The development and application of a bibliometric strength, potential and risk analysis for research strategy in a University Medical Center

Iping, R.; Leeuwen, T.N. van; Noyons, E.; Burdorf, A.; Mathijssen, I.M.J.; Leeuwen, J.P.T.M. van; Cohen, A.M.

Research Evaluation, 2025, **34(1)**, rvaf007
https://doi.org/10.1093/reseval/rvaf007
Article

The development and application of a bibliometric strength, potential and risk analysis for research strategy in a University Medical Center

Rik Iping[1,*], Thed N. van Leeuwen[2], Ed Noyons[2], Alex Burdorf[3], Irene M.J. Mathijssen[4], Johannes P.T.M. van Leeuwen[5], Adrian M. Cohen[1]

[1]Research Intelligence & Strategy Unit, Research Policy and Advice, Erasmus MC, University Medical Center Rotterdam, Rotterdam, 3015 GD, The Netherlands
[2]Centre for Science and Technology Studies (CWTS), Leiden University, Leiden, 2333 BN, The Netherlands
[3]Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, 3015 GD, The Netherlands
[4]Department of Plastic and Reconstructive Surgery, Erasmus MC, University Medical Center Rotterdam, Rotterdam, 3015 GD, The Netherlands
[5]Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, 3015 GD, The Netherlands

*Corresponding author. Research Intelligence & Strategy Unit, Research Policy and Advice, Erasmus MC, University Medical Center Rotterdam, Dr. Molewaterplein 40 3015 GD Rotterdam, The Netherlands. E-mail: r.iping@erasmusmc.nl

Abstract

This paper describes the development of a bibliometric strength, potential and risk analysis tool, and its applications for research strategy and evaluation. We focus specifically on the motivation, organizational strategic needs, the development and evaluation of the tool. Furthermore, we highlight the co-creation process of the tool and discuss the methodology behind the tool, how it works and initial feedback on how insights from the tool can be applied for research strategy and evaluation.

Keywords: bibliometrics; research policy; citation analysis; research intelligence.

1. Introduction

Erasmus MC is one of the largest University Medical Centers (UMC) in The Netherlands. There are over 45 departments covering fundamental biomedical sciences, clinical sciences and health sciences. The broad coverage of disciplines allows developments and innovations in research to be directly translated and applied into clinical practice and societal applications, and at the same time answer new questions from the clinic and society with new research. To keep track of the quality and impact of research and strategic research choices, there are several factors that are regularly monitored such as the ability to attract external funding, successful PhD tracks, talent management and succession planning, impact of research, and academic networks by using research intelligence analyses.

The (potential) impact of research can be assessed in many different ways (e.g. Penfield et al. 2014; Moed and Halevi 2015). Within the context of University Medical Centers, publication- and citation-based indicators have traditionally been emphasized in research impact assessments and supporting research evaluations.

Bibliometric analyses on research output of departments and research lines have been performed since the mid-90s in Erasmus MC. Bibliometric analyses revolve around several aspects of academic performance, such as scientific output, scientific impact as can be measured via citations, as well as visibility in the top of the literature, disciplinary embedding, and scientific cooperation. Scientific impact is expressed as comparing the received number of citations to a certain expected value, which functions as average on a global scale (Waltman et al. 2011). This comparison to an average and within a publication year is called normalization. For a long time, the Journal Subject Categories, as known from the Web of Science core collection, were used as a basis to compute a so-called world average level to compare the citations of publications to, within certain disciplines. These categories are journal based, and journals could be assigned to multiple categories, making citation-based comparisons complex. Furthermore, the journal categories are often very large and heterogeneous, covering a wide range of specialisms within a broader discipline. Because publication and citation behaviour varies greatly across specialisms subject category normalized citation analyses can lead to unfair comparisons (Ruiz-Castillo and Waltman 2015; Waltman and van Eck 2012a). This problem was especially noticeable in small or niche fields, for instance plastic surgery (Iping et al. 2021), medical ethics or rare diseases (Zampeta et al. 2022). Around 2019 researchers came up with an algorithm that could cluster publications based on direct citation relations (Waltman et al. 2020). This algorithm can make clusters of publications that have a link in their content (based on where they cite to and are cited from). These clusters are more suitable to be used as averages to compare citations of publications to, because they create a more level playing field, and reflect a more homogenous body of literature. On top of that they provide a proper context for publications in general and multidisciplinary journals.

Bibliometric analyses have been scrutinized ever since they came into existence, but they rose to great popularity because these are strong tools to quantitatively assess scientific impact and provide decision makers direct insight into their research strategy. There are, however, many examples of heavily criticised bibliometric indicators (such as h-index, and using journal-based metrics, such as the Journal Impact Factor) (e.g. Moed 2002; Waltman and van Eck 2012b; Brito and Rodríguez Navarro 2021), which gained fast popularity because of free availability and quick uptake in database systems. For some years, experts have been advocating a more responsible calculation and use of metrics, and the development of alternative metrics [e.g. Leiden Manifesto (Hicks et al. 2015); Wilsdon 2015; San Francisco Declaration on Research Assessment DORA 2012; Recognition & Rewards 2019; CoARA 2022]. A form of responsible calculation is normalization (internal contextualization), but even more important is the situatedness and interpretation of bibliometric analyses (external contextualization) (e.g. Iping et al. 2022). Analyses can only shed light on a limited part of academic research, and the insights derived from analyses are often only valuable when the context is made clear by the interpreter to the receiver.

Under the pressure of changes in the way we recognize and reward academics, and assess academic research, bibliometric analyses are scrutinized. It is sometimes advocated that they should be disregarded entirely, without offering alternatives. However, with the right context and interpretation bibliometrics still hold much value when acknowledging its limitations. The information is very usable to understand the dynamics of research fields, to understand what your colleagues and competitors are doing, who are the most interesting parties to collaborate with and simply to see how your research resonates in the scientific community, and if not, to explain why.

Our research question was how we could advance the bibliometric analyses that were traditionally provided at Erasmus MC, and transform them into a form of research intelligence, taking into account the context and external developments. Our aim is to provide intelligence to support department heads in making better-informed decisions, but also to help researchers potentially improve their research by giving them strategic leads for collaboration or research focus, while taking into account the limitations. Our methods include interviews with stakeholders, and a process of triangular co-creation (of the RI advisor, department heads, and bibliometric experts) and validation of a tool that was designed based on the needs of the stakeholders. The presented combination of methods and techniques leads to an innovative practical application of bibliometric information as Research Intelligence. This way of working towards a form of applied bibliometry, based on a clear strategic question, is important for the field of research evaluation because it studies applied bibliometry within a shifting context and helps to position it better as a supporting form of evaluation that takes into account contextual elements and can be applied to other organizations when limitations are properly addressed.

However, Research Intelligence is not only the (quantitative and/or qualitative) substance, but involves the whole process of information handling, that is the process in which the Research Intelligence advisor is the representation of both the information, tools and techniques, and the human interaction towards the users of the Research Intelligence. One of the unique aspects addressed in this paper is the co-creation and further shaping of the tool by developers and end-user. The way we have shaped this process and how it is implemented in research evaluation processes is described in this article and to our knowledge provides a unique and original contribution to the field.

2. The use of metric tools in research evaluation contexts

An analysis of the occurrence of scientometric/bibliometric tools in support of research assessment resulted in total 541 publications. These publications cover quite a long period of time, from 1993 till 2023, with a rapidly increasing number over the last couple of years, with the period 2019 till 2024 witnessing three quarters of all publications on the topic being published. So clearly we witness an increased attention for the application of metric tools. However, looking at these publications, we can make a distinction between two types of studies that use metric tools, namely the type that studies a field on a meta-level basis, describing the whereabouts and developments within a research field, next to studies that use such tools for the evaluation of research. There are only five publications using such tools for the evaluation of research, addressing citation impact to be a proxy of visibility by looking at the number of received citations among the top 100 ranked documents. In addition to this search, we conducted a search for the usage of proprietary metric tools, such as InCites from Clarivate and Scival from Elsevier. For these two tools, we retrieved 80 unique publications, covering the time period 2009 till 2024, of which 68 were published between 2019 and 2024. Given that these are in-house facilities, and not open for the general (academic) public, we can clearly observe an increase of an information need regarding the actual assessment of research, a need for which at least these two proprietary metric tools offer answers. Furthermore, we might assume that not all internal usages of these two systems for assessment purposes lead to publications in scholarly journals, so we are probably looking at the tip of the iceberg here, when it is about visibility and reporting of the use of SciVal and InCites for assessment purposes.

3. The co-creation of a bibliometric strength, potential and risk analysis for research strategy

3.1 Analysing the needs—structured interviews with department heads

The way we were able to inform the heads of department at Erasmus MC about bibliometric analyses and their value in the past was limited. We could only provide them basic bibliometric statistics at department level (publication output, normalized citation scores and collaboration metrics) which they used to roughly monitor trends in overall citation impact. These statistics were now calculated using the aforementioned cluster algorithm. In advisory talks with the department heads they indicated that having the overall statistics could not provide them the detailed insights required to use this information to take informed strategic decisions. They lacked a general understanding on how their department scores were constructed. The research intelligence expert was dealing with the same problem, not having enough in-depth information to provide the department heads

strategic insights. Driven by these restrictions and the developments outlined earlier about the changing landscape of applied bibliometrics, structured separate in-depth interviews to gather the needs were carried out among four department heads representing different domains: biomedical, clinical and health sciences. In these interviews we used a standard set of questions probing into the current uses of the citation data at the level of the departments, and the demands and potential use of more in-depth data to support strategic decision-making. Most of the feedback concerning the available citation data focused on the limited information it provided on the dynamics behind the citation score of a department. This score (the mean normalized citation score) is calculated by comparing the actual number of citations of an article to the mean number of citations of all articles from the same year in the same cluster (the expected value or world average). Though this score allows monitoring the citation impact trend of the total body of publications of a department, it lacks the power to show why a score is going up or down. The department heads unanimously agreed that citation-based information on research performance could be used to support strategic decision-making in research, but only if it could show the dynamics underlying the overall scores of the departments. The main questions that could be distilled from the first round of interviews were:

- Can we break down the citation scores of a department to a lower level thereby showing the performance of groups in unbiased clusters?
- Can we use this information to identify strengths and weaknesses?
- Can we identify potentially interesting collaboration partners based on their impact in the clusters in which we are active?

In all the interviews, it was stressed by the heads of department that the information coming from the bibliometric analyses is just one dimension of the total portfolio of research performance of groups and individuals, and it should be treated as such. Regarding the use of the concept performance, we follow the general line of thinking in quantitative science studies, namely that production is the actual number of outputs (of any type, or summed up), while productivity is the production related to the financial input necessary to create that outputs (e.g. the money necessary to fund a certain number of FTEs, possible involving also money forthcoming from external funding) (Aksnes et al. 2016). Having clarified this first, we consider performance as the overall accomplishment in terms of either production or productivity, both are possible, and potentially also involving reception/perception information, as attributed to citation impact measurements.

The second stage of the co-creation process consisted of discussing the strategic questions we identified with the bibliometric experts who subsequently transformed the citation analyses into a bibliometric Strength, Potential and Risk dashboard (SPR).

3.2 The methodology behind the bibliometric Strength, Potential & Risk analysis

The Strength, Potential and Risk (SPR) analysis tool aims at classifying research foci of a unit to potentially feed into strategic decisions. The SPR analysis uses the publication-level classification system of an entire bibliographic database, for example Web of Science (WoS). In this classification, publications are grouped in clusters based on direct citation data, that is publications citing each other. Hence, it is independent from any journal classification. This allows publications to be connected based on their content (assuming that citing to publications is based on matching content) rather than the medium or source (Traag, Waltman and van Eck 2019). This classification can be created on different levels (resolutions) and when based on multi-disciplinary databases like WoS or Scopus, the result is often referred to as the map or landscape of all science. The resolution we apply contains ~4,000 disjoint clusters (research areas, also referred to as micro-clusters).

For the SPR analysis, each research area contains publications from around the globe within a number of years (in our case 2000 onwards). The publication set of a research unit (e.g. department, research group) is projected in this landscape of science. This means that a unit's publications are positioned across research clusters. Consequently, research clusters may contain many, few or no publications from this unit. For each cluster some general statistics are available such as total number of publications, growth over the years (volume per year), citation statistics and the average number of authors per paper. Besides these general statistics, specific information can be calculated for subsets, for example a unit's output. The characteristics at the cluster level at large can be considered as a benchmark to assess the performance of a unit within (like the way the mean normalized citation score mentioned before is calculated).

In the SPR analysis, we relate statistics of a unit to the statistics overall within each area in which it has publications. For instance, a unit may show an increase of output in area A, whereas the area at large shows a saturation (stabilization of publication numbers per year) since recently. Such a result may be interesting to further explore and interrogate. The statistics at the area level plus the results of the unit within the areas in which it is active from the basis of the SPR analysis. SPR areas may be identified by relating the unit's performance to the overall statistics. The definition of a unit's Strength, Potential or Risk area, is not carved in stone. It is to be defined by the unit itself. Examples are presented later on in this paper. Having many publications in an area, for instance, with high citation-based impact is often referred to as a Strength area, but if all unit's researchers are close to retirement, it may also be a Risk area. Once, there is an agreement between the head of the unit and the research intelligence expert on definition of Strength, Potential and Risk areas, taking into account the relevant contextual factors, we define thresholds to identify S, P and R areas. Such thresholds can for instance be based on:

- Output volume
- Citation-based impact
- Growth over years
- Impact of journals used
- Collaboration statistics (international, with private sector)

Once indicators and thresholds are defined they can be incorporated in a web-based tool (dashboard) for the unit to explore the results. Such results focus on the areas in which

the unit is active (i.e. peer-reviewed publication output). However, by providing statistics and characteristics of all areas, one may explore opportunities and potential collaborations in other areas as well.

4. The application of the bibliometric strength, potential and risk analysis for research strategy

4.1 Description of the SPR dashboard

The outcomes of the SPR analyses for all departments at Erasmus MC are presented in a dashboard, and the thresholds to define the S, P and R can be adjusted for each department separately. The first part of the SPR dashboard contains the basic bibliometric indicators that the department heads previously received. The user can select any of the departments that are analysed and can select to filter only publications with a primary author from the department. Primary authorship here is defined as either first, last or corresponding author. The included bibliometric indicators are: number of publications, percentage of publications with external collaboration and international collaboration, mean normalized citation score (MNCS) and percentage of publications belonging to the top 10% most cited publications in their respective clusters (PPtop10%). These are presented both with full counting and fractional counting. Full counting does not apply a weighing factor for number of authors or author position, while factional counting does. All indicators are calculated in periods of 4 years moving up one publication year to show the developments, in a rooftile like fashion, to not lose any sight on particular developments. The indicators are also available per year, to be able to pinpoint developments more accurately. The trends in publication output and MNCS are visualized in a graph (Fig. 1).

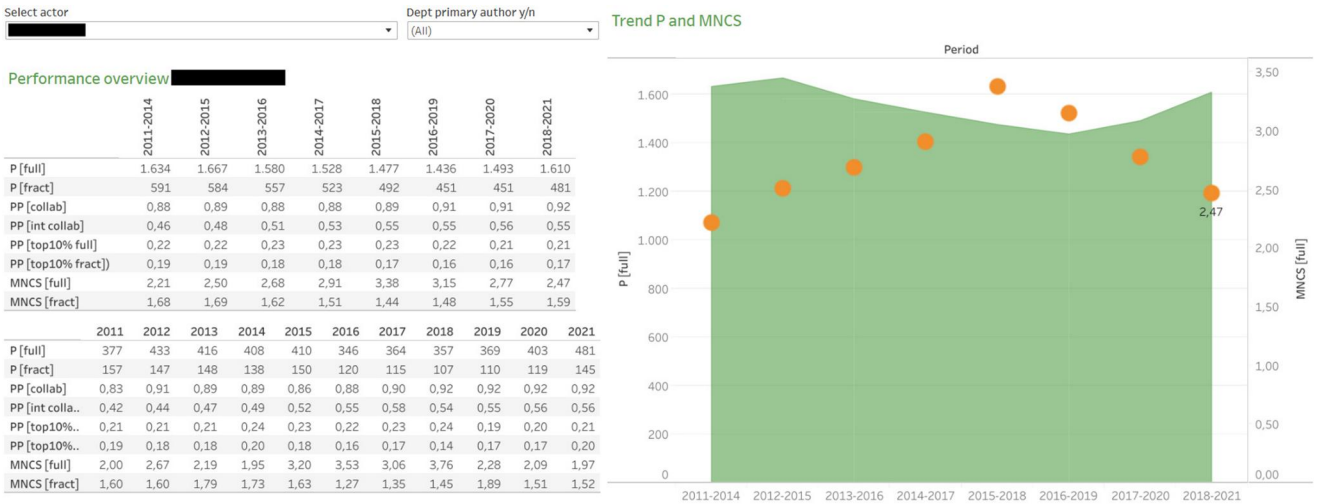
The second part of the dashboard introduces the division of publications in clusters. For the selected department, and selected time period the users see the most prominent clusters in which publications of the department are assigned to, based on algorithmic clustering (Fig. 2).

In this way the user immediately gets an impression of the research portfolio of the department and the size of the different research areas, and it answers the first question we identified from our interviews ‘Can we break down the citation scores of a department on a lower level, showing the performance of groups in unbiased clusters?’. Because the clustering is algorithmic the list of clusters to which publications of a department are assigned to can be long. Therefore, it makes sense to create a cut-off point, for example to only look at clusters with more than 5 publications in 4 years to count as substantial research area for a small or middle-size department, and maybe 10 or more for a large department. For each individual department parameters can be chosen to identify Strength, Potential and Risk areas using filters. For instance, a strength can be a research cluster in which a large department has over 25 publications in 4 years and has an MNCS above 2.0 (twice world average), or a PP(top10%) of above 20%, or both. This functionality provides an answer to the second question identified from the interviews ‘Can we use this information to identify strengths and weaknesses?’. The parameters to choose from may differ per department, because they all differ in size and publication culture. The parameters can best be discussed together with the head of the department who has the most in-depth knowledge of the own department and how they would define a strength, potential or risk. A ‘potential’ can be defined as a cluster in which a department is increasingly publishing and where the citation-based impact is relatively high to very high. And a ‘risk’ could be a cluster in which a department publishes substantially, but the citation impact is lagging behind.

The SPR tool subsequently also helps in putting the information into context by providing an overview of the most prominent institutions world-wide also active in a cluster, the size of their output and PPtop10%, and if there is already collaboration with your own institution. This information provides the answer to the final question we identified from our interviews (Can we identify potentially interesting collaboration partners based on their impact in the clusters in which we are active?). Also, the authors from the department active in the cluster are listed, and the individual publications assigned to the cluster and their citations are listed and can be accessed directly for the tool using the DOI (which stands for Digital Object Identifier).

4.2 Translation to insights for research strategy

The context that all this information described above provides is essential to interpret a Strength, Potential or Risk.



	2011-2014	2012-2015	2013-2016	2014-2017	2015-2018	2016-2019	2017-2020	2018-2021
P [full]	1.634	1.667	1.580	1.528	1.477	1.436	1.493	1.610
P [fract]	591	584	557	523	492	451	451	481
PP [collab]	0,88	0,89	0,88	0,88	0,89	0,91	0,91	0,92
PP [int collab]	0,46	0,48	0,51	0,53	0,55	0,55	0,56	0,55
PP [top10% full]	0,22	0,22	0,23	0,23	0,23	0,22	0,21	0,21
PP [top10% fract])	0,19	0,19	0,18	0,18	0,17	0,16	0,16	0,17
MNCS [full]	2,21	2,50	2,68	2,91	3,38	3,15	2,77	2,47
MNCS [fract]	1,68	1,69	1,62	1,51	1,44	1,48	1,55	1,59

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
P [full]	377	433	416	408	410	346	364	357	369	403	481
P [fract]	157	147	148	138	150	120	115	107	110	119	145
PP [collab]	0,83	0,91	0,89	0,89	0,86	0,88	0,90	0,92	0,92	0,92	0,92
PP [int colla..	0,42	0,44	0,47	0,49	0,52	0,55	0,58	0,54	0,55	0,56	0,56
PP [top10%..	0,21	0,21	0,21	0,24	0,23	0,22	0,23	0,24	0,19	0,20	0,21
PP [top10%..	0,19	0,18	0,18	0,20	0,18	0,16	0,17	0,14	0,17	0,17	0,20
MNCS [full]	2,00	2,67	2,19	1,95	3,20	3,53	3,06	3,76	2,28	2,09	1,97
MNCS [fract]	1,60	1,60	1,79	1,73	1,63	1,27	1,35	1,45	1,89	1,51	1,52

Figure 1. Bibliometric indicators of the SPR dashboard.

Selected clusters: P between 1 and 1.000, MNCS between 0 and 100 and PP[top10%] between 0 and 1

Id	cluster label	P [full]	P [fract]	PP [int collab]	MNCS [full]	PP [top10% full]
448	unemployment; socioeconomic inequality; he..	187	77	0,63	1,84	0,17
62	euthanasia; advance directive; life care; adva..	169	61	0,44	2,03	0,22
287	colorectal cancer screening; colonoscopy; ct c..	140	42	0,54	2,70	0,30
161	concussion; mild traumatic brain injury; sport..	112	23	0,86	3,02	0,37

Figure 2. Clusters of research output and the corresponding bibliometric indicators.

When a Risk or Potential is identified a head of department wants to know what other institutions are doing in this cluster, and if their researchers collaborate with the right partners. They can also find institutions with a similar profile (for instance other Dutch UMCs) in the list and see if how their activity is compared to their own. A citation score can be lower if a cluster is very large and dominated by exceptionally active and/or influential institutions. Then a Risk can be put into context, and it can be concluded that, compared to that highly competitive field, the impact is still significant. On the other hand, the information can also help to define a Potential area or sharpen the parameters comparing it to the total activity in the cluster. If a cluster remains a Risk area the next step can be to discuss this with the researchers involved, by looking in depth at what their competitors are doing, if they are active in the right network, or if their focus can be adjusted. Also, an analysis can be made as to why certain publications are limited or not cited. A Risk area can be a reflection of a research line within a department which is discontinued, or an underperforming research line. This should be added to other indicators (qualitative and quantitative), when a head of department evaluates her or his research groups. We have seen examples of an area that came out as a Risk, but the work was extremely important for clinical or medical practitioners in the field specifically in the Netherlands. This explained the limited number of (international) citations, but because of the context and interpretation this research line would never be actually labelled a Risk.

The four department heads involved in the in-depth interviews were also involved in the evaluation of the first version of the SPR dashboard. We asked them up-front what their expectations were, and which research lines they would list as Strengths, Potential and Risks. In most cases the information from the SPR tool corresponded well with the image that the heads of department have of their research lines, but certainly not in all cases. Sometimes they were surprised by a cluster in which the citation-based impact was high. A concrete example was a research area in which a professor was active. The head of department was not directly aware of the high impact of their work in the field and decided to assign two researchers to expand the research in this line and embed it better in the department. Another example concerned an associate professor with two research lines. When discussing an academic promotion for the associate professor, the head of department wanted him to focus his research more in one direction. The tool clearly showed one research line had significantly higher citation impact. The associate professor confirmed that this line was also more feasible based on the funding and PhD students he had. Together they decided that he should focus his research on this specific topic, and his chair was also aimed in this direction, and finally approved by the Executive Board of the hospital. A third example came from a research group that wanted to organize an international scientific meeting and invite not only the established groups but also emerging groups with potential. We used the SPR tool, together with network analyses to find interesting groups. Several new collaborations came forth from this meeting.

A very clear additional benefit of the SPR tool we observed is that it provides insight in citation dynamics between fields. The clustering methodology was designed to correct for field heterogeneity. Even though this makes the analyses much more meaningful, certain differences even within clusters still exist. The tool helps to understand this. An example of this dynamic is a very concise and demarcated cluster concerning research on a rare disease. In this field roughly two streams of research can be identified: clinical and genetic. The researchers active in genetics dominate this field. This is because the genes they discover that play a role in the origination of the disease are often associated with other diseases, meaning that they often cross-over into other fields allowing their work to receive more citations from outside their own domain. On the other hand, clinical researchers publishing about certain techniques or treatments often only get cited by the small group of their peers that are active worldwide on this rare disease. Intrinsically, a high citation score is improbable. These things are very hard to correct for, but knowing why this happens, because of the insights of the SPR tool, helps researchers to understand the reasons, and shows a head of department that a lower citation score is never direct evidence for lacking research quality. The way to unravel these dynamics is by jointly exploring the rich citation based information that the SPR tool provides: the research intelligence expert understands and explains the bibliometric mechanics, a head of department knows the research field, the active groups and the content of the work that they and others are doing. Together they can combine all insights and draw evidence-based conclusions. The head of department can use this information, combined with other aspects of the work their researchers do to make evidence based strategic decisions. In their reflections the four department heads were very positive about the insights provided by the tool and potential applications to support their strategic decision-making. One called it 'The richest source of information about research they had ever seen'. It is an advanced tool that allows the user to 'ask' questions and adjust parameters to suit their situation. For the research intelligence expert, the tool provides many novel insights they can share with the department heads, and tailor advice to their specific situation. Because advanced bibliometric analyses are often overwhelming to comprehend for an inexperienced user, the role of the expert becomes much more valuable in making the translation of the available information to the strategic demands.

4.3 Broader usability evaluation

After the development with feedback of the pilot departments, we evaluated the usability and insights of the tool by demonstrating it to a group of 15 department heads.

The first question we asked was whether they recognized the information presented in the overview of the SPR dashboard. Overall, the heads of departments found the information presented very recognizable ('*Though the labels can be a bit confusing, the research lines of the department are reflected nicely*'). They indicated that they were impressed by the richness of the information in the dashboard. Almost all of them felt that the explanation provided by the Research Intelligence expert was indispensable, and without it they would not have been able to interpret all the information in the right way and get the most relevant information out. Most questions arose about the SPR clusters. How were they defined? What do the labels mean? Can they be made more flexible to reflect research lines? The overview of clusters was often more accurate in departments with more clearly constrained diseases areas. Research of departments focussing more on techniques and/or methodology spread out over more clusters, making it seem like there is less focus. Most of the performance indicators are in line with gut feelings, but some are surprising. Comparisons within clusters felt more even and just because of a more level playing field.

The second and third question revolved around the value of the presented information, and specific strategic purposes they could think of where they could use the information for.

All heads of department expressed the need for concrete steering information. The types of insights offered by the dashboard can either confirm or deny a gut feeling or paint a more objective picture of something that otherwise remains narrative. It can help in making strategic decisions about funding certain groups, continuing research lines, academic promotions and choosing the right collaborators ('*This information gives an impetus to think about our departmental strategy, it provides handholds to steer and navigate on*'). At this point a lot of decisions are based on rather haphazardly collected information, the need is felt to work towards more data-driven decision-making ('*The presented information is very valuable to me, as it provides a vision on my department from a data-driven perspective*'). Also, it is valuable to be able to steer people earlier on. It can help researchers in their focus, knowing their research cluster(s) and how they collaborate. It informs them about potential opportunities ('*For my researches it is crucial to identify the right collaboration partners. Knowing your field is essential, and this tool can be an eye-opener to some*'). The information also helps to understand field or cluster dynamics and how the principal researchers operate within their field. Bigger clusters with a lot of dynamic are harder to grasp and the relevance of these type of analysis declines. The information can be used to identify hidden gems in a department, or to make impact of groups working in niches more visible. It confirms that a lot of impact is generated through academic networks. Co-publishing, working on guidelines and standards and setting up larger clinical trials takes a lot of effort, but it tends to generate a lot of attention in your scientific field looking at the uptake and citation impact. The need is felt to use this information very carefully and know what it can and cannot tell. It is a good support tool to get an objective look on the research portfolio ('*When it comes to tough decisions, like discontinuing a research line, or deciding where to invest it is very important that it is not based just on a certain gut feeling but that it can be supported from multiple data sources, for instance with this tool*').

We also asked the heads of department about the usefulness of this tool for understanding citation analyses better. They felt that the tool in itself is very rich with information, but only with the right guidance and explanations one is able to extract all useful information ('*I understand it a lot better now, especially with the explanation. There was a lot more information available than I thought upfront*'). While talking and explaining the understanding of advanced citation analyses grew. It helps in understanding dynamics in clusters, get a grip on citation behaviour, where comparisons are based upon. Shortcomings are more visible but understanding those also helps to pinpoint what the actual worth is ('*It definitely helps me to understand better how citation analyses work, but it also makes its shortcomings more visible*').

As a final question we asked them about the potential limitations of the tool, and which information they felt was missing. A number of them commented on the looking-back aspect of the analyses. The articles that are taken into account are results from research years ago. Things have often already changed and results can sometimes only confirm or deny certain choices that were made in the past ('*To be able to steer, I would need more up-to-date information. Even though it might be premature, you should be able to say something about the direction a field is moving in*'). Also, it was mentioned that citations are a very limited view on scientific impact, and they advised us to continue working on tools that take more elements into account. It was mentioned by two department heads that in their opinion, certain document types (reviews) accumulate (too) many citations compared to their actual worth.

Overall, we conclude that the dashboard is a valuable research intelligence tool, providing in depth insights. The value only truly comes out with the right guidance and explanations of a research intelligence expert.

4.4 Embedding the SPR tool in Erasmus MC

The information that can be derived from the SPR tool is quite rich but can also be misinterpreted when not properly explained. We found that the guidance on the interpretation of the collected intelligence by the Research intelligence expert who performed all advisory talks was essential for the proper application of the new insights. The research intelligence expert has an individual meeting with all department heads at Erasmus MC once every two to three years, aligning with the external evaluation cycle (once every six years with a midterm evaluation), and on request when required for strategic discussions or decisions regarding academic promotions. In these meetings the research intelligence analyses are discussed using the SPR dashboard. Trends are discussed, and the research groups are discussed. The heads of department take this information into account when drafting and updating their strategic research plans and talent management plans. Specific analyses can be requested for research groups in the department when deemed valuable. Also, the analyses are often presented at strategy meetings within the departments to serve specific purposes. The SPR tool is not openly accessible for the heads of department because the correct context and interpretation is essential. The basic bibliometric statistics of all departments are openly available for everyone within Erasmus MC. The development of citation impact scores is part of the annual appraisal talks of department heads with the Executive Board. With the information from SPR and explanations of the research intelligence expert

they are now able to put these scores into context and explain certain trends.

Given the relative recent implementation of the tool in the Research Intelligence usages within Erasmus MC, a quantitative assessment of the impact of the tool (effects, usefulness, etc) is not yet possible. However, this is certainly something the Research Intelligence advisor aims to implement in the next few years in order to continuously hone the tool according to the information needs of the heads of department and their teams.

5. Discussion

In this paper we investigated a way to advance bibliometric analyses into research intelligence. Advances in bibliometry allow for more in depth analyses based on citations and content-based analyses. They can provide insight in topical focus, collaborations and comparisons. Information on this level can help decision makers, and researchers to understand the dynamics of the fields they are active in, and give them concrete hand-holds. In order to convey this information, we decided on developing a tool that visualizes and summarizes the information, and is intuitive and easy to use. We approached this process as a unique co-creation between bibliometric experts, a research intelligence advisor and with input directly from decision makers. In this way we could design a tool that was suitable for its target audience and ready for direct implementation.

The tool we developed in this process helps to support informed decision-making in a number of ways:

- The tool assists users in identifying their research strengths, risks, and potential areas to develop, from a publication and citation based viewpoint.
- By clustering publications algorithmically, the tool provides insights into the research portfolio, allowing users to understand citation scores and research dynamics better.
- The context is of key importance in interpreting data from the SPR tool. For instance, a low citation score may not necessarily indicate poor research quality, especially in fields dominated by highly influential institutions.
- Insights from the tool can lead to informed discussions about the future direction of research lines and collaborations, enhancing strategic decision-making.
- The ability to recognize 'hidden gems' or underperforming research lines allows for more focused efforts in enhancing research impact.

Furthermore in this process we learned some valuable lessons about the how to embed Research Intelligence in our organization:

- User feedback indicates that while the tool is rich in information, proper interpretation often requires expert explanation to fully leverage the insights provided.
- The role of research intelligence experts is crucial for making sense of complex bibliometric data and guiding department heads in their strategic planning.
- The effectiveness of the tool is contingent upon expert guidance, suggesting that ongoing support and training may be essential for maximizing its benefits.

In summary, the SPR tool provides a comprehensive and customizable approach to bibliometric analysis that can facilitate data-driven decision-making in research strategy. However, its utility is enhanced when combined with expert interpretation, and there is room for improvement in addressing the timeliness and extent of the data analysed.

5.1 Limitations & Pitfalls

In our evaluations we have come across a number of limitations and pitfalls. The most notable limitations we have come across regarding the content of the dashboard were the looking-back aspect of the data making the insights not always useful for decision-making and the specific dynamics within clusters making comparisons even within departments tough. We also conclude that a small number of highly cited papers can influence the MNCS indicator significantly, especially in departments with smaller numbers of publications. For that reason, we prefer to take into account both MNCS and the PPtop10% indicator, which can together provide the best perspective on the citation based performance of a department. Finally, we found that it is sometimes not intuitive why a paper is assigned to a certain cluster, and why there can be clusters that are very similar. There is no flexibility in merging these. This can have consequences for the recognizability of the data. Finally, the information we use to build the tool is not freely available and needs to be purchased from a provider (in this case Clarivate), and that the tool itself is also a paid-for product. However, the methodology to calculate clusters is openly available and can be reproduced with the right knowledge. A similar tool can be built using open research information.

5.2 Recommendations for future research

In order to assess the added value and the use and implications of this tool, a long term evaluation should be performed to see if users were able to make strategic choices aided by information from this tool, and how it was used in evaluative procedures. Another interesting study would be to apply such a tool in different contexts and disciplines.

A big challenge in the landscape of research information is the aspect of openness. It would be an interesting test case to see if a similar tool can be designed based on openly available metadata. Finally, another future perspective for the tool can be to use it to perform blind-spot analyses to see where research trends are picking up elsewhere and where it might be promising to initiate research for a department.

6. Conclusions

Within a changing landscape of how research evaluation is shaped and embedded in organizations, this study demonstrates a novel application of sophisticated bibliometry. We identified a clear demand by decision makers for quantitative information that could support them in evaluating the research in their department as addition to a more qualitative approach. Their information-needs were on a much more detailed level than we could previously offer, looking deeper into the dynamics of specific fields. In order to get to the level of desired information it was crucial to work in a co-creation process together with bibliometric experts. This way we were able to create a tailor-made tool to suit our needs, but in principle this method can also be easily applied by others to create similar tools.

6.1 The role of the research intelligence expert

In the process of developing the SPR tool, the role of the research intelligence expert was crucial. The expert functioned as a broker between the layer of management and the bibliometrics experts to translate strategic demands into a technical description of a tool to be designed. And in the end, it is up to the research intelligence expert to translate the advanced bibliometric information from the dashboard back in the form of answers on specific questions of the heads of department and the Dean. The research intelligence expert should be someone that understands both worlds. Someone who is knowledgeable about bibliometrics, about the entire research process, about the context in which research is performed and should be valued and evaluated, and about overall department research strategy and the choices department heads are faced with distributing scare resources. Without this full perspective there is a risk of not sending the right information or not receiving it the right way. Finally, the Research intelligence expert is someone that is aware of the broad context of research evaluation, science policy initiatives like 'recognition and reward' and the wider academic drive towards 'Open science', to be able to frame their strategic advice knowing there are many more dimensions of research than just publishing and citing.

6.2 General recommendations and responsible use

The general problem with using bibliometric tools lies in their reliance on quantitative metrics to assess research quality, productivity, and impact, which can often lead to misleading results. These tools prioritize measurable outputs, such as citation counts, impact factors, or h-index, while neglecting qualitative factors like originality, societal relevance, or long-term significance. Data accuracy and completeness pose further challenges, as bibliometric databases may suffer from errors, omissions, or biases, particularly when dealing with interdisciplinary, non-English, or humanities research. Additionally, the norms of publication and citation vary greatly across disciplines, making direct comparisons problematic. Over-reliance on bibliometric indicators can thus oversimplify complex scholarly contributions, encourage superficial research strategies aimed at boosting metrics, and unfairly disadvantage fields or researchers that do not conform to dominant publication practices. Careful and contextualized use of these tools, alongside qualitative assessments, is essential to avoid misrepresenting research performance or value.

Therefore, we recommend to:

- Always apply proper contextualization while using citation-based insights leading to the proper interpretation.
- Always provide such contextualized citation-based insights to support decision-making, and avoid providing research metrics as sole ground for decision-making.
- Always provide the guidance of a research intelligence expert, who understands the context in which the SPR tool is used, to interpret the information and translate it into usable insights.

Author contributions

Rik Iping conceptualized the idea, performed the interviews and guided the development. Thed N. van Leeuwen and Adrian M. Cohen supervised the development. Ed Noyons and Thed N. van Leeuwen developed the SPR methodology and tool. Rik Iping wrote the original draft. Thed N. van Leeuwen, Adrian M. Cohen, Ed Noyons, Alex Burdorf, Irene M.J. Mathijssen and Johannes P.T.M. van Leeuwen reviewed and edited the manuscript.

Funding

This work was supported by Erasmus MC, which funded the development of the tool that was described in this paper.

References

Aksnes, D. et al. (2016) 'Measuring the Productivity of National R&D–Systems: Challenges in Cross-National Comparisons of R&D-Input and Publication-Output Indicators', *Science and Public Policy*, 44: scw058. https://doi.org/10.1093/scipol/scw058

Brito, R., and Rodríguez Navarro, A. (2021) 'The Inconsistency of h-Index: A Mathematical Analysis', *Journal of Informetrics*, 15: 101106. https://doi.org/10.1016/j.joi.2020.101106

CoARA. (2022). Agreement on Reforming Research Assessment. https://coara.eu/agreement/the-agreement-full-text/

Hicks, D. et al. (2015) 'Bibliometrics: The Leiden Manifesto for Research Metrics', *Nature*, 520: 429–31. https://doi.org/10.1038/520429a

Iping, R. et al. (2021) 'A Bibliometric Overview of Craniosynostosis Research Development', *European Journal of Medical Genetics*, 64: 104224. https://doi.org/10.1016/j.ejmg.2021.104224

Iping, R. et al. (2022) 'A Research Intelligence Approach to Assess the Research Impact of the Dutch University Medical Centres', *Health Research Policy & Systems*, 20: 118. https://doi.org/10.1186/s12961-022-00926-y

Moed, H. F. (2002) 'The Impact-Factors Debate: The ISI's Uses and Limits', *Nature*, 415: 731–2. https://doi.org/10.1038/415731a

Moed, H. F., and Halevi, G. (2015) 'Multidimensional Assessment of Scholarly Research Impact', *Journal of the Association for Information Science and Technology*, 66: 1988–2002.

Penfield, T. et al. (2014) 'Assessment, Evaluations, and Definitions of Research Impact: A Review', *Research Evaluation*, 23: 21–32.

Recognition & Rewards. (2019). Position Paper. https://recognitionrewards.nl/

Ruiz-Castillo, J., and Waltman, L. (2015) 'Field-Normalized Citation Impact Indicators Using Algorithmically Constructed Classification Systems of Science', *Journal of Informetrics*, 9: 102–17. https://doi.org/10.1016/j.joi.2014.11.010

San Francisco Declaration on Research Assessment. (2012). https://sfdora.org/read/

Traag, V. A., Waltman, L., and van Eck, N. J. (2019) 'From Louvain to Leiden: guaranteeing Well-Connected Communities', *Scientific Reports*, 9: 5233. https://doi.org/10.1038/s41598-019-41695-z

Waltman, L., and van Eck, N. J. (2012a) 'A New Methodology for Constructing a Publication-Level Classification System of Science', *Journal of the American Society for Information Science and Technology*, 63: 2378–92. https://doi.org/10.1002/asi.22748

Waltman, L., and van Eck, N. J. (2012b) 'The Inconsistency of the h-Index', *Journal of the American Society for Information Science and Technology*, 63: 406–15. https://doi.org/10.1002/asi.21678

Waltman, L. et al. (2011) 'Towards a New Crown Indicator: Some Theoretical Considerations', *Journal of Informetrics*, 5: 37–47. https://doi.org/10.1016/j.joi.2010.08.001

Waltman, L. et al. (2020) 'A Principled Methodology for Comparing Relatedness Measures for Clustering Publications', *Quantitative Science Studies*, 1: 1–23. https://doi.org/10.1162/qss_a_00035

Wilsdon, J. (2015) *The Metric Tide*. SAGE Publications Ltd. https://doi.org/10.4135/9781473978782

Zampeta, F. I. et al. (2022) 'From First Report to Clinical Trials: A Bibliometric Overview and Visualization of the Development of Angelman Syndrome Research', *Human Genetics*, 141: 1837–48. https://doi.org/10.1007/s00439-022-02460-x

Research Evaluation, 2025, 34, 1–8
https://doi.org/10.1093/reseval/rvaf007
Article