



The influence of research collaboration on citation impact: the countries in the European Innovation Scoreboard

Elizabeth S. Vieira¹

Received: 6 September 2022 / Accepted: 19 April 2023 / Published online: 8 May 2023
© The Author(s) 2023

Abstract

The effect of research collaboration (RC) on citation impact (the citations that the scientific knowledge has received) has been studied over the years. However, the extant literature has not considered that the effect of RC can not be efficiently assessed if the level of development of the research and innovation system (RIS) of the collaborating countries is disregarded. I have addressed this gap by looking at the articles published by 36 countries and controlling for other variables that also influence citation impact. The results show that the higher the development of the RIS, the higher the citation impact, as measured by the percentile each article occupies in the respective Web of Science subject area, of single-authored articles and of articles with domestic research collaboration (DRC). The articles representing DRC always have a higher citation impact than single-authored articles, regardless of the level of development of RIS of the countries. Even the articles representing an international research collaboration do not always have a higher impact than single-authored articles; (1) there is no statistically significant difference between the citation impact of articles published in collaboration between scientists from countries considered strong innovators and modest innovators and the citation impact of single-authored articles of scientists from countries considered strong innovators; and (2) articles published by scientists from countries considered innovation leaders in collaboration with scientists from countries considered modest innovators have a lower impact than single-authored articles published by scientists from countries considered innovation leaders. In general, collaboration between scientists from countries with a well-developed RIS and those with a less developed RIS results in articles with a lower citation impact than articles with DRC originating from scientists from countries with a well-developed RIS and higher than articles with DRC originating from scientists from countries with a less developed RIS.

Keywords Citation impact · Research collaboration · Domestic · International · Research and innovation systems · Europe

✉ Elizabeth S. Vieira
elizabeth.vieira@fc.up.pt

¹ LAQV/REQUIMTE, Departamento de Ciências dos Computadores, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal

Introduction

Policy actors see science, more specifically scientific knowledge, as a valuable instrument in giving numerous opportunities to a country concerning its innovation and economic development. Studies have suggested that scientific knowledge, generated through research activities and disseminated in publications, is positively associated with economic growth (e.g. Inglesi-Lotz & Pouris, 2013; Inglesi-Lotz et al., 2014; Ntuli et al., 2015). Given the role of science, several policy instruments have been developed and implemented to foster the RIS (Boekholt et al., 2009). Well-known instruments are those aimed to increase RC, mainly in an international environment (Boekholt et al., 2009).

RC is likely to provide scientists and consequently the RIS where they belong to with numerous advantages. Among the many benefits of RC are: access to tacit and codified knowledge and to costly and complex equipment, the opportunity to expand and diversify research networks, the chance to explore new scientific problems and obtain funding (e.g. Katz & Martin, 1997; Melin, 2000; Owusu-Nimo & Boshoff, 2017; Thorsteinsdottir, 2000; Zdravkovic et al., 2016). Another benefit frequently addressed by the literature is the increasing citation impact, as measured by the number of citations received, of scientific knowledge (throughout the paper, I use the terms citation impact, scientific impact, the impact of the scientific articles, and the impact of scientific knowledge interchangeably). The empirical literature has shown that the impact, as measured by citations, of publications resulting from RC, is likely to be higher than that of publications without collaboration (e.g. Beaver, 2004; Frenken et al., 2010; Sooryamoorthy, 2009; Van Raan, 1998). The studies examining the relationship between RC and impact have been done at the: micro-level (individual publications or individuals scientists), meso-level (universities), and national-level (countries) (e.g., Aldieri et al., 2018; Didegah & Thelwall, 2013b; Katz & Hicks, 1997; Lee & Bozeman, 2005; Leydesdorff et al., 2019; Peters & van Raan, 1994; Puuska et al., 2014) and controlling for other factors that might also influence the scientific impact of knowledge. Such factors might be the attributes of the publications and authors (e.g., Bornmann et al., 2012a, 2012b; Didegah & Thelwall, 2013b; Onodera & Yoshikane, 2015) or the expenditure on research and development (R&D) (Bordons et al., 2015; Chinchilla-Rodriguez et al., 2019; Leydesdorff et al., 2019) considered in some studies as a proxy for scientific capacity (e.g., Chinchilla-Rodriguez et al., 2019).

In this paper, I have examined the effect of RC on the impact of the articles of the countries that constitute the EU-27 plus The United Kingdom (UK), Norway, Switzerland, Serbia, Turkey, Iceland, North Macedonia, Ukraine and Israel taking into account that this effect is related to the development of the RIS of each country involved in the collaboration and also controlling for other variables of interest. I contribute to the literature on this stream in several ways. First, differently from the extant literature on the area (e.g., Bordons et al., 2015; Chinchilla-Rodriguez et al., 2019), I have looked at the effect of RC on scientific impact of countries considered to be modest innovators, moderate innovators, strong innovators, and innovation leaders considering the interaction of RC with the level of development of the RIS of these countries. As for RC, I have disaggregated the information and considered five types of collaboration—DRC and collaborations between countries with RIS at different levels of development—and estimated the incremental effect for a given type of collaboration taking into account different baseline collaborations. Second, in contrast to the literature that has used the expenditure on R&D as a proxy for scientific capacity (e.g., Bordons et al., 2015; Chinchilla-Rodriguez et al., 2019), I have used a more comprehensive indicator to describe the development of the RIS of a country, the

Summary Innovation Index used in the European Innovation Scoreboard (EIS). This allows for a better characterisation of the RIS as it takes into account the main drivers of innovation; the investments in the public and private sectors, the different aspects of innovation in the business sector and the effects of firms' innovation activities (Hollanders et al., 2020). Third, I have included another variable that has been shown to affect citations of publications, but which has not been explored in the presence of RC at the time of this study: the fact that publications may be open access. Fourth, I have demonstrated the applicability of fractional regression models, which are widely applied in econometrics, in the scientometric field.

The paper is organised as follows: in the next section, I review the literature that discusses the influence of RC and other variables on scientific impact and present the framework supporting the formulated hypothesis. The section methodology includes the description of the dataset, variables, and specification of the model. In the following section, it is presented and discussed descriptive statistics and the regression results. The final section summarises the main findings and limitations of the study and suggest some future research.

Literature review and hypothesis

RC is a complex phenomenon due to the nature of human interactions among scientists who have been collaborating (Subramanyam, 1983). Thus, there is no single concept of RC. Katz and Martin, in their attempt to define RC, ended up with suggestions of criteria for identifying collaborators (Katz & Martin, 1997). Bozeman and Boardman defined RC as “social processes whereby human beings pool their experience, knowledge and social skills with the objective of producing new knowledge, including knowledge as embedded in technology” (Bozeman & Boardman, 2014).

It has been claimed that RC allows scientists to produce knowledge with higher quality (Melin, 2000). Indeed, several empirical settings have shown that papers with the participation of diverse scientists attract more citations (citations have been frequently used as a proxy for quality) than single-authored papers (e.g. Beaver, 2004; Frenken et al., 2010; Puuska et al., 2014; Sooryamoorthy, 2009; Van Raan, 1998). Why does RC lead to higher quality work?

Without claiming to be exhaustive, I describe some scenarios in which collaboration may foster outputs with results of higher quality.

Scientific problems are becoming more complex because scientists need to deal with more and more data-rich and computationally intensive projects (the societal challenges are examples of such projects). Most of these problems call for interdisciplinary research as it allows for holistic integrative approaches from a variety of disciplines (Morillo et al., 2003; NSF, 2009). It is, therefore, necessary to combine the skills and expertise of different scientists. This combination will promote a more rigorous review of research, thus increasing the quality of the final outputs.

The knowledge shared by scientists in the form of publications is essential for the progress of science. However, tacit knowledge can not be learned through publications. Here, collaboration plays an important role, as this type of knowledge can be better transferred and combined through this mechanism (Gertler, 2003; Storper & Venables, 2004). Therefore, we expect research results of higher quality.

PhD students usually do not have sufficient experience, context and knowledge to conduct independent research in the early stages of their training (Bozeman & Corley, 2004). Therefore, their collaboration with senior scientists (e.g., in the form of mentoring) is essential to achieve high-quality results.

The exchange of knowledge and regular discussions among scientists about the work can enable the identification of problems that were not considered when planning the research pathway, but which are important for improving the quality of the research activities.

From the previous points, I anticipate that:

H1 Collaborative activities lead to scientific knowledge of higher impact than non-collaborative activities.

As far as the spatial level is concerned, RC has been mainly studied at the domestic and international scales. DRC is usually defined as research activities between scientists working at different institutions located in the same country, while IRC includes research activities between scientists working in different countries (Katz & Hicks, 1997).

Various studies have shown that papers with IRC tend to receive more citations than those with DRC (e.g., Adams, 2013; Katz & Hicks, 1997; Potter et al., 2020). This is understandable, as research with scientists from another country significantly increases the likelihood of a complementary knowledge base compared to DRC and therefore the quality of the scientific knowledge. On the other hand, at this level, the geographical reach of the dissemination of the knowledge generated may increase, as knowledge can flow through the different networks of co-authors (Goldfinch et al., 2003) and therefore the likelihood of being used and cited.

However, I believe that the impact of IRC and DRC depends on the development of the RIS. RIS with a long research tradition and therefore high accumulated scientific knowledge, well-equipped scientific facilities, stable and well-developed research networks and rich human capital in R&D, will benefit less from IRC. In this situation, there are more opportunities to find a partner with complementary knowledge and other resources (e.g., scientific equipment) within the borders (Frame & Carpenter, 1979; Narin et al., 1991). The dissemination of new knowledge will also be higher if we take into account that there are well-established national and international networks of scientists in these countries, resulting from their central position in the network of international research and the large number of countries with which they are involved in joint publications (Chinchilla-Rodríguez et al., 2018; Vieira & Cerdeira, 2022; Vieira et al., 2022). If the IRC offers only a small increase in knowledge compared to DRC, scientists are likely to opt for DRC. The costs and risks of IRC are higher than for DRC due to differences in institutional governance, culture, and geographical distance that might separate collaborators (e.g., Hoekman et al., 2010; Vieira et al., 2022b). This might justify the lower rates of mobility and international collaboration compared to countries with less developed RIS (Chinchilla-Rodríguez et al., 2018). The high overlap of the interquartile ranges of the category normalised citation impact for each collaboration type as regards the United States of America and the UK in the study of Potter et al. 2020 is a first indication of the veracity of the above arguments. However, we should consider that the effect of IRC on the scientific impact of a country i also depends on the RIS of the collaborating country. If the collaborating country has a RIS more developed than country i , then we should expect a higher effect on the scientific impact of IRC than of DRC for country i . A question that inevitably emerges from the previous statement is: Why do the scientists working in countries with well-developed RIS

collaborate with scientists in countries with less–developed RIS? The answer to this question can be found in the internationalization of the RIS, namely the promotion of institutions of higher education abroad, the need of tackling societal issues and challenges with research, and the necessity of maintaining good and stable diplomatic relations (Boekholt et al., 2009).

Therefore, we expect that:

H2 if the scientists of a country with a well–developed RIS collaborate with scientists from a country with a less–developed RIS, then the scientific impact of these publications is lower than the impact of the publications with DRC of the country with a well–developed RIS.

H3 if the scientists of the country with a less–developed RIS collaborate with scientists from a country with a higher developed RIS, then the scientific impact of these publications is greater than the impact of the publications with DRC of the country with a less–developed RIS.

At this point, we have discussed the role of research collaboration on the impact of scientific knowledge. However, studies have shown that the impact may depend on other factors: scientists’ scientific and non–scientific motivations (Baldi, 1998; Merton, 1973; Shadish et al., 1995; Vinkler, 1987), papers’ attributes as the number of authors, number of pages, number of references, journals’ prestige, accessibility, quality and novelty, scientific discipline, type of paper (e.g., Basson et al., 2021; Bornmann et al., 2014; Breugelmans et al., 2018; Buela-Casal & Zych, 2010; Chen, 2012; Craig et al., 2007; Didegah & Thelwall, 2013a, 2013b; Eysenbach, 2006; Gargouri et al., 2010; Haslam & Koval, 2010; McCabe & Snyder, 2014; Moed, 2007; Peng & Zhu, 2012; Peters & van Raan, 1994; van Dalen & Henkens, 2001; Vieira & Gomes, 2010) and authors’ networks positions (Biscaro & Giupponi, 2014; Colebunders et al., 2014; Miranda & Garcia-Carpintero, 2018; Peters & Vanraan, 1994; Tahmooresnejad et al., 2021; Uddin et al., 2013; Yan & Ding, 2009; Zhang et al., 2021).

The scientists’ scientific motivations are related to the normative theory of citing behaviour as defined by Merton, 1973. This theory assumes that authors select papers to cite based on their intellectual relevance. Scientists’ non–scientific motivations are associated with the social constructivist view of citing behaviour (Baldi, 1998; Case & Higgins, 2000; Gilbert, 1977; Shadish et al., 1995; Vinkler, 1987). According to this view, scientists choose to cite a particular paper because it was published in an important and respected journal, the paper was written by a widely known, respected author(s) with absolute professional recognition, the professional connection is maintained with the cited author (s) or one wishes to establish such a connection, and to convince the audience in a scientific community to share their opinions among other reasons.

The length of the paper may affect the scientific impact as longer papers may contain more findings that contribute to the advancement of knowledge. On the other hand, I believe that journal editors and reviewers will only permit papers that are longer than the threshold established if the contribution is relevant. There are some prestigious journals (e.g., Nature and Science) whose papers are usually short but highly cited, so the previous arguments do not apply in this situation. However, we should bear in mind that these papers are often supported by supplementary material with a high number of pages. The literature looking at the association between the impact of scientific knowledge and the size

of the paper has shown mixed results: positive or no association between the two variables (e.g., Didegah & Thelwall, 2013b; Gargouri et al., 2010; Peng & Zhu, 2012; Peters & van Raan, 1994; van Dalen & Henkens, 2001; Vieira & Gomes 2010).

The larger the list of references, the greater the impact of scientific knowledge can be. There are at least two reasons for this argument: the intellectual content and citation-based searching engines. Long reference lists may be associated with interdisciplinary research (the need to integrate knowledge from different scientific disciplines), and this may increase the likelihood that a paper will be cited (Chen et al., 2021; Lariviere & Gingras, 2010). The greater the diversity of scientific disciplines cited, the greater the community that may be interested in. Several algorithms have been developed (such as those in the Web of Science or Scopus) that allow identifying the number of citations that a paper can attract and the respective citing papers. Therefore, it is likely that papers with long reference lists will be found multiple times by these algorithms. Of course, this is not the reason why the paper will be cited more often, but it makes the paper more visible to the scientific community, which ultimately decides whether to cite it or not. In the literature addressing the relationship between the number of references and the scientific impact, a positive correlation between the two has been found (e.g. Bornmann et al., 2014; Chen, 2012; Didegah & Thelwall, 2013a; Haslam & Koval, 2010).

The number of authors in a paper can influence its impact due to multidisciplinary teams and network effects. First, the larger the number of authors, the more likely a multidisciplinary team, thus the paper may attract the attention of scientists working in a wide range of scientific disciplines. Finally, the communication of scientific knowledge is not restricted to journals; the same scientific knowledge can be presented at several conferences for example, especially when multidisciplinary teams are involved, which may also help to increase its visibility. Second, papers with a high number of authors are likely to be cited by a large network of colleagues (Valderas, 2007).

In addition to the number of authors, the authors' positions in the collaboration network can also influence the scientific impact. The more central the authors of a particular paper are in the collaborative network to which they belong to, the greater the possibilities for knowledge dissemination. The more scientists are directly or indirectly (colleagues of directly connected scientists) connected to the paper's authors, the higher the probability that a paper will receive more citations. Therefore, it is not surprising that several studies have demonstrated a positive association between authors' position in the collaborative network and scientific impact (e.g. Biscaro & Giupponi, 2014; Tahmooresnejad et al., 2021; Uddin et al., 2013; Yan & Ding, 2009; Zhang et al., 2021).

The open access (OA) citation postulate is based on the idea that for two articles with the same quality the one that is OA may receive more citations than the non-OA because it is available to a large audience (Gargouri et al., 2010). Access to articles that is not limited by a "paywall" allows for large audiences since the internet is available (Harnad et al., 2008). However, the studies addressing the subject have shown contradictory and inconclusive findings as regards the effect of OA on the impact of scientific knowledge. Some studies verified the existence of the OA citation postulate, while others did not confirm this (e.g., Basson et al., 2021; Breugelmans et al., 2018; Craig et al., 2007; Eysenbach, 2006; McCabe & Snyder, 2014; Moed, 2007).

A prestigious journal signals to the readers that idea of high quality. In some fields, high-quality articles tend to be sent first to prestigious journals, and if rejected, they are submitted to a journal with lower impact (Oster, 1980). The higher the quality of the editorial board, the higher the expected prestige of the journal because editors have proven credentials in most cases. Also, they are in part responsible for the outputs of their journals,

thus they have strong motivations to perform good quality control services. The literature has shown that scientific impact is associated with the prestige of the journals (e.g., Callahan et al., 2002; Didegah & Thelwall, 2013a; Peters & van Raan, 1994).

Citation culture varies across scientific disciplines and over time (Dorta-Gonzalez et al., 2014; Moed et al., 1985). The frequency with which papers cite others varies across scientific disciplines (Moed, 2010; Zitt & Small, 2008). In Biology & Biochemistry, the average number of references per paper is higher than in Mathematics (Vieira & Gomes 2010), so it is to be expected that the number of citations is higher in the first scientific discipline.

Finally, documents of the type review evaluate results and approaches from several scientific fields and consider the state of the art in each scientific field which could explain why this type of document is more frequently cited than original articles and other types of documents (Colebunders et al., 2014; Miranda & Garcia-Carpintero, 2018; Peters & van Raan, 1994).

Given the multiple variables that may influence the scientific impact, I studied the influence of RC looking at the interaction of this variable with the development of the RIS of each country, and articles features (the number of authors, pages and references, the journals prestige and the accessibility of the documents) on citation impact.

Variables like those related to persuasion and novelty of scientific knowledge are intangible and unmeasurable (persuasion) or difficult to measure (novelty), so I have not examined the influence of these dimensions. Some measures have been developed to measure novelty based on citation and text data (Shibayama et al., 2021; Uzzi et al., 2013; Wang et al., 2017). However, it has been shown that these measures are not suitable for measuring novelty (Bornmann et al., 2019; Fontana et al., 2020). Novelty as the first appearance of a knowledge combination captures the main structural properties of the citation network and it is difficult to distinguish novel from non-novel articles (the novelty measure defined in Wang et al., 2017), while novelty as an atypical knowledge combination overlaps with interdisciplinarity (the novelty measure defined in Uzzi et al., 2013). As for the measures based on citation and text data simultaneously (Shibayama et al., 2021), the limitations are due to the lack of publicly available word-embedding libraries for various scientific fields. We may be able to find these libraries for some scientific fields, but not for a large number. On the other hand, the validation process of this measure has shown a weak correlation with self-reported novelty scores from a questionnaire survey, although the correlations are positive and significant.

An additional variable could be the position of the authors of a document in the collaboration network. However, I excluded this variable due to the complexity associated with its calculation (author's name disambiguation). Therefore, these limitations of the study should be taken into account in the interpretation of the results.

Methodology

Data

I have studied the influence of RC on the impact of scientific knowledge of the publications classified as articles, (if an article received other document type classification, it was not considered) considering all Web of Science categories, published by the scientists working in countries belonging to the EU–27 plus the UK, Norway, Switzerland, Serbia, Turkey, Iceland, North Macedonia, Ukraine and Israel. The additional countries (those not

belonging to EU–27) have been chosen because we have information in the EIS 2019 about the development of the RIS. The publications of these countries are the observations used in the analysis presented below.

I have retrieved the articles published in 2017 (361 616 articles) by the scientists of each country from the Web of Science Core Collection (WoS). Then, for each article, I have identified the type of collaboration, and in the case of IRC, I have chosen only those articles representing a collaboration between scientists of two of the countries considered (this choice is related to the availability of information regarding the development of the RIS in the ESI). Also, for each article, I have looked at the number of pages, number of references, number of authors, quartile that the journal in which the article had been published belongs to, and accessibility of the article (OA or non–OA).

Variables and model specification

Dependent variable

The dependent variable is the citation impact, and it is represented by the percentile (*percentile*) occupied by each article in the respective Web of Science subject area. The higher the *percentile*, the higher the impact of the article. As the number of citations depends on the citation culture of each scientific discipline, the type of document and the publication year, I have selected a normalized indicator to represent the citation impact. The *percentile*, available through InCites, of a publication, is determined by creating a citation frequency distribution for all the publications in the same year, subject category and of the same document type, and then the percentage of papers at each level of citations.¹ The citation window is an open window, i.e. the count of citations represents the cumulative citations between the publication year and March 2022. It has been argued that the number of self–citations increases with the number of authors or with RC (Glanzel & Thijs, 2004; Van Raan, 1998). However, it has been concluded that the increase in foreign citations is much higher than that of self–citations (Costas et al., 2010; Glanzel & Thijs, 2004). On the other hand, it is common practice for scientists to follow a given research line taking into account their previous research. Considering this and the fact that removing self–citations is costly, time–consuming and computationally complex, I have considered this type of citation. As mentioned in the previous section, the citation culture varies among scientific disciplines and over time. Also, the scientific impact is related to the type of document. Therefore, I am controlling to some extent for the influence of these factors when using a normalized indicator and considering only documents of the type of article.

Other normalized indicators could be used to measure the scientific impact; the most known are those that use the citing–side or cited–side normalization approaches and that are based on average values (Bornmann, 2020; Bornmann & Marx, 2015; Waltman, 2016). The literature comparing the indicators based on these two approaches has led to contradictory findings: Leydesdorff et al., 2013 observed that cited–side normalization outperforms citing–side normalization, Waltman & van Eck, 2013 and Bornmann & Marx, 2015 concluded the opposite. Given the contradictory findings and the fact that indicators that compare the number of citations with the mean expected number of citations may be influenced

¹ <http://help.prod-incites.com/inCites2Live/indicatorsGroup/aboutHandbook/usingCitationIndicatorsWisely/averagePercentile.html>.

by one or a few publications with a very large number of citations (the citations distribution is highly skewed (Vieira & Gomes, 2010)), I have chosen to use percentile indicators.

Independent variables

In understanding the influence of RC, I considered for each article the development of the RIS of the involved countries (*EIS*), the number of pages (*pages*), references (*references*), and authors (*authors*), the accessibility (*access*), the quartile (*quartile*) to which the journal where the published article belongs to according to the Journal Impact Factors (JIF), and the type of collaboration (*collaboration*). As for the JIF, the literature is rich in discussing the limitations of this indicator (Bornmann et al., 2012a, 2012b). An interesting indicator that overcomes many of the caveats of the JIF is the recently introduced Journal Citation Indicator (JCI), a field-normalised indicator. In future studies, it will be interesting to use this indicator, although I do not expect it to behave significantly differently, as there is a 67% overlap between the sources in the first quartile according to the JIF and JCI.

As for the development of RIS, I have used as a proxy the classification of each country in EIS 2019 (the most recent report at the time I have started this study) for the year 2017. The annual EIS provides a comparative assessment of the research and innovation performance of EU Member States through a composite indicator (Summary Innovation Index) that takes into account the main drivers of innovation, public and private sector investment, the different aspects of innovation in the business sector and the impact of business innovation activities (Hollanders et al., 2020). In short, the indicator assesses the relative strengths and weaknesses of national innovation systems and helps countries identify areas they need to address. Four main types of indicators are considered in the calculation of the Summary Innovation Index—Framework conditions, Investments, Innovation activities, and Impacts (Hollanders, 2019). Framework conditions capture the key drivers of innovation performance, taking into account human resources, the attractiveness of research systems and the environment in which enterprises operate measuring, for example, the extent to which individuals pursue entrepreneurial activities when they see new opportunities arising from innovation. The Investments capture investments made in the public and business sectors, looking at the availability of finance for innovation projects by Venture Capital Expenditures, government support for research and innovation activities through R&D expenditures in universities and government research institutions, and the firms investments in R&D, non-R&D and efforts to improve the ICT skills of their employees. Innovation activities capture the share of firms that have introduced innovation into the market, collaboration efforts between innovative firms, research collaboration between the private and public sectors, extent to which the private sector funds public R&D activities, and intellectual property rights generated. Finally, Impacts capture the effects of firms' innovation activities on employment and sales. The list of 27 indicators, not described here due to space limitations, and their calculation can be found in Hollanders, 2019.

The variable *EIS* is equal to 1, 2, 3, or 4 if the country is classified as a modest innovator (Bulgaria, Croatia, Macedonia, Romania, Turkey, Ukraine), moderate innovator (Cyprus, Czechia, Estonia, Greece, Hungary, Italy, Latvia, Lithuania, Malta, Poland, Portugal, Serbia, Slovakia, Slovenia, Spain), strong innovator (Austria, Belgium, France, Germany, Iceland, Ireland, Israel, Luxembourg, Norway, UK), and innovation leader (Denmark, Finland, Netherlands, Sweden, Switzerland), respectively.

Access is 1 if the article is OA and 0 otherwise.

Quartile is 1 if the article was published in a journal that belongs to the first quartile, given the year of publication of the article, and 0 otherwise. If a given journal is classified in more than one Web of Science category and has different positions concerning *quartile*, I have considered the best position according to this measure.

Collaboration is 0, 1, 2, 3, 4, 5 if the article has one author (no collaboration), a domestic collaboration (a collaboration between two scientists working at different institutions located in the same country), a collaboration with a modest innovator ($EIS = 1$), a collaboration with a moderate innovator ($EIS = 2$), a collaboration with a strong innovator ($EIS = 3$), and a collaboration with an innovation leader ($EIS = 4$), respectively.

As for the identification of articles with DRC, I disregarded all the articles that respect both conditions: (1) have collaborations between scientists working in different institutions located in the same country and (2) have foreign scientists. I have adopted this procedure aimed to avoid impurities in the estimated effect of DRC due to the presence of IRC. The articles with DRC were identified using the InCites platform. As for the articles with IRC and given that I wanted to control for the effect of the development of the RIS, I have only considered as IRC those articles that have scientists from only two countries and the collaborating country is one of the 36 mentioned above. I adopted this procedure because data for EIS is only available for a few countries.

Model specification

The dependent variable takes values in the unit interval [0 1], therefore, I have resorted to fractional regressions that have been used in many economic settings for example in modelling the employee participation rates in pension plans (Papke & Wooldridge, 1996). Ordinary least square (OLS) regression might seem attractive but a strong drawback of this methodology is that the predicted values can never lie in the interval [0 1] (Papke & Wooldridge, 1996). Thus, the log-likelihood function for fractional models is:

$$\text{Ln}L = \sum_{j=1}^N w_j y_j \ln \left\{ X_j' \beta \right\} + w_j (1 - y_j) \ln \left\{ 1 - G \left(X_j' \beta \right) \right\}$$

where N is the sample size, y_j the dependent variable, w_j the optional weights, and $G(\bullet)$ is a probit model. The functional form for $G(X_j' \beta)$ is $\Phi(X_j' \beta)$ where X_j are the independent variables for publication j and Φ the standard normal cumulative density function. For more details on the model see (Papke & Wooldridge, 1996; Wooldridge, 2010).

$$\begin{aligned} \Phi(\beta_0 + \beta_1 * EIS_j + \beta_2 * Pages_j + \beta_3 * References_j + \beta_4 * Authors_j \\ + \beta_5 * Access_j + \beta_6 * Quartile_j + \beta_7 * EIS_j * Collaboration_j) \end{aligned}$$

In determining the model, I have used the command `fracreg` from Stata version 16.0 and robust standard errors.

To interpret the results from the regression model, I have resorted to predictive margins and marginal effects. Predictive margins are statistics calculated from a fitted model at fixed values of some independent variables and averaging or otherwise integrating over the remaining independent variables. Marginal effects are partial derivatives of the

Table 1 Descriptive statistics for pages, references, authors, and percentile

Statistics	Pages	References	Authors	Percentile
Average	12.58	41.60	4.74	0.50
Median	11	37	4	0.52
St. Deviation	8.56	26.68	4.86	0.31
Maximum	567	1476	1230	1.00
Minimum	1	0	1	9.00E-06

Table 2 The proportion of articles by levels of access, quartile, EIS and collaboration. For each variable is also presented the average percentile

Variable	Proportion (%)	Percentile (average)
Access		
Non-OA	55.5	0.46
OA	44.5	0.55
Quartile		
Other	64.3	0.41
First	35.7	0.65
EIS		
Modest (1)	8.7	0.33
Moderate (2)	31.1	0.47
Strong (3)	45.5	0.50
Leader (4)	14.6	0.57
Collaboration		
No collaboration (0)	21.5	0.40
DRC (1)	43.1	0.50
Modest (2)	1.4	0.45
Moderate (3)	10.3	0.60
Strong (4)	16.3	0.58
Leader (5)	7.4	0.61

fitted model concerning each variable in the model for each unit in the dataset (Stata-Corp., 2019).

Results

Descriptives

In the dataset, the average number of pages, references and authors per article is 12.58, 41.60 and 4.74, respectively, and the medians represent lower values, indicating that the distributions are positively skewed (Table 1). As for the number of pages and references, there are articles with a high number of pages and references. I analysed the articles with a high number of pages and found no particular reason for the high number of pages, although in some cases the high number of references takes up several pages (for example, the article with 1476 references has 60 pages and 48 pages have only references). However,

this is not the case for all articles with hundreds of pages (the article with 567 pages has 120 references). Finally, the hyperauthorship (Cronin, 2001) of the article with 1230 authors relates to the collaboration of the teams JET, EUROfusion MST1 and ASDEX-Upgrade, i.e., collaboration on a macroscopic scale. The dataset is also very heterogeneous concerning the variable *percentile* (coefficient of variation of 62%). Besides the presence of these special observations in the dataset, they have not been removed because they are true outliers, i.e., they represent natural variations in the sample and did not result from measurement, data entry, or processing errors.

More than 55% of the articles are non-OA and about 36% were published in journals that were in the first quartile in 2017 according to the JIF (Table 2). The distribution of articles according to the development of the RIS is not balanced and there are more articles of countries classified as strong innovators (Table 2). The same pattern is observed for the different types of collaboration being the articles with DRC the most common in the data (43%). The values for the *percentile* are very interesting as they show that the impact is greater for OA articles (on average, the *percentile* is 0.55) than for non-OA (Table 2). Articles in sources classified in the first quartile are cited more frequently than articles published in journals in the other quartiles (on average, the *percentile* is 0.65). As expected, countries with more developed RIS generate articles with greater impact (Table 2). As far as the type of collaboration is concerned, the behaviour in terms of scientific impact is mixed. For a detailed explanation of the expected behaviours, see Sect. “Model”.

Model

In Table 3, I present the results of the regression for three models: the model with the variables representing the articles’ features (model 1), the model with articles’ features and the level of development of the RIS of each country analysed (model 2), and the full model taking into account all the variables (model 3). Among all models, the Akaike’s information criteria indicates that the full model better describes the relationship between the dependent and independent variables.

The higher the number of pages, references and authors, the higher the scientific impact of articles, as shown by the positive and statistically significant coefficients for these variables. Furthermore, articles that are OA have a higher scientific impact than non-OA articles (positive and statistically significant coefficient). Articles published in journals in the first quartile have a higher scientific impact than articles published in journals in the other quartiles (positive and statistically significant coefficient). These results are observed in all the models.

Regarding the effect of *EIS* and *collaboration*, the results are more difficult to interpret. In general, we observe positive and statistically significant coefficients. Exceptions are (1) the interaction representing a collaboration between an innovation leader and a modest innovator, which has a negative and statistically significant coefficient, and (2) the interaction between a strong innovator and a modest innovator, which has a positive coefficient but is not statistically significant.

To compare and discuss the effects of each variable and the interaction term, I have resorted to the concept of predictive and marginal effects.

In Table 4, the gains in impact for a small increase in *pages*, *references* and *authors* are 0.00172, 0.00256 and 0.00488 respectively. We also see that the *percentile* (conditional mean) increases by 0.0155 when we move from a non-OA article to an OA article. For articles published in journals in the first quartile, the impact increases by 0.163 compared

Table 3 Fractional regression model

Variables	Percentile Model 1	Percentile Model 2	Percentile Model 3
2.EIS (Moderate)		0.198*** (0.00495)	0.0431*** (0.0118)
3.EIS (Strong)		0.260*** (0.00484)	0.344*** (0.0110)
4.EIS (Leader)		0.358*** (0.00547)	0.470*** (0.0137)
Pages	0.00400*** (0.000258)	0.00349*** (0.000246)	0.00464*** (0.000320)
References	0.00760*** (7.52e-05)	0.00739*** (7.39e-05)	0.00696*** (7.91e-05)
Authors	0.0331*** (0.000509)	0.0324*** (0.000503)	0.0132*** (0.00346)
Access	0.0594*** (0.00255)	0.0496*** (0.00255)	0.0416*** (0.00266)
Quartile	0.477*** (0.00266)	0.446*** (0.00270)	0.428*** (0.00327)
1.Collaboration#1b.EIS			0.200*** (0.0166)
1. Collaboration #2.EIS			0.407*** (0.0177)
1 Collaboration #3.EIS			0.114*** (0.0160)
1. Collaboration #4.EIS			0.108*** (0.0178)
2. Collaboration #1b.EIS			0.443*** (0.0402)
2. Collaboration #2.EIS			0.310*** (0.0211)
2. Collaboration #3.EIS			0.00665 (0.0201)
2. Collaboration #4.EIS			- 0.146*** (0.0351)
3. Collaboration #1b.EIS			0.339*** (0.0227)
3. Collaboration #2.EIS			0.489*** (0.0171)
3. Collaboration #3.EIS			0.165*** (0.0164)
3. Collaboration #4.EIS			0.0727*** (0.0201)
4. Collaboration #1b.EIS			0.356*** (0.0221)
4. Collaboration #2.EIS			0.502*** (0.0169)
4. Collaboration #3.EIS			0.186***

Table 3 (continued)

Variables	Percentile Model 1	Percentile Model 2	Percentile Model 3
			(0.0155)
4. Collaboration #4.EIS			0.111*** (0.0188)
5. Collaboration #1b.EIS			0.361*** (0.0357)
5. Collaboration #2.EIS			0.554*** (0.0186)
5. Collaboration #3.EIS			0.245*** (0.0165)
5. Collaboration #4.EIS			0.168*** (0.0192)
Constant	- 0.723*** (0.00470)	- 0.923*** (0.00569)	- 1.018*** (0.0117)
Observations	361,616	361,616	361,616
Akaike's information criteria	470,734.9	469,161.1	467,107.5

Robust standard errors in parentheses
 *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4 Marginal effects for pages, references, author, access, quartile. For the dummies, the base level is 0

Variables	Effect
Pages	0.00172*** (0.000118)
References	0.00256*** (0 0,000,289)
Authors	0.00488*** (0.00128)
Access	0.0155*** (0.000990)
Quartile	0.163*** (0.00129)
Observations	361,616

to articles published in journals in the other quartiles. The quartile has the highest effect on scientific impact. More detailed information about the gains for pages, references and authors can be obtained when we look at different values for these variables.

As the number of pages, references and authors per publication increases, the conditional mean of *percentile* also increases (Figs. 1, 2, 3, left hand), which we had already expected given the positive and statistically significant coefficient in Table 3. However, we get the most interesting results when we look at the marginal effects (right in Figs. 1, 2, 3). For *pages*, the effect on the scientific impact of articles with six pages is the same (increase in scientific impact by 0.00172) as the effect of articles with eleven pages. For *references*, the effect increases up to about 40–50 references per article (scientific impact

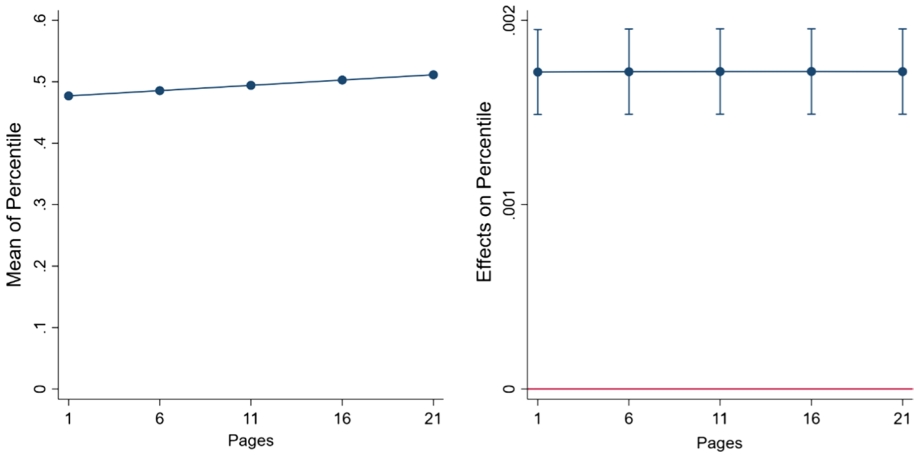


Fig. 1 The conditional mean of *percentile* (left hand) and effects on conditional mean of *percentile* (right hand) of *pages*. The bars represent confidence intervals (CIs) at 95%. I chose to limit the number of pages to 21 because 88% of the articles have 21 or fewer pages

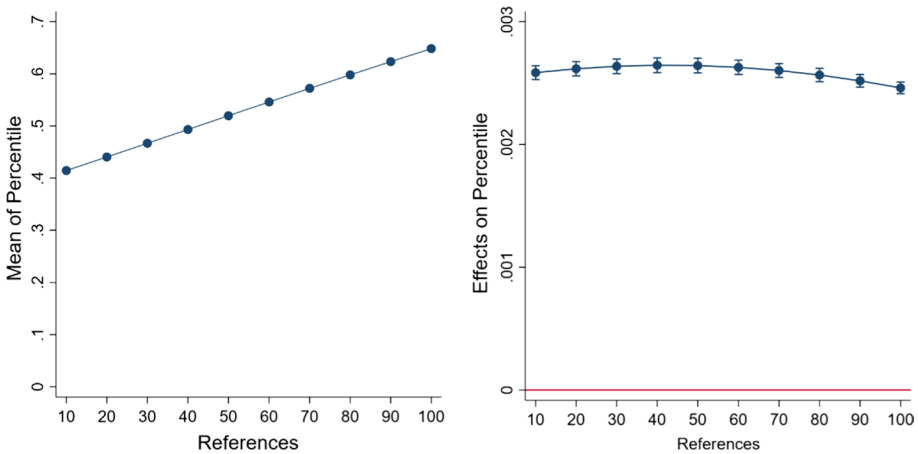


Fig. 2 The conditional mean of *percentile* (left hand) and effects on conditional mean of *percentile* (right hand) of *references*. The bars represent CIs at 95%. I chose to limit the number of references to 100 because 97% of the articles have 100 or fewer references

increases by 0.00258 for articles with 10 references and by 0.00264 for articles with 40 references) and then starts to decrease (for an article with 100 references, the impact increases by 0.0024). The same pattern is observed for *authors*, the effect on impact increases up to six authors per article (by 0.004906 for an article with two authors and by 0.004911 for an article with six authors, although we have a slight increase) and then starts to decrease. For an article with 30 authors, the impact increases by 0.00470.

The effects of *EIS* and *collaboration* are easily understood by looking at Figs. 4 and 5. The higher the development of the RIS, the higher the scientific impact of articles without collaboration or articles with DRC; the scientific impact of single-authored articles of

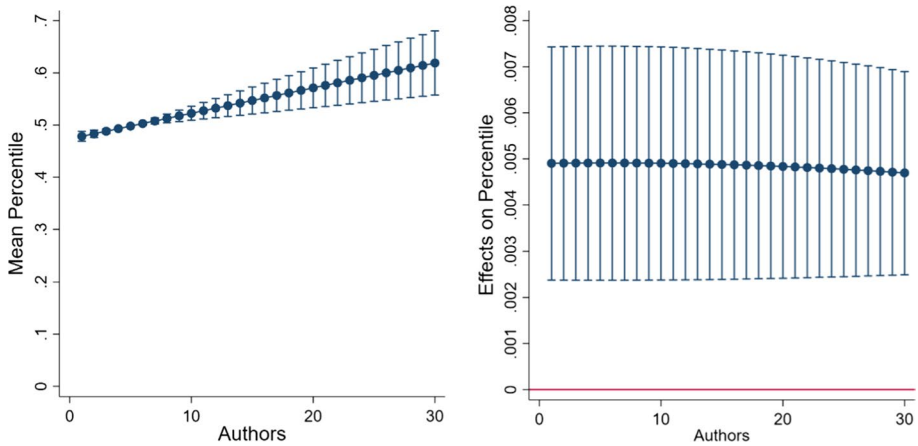


Fig. 3 The conditional mean of *percentile* (left hand) and effects on conditional mean of *percentile* (right hand) of *authors*. The bars represent CIs at 95%. I chose to limit the number of authors to 30 because 99.9% of the articles have 30 or fewer authors

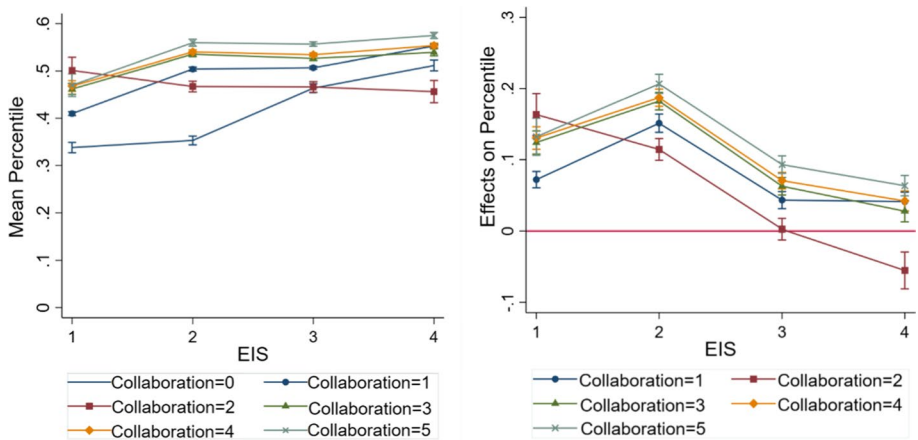
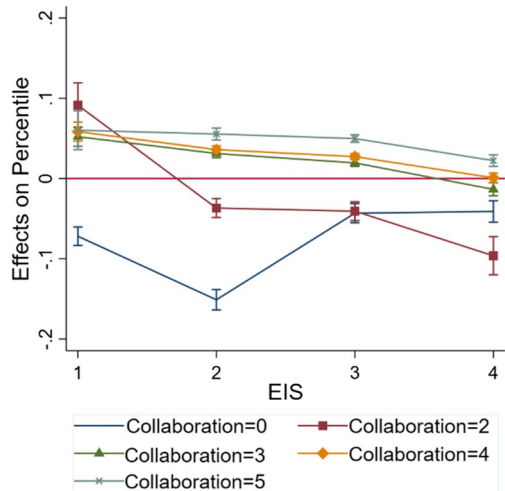


Fig. 4 The conditional mean of *percentile* (left hand) and effects on conditional mean of *percentile* (right hand) of *collaboration* according to *EIS*. For the graph on the effects, the base category is *collaboration* equal to 0, i.e. no collaboration. The bars represent CIs 95%

scientists from modest innovators is 0.338 and from innovation leaders is 0.511 (Fig. 4, left hand). In Fig. 4, we also see that collaborative activities do not always lead to articles with higher scientific impact than single-authored articles (H1 is partially confirmed). For scientists from strong innovators collaborating with scientists from modest innovators, we have already seen that the coefficient is not statistically significant, and in Fig. 4, we see the large overlap of the CI with that of articles without collaboration. For scientists from innovation leaders collaborating with scientists from modest innovators, the impact of the articles resulting from this collaboration is below (equal to 0.456) of the impact (equal to 0.511, Fig. 4, left hand) of the single-authored articles originated by the scientists in

Fig. 5 The effects on conditional mean of percentile of collaboration according to *EIS*. For the graph on the effects, the base category is collaboration equal to 1, i.e. DRC. The bars represent CIs 95%



the innovation leaders (see also the effect below 0 in Fig. 4, right hand, and the statistically significant coefficient in Table 3). Similar results have been found in a previous study (Chinchilla-Rodriguez et al., 2019).

For scientists from country *i* (independently of its RIS), the effect of collaboration with scientists from innovation leaders on scientific impact is always positive compared to articles without collaboration, and the effect decreases as the *EIS* of country *i* increases from moderate innovator to innovation leader (increases of 0.207 and 0.063, respectively, Fig. 4, right hand).

In short, for two scientists from two different countries (*i* and *j*) collaborating with scientists from the same country, saying country A, the scientists working at the country (between *i* and *j*) with the more developed RIS benefits less from IRC. An exception are scientists from modest innovators and moderate innovators; for the latter, the gain from working with scientists from strong innovators or innovation leaders is always higher than that of collaborating with scientists from modest innovators. Also, for scientists from modest innovators, the collaboration with scientists from a country with a RIS in the same category leads to articles with the highest impact (0.501 Fig. 4, left hand).

For scientists from moderate innovators, strong innovators and innovation leaders collaborations with scientists from modest innovators lead to articles with a lower scientific impact than the collaboration of type 1, i.e. DRC, and the effect tends to be higher (more negative) as the *EIS* increases (e.g., for scientists from innovation leaders collaborating with scientists from modest innovators, the scientific impact decreases by 0.096 in relation to the articles with DRC, Fig. 5). Also, a collaboration between scientists from innovation leaders and scientists from moderate innovators leads to articles with a lower impact than those representing a DRC published by scientists from innovation leaders (0.539 and 0.553, respectively, Fig. 4). For scientists from innovations leaders, but this time for a collaboration with scientists from strong innovators, the impact is slightly higher than that of their articles with DRC, but the effect is not statistically significant (Fig. 5). Finally, the impact of articles generated in a collaboration between scientists from strong innovators and scientists from moderate innovators increases by 0.0195 compared to the articles involving DRC of the scientists from strong innovators (Fig. 5). Therefore, H2 is partially verified.

For scientists that collaborate with others from countries with more developed RIS, the scientific impact of the articles resulting from this collaboration is higher than the impact of their articles involving DRC ($H3$ is true, Fig. 4, left hand). The magnitude of this effect tends to decrease as we move from a collaboration with scientists from modest innovators to collaboration with scientists from strong innovators (Fig. 5). For scientists from modest innovators that collaborates with scientists from innovation leaders, the impact of the articles increases by 0.0602 relative to their articles with DRC, while for the articles of scientists from strong innovators working with those from innovation leaders, the impact increases by 0.0498 (Fig. 5).

Finally, and although I have not formulated a hypothesis regarding the expected behaviour when scientists collaborate with others from a country that has *EIS* in the same category, the results show that the scientific impact is always higher than that of the articles with DRC; for example, for scientists from moderate innovators collaborating with those from countries in the same category, the impact is 0.535 and for their articles with DRC it is 0.504 (Fig. 4).

Conclusions

The benefits of RC have been widely explored by the scientific community. Of the several benefits, the effects on the scientific impact, i.e., the citations that the scientific knowledge has attracted, have received great attention. However, the extant literature has not considered that the effect of the different types of RC on the scientific impact depends on the development of the RIS of each country. I have explored the issue for the articles of 36 countries, the countries belonging to the EU–27 as well as the UK, Norway, Switzerland, Serbia, Turkey, Iceland, North Macedonia, Ukraine and Israel, and taking into account that other variables (number of pages, references and authors, the accessibility of the article and the quartile to which the journal in which a given article was published belongs to) also have influence. The findings of this study lead to a number of implications for both theory and practise, although they must be interpreted in light of the limitations associated with this study, which are discussed in the Limitations and Future Research section.

Contrary to much of the literature, which states that RC originates publications with higher scientific impact than non–collaborative research, the findings in this study clearly show that the impact depends on the development of the RIS of the collaborating countries. RC between scientists from innovation leaders and those from modest innovators results in articles with lower impact than the single–authored articles of the scientists from innovation leaders. Others have also shown that for some countries with high expenditure on R&D, DRC leads to lower scientific impact than single–authored papers (Fig. 3 e.g., countries with expenditure on R&D equal or above 2% of the GDP, Chinchilla-Rodriguez et al., 2019).

It has been argued that IRC generates publications of a higher impact than DRC. In the same line as the previous comment, the results show that the development of the RIS of the collaborating countries should be taken into account. For example, a collaboration between scientists from moderate innovators, (or strong innovators or innovation leaders) and scientists from modest innovators originates articles with lower impact than the articles with DRC of scientists from moderate innovators (the strong innovators or innovation leaders). A collaboration between scientists from modest innovators (or moderate innovators or

strong innovators) and scientists from innovation leaders generally leads to articles with higher impact than the articles with DRC of the scientists from modest innovators (moderate innovators or strong innovators).

For scientists from two countries (i and j) collaborating with scientists from the same country (the development of RIS is higher than that of countries i and j), the scientists from the country (between i and j) with the more developed RIS benefit less from IRC when we compare the increase in impact relatively to their single-authored articles or articles with DRC in general. For example, the gains (marginal effects in scientific terminology) are higher for articles written by scientists from modest innovators in collaboration with scientists from innovation leaders than for articles written by scientists from strong innovators in collaboration with scientists from innovation leaders. The exception is scientists from modest innovators and moderate innovators, because for scientists from moderate innovators, collaboration with scientists from countries with a more developed RIS always leads to higher gains in the scientific impact of their articles than for articles written by scientists from modest innovators when the baseline is single-authored articles. An interesting case for further study.

As for the other variables, it has been stated that the higher the number of pages, references and authors per article, the higher the scientific impact. The findings from previous studies are in concordance with this study. However, we can now draw broader conclusions as the study shows that the gain in impact is not necessarily the same for a small change (increase) in these variables. As for the number of pages, the gains are more or less the same when increasing the number of pages per article. In the case of references, the gains increase as we go from an article with 10 references to another with 40 references, and then, start to decrease. Concerning the number of authors, the gains on impact increase up to six authors per article and decreases for higher values.

Regarding the variables *access* and *quartile*, I have found that OA articles and articles published in journals in the first quartile have a higher scientific impact than those that are non-OA or published in journals in other quartiles. Finally, among *pages*, *references*, *authors*, *access* and *quartile*, *quartile* has the highest effect on scientific impact. A comparison of these effects with those of the *EIS* and *collaboration* is not straightforward, as the patterns observed are multiple. Each scenario must be evaluated individually.

These findings are relevant for scientists and policymakers who design and implement policies to promote RC. We would say that in a country with a well-developed RIS, policies that encourage collaboration with countries with less-developed RIS should be avoided. However, this is not possible as the goals of RC are not only to increase scientific impact. In many cases, these policies aim to address societal issues and challenges with research, maintain good and stable diplomatic relations, and promote higher education institutions abroad. Therefore, the strategies to achieve these goals may require collaboration with scientists in less-developed RIS. Thus, the strategies should aim at a balance, i.e. if the policies related to the above goals are in place with less-developed RIS, then others should compensate for the less positive effect of this type of collaboration on scientific impact of a country with a well-developed RIS.

The results for *access* could lead to policies that reward scientists who publish open research. Indeed open science is increasingly embedded in policies and expected in practice for a variety of reasons (transparency, accountability, equity and collaboration in knowledge production by increasing access to research results, Cole et al., 2022). However, we have to look at the findings of studies suggesting that the participation in article processing charges (APCs) of OA is skewed toward scientists with greater access to resources

(scientists in high-income countries) and job security (Olejniczak & Wilson, 2020), and that scientists in the Global South frequently cite APCs as a financial barrier (Smith et al., 2022).

Limitations and future research

Regarding the limitations of this study, we have those related to the data and the variables.

As for the data, the reader should bear in mind that I have analysed a number of specific countries (EU-27 plus the UK, Norway, Switzerland, Serbia, Turkey, Iceland, North Macedonia, Ukraine and Israel), as the Summary Innovation Index is only available for these countries. Therefore, it is not possible to generalise the results. In the data, I have retained those articles that represent hyperauthorship or collaboration at a macroscopic level, although I am aware that these are special cases and future studies on this topic should question whether these articles are used. The number of observations in the different categories of the variables EIS and collaboration (Table 1, supplementary material) varies considerably among the many possible combinations of these variables. However, a more balanced distribution is not easy to achieve. Each observation refers only to bilateral collaborations within the set of countries considered (36 countries) which in turn limits the scope of the conclusions drawn in this study. For publications with multilateral collaboration (more than two countries), I find it difficult to get a clear picture of the influence of the type of collaborating country by level of development of the RIS. On the other hand, the number of papers involving scientists from more than two countries was not taken into account, and these could represent a considerable number of publications.

As for the variables, it should be noted that some variables (persuasion, novelty and authors' position in the research network) that could influence the citation impact were not considered in this study for several reasons: they are intangible and unmeasurable (persuasion), difficult to measure given the underlying concept (novelty), and difficult to determine given the need to access large amounts of data and the complexity of the algorithms to be developed (novelty and authors' position in the research network).

As for further research, it would be very interesting to expand the frontiers of knowledge on this subject through studies that analyse bilateral collaborations between countries other than European ones. However, I realise that in order to do this, we need to develop an indicator to characterise the level of development of the RIS that can be applied to all countries, and this is a very challenging task. In this way, it would be possible to determine the extent to which the results of this study can be universal.

The results have shown that the more similar the level of development of the RIS of the two collaborating countries, the smaller the gain in impact of the articles. The exception are the articles from scientists from countries that are modest innovators. For them, the gain in impact is lower than that of the articles of the scientists from the countries considered moderate innovators. An interesting research question could be: Is there a maximum distance between the levels of development of RIS of two collaborating countries beyond which the gains start to diminish?

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11192-023-04715-4>.

Acknowledgements Elizabeth S. Vieira thanks FCT (Fundação para a Ciência e Tecnologia) for funding through program DL 57/2016—Norma transitória (DL 57/2016/CP1346/CT0017). This work received financial support from PT national funds (FCT/MCTES, Fundação para a Ciência e Tecnologia and

Ministério da Ciência, Tecnologia e Ensino Superior) through the project UIDB/50006/2020. I thank Jorge Cerdeira for the helpful comments on the regression model used in this paper.

Funding Open access funding provided by FCTIFCCN (b-on).

Declarations

Conflicts of interest The author has no potential conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adams, J. (2013). The fourth age of research. *Nature*, 497(7451), 557–560. <https://doi.org/10.1038/497557a>
- Aldieri, L., Kotsemir, M., & Vinci, C. P. (2018). The impact of research collaboration on academic performance: An empirical analysis for some European countries. *Socio-Economic Planning Sciences*, 62, 13–30. <https://doi.org/10.1016/j.seps.2017.05.003>
- Baldi, S. (1998). Normative versus social constructivist processes in the allocation of citations: A network-analytic model. *American Sociological Review*, 63(6), 829–846. <https://doi.org/10.2307/2657504>
- Basson, I., Blanckenberg, J. P., & Prozesky, H. (2021). Do open access journal articles experience a citation advantage? Results and methodological reflections of an application of multiple measures to an analysis by WoS subject areas. *Scientometrics*, 126(1), 459–484. <https://doi.org/10.1007/s11192-020-03734-9>
- Beaver, D. D. (2004). Does collaborative research have greater epistemic authority? *Scientometrics*, 60(3), 399–408. <https://doi.org/10.1023/B:SCIE.0000034382.85360.cd>
- Biscaro, C., & Giupponi, C. (2014). Co-authorship and bibliographic coupling network effects on citations. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0099502>
- Boekholt, P., Edler, J., Cunningham, P., & Flanagan, K. (2009). *Drivers of international collaboration in research*. Retrieved from <https://op.europa.eu/en/publication-detail/-/publication/712e874d-4f61-4977-9512-3bb326c2ce63/language-en>
- Bordons, M., Gonzalez-Albo, B., Aparicio, J., & Moreno, L. (2015). The influence of R&D intensity of countries on the impact of international collaborative research: Evidence from Spain. *Scientometrics*, 102(2), 1385–1400. <https://doi.org/10.1007/s11192-014-1491-4>
- Bornmann, L. (2020). How can citation impact in bibliometrics be normalized? A new approach combining citing-side normalization and citation percentiles. *Quantitative Science Studies*, 1(4), 1553–1569. https://doi.org/10.1162/qss_a_00089
- Bornmann, L., Leydesdorff, L., & Wang, J. (2014). How to improve the prediction based on citation impact percentiles for years shortly after the publication date? *Journal of Informetrics*, 8(1), 175–180. <https://doi.org/10.1016/j.joi.2013.11.005>
- Bornmann, L., & Marx, W. (2015). Methods for the generation of normalized citation impact scores in bibliometrics: Which method best reflects the judgements of experts? *Journal of Informetrics*, 9(2), 408–418. <https://doi.org/10.1016/j.joi.2015.01.006>
- Bornmann, L., Marx, W., Gasparyan, A. Y., & Kitas, G. (2012a). Diversity, value and limitations of the journal impact factor and alternative metrics. *Rheumatology International*, 32(7), 1861–1867. <https://doi.org/10.1007/s00296-011-2276-1>
- Bornmann, L., Schier, H., Marx, W., & Daniel, H. D. (2012b). What factors determine citation counts of publications in chemistry besides their quality? *Journal of Informetrics*, 6(1), 11–18. <https://doi.org/10.1016/j.joi.2011.08.004>
- Bornmann, L., Tekles, A., Zhang, H. H., & Ye, F. Y. (2019). Do we measure novelty when we analyze unusual combinations of cited references? A validation study of bibliometric novelty indicators based on F1000Prime data. *Journal of Informetrics*. <https://doi.org/10.1016/j.joi.2019.100979>

- Bozeman, B., & Corley, E. (2004). Scientists' collaboration strategies: Implications for scientific and technical human capital. *Research Policy*, 33(4), 599–616. <https://doi.org/10.1016/j.respol.2004.01.008>
- Bozeman, B., & Boardman, C. (2014). Assessing research collaboration studies: A framework for analysis. In D. B. Audretsch & A. N. Link (Eds.), *Research collaboration and team science. A State-of-the-Art Review and Agenda*. Springer.
- Breugelmans, J. G., Roberge, G., Tippett, C., Durning, M., Struck, D. B., & Makanga, M. M. (2018). Scientific impact increases when researchers publish in open access and international collaboration: A bibliometric analysis on poverty-related disease papers. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0203156>
- Buela-Casal, G., & Zych, I. (2010). Analysis of the relationship between the number of citations and the quality evaluated by experts in psychology journals. *Psicothema*, 22(2), 270–275.
- Callaham, M., Wears, R. L., & Weber, E. (2002). Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *Jama-Journal of the American Medical Association*, 287(21), 2847–2850. <https://doi.org/10.1001/jama.287.21.2847>
- Case, D. O., & Higgins, G. M. (2000). How can we investigate citation behavior? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science and Technology*, 51(7), 635–645. [https://doi.org/10.1002/\(SICI\)1097-4571\(2000\)51:7%3c635::AID-ASI6%3e3.0.CO;2-H](https://doi.org/10.1002/(SICI)1097-4571(2000)51:7%3c635::AID-ASI6%3e3.0.CO;2-H)
- Chen, C. M. (2012). Predictive effects of structural variation on citation counts. *Journal of the American Society for Information Science and Technology*, 63(3), 431–449. <https://doi.org/10.1002/asi.21694>
- Chen, S., Qiu, J., Arsenaault, C., & Larivière, V. (2021). Exploring the interdisciplinarity patterns of highly cited papers. *Journal of Informetrics*. <https://doi.org/10.1016/j.joi.2020.101124>
- Chinchilla-Rodríguez, Z., Miao, L., Murray, D., Robinson-García, N., Costas, R., & Sugimoto, C. R. (2018). A global comparison of scientific mobility and collaboration according to national scientific capacities. *Frontiers in Research Metrics and Analytics*. <https://doi.org/10.3389/frma.2018.00017>
- Chinchilla-Rodríguez, Z., Sugimoto, C. R., & Larivière, V. (2019). Follow the leader: On the relationship between leadership and scholarly impact in international collaborations. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0218309>
- Cole, N. L., Reichmann, S., & Ross-Hellauer, T. (2022). Global Thinking. ON-MERRIT recommendations for maximising equity in open and responsible research. Zenodo. <https://doi.org/10.5281/zenodo.6276753>
- Colebunders, R., Kenyon, C., & Rousseau, R. (2014). Increase in numbers and proportions of review articles in tropical medicine, infectious diseases, and oncology. *Journal of the Association for Information Science and Technology*, 65(1), 201–205. <https://doi.org/10.1002/asi.23026>
- Costas, R., van Leeuwen, T. N., & Bordons, M. (2010). Self-citations at the meso and individual levels: Effects of different calculation methods. *Scientometrics*, 82(3), 517–537. <https://doi.org/10.1007/s11192-010-0187-7>
- Craig, I. D., Plume, A. M., McVeigh, M. E., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact? A critical review of the literature. *Journal of Informetrics*, 1(3), 239–248. <https://doi.org/10.1016/j.joi.2007.04.001>
- Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, 52(7), 558–569. <https://doi.org/10.1002/asi.1097>
- Didegah, F., & Thelwall, M. (2013a). Determinants of research citation impact in nanoscience and nanotechnology. *Journal of the American Society for Information Science and Technology*, 64(5), 1055–1064. <https://doi.org/10.1002/asi.22806>
- Didegah, F., & Thelwall, M. (2013b). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, 7(4), 861–873. <https://doi.org/10.1016/j.joi.2013.08.006>
- Dorta-Gonzalez, P., Dorta-Gonzalez, M. I., Santos-Penate, D. R., & Suarez-Vega, R. (2014). Journal topic citation potential and between-field comparisons: The topic normalized impact factor. *Journal of Informetrics*, 8(2), 406–418. <https://doi.org/10.1016/j.joi.2014.01.013>
- Eysenbach, G. (2006). Citation advantage of open access articles. *Plos Biology*, 4(5), 692–698. <https://doi.org/10.1371/journal.pbio.0040157>
- Fontana, M., Iori, M., Montobbio, F., & Sinatra, R. (2020). New and atypical combinations: An assessment of novelty and interdisciplinarity. *Research Policy*. <https://doi.org/10.1016/j.respol.2020.104063>
- Frame, J. D., & Carpenter, M. P. (1979). International research collaboration. *Social Studies of Science*, 9(4), 481–497.

- Frenken, K., Ponds, R., & van Oort, F. (2010). The citation impact of research collaboration in science-based industries: A spatial-institutional analysis. *Papers in Regional Science*, 89(2), 351–371. <https://doi.org/10.1111/j.1435-5957.2010.00309.x>
- Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *Plos One*. <https://doi.org/10.1371/journal.pone.0013636>
- Gertler, M. S. (2003). Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography*, 3(1), 75–99. <https://doi.org/10.1093/jeg/3.1.75>
- Gilbert, G. N. (1977). Referencing as persuasion. *Social Studies of Science*, 7(1), 113–122.
- Glanzel, W., & Thijs, B. (2004). Does co-authorship inflate the share of self-citations? *Scientometrics*, 61(3), 395–404. <https://doi.org/10.1023/B:SCIE.0000045117.13348.b1>
- Goldfinch, S., Dale, T., & DeRouen, K. (2003). Science from the periphery: Collaboration, networks and “Periphery Effects” in the citation of New Zealand Crown Research Institutes articles, 1995–2000. *Scientometrics*, 57(3), 321–337. <https://doi.org/10.1023/a:1025048516769>
- Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., Stamerjohanns, H., & Hilf, E. R. (2008). The access/impact problem and the green and gold roads to open access: An update. *Serials Review*, 34(1), 36–40. <https://doi.org/10.1016/j.serrev.2007.12.005>
- Haslam, N., & Koval, P. (2010). Predicting long-term citation impact of articles in social and personality psychology. *Psychological Reports*, 106(3), 891–900. <https://doi.org/10.2466/pr0.106.3.891-900>
- Hoekman, J., Frenken, K., & Tijssen, R. J. W. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy*, 39(5), 662–673. <https://doi.org/10.1016/j.respol.2010.01.012>
- Hollanders, H. (2019). *European Innovation Scoreboard 2019 – Methodology Report*. Retrieved from
- Hollanders, H., Es-Sadki, N., Merkelbach, I., & Khalilova, A. (2020). *European Innovation Scoreboard 2020*. Retrieved from https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_1150
- Inglesi-Lotz, R., Balcilar, M., & Gupta, R. (2014). Time-varying causality between research output and economic growth in US. *Scientometrics*, 100(1), 203–216. <https://doi.org/10.1007/s11192-014-1257-z>
- Inglesi-Lotz, R., & Pouris, A. (2013). The influence of scientific research output of academics on economic growth in South Africa: An autoregressive distributed lag (ARDL) application. *Scientometrics*, 95(1), 129–139. <https://doi.org/10.1007/s11192-012-0817-3>
- Katz, J. S., & Hicks, D. (1997). How much is a collaboration worth? *A Calibrated Bibliometric Model*. *Scientometrics*, 40(3), 541–554. <https://doi.org/10.1007/bf02459299>
- Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26(1), 1–18. [https://doi.org/10.1016/s0048-7333\(96\)00917-1](https://doi.org/10.1016/s0048-7333(96)00917-1)
- Larivière, V., & Gingras, Y. (2010). On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology*, 61(1), 126–131. <https://doi.org/10.1002/asi.21226>
- Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, 35(5), 673–702. <https://doi.org/10.1177/0306312705052359>
- Leydesdorff, L., Bornmann, L., & Wagner, C. S. (2019). The relative influences of government funding and international collaboration on citation impact. *Journal of the Association for Information Science and Technology*, 70(2), 198–201. <https://doi.org/10.1002/asi.24109>
- Leydesdorff, L., Radicchi, F., Bornmann, L., Castellano, C., & de Nooy, W. (2013). Field-normalized impact factors (IFs): A comparison of rescaling and fractionally counted IFs. *Journal of the American Society for Information Science and Technology*, 64(11), 2299–2309. <https://doi.org/10.1002/asi.22911>
- McCabe, M. J., & Snyder, C. M. (2014). Identifying the effect of open access on citations using a panel of science journals. *Economic Inquiry*, 52(4), 1284–1300. <https://doi.org/10.1111/eicin.12064>
- Melin, G. (2000). Pragmatism and self-organization - Research collaboration on the individual level. *Research Policy*, 29(1), 31–40. [https://doi.org/10.1016/s0048-7333\(99\)00031-1](https://doi.org/10.1016/s0048-7333(99)00031-1)
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Berlin: University of Chicago Press.
- Miranda, R., & Garcia-Carpintero, E. (2018). Overcitation and overrepresentation of review papers in the most cited papers. *Journal of Informetrics*, 12(4), 1015–1030. <https://doi.org/10.1016/j.joi.2018.08.006>
- Moed, H. F. (2007). The effect of “Open access” on citation impact: An analysis of ArXiv’s condensed matter section. *Journal of the American Society for Information Science and Technology*, 58(13), 2047–2054. <https://doi.org/10.1002/asi.20663>
- Moed, H. F., Burger, W. J. M., Frankfort, J. G., & Vanraan, A. F. J. (1985). The use of bibliometric data for the measurement of university-research performance. *Research Policy*, 14(3), 131–149. [https://doi.org/10.1016/0048-7333\(85\)90012-5](https://doi.org/10.1016/0048-7333(85)90012-5)

- 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>
- Morillo, F., Bordons, M., & Gomez, I. (2003). Interdisciplinarity in science: A tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology*, 54(13), 1237–1249. <https://doi.org/10.1002/asi.10326>
- Narin, F., Stevens, K., & Whitlow, E. S. (1991). Scientific cooperation in Europe and the citation of multinationally authored papers. *Scientometrics*, 21(3), 313–323. <https://doi.org/10.1007/bf02093973>
- NSF. (2009). *Impact of transformative interdisciplinary research and graduate education on academic institutions*. Retrieved from http://www.nsf.gov/pubs/2009/nsf0933/figert_workshop08.pdf
- Ntuli, H., Inglesi-Lotz, R., Chang, T. Y., & Pouris, A. (2015). Does research output cause economic growth or vice versa? Evidence from 34 OECD countries. *Journal of the Association for Information Science and Technology*, 66(8), 1709–1716. <https://doi.org/10.1002/asi.23285>
- Olejniczak, A. J., & Wilson, M. J. (2020). Who's writing open access (OA) articles? Characteristics of OA authors at Ph.D.-granting institutions in the United States. *Quantitative Science Studies*, 1(4), 1429–1450. https://doi.org/10.1162/qss_a_00091
- Onodera, N., & Yoshikane, F. (2015). Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology*, 66(4), 739–764. <https://doi.org/10.1002/asi.23209>
- Oster, S. (1980). The optimal order for submitting manuscripts. *American Economic Review*, 70(3), 444–448.
- Owusu-Nimo, F., & Boshoff, N. (2017). Research collaboration in Ghana: Patterns, motives and roles. *Scientometrics*, 110(3), 1099–1121. <https://doi.org/10.1007/s11192-016-2221-x>
- Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11(6), 619–632. [https://doi.org/10.1002/\(sici\)1099-1255\(199611\)11:6%3c619::Aid-jae418%3e3.0.Co;2-1](https://doi.org/10.1002/(sici)1099-1255(199611)11:6%3c619::Aid-jae418%3e3.0.Co;2-1)
- Peng, T. Q., & Zhu, J. J. H. (2012). Where you publish matters most: A multilevel analysis of factors affecting citations of internet studies. *Journal of the American Society for Information Science and Technology*, 63(9), 1789–1803. <https://doi.org/10.1002/asi.22649>
- Peters, H. P. F., & van Raan, A. F. J. (1994). On the determinants of citation scores - a case-study in chemical-engineering. *Journal of the American Society for Information Science*, 45(1), 39–49. [https://doi.org/10.1002/\(sici\)1097-4571\(199401\)45:1%3c39::Aid-asi5%3e3.0.Co;2-q](https://doi.org/10.1002/(sici)1097-4571(199401)45:1%3c39::Aid-asi5%3e3.0.Co;2-q)
- Potter, R. W. K., Szomszor, M., & Adams, J. (2020). Interpreting CNCIs on a country-scale: The effect of domestic and international collaboration type. *Journal of Informetrics*. <https://doi.org/10.1016/j.joi.2020.101075>
- Puska, H. M., Muhonen, R., & Leino, Y. (2014). International and domestic co-publishing and their citation impact in different disciplines. *Scientometrics*, 98(2), 823–839. <https://doi.org/10.1007/s11192-013-1181-7>
- Shadish, W. R., Tolliver, D., Gray, M., & Sengupta, S. K. (1995). Author judgments about works they cite - 3 studies from psychology journals. *Social Studies of Science*, 25(3), 477–498. <https://doi.org/10.1177/030631295025003003>
- Shibayama, S., Yin, D. Y., & Matsumoto, K. (2021). Measuring novelty in science with word embedding. *Plos One*. <https://doi.org/10.1371/journal.pone.0254034>
- Smith, A. C., Merz, L., Borden, J. B., Gulick, C. K., Kshirsagar, A. R., & Bruna, E. M. (2022). Assessing the effect of article processing charges on the geographic diversity of authors using Elsevier's "Mirror Journal" system. *Quantitative Science Studies*, 2(4), 1123–1143. https://doi.org/10.1162/qss_a_00157
- Sooryamoorthy, R. (2009). Do types of collaboration change citation? Collaboration and citation patterns of South African science publications. *Scientometrics*, 81(1), 177–193. <https://doi.org/10.1007/s11192-009-2126-z>
- StataCorp. (2019). *Stata 16 base reference manual*. Stata Press.
- Storper, M., & Venables, A. J. (2004). Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography*, 4(4), 351–370. <https://doi.org/10.1093/jnlecg/lbh027>
- Subramanyam, K. (1983). Bibliometric studies of research collaboration - a review. *Journal of Information Science*, 6(1), 33–38. <https://doi.org/10.1177/016555158300600105>
- Tahmoorenejad, L., Beaudry, C., & Mirnezami, S. R. (2021). The study of network effects on research impact in Africa. *Science and Public Policy*, 48(4), 462–473. <https://doi.org/10.1093/scipol/scab030>
- Thorsteinsdottir, O. H. (2000). External research collaboration in two small science systems. *Scientometrics*, 49(1), 145–160. <https://doi.org/10.1023/a:1005617426557>
- Uddin, S., Hossain, L., & Rasmussen, K. (2013). Network effects on scientific collaborations. *Plos One*. <https://doi.org/10.1371/journal.pone.0057546>

- Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468–472. <https://doi.org/10.1126/science.1240474>
- Valderas, J. M. (2007). Why do team-authored papers get cited more? *Science*, *317*(5844), 1496–1496. <https://doi.org/10.1126/science.317.5844.1496b>
- van Dalen, H. P., & Henkens, K. (2001). What makes a scientific article influential? *The Case of Demographers*. *Scientometrics*, *50*(3), 455–482. <https://doi.org/10.1023/a:1010510831718>
- Van Raan, A. F. J. (1998). The influence of international collaboration on the impact of research results - Some simple mathematical considerations concerning the role of self-citations. *Scientometrics*, *42*(3), 423–428. <https://doi.org/10.1007/bf02458380>
- Vieira, E. S., & Cerdeira, J. (2022). The integration of African countries in international research networks. *Scientometrics*, *127*(4), 1995–2021. <https://doi.org/10.1007/s11192-022-04297-7>
- Vieira, E. S., Cerdeira, J., & Teixeira, A. A. C. (2022). Which distance dimensions matter in international research collaboration? A cross-country analysis by scientific domain. *Journal of Informetrics*, *16*(2), 16. <https://doi.org/10.1016/j.joi.2022.101259>
- Vieira, E. S., & Gomes, J. (2010). Citations to scientific articles: Its distribution and dependence on the article features. *Journal of Informetrics*, *4*(1), 1–13. <https://doi.org/10.1016/j.joi.2009.06.002>
- Vinkler, P. (1987). A quasi-quantitative citation model. *Scientometrics*, *12*(1–2), 47–72. <https://doi.org/10.1007/bf02016689>
- Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 365–391. <https://doi.org/10.1016/j.joi.2016.02.007>
- Waltman, L., & van Eck, N. J. (2013). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics*, *96*(3), 699–716. <https://doi.org/10.1007/s11192-012-0913-4>
- Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, *46*(8), 1416–1436. <https://doi.org/10.1016/j.respol.2017.06.006>
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed., Vol. 1). Berlin: MIT Press.
- Yan, E. J., & Ding, Y. (2009). Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science and Technology*, *60*(10), 2107–2118. <https://doi.org/10.1002/asi.21128>
- Zdravkovic, M., Chiwona-Karltun, L., & Zink, E. (2016). Experiences and perceptions of South-South and North-South scientific collaboration of mathematicians, physicists and chemists from five southern African universities. *Scientometrics*, *108*(2), 717–743. <https://doi.org/10.1007/s11192-016-1989-z>
- Zhang, X. Y., Xie, Q., & Song, M. (2021). Measuring the impact of novelty, bibliometric, and academic-network factors on citation count using a neural network. *Journal of Informetrics*. <https://doi.org/10.1016/j.joi.2021.101140>
- Zitt, M., & Small, H. (2008). Modifying the journal impact factor by fractional citation weighting: The audience factor. *Journal of the American Society for Information Science and Technology*, *59*(11), 1856–1860. <https://doi.org/10.1002/asi.20880>