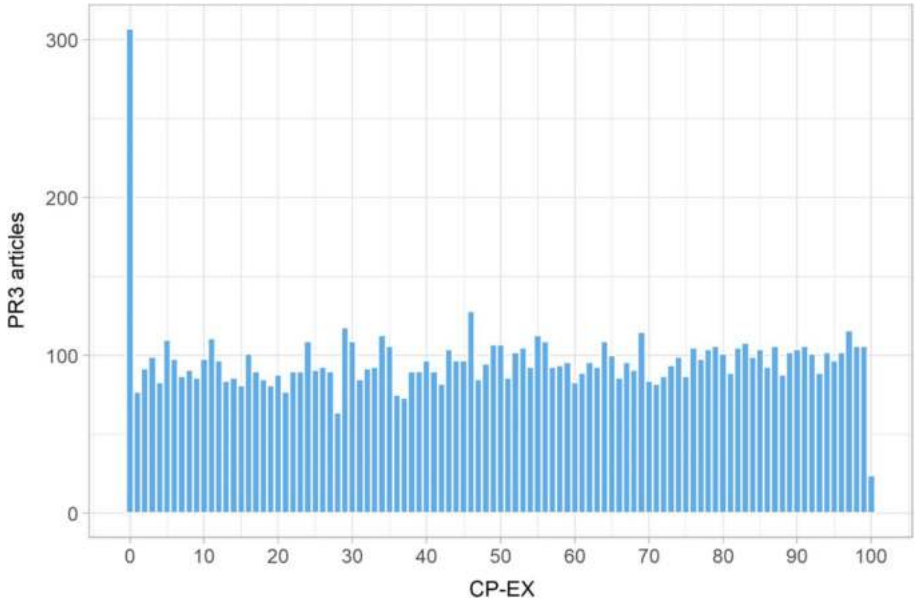

Fig. 1 Histogram of CP-EX values for 9812 PR3 articles (the bin width used in the figure is 1). All figures in the paper were created in R (R Core Team, 2020) with the *ggplot2* package (Wickham, 2016)

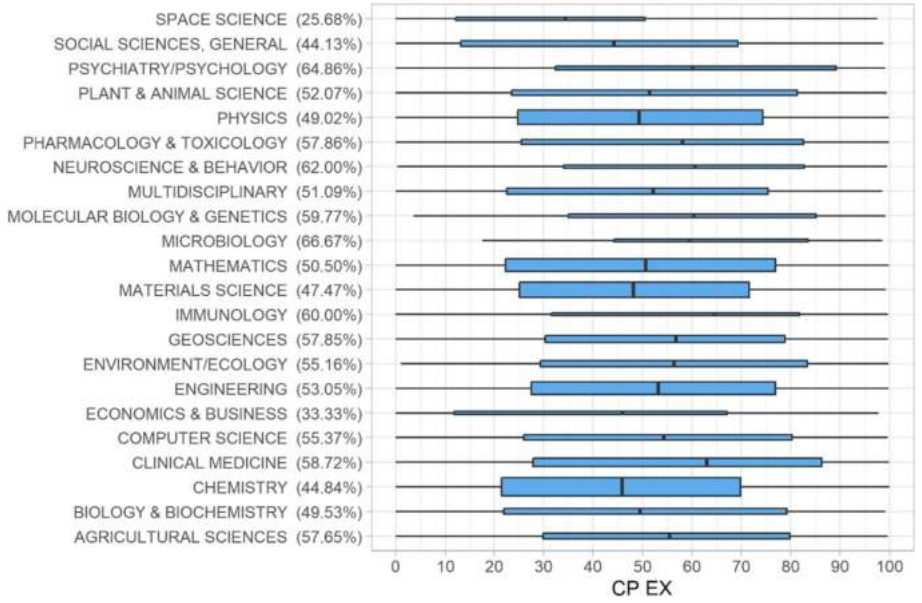

Fig. 2 Distribution of CP-EX values of PR3 articles within each broad field of science. The width of the boxplots reflects the number of papers in each field. Percentages in parentheses on the *y* axis indicate the share of papers with CP-EX over 50

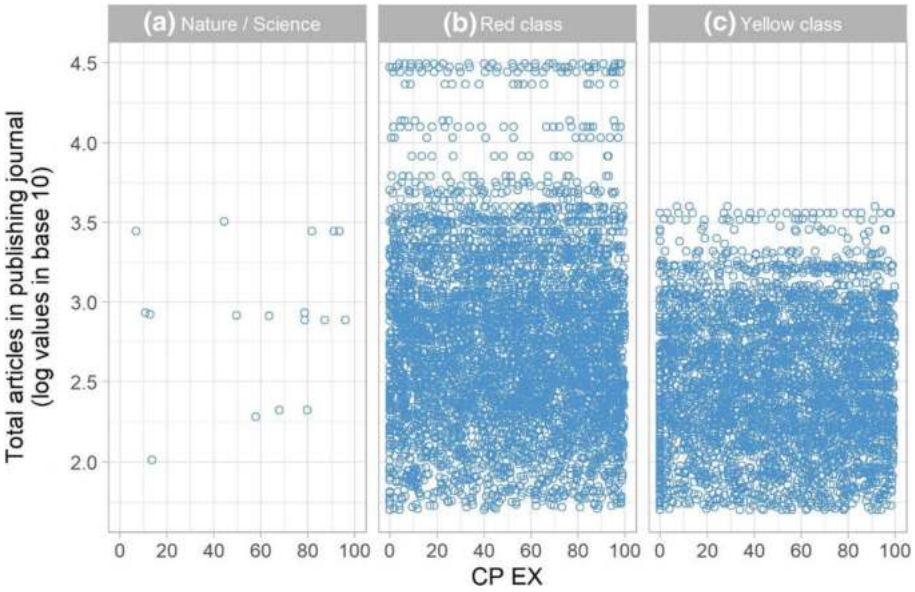

Fig. 3 Scatterplots of the CP-EX values of PR3 papers and the total number of articles in the publishing journals within each of the three main reward classes used in PR3

lower values (e.g.: *space sciences*, *economics & business*) or towards higher ones (e.g.: *psychiatry/psychology*, *microbiology*).

Finally, if one takes into account the number of papers published in the reference journal as well as the distinction between the main PR3 reward classes, a virtually uniform distribution of PR3 papers across the range of CP-EX values (similar to the findings in Fig. 1) seems to emerge once more. We document these aspects in Fig. 3 which also highlights the fact that PR3 articles in the red class (which represent about 56% of the total) are more numerous than those in the yellow class (about 44% of the total). Points further to the right in each subplot indicate higher citation performance and higher positions on the *y*-axis indicate larger journals. The figure clearly shows that there are many more red class PR3 articles in larger journals (more than 1000 items), whereas the articles from the yellow class typically appear in journals with fewer than 1000 publications in a given year.

The citation impact of PR3 articles within global scientific fields

The previous comparison of PR3 articles with all their similar journals articles provides a useful but nonetheless restricted image of citation performance since the journal articles supersets do not represent the entire, worldwide sets of comparable articles published in international journals indexed in WoS. However, citation performance in the global sets is captured in the ESI baselines described in previous paragraphs. By comparing the citation counts of PR3 articles to the global baselines in their broad scientific fields and by further determining their placement in the percentile rank classes in those fields we move beyond the confines of single journals and obtain a better representation of the worldwide citation impact of the rewarded articles.

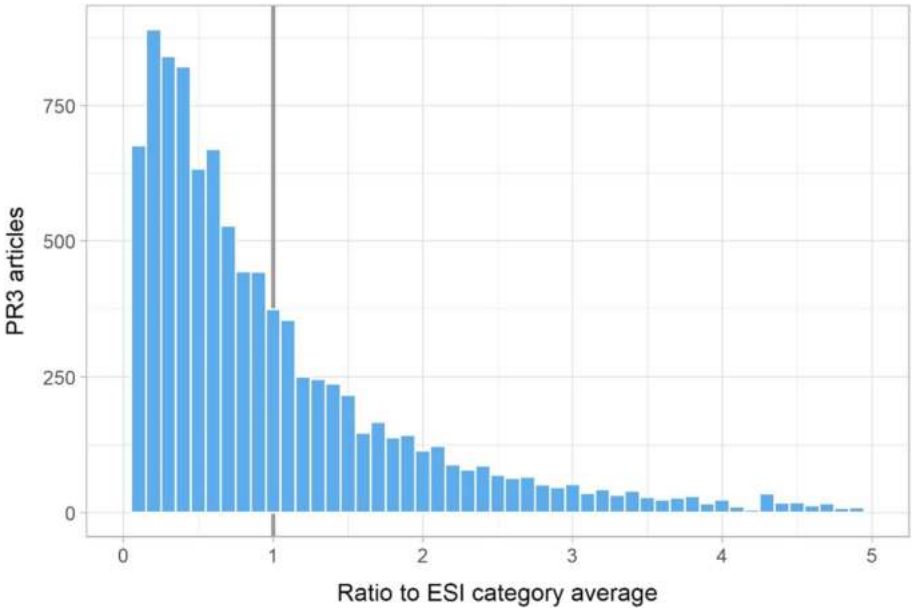

Fig. 4 General view of the distribution of PR3 articles by the ratio of their citations to the average in their broad field of science and publication year. The bin width used is 0.1 and the figure is truncated for legibility at 5; 359 sparsely distributed cases above 5 are not shown

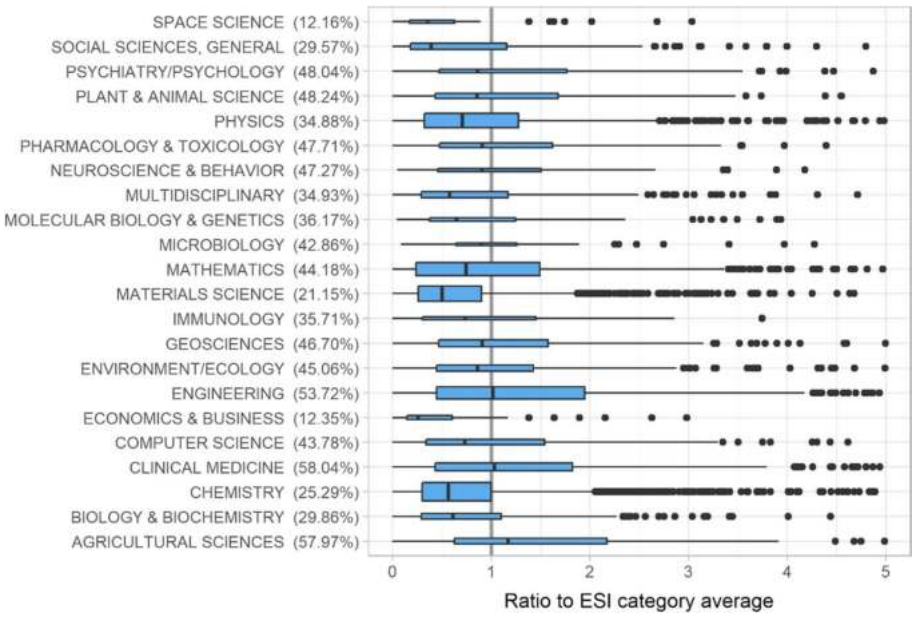

Fig. 5 Field view of the distribution of PR3 articles by the ratio of their citations to the average in their broad field of science. The *x* axis is truncated for legibility at 5 and does not show all outliers in each field. Percentages in parentheses on the *y* axis indicate the exact share of papers above the field average (and implicitly show the share below)

Figure 4 presents the overall performance of all PR3 papers relative to the global averages in their distinct fields. The figure represents the ratio of the PR3 articles' total citations to the average citations in their field and publication year and it reveals that the better part of the rewarded articles, 63.8% of the 10,281, have a below-average performance in a global context, whereas only the remaining 36.2% have an above-average performance. Variation from this overall result are of course common considering each distinct broad field and Fig. 5 documents this in more detail: in the fields with the highest shares of papers, *chemistry* and *materials science*, only about a quarter of PR3 articles have above-average performance, in *physics* only about a third and in mathematics 44%. In the only other field with more than 1000 PR3 articles, *engineering,* 53.72% of the articles have above-average impact. The highest shares of papers with above-average citation performance are in *agricultural sciences* (58% out of 207) and in *clinical medicine* (58% of 448 articles). In contrast, in some fields with few PR3 papers (*economics & business*, *space science*) there is an overwhelming share (close to 90%) of articles with below-average impact.

As a more comprehensive rendition of the global international impact of PR3 papers we present in Table 3 their distribution in each broad field across the ESI percentile rank classes. We present both the shares of PR3 papers from the six explicit classes as well as the share in the implicit rank class of papers with citation performance below the world median in the field (labelled "Lower 50%" in our table). This helps to create a general profile of citation performance for the rewarded papers in each field and is especially useful for identifying papers in the higher echelons. For example, the table shows that in *agricultural sciences*, *psychiatry/psychology*, *clinical medicine*, *computer science* and *engineering* more than 20% of PR3 papers are in the top 10% class. Furthermore, in *psychiatry/psychology* and in *clinical medicine* about 6% of the papers are in the top 1% class. With the exception of *mathematics*, *engineering*, *clinical medicine* and *pharmacology & toxicology* there are no PR papers in the top 0.1% in their field and there is only a single article in the top 0.01% rank class, in *clinical medicine*.

One additional aspect that deserves attention is whether or not visible differences in global impact may be found when considering the distinction between the main rewards classes used in PR3. Since the red class articles have been published in journals in the top quartile of (at least) a WoS subject category, while the yellow ones have been published in journals from the second quartile, we would expect to see contrasting distributions of the papers from these distinct reward groups across the ESI percentile rank classes. In particular, we would expect better performance from the papers in the red class. We present a comprehensive comparison of red and yellow PR3 articles in Fig. 6. Although in most of the 22 broad fields at least 20% of all the red class rewarded papers turn out to have a below world median impact, the papers in the red group are typically found above this threshold and, in many fields, a substantial share of them belong to the top percentile rank classes. For example, about 34% of the 137 PR3 papers in the red group in *agricultural sciences* belong to the top 10% of all papers published in this broad field. Similarly, between about 25% and 30% of the papers in the red group in *clinical medicine*, *computer science*, *engineering*, *environment/ecology*, *psychiatry/psychology* and *immunology* also belong to the top 10% in their fields. For the fields just mentioned higher shares of red papers in the top 1% class are also noticeable.

In contrast to PR3 papers in the red group, for those from the yellow group it is typical to see the greater part—i.e. between about 70% and 80% of the articles in each field—below the world median or in the lowest explicit percentile rank class of top 50% (which effectively means above the world median impact but not in the top 20%). Furthermore, it is rather uncommon to find yellow group papers in the top 10% or top 1% classes, although

Table 3 Distribution of PR3 articles (percent shares) across the ESI percentile rank classes determined by the worldwide citation thresholds corresponding to each broad field of science (rows are sorted by frequency of PR3 articles)

ESI category	PR3 articles	Lower 50%	Top 50%	Top 20%	Top 10%	Top 1%	Top 0.10%	Top 0.01%
Chemistry	2234	44.32	38.63	10.43	6.36	0.27		
Materials science	1461	44.56	42.51	8.69	4.18	0.07		
Physics	1376	35.10	39.17	15.99	9.01	0.73		
Mathematics	1168	34.25	32.71	16.01	14.98	1.88	0.17	
Engineering	1048	24.52	35.31	17.37	20.61	2.00	0.19	
Clinical medicine	448	25.00	26.79	17.63	22.99	6.03	1.34	0.22
Geosciences	349	28.94	38.40	15.76	13.47	3.44		
Social sciences, general	257	54.09	26.07	10.51	8.17	1.17		
Environment/ecology	233	31.76	36.91	11.59	16.31	3.43		
Multidisciplinary	229	31.00	41.92	10.48	14.41	2.18		
Biology & biochemistry	221	46.15	34.39	10.41	8.60	0.45		
Agricultural sciences	207	21.26	38.16	14.49	25.60	0.48		
Computer science	201	26.37	35.82	13.93	20.40	3.48		
Plant & animal science	199	32.66	31.66	17.59	17.09	1.01		
Pharmacology & toxicology	153	31.37	33.99	17.65	15.69	0.65	0.65	
Psychiatry/psychology	102	29.41	34.31	6.86	23.53	5.88		
Molecular biology & genetics	94	36.17	34.04	11.70	14.89	3.19		
Economics & business	81	72.84	20.99	3.70	2.47			
Space science	74	66.22	22.97	5.41	4.05	1.35		
Neuroscience & behavior	55	34.55	27.27	20.00	14.55	3.64		
Microbiology	49	20.41	57.14	6.12	16.33			
Immunology	42	35.71	33.33	16.67	11.90	2.38		
All PR3 articles	10,281	37.02	36.75	13.13	11.62	1.36	0.11	0.01

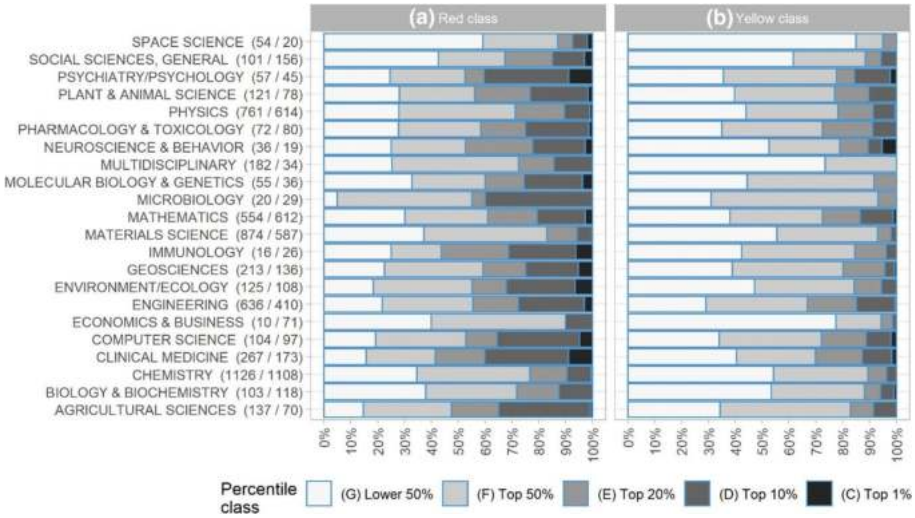

Fig. 6 Distribution of PR3 articles across the percentile rank classes for each broad field of science within the red and yellow PR3 reward groups. Numbers in parentheses on the *y* axis indicate the papers in the red/yellow reward groups (in this order) for each field. The very few papers (12) in the Top 0.1% and Top 0.01% classes are omitted for legibility, as are the 18 papers in the Nature/Science group which, with one exception, are all in the top 10% or top 1% ranking classes

some notable exceptions occur in a few fields (e.g.: *engineering*, *mathematics*, *psychiatry/psychology, neurosciences & behavior*).

A dual view of citation impact within journals and broad fields

As a corollary to the distinct appraisals of PR3 papers presented in the previous two sections a natural question to ask is how within-journal performance is connected to the overall citation performance in the global context of broad fields of science. As a final analysis we therefore present in Fig. 7 a combined view of the citation performance of PR3 articles within journals (as captured by the CP-EX indicator) and of their performance within global fields (captured by the ratio of their citations to broad field averages and by their placement in global percentile rank classes), separately for the red and yellow PR3 reward groups. We also present the results for the six broad fields with the most PR3 papers (discounting the distinction between red and yellow reward groups) in Fig. 8.

There is a positive overall association between the journal performance of articles and their wider performance in broad fields of science. In the yellow group most articles that outperform at least 85% of their journal peers are in the top 10% ranking class in their global field; for the red group papers that outperform at least 70% of their journal peers typically find themselves in the top 10% articles in their global field. Similarly, articles from the yellow group that have a citation performance that is better than that of at least 70% of papers in their journal tend to be present in the global top 20% in their field, while articles from the red group that outperform at least 50% of the articles in their journal tend to be present in the same global top 20% class for their category. Articles that have midrange CP-EX values (between about 30 and 70) are mostly found in the top 50% class in their field but are also common in the class with below world-median citation impact in

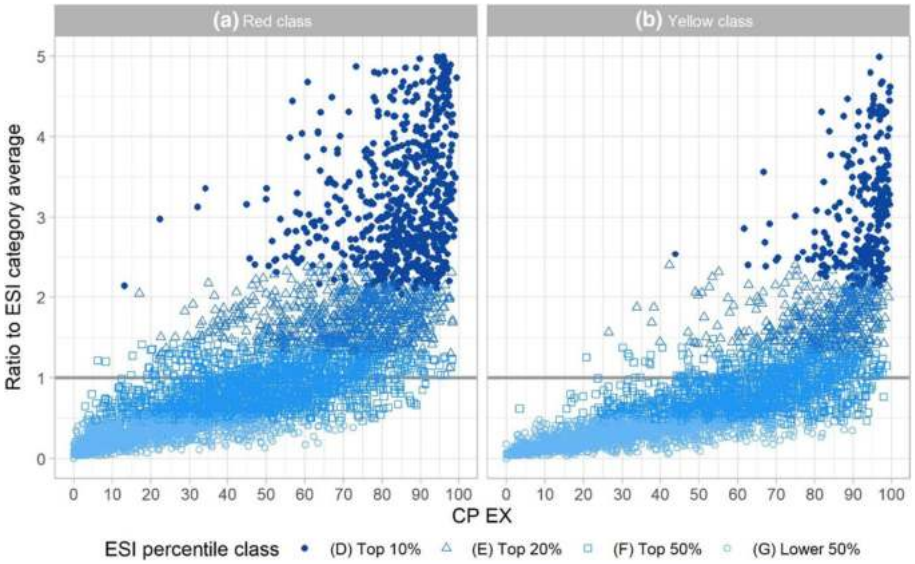

Fig. 7 The citation performance of 9812 PR3 articles within their journals plotted against their global performance in broad fields of science within the two main reward classes used in PR3. The *y* axis is truncated at 5 for legibility. Papers from the top 1%, 0.1% and 0.01% classes do not appear in the plot

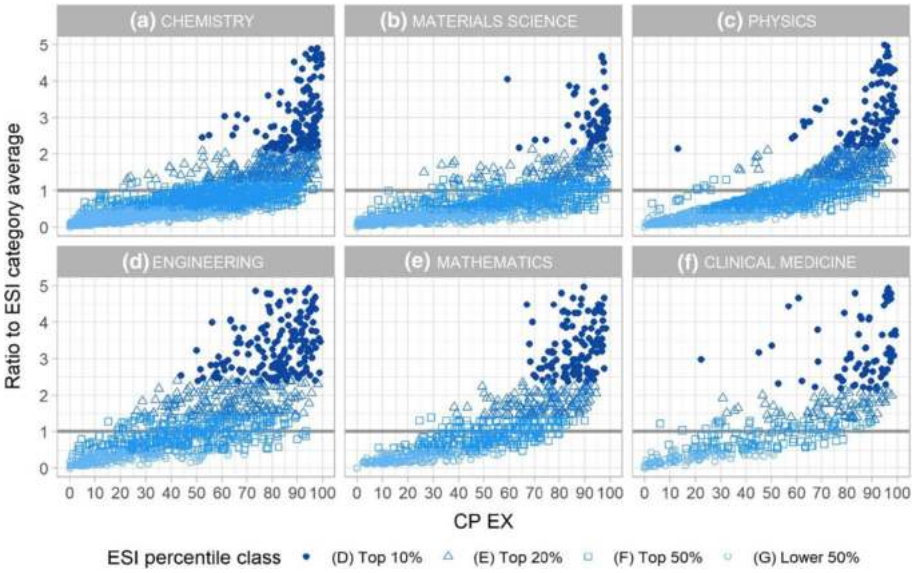

Fig. 8 The journal and global citation performance of PR3 articles for the six densest PR3 fields. The *y* axis is truncated at 5 for legibility. Papers from the top 1%, 0.1% and 0.01% classes do not appear in the plot

both PR3 reward groups. Finally, most of the PR3 articles that have a below median performance in their individual journals (CP-EX less than 50) also have a below median performance in their global fields (i.e. they belong to the lower 50% ranking class).

Although infrequent, there are some papers that perform well in their broad field but which seem very modest in the context of their journal. These are typically articles published in highly prestigious journals with a strong position in their field. For example, the article in the global top 10% and cited three times above its global field average in *clinical medicine* that has a CP-EX of only 22 (see Figs. 7a and 8f) is from the *New England Journal of Medicine*. Similar examples are found for articles published in *Lancet Oncology*, *Nature Genetics* or *Nature Physics*.

Discussion: a methodological critique of PR3 from a scientometric perspective

In the preceding sections we have presented PR3 and we have shown—based on an extensive empirical analysis of the citation counts of the individual papers—that the wholesale application of journal metrics to guide the rewards has not guaranteed for all papers the impact which the program means to support. Indeed, this could not have been reasonably expected, if one considers the breadth of empirical and theoretical knowledge that has accumulated in the field of scientometrics over the years. We will now comment a few selected issues that have greater salience, especially in relation to our own empirical analyses, thus offering a (necessarily partial) critique of the methodological underpinnings of PR3.

We must begin with the fundamental aspect which is at the very core of the PR3 process (but which also underlies output-oriented evaluation systems mentioned in previous sections), namely the questionable idea that journal-level metrics are adequate devices for the appraisal of individual articles and, by extension, of the scholars creating them. There is an abundant and still growing literature stretching across decades that specifically and repeatedly warns against this very practice, especially when JIFs are involved (Curry, 2018; Hicks et al., 2015; Larivière & Sugimoto, 2019; Larivière et al., 2016; Lozano et al., 2012; Moed, 2010; Osterloh & Frey, 2020; Pendlebury, 2009; Seglen, 1992, 1997; Wilsdon et al., 2015). Inter alia this literature—and much further work referred to therein—explicitly draws attention to the following essential fact often completely overlooked in the actual process of research evaluation policymaking: citation distributions, including those in distinct journals, are very skewed, with only a reduced number of papers receiving a disproportionately high number of citations, while the rest receive far less and many none at all. For instance, using data on the papers published in 2014 and 2015 by four biochemistry and molecular biology journals Larivière and Sugimoto (2019) noted that only a third obtain the number of citations described by the JIF values. In earlier work Larivière et al. (2016) found that even for the very influential *Nature* and *Science* about ¾ of the published papers achieve an actual citation impact *below* the JIF. The clear implication of these facts is that there is only a very weak relation between JIFs and actual citations of individual papers and that it is completely inadequate to ascribe to any single paper the average impact captured by the publishing journal's JIF. Such a practice has been argued (Leydesdorff et al., 2016) to represent a clear example of an ecological fallacy, i.e. an error which essentially consists in misattributing the property of a collective entity to its constituent parts. This fallacy is the very cornerstone of the PR3 methodology. Our own findings regarding the achieved citation impact of rewarded papers further confirm that evaluations based on journal metrics do not unequivocally and universally yield an increased impact of individual research publications.

The known skewness of citation distributions is the most important argument against the use of JIF-based evaluation—and, more generally, any journal metrics-based evaluation—for individual papers. There are many other widely-recognized deficiencies specifically attributed to the JIF, including[16] (1) the arbitrary restriction of the JIF citation window to a narrow period of only two years, a choice that favors rapidly moving fields especially in the sciences against those from the social sciences and humanities, (2) the inclusion of journal self-citations in the JIF calculation, a choice which creates opportunities for manipulation, (3) the equal consideration given to each citation, regardless of its origin. The article influence score or AIS (Bergstrom, 2007; Bergstrom et al., 2008), which is also typically used in PR3, is an improvement on the JIF as it addresses the three previous issues by considering a longer citation window of five years, by excluding self-citations and, most importantly, by weighting citations differently in accordance with the prestige of the citing journal. Prestige is essentially equated with network centrality: citations from larger, more central journals are weighed more than those from outlying journals. Though it is much more complicated than the JIF Franceschet (2010) offers some additional arguments in favor of the AIS, including some good axiomatic and mathematical foundations. However, in a much more analytical work Waltman and van Eck (2010) point out that the AIS is a parameterized indicator whose properties and outcomes can be fundamentally altered by the specific choice for the parameter (α) embedded in its calculations.

Without delving into technicalities we must stress the essential issue that the AIS is also an average, per-publication indicator describing journals. Therefore, using it as a substitute for the individual citation performance and impact of an article is, in principle, no more justified than the use of the JIF towards the same end. In essence the evaluation of an article in accordance with the overall AIS of its publishing journal is also an ecological fallacy. Though a considerably more sophisticated indicator is used, the fundamental problem is the same.

An additional topic that we must devote attention to due to its central importance in the PR3 process concerns the subject category assignment of journals and papers in WoS. As explained above, PR3 typically makes use of both the JIF and AIS quartile classification of journals within each WoS subject category to distinguish between the red (JCR Q1) and yellow (JCR Q2) regions corresponding to different rewards. The most favorable ranking by either indicator, in any category is taken into account. There are four aspects we wish to note in connection to this issue.

First, the WoS subject categories are known to be very problematic as a contextual reference for research evaluation, especially at low levels of aggregation, with publication-level classifications based on algorithmic procedures (e.g.: Perianes-Rodriguez & Ruiz-Castillo, 2017; Ruiz-Castillo & Waltman, 2015) increasingly advocated as an alternative. The essential problem is that the WoS categories are not always internally coherent, homogeneous reflections of scientific fields. The following remarks are quite revealing in this regard: "Journals are assigned to categories by subjective, heuristic methods. In many fields these categories are sufficient but in many areas of research these 'classifications' are crude" (Pudovkin & Garfield, 2002, p. 1113). A recent discussion of the problems plaguing the WoS subject categories is available in Milojević (2020) who mentions that a known

[16] We will not address all of them here as this would itself be a very lengthy undertaking. Bullet point accounts—each noting more than a dozen aspects—can be found, for example, in Pendlebury (2009, p. 3) and Seglen (1997, p. 499).

problem with the WoS system is its "erroneous lumping of unconnected journals into a single category" (p. 194).

A second remark we deem necessary in connection to the WoS category assignment of journals and papers starts from the fact that about 40% of all WoS-indexed journals are assigned to more than one category (Wang & Waltman, 2016) which means that all their corresponding articles also have non-unique assignment. In this context the PR3 methodological choices of allowing both JIF and AIS rankings and of considering the most favorable standing in any indicator–category combination artificially enlarge the scope of research excellence meant to be rewarded: more journals are effectively allowed to count as being in the top quartiles and authors publishing in journals with multiple category assignment have better chances of receiving rewards.

Third, work focusing on the JIF quartile classification of journals has found a very unequal share of actual publications across the quartile classes: Liu et al. (2016, p. 1273) note that "at least one-third of Web of Science publications are actually published in the first quartile (high impact factor journals)" and Miranda and Garcia-Carpintero (2019) show that on average across the largest 25 WoS categories in SCIE Q1 journals actually publish 38.4% of articles and reviews, while Q2 journals publish 27.5% of these items. Though variation is present across the fields, the fundamental empirical fact is that Q1 and Q2 publications are in fact the most typical publications in any WoS category. If the ultimate aim of PR3 is that of rewarding excellence—which, by definition, is *not* common—then using the WoS quartile classification is an inadequate methodological choice. Recall that our findings show that articles in the red reward group are the most numerous of PR3 papers (56%).

Finally, as we show in a very recent comprehensive analysis presented in a distinct work (Vîiu & Păunescu, 2021), the quartile classification of journals based on JIFs is problematic for research assessment at the micro level due to the non-meaningful impact differences which can separate journals from different quartile classes. In the context of PR3 this means that the differences underlying the yellow/red class distinction might not always justify the disparity between the rewards corresponding to the two groups.

There are additional issues worthy of critical reflection in the PR3 regulations for the allocation of monetary rewards which we do not touch upon in detail as they manifest in specific fields or journals and cannot be generalized to the entire set of articles. Nonetheless, noting that the *per-publication, per-Romanian author* approach stipulated in PR3 disregards important disciplinary differences, for instance discipline-specific productivity and collaborative practices, we point out the complex issue of multiple authorship and proper credit allocation between authors. This is very problematic in some fields, especially in intensively collaborative research areas such as high-energy physics and clinical medicine. Thelwall (2020) has recently shown that collaborative publishing consortia usually yield research with very significant citation impact but has also argued that the large number of authors can make individual credit assignment very difficult. On the other hand, in high-energy and particle physics "all participants are listed as authors as a mark of membership of the team, not for writing or revising the papers" (Ioannidis et al., 2018, p. 167). Since in the case of papers with hundreds of authors "one wonders if these reports are really attributable to any scientists or any institutions" (Pendlebury, 2009, p. 7) we must also wonder if a single Romanian author in such a group should be entitled the full PR3 reward for each collaborative paper. Our data show for instance 387 papers that have at least ten authors in total but a single Romanian one, including 32 produced in the large CMS Collaboration at CERN and three in the *Nature/Science* reward category with more than 60 authors each. We would debate that these papers substantively reflect *Romanian* research.

Furthermore, the true contribution to the core of such works is ultimately only know to the authors themselves. These considerations also apply in circumstances in which supervisors act as coauthors for young scholars who "lack the negotiating power and freedom to influence decisions about co-authorship" (Cutas & Shaw, 2015, p. 1322): there is no guarantee of the supervisors' substantive contribution. More generally, it is reasonable to presume that unethical publication practices such as gift or ghost authorship are strongly encouraged when specific financial rewards are involved.

The previous points raise the larger issue of the overall advantages and disadvantages of rewarding individual researchers for their research output. The advantages include clear incentives for quality research articles (as these are typically found in prestigious journals) and increased objectivity and procedural efficiency in assessment (in contrast to subjective and time-consuming peer review). On the other hand, we believe that cash-per-publication systems like PR3 are likely to engender unintended deleterious outcomes similar to (and, probably more so than) general performance based funding systems for research. These include (Geuna & Martin, 2003): discouraging innovative research in favor of mainstream research, publication inflation or "salami publication", incentivizing research at the expense of teaching, reinforcing the status quo in which current research elites accrue the resources needed by others to achieve their own potential. Aside from these issues it is clear from the aforementioned PR3 regulations that collegiality is also inherently undermined, at least (ironically) among Romanian scholars. A purely rational researcher seeking to maximize his/her rewards has every incentive to publish single-author papers or to selectively engage in publication with coauthors outside of Romanian institutions who would therefore not count toward the division of rewards. In this latter case there are obvious reasons to debate the fairness and morality of fully rewarding only one or some Romanian individual researchers for the joint results produced in an international collaboration.

Summary and concluding remarks

In this paper we have presented the Romanian Program for Rewarding Research Results, a national cash-per-publication policy directed at individual scholars now implemented for 14 consecutive years. To measure the impact of papers (rewarded in the program based solely on journal metrics) we have studied the actual citation performance of about 10,000 articles from five consecutive PR3 editions. We have studied the citations of PR3 articles relative to all the articles from their relevant publishing journals but also relative to world citation standards in broad scientific fields. We have found that six broad fields—*chemistry*, *materials science*, *physics*, *engineering*, *mathematics* and *clinical medicine*—account for almost 80% of the rewarded articles and that more papers were attributed to the higher (red) reward class than to the lower (yellow) one, possibly as a consequence of the inclusive regulations that mandate the use of the most favorable ranking in any combination of journal indicator and WoS subject category. We have shown that rewards based on journal metrics do not seem to increase the impact of research.

The analysis of the citation impact of PR3 papers *relative to their publishing journals* offers a kaleidoscopic picture. If one looks at the overall population of PR3 papers they are almost similarly distributed across the various citation percentiles as the larger population of all articles published in the same journals and years. Thus, with regard to their citation impact, one can consider PR3 papers to be a representative sample of the whole population of articles in the publishing journals. PR3 papers are not essentially better or worse than

the population they are extracted from. They also exhibit the same skewness as the larger reference population and this leads to a very important conclusion policymakers in particular should contemplate: PR3 has essentially offered the same reward for articles with markedly different impact. For instance, the rough division of PR3 reward groups notwithstanding, articles with a CP-EX close to 100 made their authors eligible for the exact same reward as articles with a CP-EX close to 0. On the other hand, if one looks at PR3 papers relative to their publishing journals but grouped by ESI categories, then there is considerable variation with regard to the articles' impact between these categories. For example, psychology, ecology and clinical medicine appear to be fields where Romanian PR3 articles fare considerably better than their counterparts in the same publishing journals. In contrast, PR3 articles in chemistry, space science or economics seem to have a weaker impact compared to their journal peers.

Moving beyond the reference set of journals and considering the position *within global broad fields of science* we find that the performance of PR3 articles is quite mixed: there are fields where the Romanian PR3 articles manage to pass the higher benchmark levels (top 20%, 10%, 1%, 0.1%) in significantly higher proportions but also fields where a significantly lower proportion of PR3 articles is found in each class. However, just as it is insensitive to the "local", within-journal performance, the PR3 reward mechanism is also indifferent to the merits of the articles in the global fields of science: within the red or yellow reward groups a worldwide top 1% article is valued identically to one that is not even in the top 50% class. In the aggregated PR3 set about 13% of the articles are in the worldwide top 10% class—which is now routinely acknowledged in scientometrics as representing highly cited research—or even in the higher classes. This overall positive share of top performing papers masks the important variation among the articles in the main scientific fields but it makes clear that PR3 has overwhelmingly rewarded articles characterized by a rather commonplace performance relative to worldwide benchmarks. Recently Leydesdorff et al. (2019) have suggested that top 1% and top 10% papers be considered excellence indicators, while top 50% and bottom 50% be considered output indicators. By the top 10% token, PR3 may be viewed as having mainly rewarded a certain type of productivity, rather than only research excellence.

The preceding remarks seem to indicate that article-level citation-based appraisal may be preferable to the current journal-based PR3 methodology. However, citations to articles may also not be universally appropriate proxies for impact. Profound and credible objections against the use of citations for evaluating individual articles have existed for a long time, formulated for example from a social constructivist perspective on science (MacRoberts & MacRoberts, 1996). Without delving into the subtle limitations of citation analysis we wish to point out that there is a more pragmatic restriction of citations which manifests especially in the case of very recently published research: the reliability of citations and their viability as a proxy for impact require the passage of considerable spans of time. This creates an important weakness for the article citations approach (and a corresponding strength of the journal-based approach) in the arena of science policy because practical decisions cannot typically afford to wait years upon years for the proper circulation and maturation of scholarly records, visible in multi-annual citation histories. Real-time "actionable" information, although not at the ideal level of detail, will likely be preferred by decision makers for practical matters of policy concerning current research. Even though a *retrospective* assessments based on article citations may be fairer, it comes at the cost of infeasible time lags, unlike *prospective* evaluations based on journals. An additional advantage of the latter approach is that it nonetheless typically incentivizes scholars to submit their work to high quality outlets. These points make the overall PR3 philosophy

defendable. They do not, however, legitimize all of the particular methodological choices of the program.

With regard to the operation of PR3 a break from inertia—if not a comprehensive revision and/or discontinuation—seems appropriate. At the very least the information package should be amended to acknowledge that the aim of the program is to increase productivity in top international outlets, rather than to reward research impact. Our findings clearly show that rewarding impact is not what the program has achieved. In addition, assuming the likely further recourse to journal metrics on grounds of convenience and efficiency in future PR3 editions, it might be preferable to resort only to metrics that encompass longer journal performance histories that signal the continued relevance and maturity of a journal within a field. The AIS relies on five years of data and also has additional strengths relative to the standard two-year JIF; unfortunately, its complexity makes it an irreproducible black-box and this might inherently undermine its trust and acceptance. The current discriminatory fractional approach to the allocation of rewards should also be revised: the fact that it completely ignores the contribution of (sometimes many) non-Romanian coauthors in the case of internationally-produced multi-author papers is morally dubious, to say the least. A simple remedy would be to consistently apply the same provisions directed towards papers authored exclusively by Romanian scholars, i.e. divide the reward by the full number of authors and grant the Romanian one(s) only the resulting fraction of a reward.

We conclude by reiterating that taking the publishing journal as the sole reference for appraising a publication's impact is highly problematic. As one shifts the reference from the journal to the broader scientific field, the lens of the kaleidoscope shows a different picture of the same article's influence. While citation counts change in time and any form of citation analysis offers only a snapshot of performance at a certain moment, we believe the main trends uncovered in this paper are unlikely to change fundamentally. Our results ultimately show that direct monetary incentives for articles may support productivity but they cannot guarantee impact. Administrators and policymakers employing journal-level metrics for research evaluation should consider more nuanced forms of appraisal if their intention is to reward excellence. Important aspects of such an appraisal might include the collaborative nature of research in some scientific fields and the career stage of the authors making the submissions. Furthermore, the appraisal should involve peer review rather than rely exclusively on simple administrative checks. Decision makers should ultimately be mindful that any form of citation impact (even at the level of individual articles considered in relevant reference sets) is only a facet of the wider array of influences that research articles can have in academia and in society.

Acknowledgements Anonymous reviewer comments that helped to clarify and improve aspects of the initial manuscript are gratefully acknowledged by the authors. This paper was financially supported by the Human Capital Operational Program 2014-2020, co-financed by the European Social Fund, under the project POCU/380/6/13/124708 No. 37141/23.05.2019 with the title "Researcher-Entrepreneur on Labour Market in the Fields of Intelligent Specialization (CERT-ANTREP)", coordinated by the National University of Political Studies and Public Administration.

References

Aagaard, K. (2015). How incentives trickle down: Local use of a national bibliometric indicator system. *Science and Public Policy, 42*(5), 725–737. https://doi.org/10.1093/scipol/scu087

Aboal, D., & Tacsir, E. (2017). The impact of subsidies on researcher's productivity: Evidence from a developing country. *Research Evaluation, 26*(4), 269–283. https://doi.org/10.1093/reseval/rvx031

Abritis, A., & McCook, A. (2017). Cash incentives for papers go global. *Science, 357*(6351), 541–541. https://doi.org/10.1126/science.357.6351.541

Albarrán, P., Crespo, J. A., Ortuño, I., & Ruiz-Castillo, J. (2011). The skewness of science in 219 subfields and a number of aggregates. *Scientometrics, 88*(2), 385–397. https://doi.org/10.1007/s11192-011-0407-9

Bak, H.-J., & Kim, D. H. (2019). The unintended consequences of performance-based incentives on inequality in scientists' research performance. *Science and Public Policy, 46*(2), 219–231. https://doi.org/10.1093/scipol/scy052

Bergstrom, C. (2007). Eigenfactor: Measuring the value and prestige of scholarly journals. *College & Research Libraries News, 68*(5), 314–316. https://doi.org/10.5860/crln.68.5.7804

Bergstrom, C., West, J., & Wiseman, M. (2008). The EigenfactorTM metrics. *Journal of Neuroscience, 28*(45), 11433–11434. https://doi.org/10.1523/JNEUROSCI.0003-08.2008

Bornmann, L., & Pudovkin, A. I. (2017). The journal impact factor should not be discarded. *Journal of Korean Medical Science, 32*(2), 180–182. https://doi.org/10.3346/jkms.2017.32.2.180

Bornmann, L., & Williams, R. (2020). An evaluation of percentile measures of citation impact, and a proposal for making them better. *Scientometrics, 124*(2), 1457–1478. https://doi.org/10.1007/s11192-020-03512-7

Bornmann, L., Leydesdorff, L., & Mutz, R. (2013). The use of percentiles and percentile rank classes in the analysis of bibliometric data: Opportunities and limits. *Journal of Informetrics, 7*(1), 158–165. https://doi.org/10.1016/j.joi.2012.10.001

Braun, T., & Glänzel, W. (1996). International collaboration: Will it be keeping alive East European research? *Scientometrics, 36*(2), 247–254. https://doi.org/10.1007/BF02017317

Clarivate Analytics. (2018). In: Cites indicators handbook. http://help.incites.clarivate.com/inCites2Live/8980-TRS/version/default/part/AttachmentData/data/InCites-Indicators-Handbook-June2018.pdf

Cleere, L., & Ma, L. (2018). A local adaptation in an output-based research support scheme (OBRSS) at University College Dublin. *Journal of Data and Information Science, 3*(4), 74–84. https://doi.org/10.2478/jdis-2018-0022

Curry, S. (2018). Let's move beyond the rhetoric: it's time to change how we judge research. *Nature, 554*(7691), 147–147. https://doi.org/10.1038/d41586-018-01642-w

Cutas, D., & Shaw, D. (2015). Writers blocked: On the wrongs of research co-authorship and some possible strategies for improvement. *Science and Engineering Ethics, 21*(5), 1315–1329. https://doi.org/10.1007/s11948-014-9606-0

Demir, S. B. (2018). Pros and cons of the new financial support policy for Turkish researchers. *Scientometrics, 116*(3), 2053–2068. https://doi.org/10.1007/s11192-018-2833-4

European Commission. (2018). Science, research and innovation performance of the Eu 2018. Strengthening the foundations for Europe's future. *Science, Research and Innovation Performance of the EU*. https://doi.org/10.2777/14136

Franceschet, M. (2010). Ten good reasons to use the Eigenfactor^{TM} metrics. *Information Processing & Management, 46*(5), 555–558. https://doi.org/10.1016/j.ipm.2010.01.001

Franzoni, C., Scellato, G., & Stephan, P. (2011). Changing incentives to publish. *Science, 333*(6043), 702–703. https://doi.org/10.1126/science.1197286

Geuna, A., & Martin, B. (2003). University research evaluation and funding: an international comparison. *Minerva, 41,* 277–304.

Gingras, Y. (2016). *Bibliometrics and research evaluation : uses and abuses*. The MIT Press.

Good, B., Vermeulen, N., Tiefenthaler, B., & Arnold, E. (2015). Counting quality? The Czech performance-based research funding system. *Research Evaluation, 24*(2), 91–105. https://doi.org/10.1093/reseval/rvu035

Hammarfelt, B., & de Rijcke, S. (2015). Accountability in context: effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University. *Research Evaluation, 24*(1), 63–77. https://doi.org/10.1093/reseval/rvu029

Hedding, D. W. (2019). Payouts push professors towards predatory journals. *Nature, 565*(7739), 267–267. https://doi.org/10.1038/d41586-019-00120-1

Heywood, J. S., Wei, X., & Ye, G. (2011). Piece rates for professors. *Economics Letters, 113*(3), 285–287. https://doi.org/10.1016/j.econlet.2011.08.005

Hicks, D. (2012). Performance-based university research funding systems. *Research Policy, 41*(2), 251–261. https://doi.org/10.1016/j.respol.2011.09.007

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature, 520*(7548), 9–11. https://doi.org/10.1038/520429a

Ioannidis, J. P. A., Klavans, R., & Boyack, K. W. (2018). Thousands of scientists publish a paper every five days. *Nature, 561*(7722), 167–169. https://doi.org/10.1038/d41586-018-06185-8

Jiménez-Contreras, E., de Moya Anegón, F., & López-Cózar, E. D. (2003). The evolution of research activity in Spain. *Research Policy, 32*(1), 123–142. https://doi.org/10.1016/S0048-7333(02)00008-2

Kim, D. H., & Bak, H.-J. (2016). How do scientists respond to performance-based incentives? Evidence from South Korea. *International Public Management Journal, 19*(1), 31–52. https://doi.org/10.1080/10967494.2015.1032460

Korytkowski, P., & Kulczycki, E. (2019). Examining how country-level science policy shapes publication patterns: The case of Poland. *Scientometrics, 119*(3), 1519–1543. https://doi.org/10.1007/s11192-019-03092-1

Kozak, M., Bornmann, L., & Leydesdorff, L. (2014). How have the Eastern European countries of the former Warsaw Pact developed since 1990? A bibliometric study. *Scientometrics, 102*(2), 1101–1117. https://doi.org/10.1007/s11192-014-1439-8

Kozlowski, J., Radosevic, S., & Ircha, D. (1999). History matters: The inherited disciplinary structure of the post-communist science in countries of central and eastern Europe and its restructuring. *Scientometrics, 45*(1), 137–166. https://doi.org/10.1007/BF02458473

Kulczycki, E. (2017). Assessing publications through a bibliometric indicator: The case of comprehensive evaluation of scientific units in Poland. *Research Evaluation, 26*(1), 1–12. https://doi.org/10.1093/reseval/rvw023

Larivière, V., & Sugimoto, C. R. (2019). The journal impact factor: A brief history, critique, and discussion of adverse effects. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 3–24). Cham: Springer. https://doi.org/10.1007/978-3-030-02511-3_1

Larivière, V., Kiermer, V., MacCallum, C. J., McNutt, M., Patterson, M., Pulverer, B., et al. (2016). A simple proposal for the publication of journal citation distributions. *bioRxiv*. https://doi.org/10.1101/062109

Leydesdorff, L., Bornmann, L., & Adams, J. (2019). The integrated impact indicator revisited (I3*): A non-parametric alternative to the journal impact factor. *Scientometrics, 119*(3), 1669–1694. https://doi.org/10.1007/s11192-019-03099-8

Leydesdorff, L., Wouters, P., & Bornmann, L. (2016). Professional and citizen bibliometrics: Complementarities and ambivalences in the development and use. *Scientometrics, 109*(3), 2129–2150. https://doi.org/10.1007/s11192-016-2150-8

Liu, W., Hu, G., & Gu, M. (2016). The probability of publishing in first-quartile journals. *Scientometrics, 106*(3), 1273–1276. https://doi.org/10.1007/s11192-015-1821-1

Lozano, G. A., Larivière, V., & Gingras, Y. (2012). The weakening relationship between the impact factor and papers' citations in the digital age. *Journal of the American Society for Information Science and Technology, 63*(11), 2140–2145. https://doi.org/10.1002/asi.22731

Ma, L. (2019). Money, morale, and motivation: A study of the output-based research support scheme in University College Dublin. *Research Evaluation, 28*(4), 304–312. https://doi.org/10.1093/reseval/rvz017

MacRoberts, M. H., & MacRoberts, B. R. (1996). Problems of citation analysis. *Scientometrics, 36*(3), 435–444. https://doi.org/10.1007/BF02129604

Milojević, S. (2020). Practical method to reclassify Web of Science articles into unique subject categories and broad disciplines. *Quantitative Science Studies, 1*(1), 183–206. https://doi.org/10.1162/qss_a_00014

Miranda, R., & Garcia-Carpintero, E. (2019). Comparison of the share of documents and citations from different quartile journals in 25 research areas. *Scientometrics, 121*(1), 479–501. https://doi.org/10.1007/s11192-019-03210-z

Miroiu, A., & Vlăsceanu, L. (2012). Relating quality and funding: The Romanian case. In A. Curaj, P. Scott, L. Vlăsceanu, & L. Wilson (Eds.), *European higher education at the crossroads* (pp. 791–807). Dordrecht: Springer, Netherlands. https://doi.org/10.1007/978-94-007-3937-6_41

Moed, H. F. (2007). The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Science and Public Policy, 34*(8), 575–583. https://doi.org/10.3152/030234207X255179

Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics, 4*(3), 265–277. https://doi.org/10.1016/j.joi.2010.01.002

Müller, R., & de Rijcke, S. (2017). Thinking with indicators. Exploring the epistemic impacts of academic performance indicators in the life sciences. *Research Evaluation, 26*(3), 157–168. https://doi.org/10.1093/reseval/rvx023

Must, Ü. (2006). "New"countries in Europe—Research, development and innovation strategies versus bibliometric data. *Scientometrics, 66*(2), 241–248. https://doi.org/10.1007/s11192-006-0016-1

Neff, M. W. (2018). Publication incentives undermine the utility of science: Ecological research in Mexico. *Science and Public Policy, 45*(2), 191–201. https://doi.org/10.1093/scipol/scx054

Osterloh, M., & Frey, B. S. (2020). How to avoid borrowed plumes in academia. *Research Policy, 49*(1), 103831. https://doi.org/10.1016/j.respol.2019.103831

Osuna, C., Cruz-Castro, L., & Sanz-Menéndez, L. (2011). Overturning some assumptions about the effects of evaluation systems on publication performance. *Scientometrics, 86*(3), 575–592. https://doi.org/10.1007/s11192-010-0312-7

Pajić, D. (2015). Globalization of the social sciences in Eastern Europe: genuine breakthrough or a slippery slope of the research evaluation practice? *Scientometrics, 102*(3), 2131–2150. https://doi.org/10.1007/s11192-014-1510-5

Pendlebury, D. A. (2009). The use and misuse of journal metrics and other citation indicators. *Archivum Immunologiae et Therapiae Experimentalis, 57*(1), 1–11. https://doi.org/10.1007/s00005-009-0008-y

Perianes-Rodriguez, A., & Ruiz-Castillo, J. (2017). A comparison of the web of science and publication-level classification systems of science. *Journal of Informetrics, 11*(1), 32–45. https://doi.org/10.1016/j.joi.2016.10.007

Pisár, P., & Šipikal, M. (2017). Negative effects of performance based funding of universities: The case of Slovakia. *NISPAcee Journal of Public Administration and Policy, 10*(2), 171–189. https://doi.org/10.1515/nispa-2017-0017

Pisár, P., Šipikal, M., Jahoda, R., & Špaček, D. (2019). Performance based funding of universities: Czech Republic and Slovakia. In M. S. de Vries, J. Nemec, & D. Špaček (Eds.), *Performance-based budgeting in the public sector* (pp. 237–254). Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-02077-4_13

Pudovkin, A. I., & Garfield, E. (2002). Algorithmic procedure for finding semantically related journals. *Journal of the American Society for Information Science and Technology, 53*(13), 1113–1119. https://doi.org/10.1002/asi.10153

Quan, W., Chen, B., & Shu, F. (2017). Publish or impoverish. *Aslib Journal of Information Management, 69*(5), 486–502. https://doi.org/10.1108/AJIM-01-2017-0014

R Core Team. (2020). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.r-project.org/.

Ruiz-Castillo, J., & Costas, R. (2018). Individual and field citation distributions in 29 broad scientific fields. *Journal of Informetrics, 12*(3), 868–892. https://doi.org/10.1016/j.joi.2018.07.002

Ruiz-Castillo, J., & Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. *Journal of Informetrics, 9*(1), 102–117. https://doi.org/10.1016/j.joi.2014.11.010

Sandoval-Romero, V., & Larivière, V. (2020). The national system of researchers in Mexico: Implications of publication incentives for researchers in social sciences. *Scientometrics, 122*(1), 99–126. https://doi.org/10.1007/s11192-019-03285-8

Sandström, U., & Van den Besselaar, P. (2018). Funding, evaluation, and the performance of national research systems. *Journal of Informetrics, 12*(1), 365–384. https://doi.org/10.1016/j.joi.2018.01.007

Schneider, J. W. (2009). An outline of the bibliometric indicator used for performance-based funding of research institutions in Norway. *European Political Science, 8*(3), 364–378. https://doi.org/10.1057/eps.2009.19

Schubert, A., & Braun, T. (1996). Cross-field normalization of scientometric indicators. *Scientometrics, 36*(3), 311–324. https://doi.org/10.1007/BF02129597

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science, 43*(9), 628–638. https://doi.org/10.1002/(SICI)1097-4571(199210)43:9%3c628::AID-ASI5%3e3.0.CO;2-0

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ, 314*(7079), 497–497. https://doi.org/10.1136/bmj.314.7079.497

Shu, F., Quan, W., Chen, B., Qiu, J., Sugimoto, C. R., & Larivière, V. (2020). The role of Web of Science publications in China's tenure system. *Scientometrics, 122*(3), 1683–1695. https://doi.org/10.1007/s11192-019-03339-x

Sivertsen, G. (2016). Publication-based funding: The Norwegian model. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research assessment in the humanities* (pp. 79–90). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-29016-4_7

Teodorescu, D., & Andrei, T. (2011). The growth of international collaboration in East European scholarly communities: A bibliometric analysis of journal articles published between 1989 and 2009. *Scientometrics, 89*(2), 711–722. https://doi.org/10.1007/s11192-011-0466-y

Thelwall, M. (2020). Large publishing consortia produce higher citation impact research but coauthor contributions are hard to evaluate. *Quantitative Science Studies, 1*(1), 290–302. https://doi.org/10.1162/qss_a_00003

Tonta, Y. (2018). Does monetary support increase the number of scientific papers? An interrupted time series analysis. *Journal of Data and Information Science, 3*(1), 19–39. https://doi.org/10.2478/jdis-2018-0002

Tonta, Y., & Akbulut, M. (2020). Does monetary support increase citation impact of scholarly papers? *Scientometrics, 125*(2), 1617–1641. https://doi.org/10.1007/s11192-020-03688-y

Trow, M. (1994). Managerialism and the academic profession: The case of England. *Higher Education Policy, 7*(2), 11–18. https://doi.org/10.1057/hep.1994.13

Vanecek, J. (2014). The effect of performance-based research funding on output of R & D results in the Czech Republic. *Scientometrics, 98*(1), 657–681. https://doi.org/10.1007/s11192-013-1061-1

Vîiu, G.-A., Păunescu, M., & Miroiu, A. (2016). Research-driven classification and ranking in higher education: An empirical appraisal of a Romanian policy experience. *Scientometrics, 107*(2), 785–805. https://doi.org/10.1007/s11192-016-1860-2

Vîiu, G.-A., & Păunescu, M. (2021). The lack of meaningful boundary differences between journal impact factor quartiles undermines their independent use in research evaluation. *Scientometrics*. https://doi.org/10.1007/s11192-020-03801-1

Vinkler, P. (2008). Correlation between the structure of scientific research, scientometric indicators and GDP in EU and non-EU countries. *Scientometrics, 74*(2), 237–254. https://doi.org/10.1007/s11192-008-0215-z

Vlăsceanu, L., & Hâncean, M.-G. (2015). Policy incentives and research productivity in the Romanian higher education. An institutional approach. In A. Curaj, L. Matei, R. Pricopie, J. Salmi, & P. Scott (Eds.), *The European higher education area* (pp. 185–203). Springer International Publishing. https://doi.org/10.1007/978-3-319-20877-0_13

Waltman, L., & van Eck, N. J. (2010). The relation between Eigenfactor, audience factor, and influence weight. *Journal of the American Society for Information Science and Technology, 61*(7), 1476–1486. https://doi.org/10.1002/asi.21354

Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics, 10*(2), 347–364. https://doi.org/10.1016/j.joi.2016.02.003

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., et al. (2015). *The metric tide : Report of the Independent review of the role of metrics in research assessment and management*. https://doi.org/10.13140/RG.2.1.4929.1363