



RESEARCH ARTICLE

# Overton: A bibliometric database of policy document citations

Martin Szomszor<sup>1</sup>  and Euan Adie<sup>2</sup> 

<sup>1</sup>Electric Data Solutions LTD, London, UK

<sup>2</sup>Overton, London, UK

an open access  journal



Citation: Szomszor, M., & Adie, E. (2022). Overton: A bibliometric database of policy document citations. *Quantitative Science Studies*, 3(3), 624–650. [https://doi.org/10.1162/qss\\_a\\_00204](https://doi.org/10.1162/qss_a_00204)

DOI: [https://doi.org/10.1162/qss\\_a\\_00204](https://doi.org/10.1162/qss_a_00204)

Peer Review: [https://publons.com/publon/10.1162/qss\\_a\\_00204](https://publons.com/publon/10.1162/qss_a_00204)

Received: 28 January 2022  
Accepted: 17 June 2022

Corresponding Author:  
Martin Szomszor  
[martin@electricdata.solutions](mailto:martin@electricdata.solutions)

Handling Editor:  
Ludo Waltman

Copyright: © 2022 Martin Szomszor and Euan Adie. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



**Keywords:** bibliometrics, citation metrics, impact assessment, Overton, policy influence, research evaluation

## ABSTRACT

This paper presents an analysis of the Overton policy document database, describing the makeup of materials indexed and the nature in which they cite academic literature. We report on various aspects of the data, including growth, geographic spread, language representation, the range of policy source types included, and the availability of citation links in documents. Longitudinal analysis over established journal category schemes is used to reveal the scale and disciplinary focus of citations and determine the feasibility of developing field-normalized citation indicators. To corroborate the data indexed, we also examine how well self-reported funding outcomes collected by UK funders correspond to data indexed in the Overton database. Finally, to test the data in an experimental setting, we assess whether peer-review assessment of impact as measured by the UK Research Excellence Framework (REF) 2014 correlates with derived policy citation metrics. Our findings show that for some research topics, such as health, economics, social care, and the environment, Overton contains a core set of policy documents with sufficient citation linkage to academic literature to support various citation analyses that may be informative in research evaluation, impact assessment, and policy review.

## 1. INTRODUCTION

The premise that academic research leads to wider social, cultural, economic, and environmental benefits has underpinned our investment in publicly funded research since the 1950s (Bush, 1945). It was broadly accepted that research leads to positive outcomes (Burke, Bergman, & Asimov, 1985), but this belief was further scrutinized as technical analyses were developed to unpick the exact nature and scale of these impacts (Evenson, Waggoner, & Ruttan, 1979). The types of evaluation became more varied and complex as the investigators focused on specific domains (Hanney, Packwood, & Buxton, 2000; van der Meulen & Rip, 2000), taking into account the myriad ways in which knowledge is generated, exchanged, assimilated, and utilized outside of academia. The general assumption holds that there is a return on investment in research through direct and indirect mechanisms (Salter & Martin, 2001) and the most recent literature reviews (Bornmann, 2013; Greenhalgh, Raftery et al., 2016; Penfield, Baker et al., 2014) provide detailed perspectives on how to identify and differentiate between outputs and outcomes across a range of settings.

Research evaluation also developed to support a greater need for accountability (Thomas, Nedeva et al., 2020): initially, by peer review (Gibbons & Georghiou, 1987), then strategic

reorientation (Georghiou, 1995), and recently using more data-driven approaches that incorporate bibliometric components (Adams, Gurney, & Marshall, 2007; Hicks, 2010; Hicks & Melkers, 2013; Martin, 1996). Despite shortcomings in their suitability to judge research quality (Moed, Burger et al., 1985; Pendlebury, 2009), citation indicators became more popular (May, 1997) due to their growing availability, relatively low cost compared with conventional peer review, and ready application to national, regional, and institutional portfolios (BEIS, 2017). Current evaluation programs that consider citation data include: Australia (ARC, 2018), EU (Reinhardt & Milzow, 2012), Finland (Lahtinen, Koskinen-Ollonqvist et al., 2005), Italy (Abramo & D'Angelo, 2015), New Zealand (Buckle & Creedy, 2019), Norway (Sivertsen, 2018), Spain (Jiménez-Contreras, Anegón et al., 2003), United Kingdom (REF2020, 2020) and United States (NIH, 2008).

However, growing use of bibliometric indicators also altered researcher behaviors via corrupted incentives, leading to a variety of negative outcomes (Abramo, D'Angelo, & Grilli, 2021; Butler, 2003; Lopez Pineiro & Hicks, 2015; Yücel & Demir, 2018) and motivating various groups to call for more nuanced and equitable research assessment, such as in the San Francisco Declaration on Research Assessment (DORA) (Cagan, 2013), Metrics Tide report (Wilsdon, Allen et al., 2015), and Leiden Manifesto (Hicks, Wouters et al., 2015). This has resulted in publishers, research organizations, and funders signing up to the aforementioned initiatives and developing their own policies to ensure metrics are deployed and used responsibly. A key aspect has been a push towards broad recognition of research contributions (Morton, 2015) and a more nuanced use of bibliometric indicators (Adams, Marie et al., 2019).

Throughout this growth and development in the use of metrics, it has become clear that standard citation indicators reflect only the strength of influence within academia and are unable to measure impact beyond this realm (Moed, 2005; Ravenscroft, Liakata et al., 2017). This has led to the exploration of adjacent data sources to provide signals of the wider impact of research, which have been collectively named altmetrics (Priem, Taraborelli et al., 2010). This term refers to a range of potential data sources that could potentially reveal educational impact (Kousha & Thelwall, 2008; Mas-Bleda & Thelwall, 2018), knowledge transfer (Kousha & Thelwall, 2017), commercial use (Orduna-Malea, Thelwall, & Kousha, 2017), public engagement (Shema, Bar-Ilan, & Thelwall, 2015), policy influence (Tattersall & Carroll, 2018), and more. With access to a broader range of indicators, it may be possible to address some contemporary research evaluation issues by increasing the scope of how research is measured and allow the full range of research outcomes to be attributed to researchers.

In the area of policy influence, the research underpinning clinical guidelines, economic policy, environmental protocols, etc. is a significant topic of interest. Analysis of the REF2014 Impact Case Study data (Grant, 2015) showed that 20% of case studies were associated with the topic *Informing government policy*, and 17% were associated with *Parliamentary scrutiny*, most frequently in Panel C (social sciences). In many cases, evidence cited in case studies included citations to the research from national and international policy organizations. In Unit of Assessment 1 (clinical medicine), 41% of case studies were allocated to the topic *Clinical guidance*, indicating some use of the academic research in policy setting.

Since 2019, a large database of policy documents and their citations to academic literature has been developed by Overton (see [overton.io](http://overton.io)). As of December 2021, it indexes publications from more than 30,000 national and international sources including governments, think tanks, intergovernmental organizations (IGOs), and charities. The focus of this paper is to evaluate Overton as a potential bibliometric data source using a series of analyses that investigate the makeup of documents indexed (e.g., by geography, language, and year of publication), the

network of citations (e.g., volume, distribution, time-lag), and how well data correlate with other impact logging processes (e.g., as reported to funders). An example analysis is also provided to show how Overton data can be used to test whether peer-review scores correlate with derived citation metrics. In doing so, it is our hope to understand more about the potential uses of policy citation data by highlighting which disciplines are most frequently cited and if citation volumes are sufficient to support the development of citation indicators.

The paper is structured as follows: Section 2 summarizes related work. Section 3 presents the methodology for each experiment and outlines the data sets used. Section 4 presents the results of each analysis before discussion in Section 5.

## 2. RELATED WORK

The traditional bibliometric databases, namely the Web of Science (Clarivate), Scopus (Elsevier), Dimensions (Digital Science), Microsoft Academic (Microsoft), and Google Scholar (Google), have been extensively evaluated (Aksnes & Sivertsen, 2019; Chadegani, Salehi et al., 2013; Falagas, Pitsouni et al., 2008; Harzing & Alakangas, 2016; Visser, van Eck, & Waltman, 2021), particularly in terms of cited references (Martín-Martín, Thelwall et al., 2021), subject coverage (Martín-Martín, Orduna-Malea et al., 2018), comparability of citation metrics (Thelwall, 2018), journal coverage (Mongeon & Paul-Hus, 2016; Singh, Singh et al., 2021), classification systems (Wang & Waltman, 2016), accuracy of reference linking (Alcaraz & Morais, 2012; Olensky, Schmidt, & van Eck, 2016), duplication (Valderrama-Zurián, Aguilar-Moya et al., 2015), suitability for application with national and institutional aggregations (Guerrero-Bote, Chinchilla-Rodríguez et al., 2021), language coverage (Vera-Baceta, Thelwall, & Kousha, 2019), regional bias (Rafols, Ciarli, & Chavarro, 2020; Tennant, 2020), and predatory publishing (Björk, Kanto-Karvonen, & Harviainen, 2020; Demir, 2020). The notion of best data source is partly subjective (i.e., depending on personal preference), but also depends on the type of use (e.g., search and discovery versus bibliometric analysis), discipline, regional focus, and time period in question, and can be influenced by the availability of metadata and links to adjacent data sets (e.g., patents, grants, clinical trials, etc.), depending on task.

Much like the preference for bibliographic data source, the choice of citation impact indicator (Waltman, 2016) is highly debatable. It is generally accepted that citations should be normalized by year of publication, discipline, and document type, although whether the calculation should be based on the average of ratios (Opthof & Leydesdorff, 2010; Waltman, van Eck et al., 2011) or ratio of averages (Moed, 2010; Vinkler, 2012) is contentious (Larivière & Gingras, 2011), as is the selection of counting methodology (Potter, Szomszor, & Adams, 2020; Waltman & van Eck, 2015). Suitable sample size is key to providing robust outcomes (Rogers, Szomszor, & Adams, 2020), and any choices made with respect to category scheme used and indicator choice should influence interpretation of results (Szomszor, Adams et al., 2021).

The potential for use of altmetric indicators was initially focused on the prediction of traditional citations (Thelwall, Haustein et al., 2013) and possible correlation with existing indicators (Costas, Zahedi, & Wouters, 2015; Zahedi, Costas, & Wouters, 2014). It was suggested that “little knowledge is gained from these studies” (Bornmann, 2014) and that the biggest potential for altmetrics was toward measurements of broader societal impact (Bornmann, 2015). At this point, the coverage of altmetrics was limited to social media attention (e.g., Twitter and Facebook mentions), usage metrics (e.g., website downloads, Mendeley readers), and online news citations (both traditional and blogs). Comparisons with peer-review assessment

(Bornmann & Haunschild, 2018a) revealed that Mendeley readership was the most strongly associated of these with high-quality research, but still much less than conventional citation indicators. More recent analysis (Bornmann, Haunschild, & Adams, 2019) have incorporated other altmetric indicators, showing Wikipedia and policy document citations to have the highest correlation with REF Impact Case study scores out of the available indicators. Bornmann, Haunschild, and Marx (2016) conclude “Policy documents are one of the few altmetrics sources which can be used for the target-oriented impact measurement.” To date, Overton data has been utilized in a small number of studies, including an investigation of how cross-disciplinary research can increase the policy relevance of research outcomes (Pinheiro, Vignola-Gagné, & Campbell, 2021), and the interactions between science and policy making during the COVID-19 pandemic (Gao, Yin et al., 2020; Yin, Gao et al., 2021). Most recently, Bornmann, Haunschild et al. (2022) explore how climate change research is cited in climate change policy, uncovering the complexities of how research is translated into policy setting.

Prior work investigating the translation of research through citations in clinical guidelines (Grant, 2000; Kryl, Allen et al., 2012; Newson, Rychetnik et al., 2018) have utilized specific data sources (often requiring significant manual curation) to show their value in evaluating research outcomes. Databases of clinical practice guidelines have emerged (Eriksson, Billhult et al., 2020) to support this specific line of inquiry, and recent work (Guthrie, Cochrane et al., 2019; Pallari, Eriksson et al., 2021; Pallari & Lewison, 2020) utilizes this information to uncover national trends and highlight relative differences in the evidence base used.

Patent data citations are another important data source that have been utilized in studies relating to the wider impact of scientific research (van Raan, 2017), usually for tracking technology transfer (Alcácer & Gittelman, 2006; Meyer, 2000; Roach & Cohen, 2013) or industrial R&D links (Tijssen, Yegros-Yegros, & Winnink, 2016), and often in the context of national assessment (Carpenter, Cooper, & Narin, 1980; Chowdhury, Koya, & Philipson, 2016; Narin & Hamilton, 1996) and convergence research (Karvonen & Kässi, 2013). Notably, recent research casts doubt on the suitability of patent data citations for these purposes (Abrams, Akcigit, & Grennan, 2018; Kuhn, Younge, & Marco, 2020) due to changes in citation behaviour and growth in the use of the patent system as a strategic instrument.

### 3. METHODOLOGY

The Overton database is the primary source of data for this study. It is created by web-crawling publicly accessible documents published by a curated list of over 30,000 organizations, including governments, intergovernmental organizations, think tanks, and charities. Each document is processed to extract bibliographic information (title, authors, publication date, etc.) along with a list of cited references, including those to academic literature as well as other policy documents. Technical details regarding the reference matching process can be found on the Overton website (Overton, 2022). A policy document itself may be composed of multiple items, referred to herein as *PDFs* because they are the majority format type, such as clinical guidelines (which contain separate documents with recommendations and evidence bases) or when language translations exist. The types of documents vary in nature and include reports, white papers, clinical guidelines, parliamentary transcripts, legal documents, and more, intended for a variety of audiences, including journalists, policy makers, government officials, and citizens. Generally speaking, Overton seeks to index materials written by a policy maker or primarily for a policy maker.

Overton classifies publication sources using a broad taxonomy that is further subdivided by type. Top-level source types are: government, intergovernmental organizations (igo), think

tank, and other. Subtypes include bank, court, healthcare agency, research center, and legislative. Each publication source is assigned a geographic location, including country and region (e.g., state or devolved territory). Some sources are classified as IGO (i.e., global reach), or EU (European Union).

For this study, 4,504,896 policy documents (made up of 4,854,919 individual PDFs) citing 3,579,710 unique articles (DOIs) were used. To integrate this data with other sources, all records were converted into Resource Description Framework (RDF) (Bizer, Vidal, & Weiss, 2018), a semantic web metadata model, and loaded into a the graph database GraphDB™. The following additional data sources were used:

- **Crossref:** Metadata for all DOIs were extracted from Crossref records providing titles, source names (i.e., journal), collection identifiers (ISSNs and ISBNs), and publication dates.
- **Scopus journal categories:** As determined by linking ISSN to Crossref records, each journal is associated with up to 13 All Science Journal Classification (ASJC) categories for journals, organized in a hierarchy under areas and disciplines. ( $n = 19,555$ ). Source: [scopus.com](http://scopus.com).
- **REF2014 Case Studies:** All publicly available case studies submitted to REF2014 and the associated DOIs mentioned in the references section. A total of 6,637 Case Studies were included, linking to 24,945 unique DOIs. Source: [impact.ref.ac.uk](http://impact.ref.ac.uk).
- **REF2014 Results:** The final distribution of scores awarded in REF2014. For each Institution and UoA, scores for Outputs and Case Studies were loaded, expressed as the percentage of outputs in categories 4\* (world-leading), 3\* (internationally excellent), 2\* (internationally recognized), and 1\* (nationally recognized). Source: [results.ref.ac.uk](http://results.ref.ac.uk).
- **Gateway to Research (GTR):** All funded projects from UKRI Research Councils ( $n = 123,399$ ), their associated publications ( $n = 1,015,664$ ), and outcomes categorized as policy outcome ( $n = 39,406$ ). Source: [gtr.ukri.org](http://gtr.ukri.org).

This combination of information allows us to investigate a range of questions that will inform the potential viability of Overton as a bibliometric data source:

1. *What is the makeup of the database in terms of sources indexed by geography, language, type, and year of publication?* This analysis will determine, by year of publication, the count of policy documents and PDFs indexed according to source type, region, country, and language. This will reveal potential biases in coverage that would inform suitability for certain types of analysis. Overton does contain locally relevant policy sources, such as regional government publications, but not for all geographies.
2. *How many scholarly references are extracted and over what time period?* This will measure the total number of references to DOIs extracted according to policy publication year and source type, and show the count of citations received to DOIs by their publication year according to broad research area. It is important to know how many citations to research articles are tracked because the volume will inform their suitability for citation-based indicator development.
3. *How long does it take research articles to accumulate policy citations and how does this vary across disciplines?* This will provide details on how long DOIs take to accumulate citations, both in absolute volume per year, and cumulatively. Research areas and disciplines will be analyzed separately to illustrate any differences and to highlight domains in which citation analysis may be fruitful.



4. *What is the time lag between the publication of scholarly works and their citation within policy literature and how does this vary between disciplines?* This will show the distribution from the perspective of citing policy document (i.e., how old are cited references?), and from cited DOI (i.e., when are citations to research articles received?). A sample of policy sources for healthcare agencies and governmental banks is also benchmarked to illustrate feasible comparisons. The range and timeliness of evidence used is an important consideration in policy evaluation and may be possible using the Overton database.
5. *What statistical distribution best models policy citation counts to research articles?* This will test the fit of various distributions (e.g., power law, lognormal, exponential) to empirical data using conventional probability distribution plots. Analysis by research discipline and subject will be used to inform potential field-based normalization techniques (i.e., appropriate level of granularity).
6. *How feasible is field-based citation normalization?* This will determine if a minimum sample size can be created for each subject category and year for DOIs published between 2000 and 2020. This analysis will highlight subjects that may be suitable for citation metrics and those where insufficient data are available to make robust benchmarks.
7. *Do the citations tracked in the policy literature correlate with policy influence outcomes attributed to funded grants?* This will test the correlation between policy influence outcomes reported against funded grants (submitted via the ResearchFish platform to UKRI), and the number of Overton policy citations from DOIs specified as outputs of these projects. Correlations will also be calculated for each subject according to the GTR classification.
8. *Does the amount of policy citation correlate with peer-review assessment scores as reported in the UK REF2014 impact case study data?* This will test size-independent correlation (Traag & Waltman, 2019) between normalized policy citation metrics (percentiles) and peer-review assessment (according to 4\* rating). Percentiles are calculated based on year of publication and Scopus ASJC subject categories.

To analyze data by research subjects, disciplines, and areas, we utilize the Scopus ASJC journal subject mapping. This is the preferred categorical system for this analysis because it is matched to the highest number of journals in the data set (compared to Web of Science journal categories or the ScienceMetrix journal classification), and offers three levels of aggregation (*areas* → *disciplines* → *subjects*).

## 4. RESULTS

### 4.1. What Is the Makeup of the Database in Terms of Sources Indexed (by Geography, Language, Type and Year of Publication)?

The growth of documents indexed in Overton is depicted in Figure 1. Four plots are included 1(a) the number of documents according to publication source type (government, think tank, igo, and other); 1(b) the number of documents indexed according to publication source region; 1(c) by publication source country (top 20); and 1(d): publication language (top 20). As mentioned earlier, a policy document may contain multiple PDFs, typically language translations or different parts of a larger report or set of guidelines. The total number of PDFs indexed is shown with a dotted line in Figure 1(a), which also corresponds to the total in Figure 1(d) because PDFs are associated with languages rather than the policy document container (i.e., a single policy document may exist in multiple languages as different PDFs). It

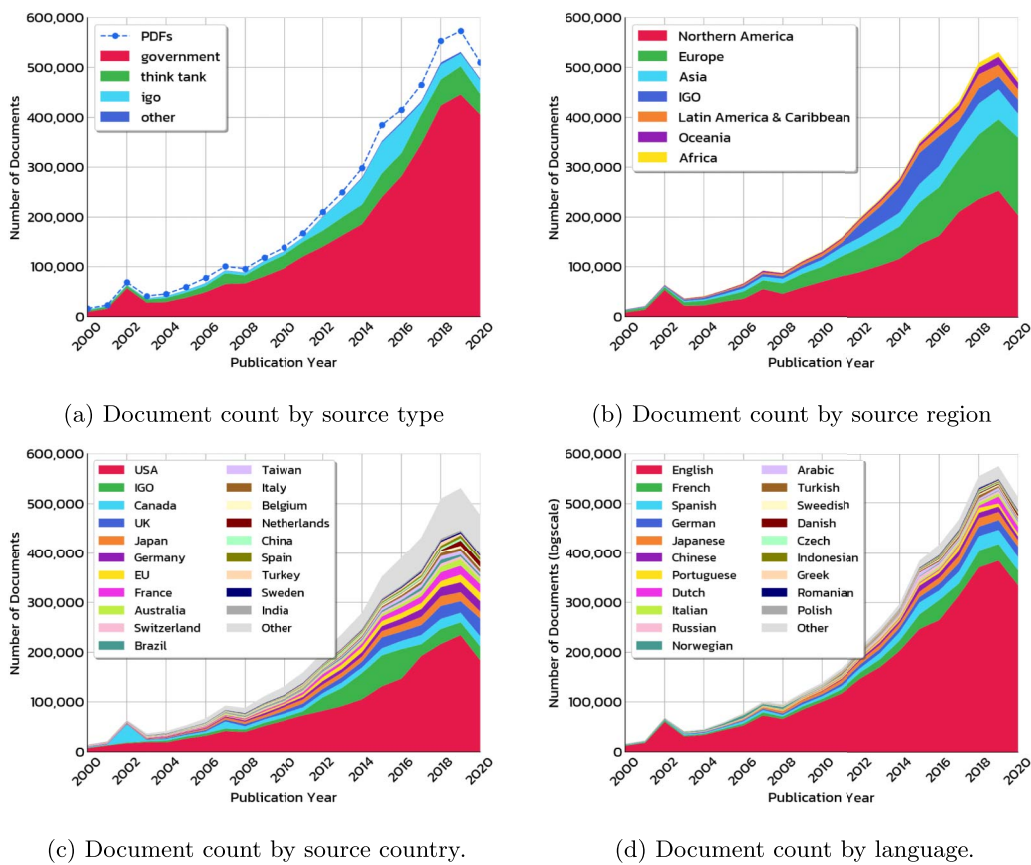


Figure 1. Time-series of Overton policy document count.

should be noted that while there is a significant growth in the total number of documents indexed, this doesn't necessarily correlate to a growth in the publication of policy documents overall—it only reflects how many resources are currently discoverable on the web. In this sense, our analysis shows that the availability of data is improving.

To illustrate global coverage, we also supply a map in Figure 2. The map includes an insert showing the number of documents indexed for the top eight regions. Due to the scale difference between the large number of documents indexed from the United States compared with other countries, four color bins are used rather than a straightforward linear gradient.

Clearly, Overton is dominated by policy documents published by sources in the United States, but it also includes significant coverage for Canada, the United Kingdom, Japan, Germany, France, and Australia, with the majority of content originating from governmental sources. The IGO grouping (including organizations such as the WHO, UNESCO, World Bank, and United Nations) and the European Union also make up a sizable portion of the database. In terms of the makeup of sources and languages, Figure 3 is included to show the percentage makeup of documents from the top 30 regions according to source type (left) and language (middle-left). For language, three values are shown: those in English, those in a local language, and those in other languages. For the regions IGO and EU, no local languages are specified. For reference, the total policy document count for each is shown (middle-right, log scale), along with the 2018 count of articles attributed to the country in the SCImago journal ranking.

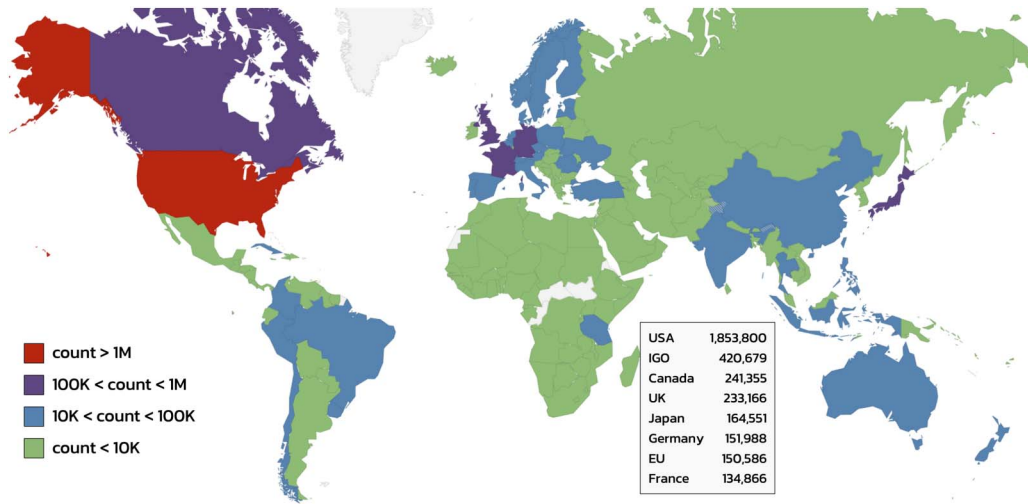


Figure 2. Map showing the volume of policy documents indexed by country.

The balance of source types in each country does vary, with some regions almost entirely represented by governmental sources, such as Japan, Taiwan, Turkey, and Uruguay. The unusually high percentage of documents from Australian sources categorized as other is due to articles indexed from the Analysis & Policy Observatory (also known as *APO*). Another large aggregator, PubMed Central, is also indexed by Overton (for practice and clinical

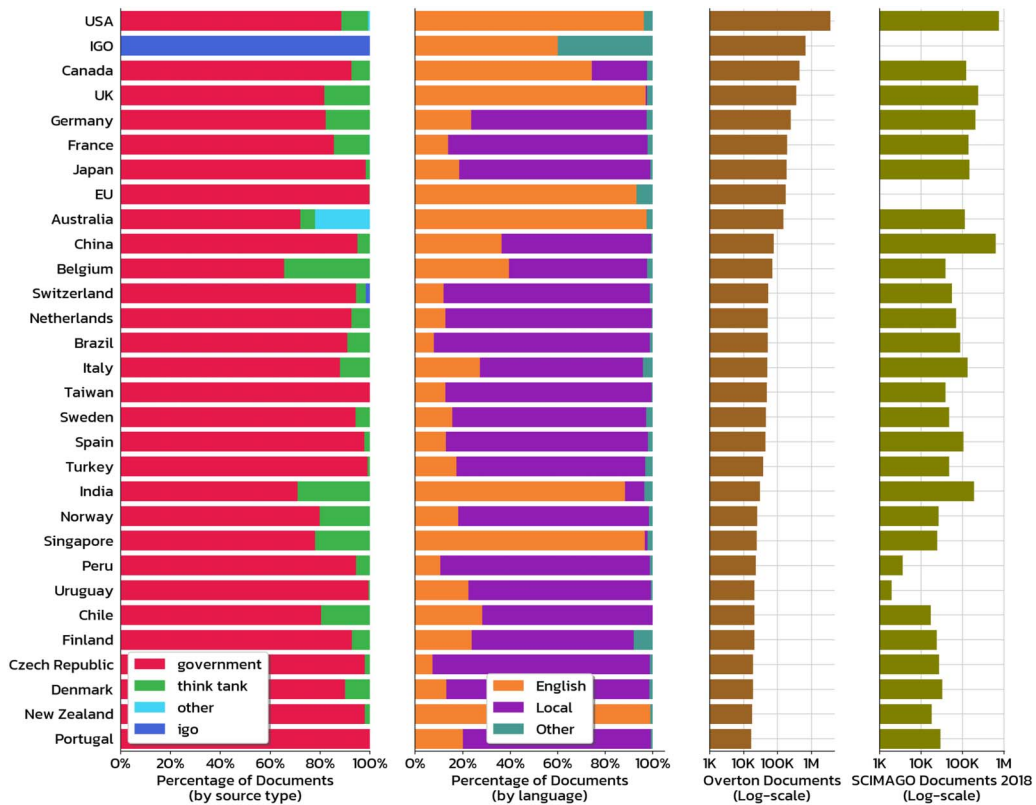


Figure 3. Make up of policy documents by country.



guidelines), but is attributed to the United States and hence only appears as a small fraction of their output, which is very large overall.

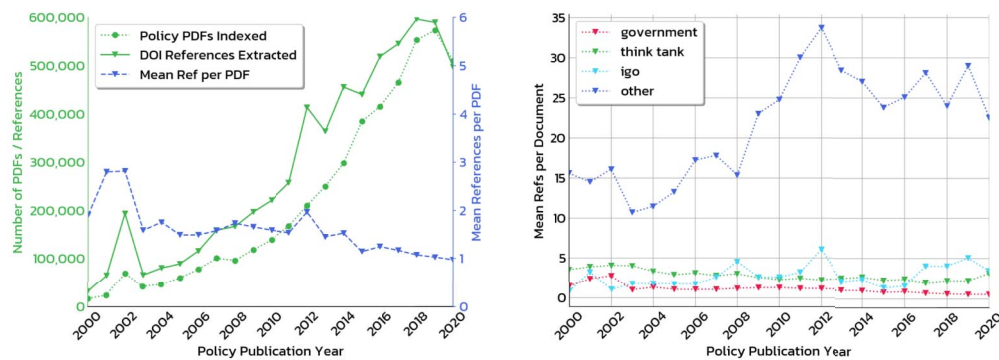
In terms of language balance, many countries have a significant proportion of content in local languages—more than 80% for France, Japan, Switzerland, Netherlands, Brazil, Taiwan, Sweden, Spain, Norway, Peru, Czech Republic, and Denmark. Those that do not are either English-speaking (United States, United Kingdom, Australia, New Zealand) or have strong colonial ties (India and Singapore).

The comparison of Overton content to SCIMago article count is included to show possible over- and underrepresentation. For example, China produces the second largest number of academic articles (after the United States) but is only the eighth most frequently indexed country (excluding IGO and EU) in Overton. In contrast, Peru and Uruguay produce a much lower number of research articles than Brazil and Chile, but a similar amount of content is indexed in Overton.

#### 4.2. How Many Scholarly References Are Extracted and Over What Time Period?

For each PDF indexed by Overton, references to research literature are identified and extracted. The number of PDFs indexed and the corresponding number of scholarly references extracted are shown for each year in the period 2000–2020 in Figure 4(a). Only references to DOIs are included in this analysis—2,027,440 references to other policy documents are excluded. The left axis (green) shows the totals and the right axis (blue) shows the average number of references per PDF. These data are also broken down by publication source type in Figure 4(b) where the average (mean) is shown for each through the period 2000–2020. The type “other” includes articles from PubMed Central, which would account for the relatively high rate of reference extraction for that source type compared to others, albeit for a small fraction of the database (about 1% of PDFs).

Data are also summarized in Table 1 where each row corresponds to a set of policy PDFs that contain a minimum number of scholarly references. For example, row  $\geq 10$  counts all PDFs that have 10 or more references to scholarly articles. There are 214,082 of these (4.4% of the corpus), accounting for 8,633,884 reference links, or 89% of references overall. The data indicate that although there are many policy documents that have no references, a core set of documents (approximately 200,000) may contain a sufficient number of references to build useful citation indicators. It is also possible that the documents that have no references



(a) Global count of references extracted from policy PDFs (left) and running average (right). (b) Average number of references extracted by source type.

Figure 4. Time-series of Policy Document references extracted.

**Table 1.** Count and percentage of scholarly references made from policy PDFs by reference count category

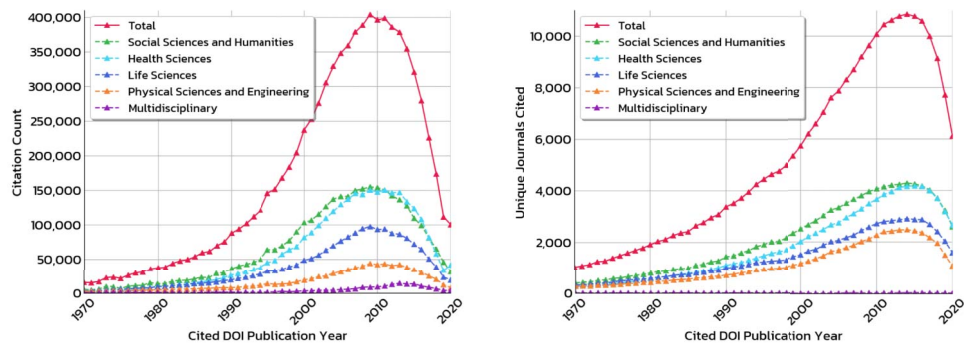
Refs. count	PDFs	% PDFs	Total refs.	% Refs.
≥ 0	4,854,919	100.00	9,747,436	100.00
≥ 1	570,830	11.76	9,747,436	100.00
≥ 5	305,637	6.30	9,248,600	94.88
≥ 10	214,082	4.41	8,633,884	88.58
≥ 50	38,235	0.79	4,772,402	48.96
≥ 100	14,162	0.29	3,139,856	32.21
≥ 500	794	0.02	725,307	7.44
≥ 1000	181	0.00	312,596	3.21

may be linked to other entities in Overton, such as researchers, institutions and topics of interest, providing other analytical value.

Perhaps of more interest from the perspective of building citation indicators, Figure 5(a) presents the number of citations received by DOIs according to their year of publication, dating back to 1970. The database total is shown in red, along with the corresponding totals for main research areas (as defined by ASJC). The data show that since 2000, publications have been cited in each year at least 200,000 times, with a maximum of 404,271 in 2009. We also use the same data to plot Figure 5(b), which shows the number of unique journals receiving citations in each year. The total maximum of around 10,000 corresponds well with the core set of global journals, for example in the Web of Science flagship collection or core publication list in the Leiden Ranking (Van Eck, 2021).

**4.3. How Long Does It Take Research Articles to Accumulate Policy Citations and How Does This Vary Across Disciplines?**

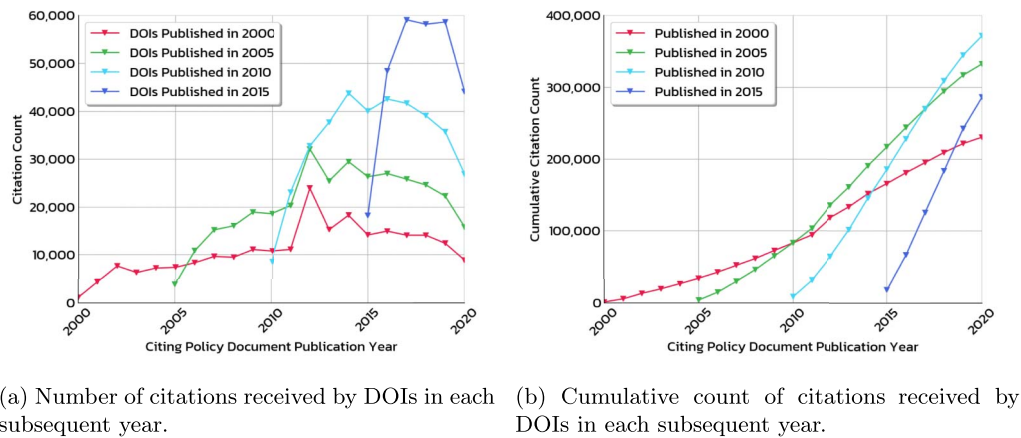
To appreciate the dynamics of how research articles accumulate citations from policy literature, we plot the number of citations received in years following original publication for DOIs published in 2000, 2005, 2010, and 2015. In Figure 6(a), the total number of citations received



(a) Number of citations to DOIs—total and by research area.

(b) Number of unique journals cited—total and by research area.

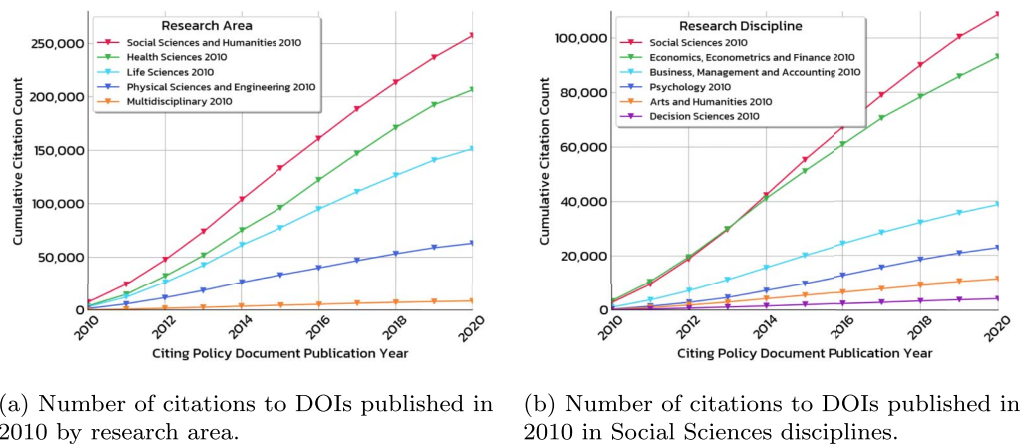
**Figure 5.** Time series of citation counts to DOIs and unique journals.



**Figure 6.** Time series (total and cumulative) of citations received to DOIs published in 2000, 2005, 2010, and 2020.

in each year is plotted, and in Figure 6(b) the cumulative total is displayed. These data indicate that the citation lifetime for DOIs is not even across years—older publications have received fewer citations overall and over a longer time period than those published more recently. Articles published in 2005 peaked 7 years after publication, those published in 2010 peaked after 4 years, and those published in 2015 after only 2 years. Further investigation is necessary to understand these differences, but it might be accounted for by the way the database is growing—an increasing number of documents indexed year-on-year could manifest as a recency bias.

Differences in the rate of citation accumulation between different disciplines were also analyzed. In terms of broad research areas, Figure 7(a) shows cumulative citation rates for articles that were published in 2010. DOIs published in journals categorized as Social Science and Humanities received the most citations, followed by Health Sciences and then Life Sciences. There is marked drop in citation rate for Physical Sciences and Engineering journals. The data for Social Science and Humanities is further decomposed into disciplines in Figure 7(b) and reveals most citations in this area are to journals in Social Sciences and Economics fields. This subject balance is in contrast to traditional bibliometric databases which tend to be dominated by citations to papers in biological and physical sciences, but could reasonably be expected given the typical domain of policy setting (e.g., social, economic, and environmental).

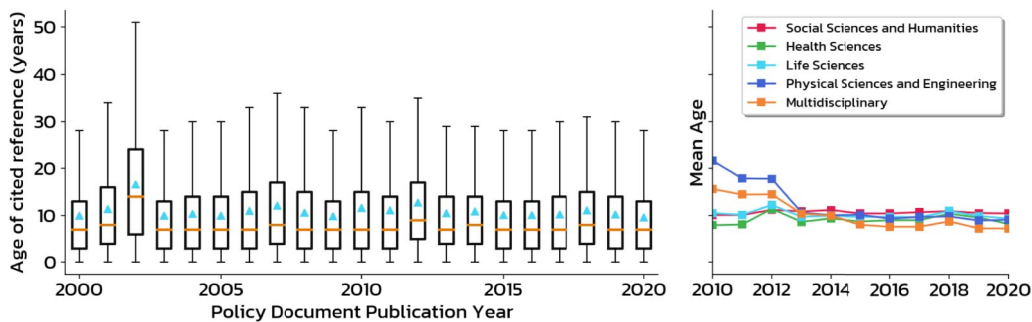


**Figure 7.** Time series of citations received to DOIs.

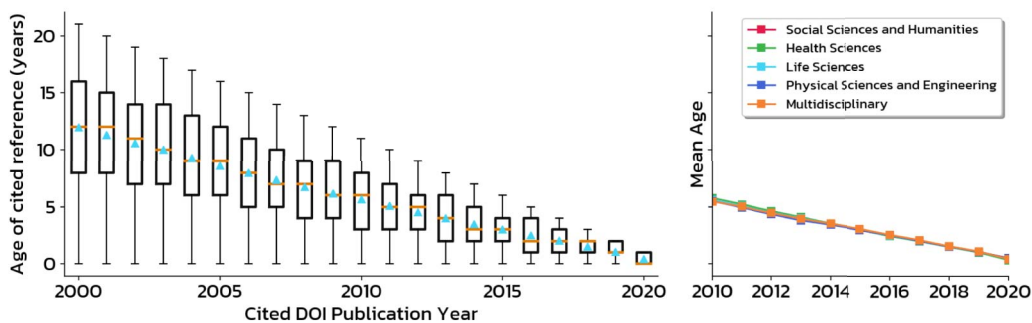
**4.4. What Is the Time Lag Between the Publication of Scholarly Works and Their Citation Within Policy Literature and How Does This Vary Between Disciplines?**

For each year between 2000 and 2020, we analyze the age of cited references in all policy documents indexed. For example, a policy document published in 2015 that references a DOI published in 2010 has a cited reference age of 5 years. For the purposes of this analysis, any reference ages that are calculated to be negative (i.e., the policy document publication date is before that of the cited reference) are removed on the assumption that they represent data errors. The distribution of these ages is displayed using standard box and whisker plots in Figure 8 (orange lines denoting median values, blue triangles for mean). The upper plot (Figure 8(a)) aggregates by the publication year of the citing policy document, and the lower plot (Figure 8(b)) aggregates by the year of publication for the cited DOI. The right insert in each shows the mean of the distribution for each of the ASJC research areas. Over the 21-year period sampled, there is little variation in the distribution of cited reference ages, with a mean of around 10 years (Figure 8(a)), and no significant differences between research areas (right plot). As a result, the distribution of reference ages aggregated by cited DOI publication year (Figure 8(b)) shows a consistent trend where the oldest publications have had the longest period to accumulate citations.

Although cited reference age appears to be consistent at a broad level, we also checked for differences in the age of references between different policy organizations. Two examples are provided in Figure 9, showing four organizations classified as either Healthcare Agency (Figure 9(a)) or Government Bank (Figure 9(b)). In both of these plots, it is apparent that different organizations cite research with different age ranges. The Canadian Agency for Drugs and

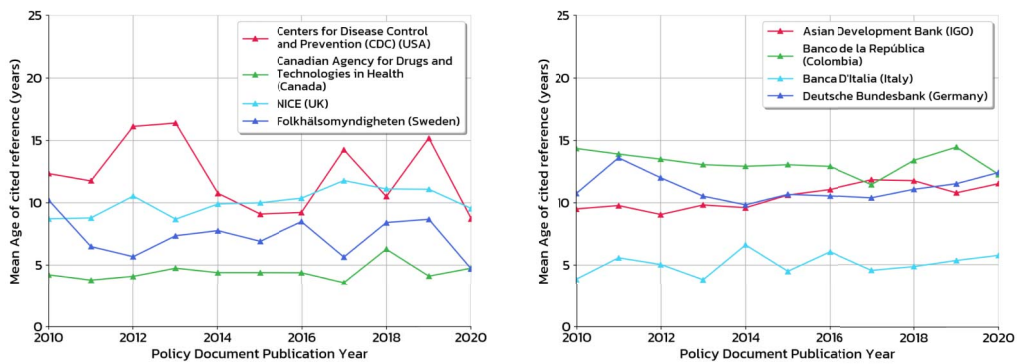


(a) Distribution of cited reference ages aggregated by policy document publication year.



(b) Distribution of cited reference ages aggregated by cited DOI publication year.

**Figure 8.** Age of publications referenced by policy documents.



(a) Cited reference age for Healthcare Agencies. (b) Cited reference age for Government Banks.

Figure 9. Age of publications referenced by policy organization type.

Technologies in Health Canada cite many more recent articles on average than the Centers for Disease Control and Prevention (United States). Of course, there are many factors that could influence such a difference, so any interpretation should be mindful of context and comparability of items.

#### 4.5. What Distribution Best Models Policy Citations Counts to DOIs?

When examining the policy citation counts of DOIs, it is apparent that the distribution is heavy-tailed (Asmussen, 2003). For example, for DOIs published between 2010 and 2014 ( $n = 731,696$ ), 425,268 are cited only once (58%), and only 25,190 are cited 10 or more times (3.4%). Prior research using conventional bibliographic databases has investigated possible statistical distributions that model citation data (Brzezinski, 2015; Eom & Fortunato, 2011; Golosovsky, 2021; Thelwall, 2016), although there is some disagreement on whether power law, log-normal, or negative binomial distributions are best. Results vary depending on time period and discipline analyzed, database used, and if documents with zero citations are included. For this analysis, uncited DOIs are not known because the database is generated by following references made at least once from the policy literature.

Figure 10 provides the probability distribution function (PDF—left), cumulative distribution function (CDF—middle), and complementary cumulative distribution function (CCDF—right) for citations received by DOIs published between 2010 and 2014. We use the Python package Powerlaw (Alstott, Bullmore, & Plenz, 2014) to fit distributions to exponential, power law, and lognormal. None of these provide an excellent fit for the data, although lognormal is the

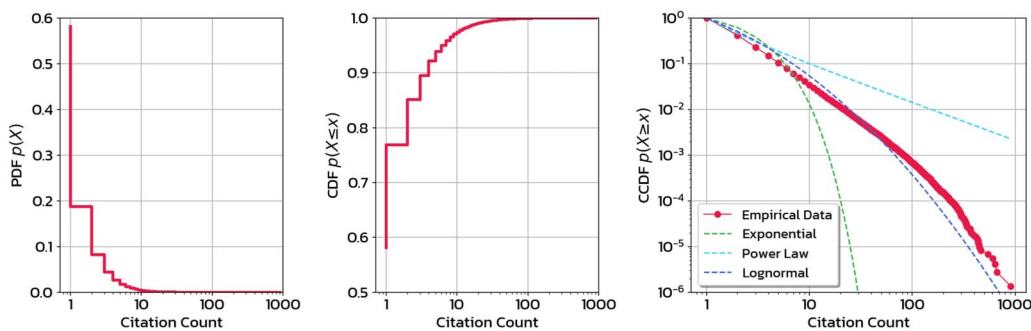
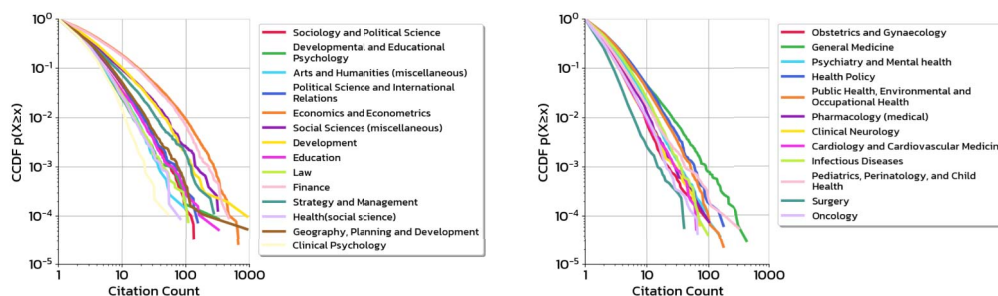


Figure 10. Probability distribution functions for DOIs published between 2010 and 2014.





(a) Citations to DOIs published 2010–2014 in Social Sciences and Humanities subjects.

(b) Citations to DOIs published 2010–2014 in Health Sciences subjects.

**Figure 11.** Time-series of citations received to DOIs.

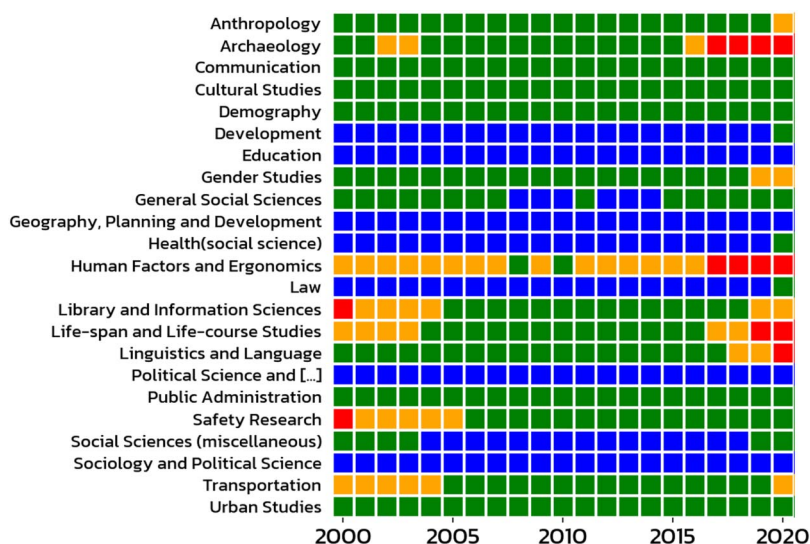
closest. In all cases, the fitted data overestimate slightly the frequency of low-cited DOIs (i.e., cited fewer than 10 times). Broadly speaking, it appears as though the distribution of policy document citations is similar in nature to that of academic citations.

As prior research has shown some variation in citation distributions according to subject (Wallace, Larivière, & Gingras, 2009), we analyzed a sample of subjects from the ASJC research areas Social Sciences and Humanities (Figure 11(a)) and Health Sciences (Figure 11(b)). In both cases, it is evident that substantial differences occur between subjects. For example, in the Social Sciences, Economics and Finance receive significantly more citations than in Clinical Psychology or the Arts. This is important to note, as it informs the selection of granularity for any field-based normalization. These findings suggest that variation at the subject level is present and therefore subject-level normalization is preferable, providing sufficiently large benchmark sets can be constructed.

#### 4.6. How Feasible Is Field-Based Citation Normalization?

As with standard citation metrics, citation counts from policy documents to DOIs also vary according to year of publication and field. Hence, we consider the feasibility of producing field-normalized citation indicators by analyzing the number of DOIs cited at least once according to subject and year. From a practical point of view, it is necessary to have a minimum number of DOIs to compare for any combination of subject and publication year. If the data are too sparse (i.e., there are only a handful of DOIs to compare for any subject-year), more specialized techniques are required to give robust results (Bornmann & Haunschild, 2018b).

To illustrate coverage, Figure 12 is provided showing a heatmap of subjects in the discipline Social Sciences in terms of the number of DOIs cited each year from 2000 to 2020. The color coding shows cases where  $n$  documents are cited where  $n < 150$  (red),  $150 \leq n < 250$  (orange),  $250 \leq n < 1,000$  (green), and  $n \geq 1,000$  (blue). According to Rogers et al. (2020), a minimum sample size of 250 is advised for bibliometric samples. The image clearly shows variation in the availability of data. In some subjects, large enough samples could be drawn throughout the study period (e.g., Development, Education, Law), but in other subjects, the data are more sparse and it would be ill-advised to construct normalized indicators (e.g., Human Factors and Ergonomics). As expected, sample sizes are much smaller in the most recent years as these articles are yet to accumulate a significant number of citations. The above analysis was carried out for all 330 ASJC subjects linked in the data, grouped into 26



**Figure 12.** Number of DOIs cited at least once in the discipline Social Sciences by subject and year:  $s < 150$  (red),  $150 \leq s < 250$  (orange),  $250 \leq s < 1,000$  (green), and  $s \geq 1,000$  (blue).

disciplines, to determine the overall spread of data availability. For each row in Table 2, a discipline is listed along with:

- **Subjects:** The total number of subjects in the discipline.
- **2000–2020%:** The percentage of subjects where  $n \geq 250$  in every year 2000–2020.
- **2000–2018%:** The percentage of subjects where  $n \geq 250$  in every year 2000–2018.
- **years%:** Across all subjects in the discipline, the percentage of subject-years where  $n \geq 250$ .
- **dois%:** Across all subjects, the percentage of DOIs that are in a subject-year where  $n \geq 250$ .

From these data, it is clear that some disciplines are well covered and others are not. The best covered (i.e., with years% > 80 and dois% > 90) are Agricultural and Biological Sciences, Economics, Econometrics and Finance, Environmental Science, Immunology and Microbiology, Medicine, and Social Sciences. The least well covered in terms of dois% are Materials Science, Dentistry, Physics and Astronomy, Health Professions, and Chemical Engineering.

Of the 2,270,711 cited DOIs that were published between 2000 and 2018, 2,009,302 (88%) are in a subject that contains at least 250 other cited articles in the same year. This means a subject-level normalization approach is practical and could be applied to a large portion of scholarly references.

#### 4.7. Do the Citations Tracked in the Policy Literature Correlate With Policy Influence Outcomes Attributed to Funded Grants?

To validate the citation data linked via the Overton database, we perform an analysis using data gathered by UK funders from the Gateway to Research (GTR) portal (UKRI, 2018). Following funding of certain grants in the United Kingdom, academics are required to submit feedback using the ResearchFish platform stating publications that resulted from the funding, as well as various research outcomes, including engagement activities, intellectual property, spin out companies, clinical trials, and more. One of these categories, policy influence, is used

**Table 2.** Completeness of disciplines and their subjects in terms of minimum sample size for normalization

Discipline	Subjects	2000–2020%	2000–2018%	years%	dois%
Agricultural and Biological Sciences	12	41.7	83.3	88.9	98.3
Arts and Humanities	14	14.3	14.3	35.7	82.8
Biochemistry, Genetics and Molecular Biology	16	31.2	68.8	73.2	95.9
Business, Management and Accounting	11	54.5	63.6	74.5	93.1
Chemical Engineering	9	0.0	0.0	16.4	59.6
Chemistry	8	12.5	25.0	41.1	83.1
Computer Science	13	7.7	7.7	27.5	63.8
Decision Sciences	5	0.0	20.0	36.2	70.6
Dentistry	5	0.0	0.0	18.1	52.4
Earth and Planetary Sciences	14	7.1	42.9	52.0	85.9
Economics, Econometrics and Finance	4	75.0	75.0	95.2	99.6
Energy	6	16.7	16.7	53.2	87.1
Engineering	17	0.0	29.4	45.4	82.3
Environmental Science	13	84.6	92.3	97.8	99.7
Health Professions	16	0.0	6.2	9.2	50.8
Immunology and Microbiology	7	57.1	71.4	81.6	98.8
Materials Science	9	0.0	0.0	23.3	51.5
Mathematics	15	0.0	6.7	15.9	62.4
Medicine	49	53.1	73.5	82.9	98.8
Multidisciplinary	1	100.0	100.0	100.0	100.0
Neuroscience	10	10.0	30.0	53.3	83.3
Nursing	23	4.3	8.7	17.2	66.6
Pharmacology, Toxicology and Pharmaceutics	6	33.3	33.3	55.6	94.5
Physics and Astronomy	11	0.0	0.0	19.0	53.4
Psychology	8	62.5	62.5	73.8	95.8
Social Sciences	23	60.9	69.6	87.8	98.1
Veterinary	5	20.0	20.0	21.0	74.9

to report various outcomes, including citations from policy documents, clinical guidelines, and systematic reviews. Data are collected at the project level, each of which is associated with various DOIs and policy outcomes. For this analysis, a data set is constructed using all funded grants with a start year between 2014 and 2020, recording the funder and research subjects specified. The funders analyzed are Arts and Humanities Research Council (AHRC), Biotechnology and Biological Sciences Research Council (BBSRC), Engineering and Physical

Sciences Research Council (EPSRC), Economic and Social Research Council (ESRC), Medical Research Council (MRC), and Natural Environment Research Council (NERC). 2014 is earliest year surveyed as it is the year in which ResearchFish was first adopted across all seven research councils.

For the analysis, data are aggregated at the project level noting the number of DOIs linked to the project, the total number of policy outcomes reported (referred to as *all policy influence*), the number of policy outcomes of the specific type citation (referred to as *citation influence*), and the total number of Overton citations. Effectively, this gives two features to compare—one, self-reported policy outcomes declared by academics, and another by tracking citations from policy documents via the Overton database. If Overton is able to index a sufficiently broad set of materials, these two features should be correlated.

Table 3 provides the correlation statistics (as measured using Pearson) for the complete data set (All row), and for each research council. Pearson measures the linear correlation between two sets of data, providing a normalized coefficient between  $-1$  (a perfect negative correlation) and  $+1$  (a perfect positive correlation). Positive effect sizes are commonly characterized as small (0.1–0.3), medium (0.3–0.5), and large ( $> 0.5$ ) (Cohen, 1988). The  $p$ -values are omitted because they will only depend on sample size, which is suitably large in this experiment. In every row, the total number of projects and DOIs they link to is reported (columns Projects and DOIs), along with two sets of statistics—one testing Overton citation counts against the total number of policy influence outcomes reported (All policy influence—middle columns), and the other testing Overton citation counts against the number of policy influence outcomes that are specifically for citations in policy documents, clinical guidelines, or systematic reviews (Citation influence only—right columns). In each case, the correlation coefficient  $r$  is provided, as well as the percentage of projected that were linked to any policy influence outcomes. This percentage figure is given to contextualize results, as for some funders the number of projects associated with any policy outcomes is low. According to these results, the correlation between the count of policy influence outcomes and the total number of citations in Overton is larger when considering all policy influence types, rather than only those specifically for citation, although for EPSRC they are similar, and for ESRC they are higher ( $r = 0.70$ ). There is a medium correlation over all funders ( $r = 0.42$ ), and large correlation for the EPSRC ( $r = 0.66$ ) and MRC ( $r = 0.63$ ).

**Table 3.** Pearson correlation between funded projects with policy influence and total policy citations in Overton

Funder	Projects	DOIs	All policy influence		Citation influence only	
			$r$	Projects%	$r$	Projects%
All	67,702	383,642	0.42	7.13	0.32	1.17
AHRC	3,902	14,254	0.26	13.84	0.19	2.26
BBSRC	9,031	40,642	0.30	7.60	0.23	0.68
EPSRC	17,799	106,312	0.66	4.72	0.65	0.51
ESRC	5,732	37,503	0.48	16.99	0.70	4.41
MRC	5,992	60,854	0.63	16.41	0.20	2.42
NERC	4,727	30,035	0.22	13.71	0.17	3.13

Downloaded from [http://direct.mit.edu/qss/article-pdf/3/3/624/2057871/qss\\_a\\_00204.pdf](http://direct.mit.edu/qss/article-pdf/3/3/624/2057871/qss_a_00204.pdf) by guest on 20 April 2025

The data are further decomposed according to subject category assigned to the grant, as depicted in Figure 13. Each grant may be assigned to multiple subjects and is considered in the calculation for each subject. For each subject (a row), three columns are used to show the correlation  $r$  (red), percentage of projects reporting any policy influence (green), and the total count of DOIs linked to projects (blue). In this plot, correlations are measured against all policy influence outcomes (i.e., corresponding to the middle columns in Table 3). When analyzed at

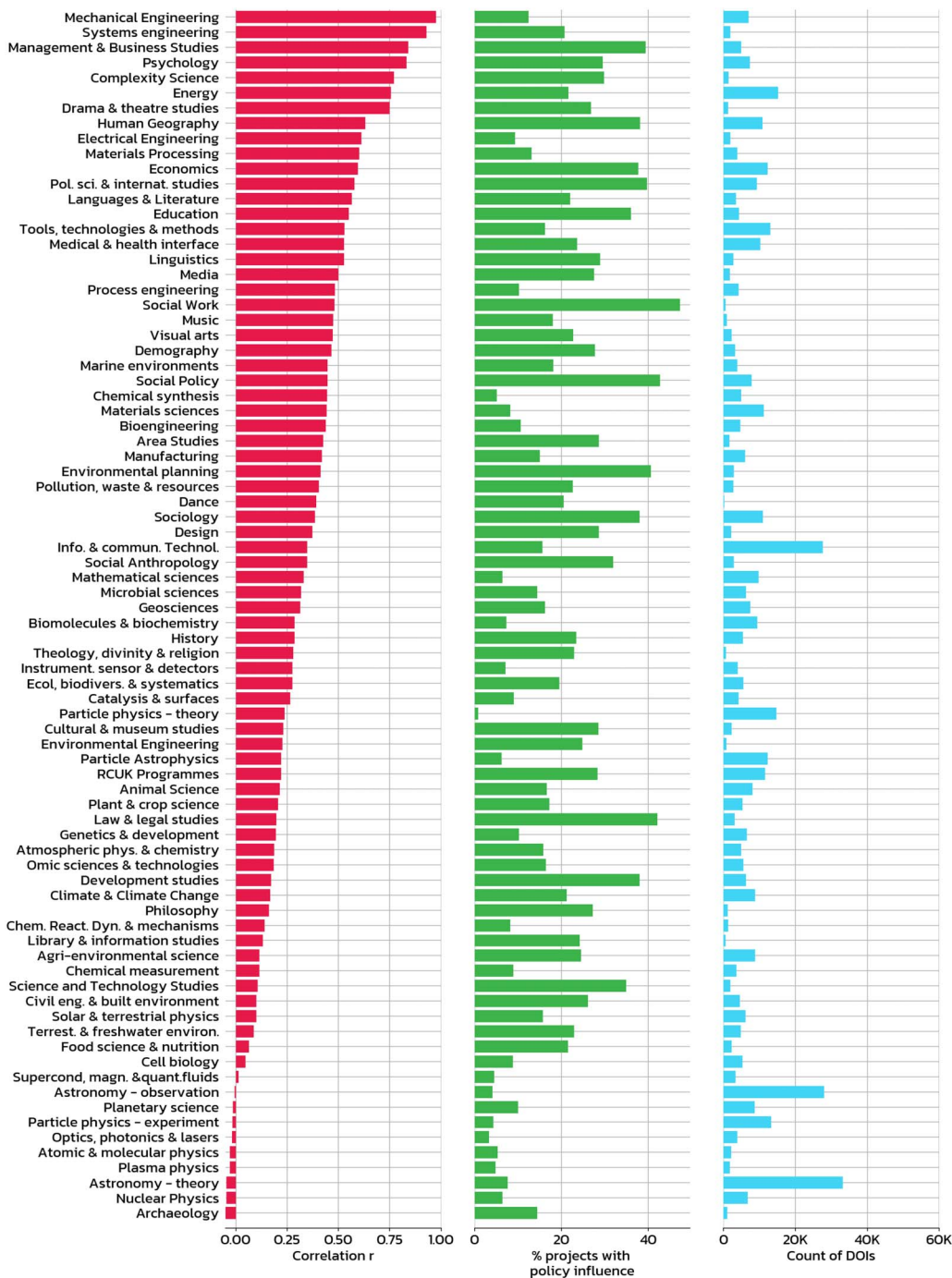


Figure 13. Number of citations to DOIs by research area.



this level of granularity, there is a large spread in the correlation values: 27 are small (0.1–0.3), 23 are medium (0.3–0.5) and 17 are large (> 0.5). Thirteen are not correlated.

These results show that for some subjects, Overton citation data correlates well with policy influence outcomes reported by academics. This occurs most in subjects that might be expected to have some policy influence, such as Management & Business Studies ( $r = 0.84$ ), Psychology ( $r = 0.83$ ), Human Geography ( $r = 0.63$ ), Economics ( $r = 0.60$ ), and Political Science & International Studies ( $r = 0.58$ ), but also in others that might not, such as Mechanical Engineering ( $r = 0.98$ ), Systems engineering ( $r = 0.93$ ), and Drama & Theatre Studies ( $r = 0.76$ ). Essentially, the analysis shows which subjects are most amenable to analysis using Overton data.

#### 4.8. Does the Amount of Policy Citation Correlate With Peer-Review Assessment Scores as Reported in the UK REF2014 Impact Case Study Data?

To test for possible correlation, we utilize the Impact Case Study database from REF2014. This contains 6,637 four-page documents that outline the wider socioeconomic and cultural impact of research attributed to a particular university and Unit of Assessment (UoA). Part of the case study document references the original underpinning research (up to six references per case study) which has been linked via DOIs. By means of peer review, each case study is scored as 4\* (world-leading), 3\* (internationally excellent), 2\* (internationally recognized), or 1\* (nationally recognized). Although the scores for individual cases studies are not known, the aggregate scores are made available as the percentage of case studies that received each score. Hence, it is possible to test possible correlations at the aggregate level (namely, institution and UoA).

For this analysis, we test the correlation between research scored as 4\* (excellent) and citations to the underpinning research as reported in the Overton database. As the assessment exercise took place in 2014, only citations from policy documents published in or earlier than 2014 are considered. Rather than test raw citation counts, we calculate a year-subject normalized citation percentile for each DOI using ASJC journal categories (i.e., all DOIs published in a certain year and subject are compared with each other). Any DOIs in a year-subject group that contain < 250 examples are marked as invalid and excluded from the analysis. Of the total 24,945 unique DOIs associated with an impact case study, 4,292 are referenced in Overton and have a valid citation percentile.

Following the methodology presented in Traag and Waltman (2019), we measure the correlation between the percentage of case studies that scored 4\* and the percentage of DOIs in the top 99, 90, and 75th Overton citation percentiles. Multiple percentiles were tested as it is not necessarily clear where the benchmark for 4\* research would lie. We evaluated 1,847 scores are evaluated—one for each university and UoA. A size-independent test measures the Pearson correlation between the percentage of research scored 4\* and the percentage of DOIs with a normalized citation percentile above the threshold.

Table 4 provides the results of this analysis. All 36 UoAs are shown along with the Pearson correlation  $r$  for three citation percentile thresholds: 99%, 90%, and 75%. In some cases (for example Classics), when no DOIs could be found exceeding the percentile threshold, the correlation is undefined and hence, left blank. Based on these results, it is apparent that different percentile thresholds yield different results depending on UoA. For example in UoAs 18—Economics and Econometrics and 25—Education, the highest correlations of 0.52 and 0.46 respectively are obtained with a threshold of 90%, but in UoA 7—Earth Systems and Environmental Sciences, a threshold of 99% yields the highest correlation of 0.52. This

**Table 4.** Pearson correlation for REF2014 impact scores marked 4\* against Overton citation percentiles above three threshold values: 99%, 90%, and 75%. Medium correlations are highlighted with †, large correlations are identified with \*

UoA	4*_top99 r	4*_top90 r	4*_top75 r
1—Clinical Medicine	0.20	0.24	0.25
2—Public Health, Health [...]	0.22	−0.20	0.24
3—Allied Health Professions, [...]	0.02	0.04	0.13
4—Psychology, Psychiatry and [...]	0.18	0.13	0.27
5—Biological Sciences	0.16	0.08	0.02
6—Agriculture, Veterinary and [...]	0.28	*0.57	*0.54
7—Earth Systems and [...]	*0.52	0.24	0.17
8—Chemistry		0.15	0.00
9—Physics	0.07	0.07	0.02
10—Mathematical Sciences	†0.32	0.11	0.13
11—Computer Science and Informatics		0.27	†0.30
12—Aeronautical, Mechanical, [...]		†0.49	*0.64
13—Electrical and Electronic [...]		†0.42	0.11
14—Civil and Construction Engineering	0.17	−0.22	−0.11
15—General Engineering		0.11	0.15
16—Architecture, Built [...]	0.13	†0.34	†0.30
17—Geography, Environmental [...]	0.21	0.16	0.20
18—Economics and Econometrics	†0.39	*0.52	†0.33
19—Business and Management Studies	0.00	0.10	0.18
20—Law		0.16	0.05
21—Politics and International Studies	0.16	0.07	0.26
22—Social Work and Social Policy	†0.47	0.24	†0.32
23—Sociology	0.01	0.07	0.08
24—Anthropology and Development [...]	0.10	0.15	0.21
25—Education	†0.33	†0.46	†0.40
26—Sport and Exercise Sciences, [...]	0.22	†0.40	†0.31
27—Area Studies		†0.44	†0.42
28—Modern Languages and Linguistics	0.12	0.24	0.24
29—English Language and Literature		0.00	0.00
30—History		0.12	0.13

Table 4. (continued)

UoA	4*_top99 r	4*_top90 r	4*_top75 r
31—Classics			
32—Philosophy		0.24	0.20
33—Theology and Religious Studies			0.41
34—Art and Design: History, [...]			-0.15
35—Music, Drama, Dance and [...]		-0.01	0.20
36—Communication, Cultural and [...]		†0.38	†0.33

suggests that the threshold for what is considered 4\* impact varies across fields in terms of policy influence.

This analysis shows that for some UoAs, Overton policy citation percentiles do correlate with peer-review assessment, but less than reported for citation data (Traag & Waltman, 2019) when compared to scoring of outputs. Ideally, the test would only be performed on the subset of case studies that might reasonably be expected to have some form of policy outcome. For example, searching the database for "policy outcome" ~ 5 OR "policy influence" ~ 5 (where the ~ 5 operator specifies that terms must be within five words of each other) returns only 406 results. Hence, our test effectively measures the correlation of impact in general against that of policy citation and could only be expected to find correlation in UoAs where the dominant form of impact is policy-related, such as in UoA 22 Social Work and Social Policy. Unfortunately, because scores are not known for individual case studies, this type of analysis is not possible.

### 5. DISCUSSION

Our analysis of the Overton policy document citation database yields a promising outlook. Using this kind of data, it is possible to link the original research published in scholarly literature to their use in a policy setting environment. The Overton database indexes a sufficient amount of content to create large volumes of citations (> 400,000 every year since 2014) across a wide range of research topics and journals. Unlike conventional bibliometric databases, citations are more focused towards social sciences, economics, and environmental sciences than to biological and physical sciences, a feature that suggests novel value in the content in terms of analytical potential.

The balance of content by region broadly follows that of other bibliometric databases, namely it is dominated by North America and Europe, but the representation of local language documents is much higher than in scholarly publishing, where English dominates (Márquez & Porras, 2020; Mongeon & Paul-Hus, 2016). Anecdotal evidence in this study hints that Overton may have more equitable coverage across some countries: Figure 3 shows that Peru and Uruguay have a similar volume of policy documents indexed to Brazil and Chile despite producing fewer scholarly works. However, more detailed analysis, drawing on other indicators (e.g., economic and industrial), is required to produce robust conclusions in relation to this question.

One important issue that is not addressed in this study is the question of coverage. By indexing as much of the freely accessible data on the web, it is possible that some countries, organization types, or document types are better represented than others. However, this is not a

straightforward question to tackle because the universe of policy documents is not well defined (i.e., what should and should not be considered a policy document?) and the only route to obtain information on missing sources requires significant manual effort. A practical approach may be to survey certain policy topics and regions to estimate coverage levels.

Although a significant proportion of the policy documents indexed are not linked to DOIs (88% of PDFs), a core set of around 200,000 contain more than 8 million references. This reflects the diverse range of material indexed including statistical reports, informal communications, proceedings, and commentary, many of which one would not expect to contain references to original research articles. A considerable pool of citations is generated—between 200,000 and 400,000 per year since 2000 across a broad set of journals. A more detailed analysis of these data could compare how citations are distributed across journals and if citations patterns from policy documents follow the same tendencies as scholarly publishing. It may be true that some journals are able to demonstrate higher utilization in policy documents relative to a citation-based ranking.

The potential for development of field-normalized citation indicators is good. When analyzed at the ASJC subject level, many fields contain a sufficient number of cited articles to create benchmarks (i.e.,  $\geq 250$ ), especially if the most recent 2 years are excluded. Overall, 88% of articles published between 2000 and 2018 that receive any policy citations could be field-normalized in this way. However, although this approach is practical, it may not be best—a more detailed analysis comparing normalization results at different levels of granularity (i.e., field-based or discipline-based) would be required to make any recommendation.

One potentially interesting line of inquiry is that of citation lag. At the macro scale, our analysis shows there is little variation in the distribution of ages, even across disciplines, but when viewed at a more granular level (such as individual policy organizations), diversity occurs. This may offer useful insights into the differences between what research is used, in terms of age and also in citation ranking. Some organizations may favor newer but less established evidence than others that prefer older but more widely recognized research.

The distribution of citations accumulated by research articles seems to follow similar trends to that seen in other altmetric indicators, especially Mendeley, Twitter, and Facebook as reported in Fang, Costas et al. (2020), and like conventional citation data, are best matched to a log normal distribution. It is interesting to note that in Fang et al. (2020), 12,271,991 articles published between 2012 and 2018 were matched to Altmetric data and yielded 156,813 citations across 137,326 unique documents. For the same time period, Overton contains 2,600,464 citations across 1,006,439 unique DOIs. These coverage statistics are not directly comparable because the original pool of articles surveyed in Fang et al. (2020) is limited to the Web of Science and Overton tracks citations to any journal. Nevertheless, it does suggest that Overton tracks substantially more citations to policy literature than Altmetric.

Possibly the most striking and encouraging result is from the analysis of policy influence outcomes reported to UK funders. Our findings show that for some subjects, correlation between self-reported data and that extracted from Overton is high. This offers additional opportunities to reduce reporting burden, through either semiautomated or automated approaches. Further, it provides a basis to benchmark funders and institutions from different regions where self-reported data may not be available, although such an analysis should consider coverage variation across geographies.

Finally, our experiment to test for possible correlation between peer-review assessment of impact and Overton policy citations hints at some utility: For certain Units of Assessment, a correlation between peer-review score of impact and citation rank does exist, although less

than that seen in other studies that assessed peer-review scores of academic impact against conventional citation data (Traag & Waltman, 2019). While the REF2014 impact case study data do provide a unique opportunity to understand how research is assessed from the perspective of wider socioeconomic impact, obfuscation of the individual scores prevents deeper analysis that is focused on research pertinent to policy outcomes. It may be more fruitful to utilize other sources to benchmark peer review, such as postpublication peer-review score (Waltman & Costas, 2014).

#### AUTHOR CONTRIBUTIONS

Martin Szomszor: Conceptualisation, Investigation, Methodology, Visualisation, Writing—original draft, Writing—review & editing. Euan Adie: Conceptualisation, Methodology, Writing—review & editing.

#### COMPETING INTERESTS

Martin Szomszor is an independent researcher. Euan Adie is Founder and CEO of Overton.

#### FUNDING INFORMATION

This research has been funded by Open Policy Ltd, which runs Overton.

#### DATA AVAILABILITY

Because Overton is a commercial database, data for this study cannot be shared publicly. For more information about access to Overton data for research purposes, please email support@overton.io.

#### REFERENCES

- Abramo, G., & D'Angelo, C. A. (2015). The VQR, Italy's second national research assessment: Methodological failures and ranking distortions. *Journal of the Association for Information Science and Technology*, 66(11), 2202–2214. <https://doi.org/10.1002/asi.23323>
- Abramo, G., D'Angelo, C. A., & Grilli, L. (2021). The effects of citation-based research evaluation schemes on self-citation behavior. *Journal of Informetrics*, 15(4), 101204. <https://doi.org/10.1016/j.joi.2021.101204>
- Abrams, D., Akcigit, U., & Grennan, J. (2018). *Patent value and citations: Creative destruction or strategic disruption?* (Tech. Rep. w19647). Cambridge, MA: National Bureau of Economic Research. <https://doi.org/10.3386/w19647>
- Adams, J., Gurney, K., & Marshall, S. (2007). Profiling citation impact: A new methodology. *Scientometrics*, 72(2), 325–344. <https://doi.org/10.1007/s11192-007-1696-x>
- Adams, J., Marie, M., David, P., & Martin, S. (2019). *Profiles, not metrics*. London: Clarivate Analytics.
- Aksnes, D. W., & Sivertsen, G. (2019). A criteria-based assessment of the coverage of Scopus and Web of Science. *Journal of Data and Information Science*, 4(1), 1–21. <https://doi.org/10.2478/jdis-2019-0001>
- Alcácer, J., & Gittelman, M. (2006). Patent citations as a measure of knowledge flows: The influence of examiner citations. *Review of Economics and Statistics*, 88(4), 774–779. <https://doi.org/10.1162/rest.88.4.774>
- Alcaraz, C., & Morais, S. (2012). Citations: Results differ by database. *Nature*, 483(7387), 36. <https://doi.org/10.1038/483036d>, PubMed: 22382969
- Alstott, J., Bullmore, E., & Plenz, D. (2014). Powerlaw: A Python package for analysis of heavy-tailed distributions. *PLOS ONE*, 9(1), e85777. <https://doi.org/10.1371/journal.pone.0085777>, PubMed: 24489671
- ARC. (2018). *ERA national report* (Tech. Rep.). Australian Research Council. <https://dataportal.arc.gov.au/ERA/NationalReport/2018/>
- Asmussen, S. (2003). Steady-state properties of *G/G/1*. In *Applied probability and queues* (pp. 266–301). New York: Springer. [https://doi.org/10.1007/0-387-21525-5\\_10](https://doi.org/10.1007/0-387-21525-5_10)
- BEIS. (2017). *International comparative performance of the UK research base—2016* (Tech. Rep.). A report prepared by Elsevier for the UK's Department for Business, Energy & Industrial Strategy (BEIS). <https://www.gov.uk/government/publications/performance-of-the-uk-research-base-international-comparison-2016>
- Bizer, C., Vidal, M.-E., & Weiss, M. (2018). Resource description framework. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of database systems* (pp. 3221–3224). New York: Springer. [https://doi.org/10.1007/978-1-4614-8265-9\\_905](https://doi.org/10.1007/978-1-4614-8265-9_905)
- Björk, B.-C., Kanto-Karvonen, S., & Harviainen, J. T. (2020). How frequently are articles in predatory open access journals cited? *Publications*, 8(2), 17. <https://doi.org/10.3390/publications8020017>
- Bornmann, L. (2013). What is societal impact of research and how can it be assessed? A literature survey. *Journal of the American Society for Information Science and Technology*, 64(2), 217–233. <https://doi.org/10.1002/asi.22803>
- Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics*, 8(4), 895–903. <https://doi.org/10.1016/j.joi.2014.09.005>



- Bornmann, L. (2015). Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics. *Scientometrics*, 103(3), 1123–1144. <https://doi.org/10.1007/s11192-015-1565-y>
- Bornmann, L., & Haunschild, R. (2018a). Do altmetrics correlate with the quality of papers? A large-scale empirical study based on F1000Prime data. *PLOS ONE*, 13(5), e0197133. <https://doi.org/10.1371/journal.pone.0197133>, PubMed: 29791468
- Bornmann, L., & Haunschild, R. (2018b). Normalization of zero-inflated data: An empirical analysis of a new indicator family and its use with altmetrics data. *Journal of Informetrics*, 12(3), 998–1011. <https://doi.org/10.1016/j.joi.2018.01.010>
- Bornmann, L., Haunschild, R., & Adams, J. (2019). Do altmetrics assess societal impact in a comparable way to case studies? An empirical test of the convergent validity of altmetrics based on data from the UK research excellence framework (REF). *Journal of Informetrics*, 13(1), 325–340. <https://doi.org/10.1016/j.joi.2019.01.008>
- Bornmann, L., Haunschild, R., Boyack, K., Marx, W., & Minx, J. C. (2022). How relevant is climate change research for climate change policy? An empirical analysis based on Overton data. *arXiv:2203.05358*. <https://doi.org/10.48550/arXiv.2203.05358>
- Bornmann, L., Haunschild, R., & Marx, W. (2016). Policy documents as sources for measuring societal impact: How often is climate change research mentioned in policy-related documents? *Scientometrics*, 109(3), 1477–1495. <https://doi.org/10.1007/s11192-016-2115-y>, PubMed: 27942080
- Brzezinski, M. (2015). Power laws in citation distributions: Evidence from Scopus. *Scientometrics*, 103(1), 213–228. <https://doi.org/10.1007/s11192-014-1524-z>, PubMed: 25821280
- Buckle, R. A., & Creedy, J. (2019). An evaluation of metrics used by the Performance-based Research Fund process in New Zealand. *New Zealand Economic Papers*, 53(3), 270–287. <https://doi.org/10.1080/00779954.2018.1480054>
- Burke, J., Bergman, J., & Asimov, I. (1985). *The impact of science on society* (Tech. Rep.). National Aeronautics and Space Administration, Hampton, VA: Langley Research Center.
- Bush, V. (1945). *Science: The endless frontier* (Tech. Rep.). [A report to President Truman outlining his proposal for post-war U.S. science and technology policy.] Washington, DC.
- Butler, L. (2003). Explaining Australia's increased share of ISI publications—The effects of a funding formula based on publication counts. *Research Policy*, 32(1), 143–155. [https://doi.org/10.1016/S0048-7333\(02\)00007-0](https://doi.org/10.1016/S0048-7333(02)00007-0)
- Cagan, R. (2013). The San Francisco Declaration on Research Assessment. *Disease Models & Mechanisms*, 6(4), 869–870. <https://doi.org/10.1242/dmm.012955>, PubMed: 23690539
- Carpenter, M. P., Cooper, M., & Narin, F. (1980). Linkage between basic research literature and patents. *Research Management*, 23(2), 30–35. <https://doi.org/10.1080/00345334.1980.11756595>
- Chadegani, A. A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., ... Ebrahim, N. A. (2013). A comparison between two main academic literature collections: Web of Science and Scopus databases. *Asian Social Science*, 9(5), 18. <https://doi.org/10.5539/ass.v9n5p18>
- Chowdhury, G., Koya, K., & Philipson, P. (2016). Measuring the impact of research: Lessons from the UK's Research Excellence Framework 2014. *PLOS ONE*, 11(6), e0156978. <https://doi.org/10.1371/journal.pone.0156978>, PubMed: 27276219
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Costas, R., Zahedi, Z., & Wouters, P. (2015). Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, 66(10), 2003–2019. <https://doi.org/10.1002/asi.23309>
- Demir, S. B. (2020). Scholarly databases under scrutiny. *Journal of Librarianship and Information Science*, 52(1), 150–160. <https://doi.org/10.1177/0961000618784159>
- Eom, Y.-H., & Fortunato, S. (2011). Characterizing and modeling citation dynamics. *PLOS ONE*, 6(9), e24926. <https://doi.org/10.1371/journal.pone.0024926>, PubMed: 21966387
- Eriksson, M., Billhult, A., Billhult, T., Pallari, E., & Lewison, G. (2020). A new database of the references on international clinical practice guidelines: A facility for the evaluation of clinical research. *Scientometrics*, 122(2), 1221–1235. <https://doi.org/10.1007/s11192-019-03318-2>
- Evenson, R. E., Waggoner, P. E., & Ruttan, V. W. (1979). Economic benefits from research: An example from agriculture. *Science*, 205(4411), 1101–1107. <https://doi.org/10.1126/science.205.4411.1101>, PubMed: 17735033
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *The FASEB Journal*, 22(2), 338–342. <https://doi.org/10.1096/fj.07-9492LSF>, PubMed: 17884971
- Fang, Z., Costas, R., Tian, W., Wang, X., & Wouters, P. (2020). An extensive analysis of the presence of altmetric data for Web of Science publications across subject fields and research topics. *Scientometrics*, 124(3), 2519–2549. <https://doi.org/10.1007/s11192-020-03564-9>, PubMed: 32836523
- Gao, J., Yin, Y., Jones, B. F., & Wang, D. (2020). Quantifying policy responses to a global emergency: Insights from the COVID-19 pandemic. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3634820>
- Georghiou, L. (1995). Research evaluation in European national science and technology systems. *Research Evaluation*, 5(1), 3–10. <https://doi.org/10.1093/rev/5.1.3>
- Gibbons, M., & Georghiou, L. (1987). *Evaluation of research: A selection of current practices*. Organisation for Economic Co-operation & Development.
- Golosovsky, M. (2021). Universality of citation distributions: A new understanding. *Quantitative Science Studies*, 2(2), 527–543. [https://doi.org/10.1162/qss\\_a\\_00127](https://doi.org/10.1162/qss_a_00127)
- Grant, J. (2000). Evaluating “payback” on biomedical research from papers cited in clinical guidelines: Applied bibliometric study. *BMJ*, 320(7242), 1107–1111. <https://doi.org/10.1136/bmj.320.7242.1107>, PubMed: 10775218
- Grant, J. (2015). *The nature, scale and beneficiaries of research impact: An initial analysis of Research Excellence Framework (REF) 2014 impact case studies* (Tech. Rep.). Prepared for the Higher Education Funding Council of England, Higher Education Funding Council for Wales, Scottish Funding Council, Department of Employment, Learning Northern Ireland, Research Councils UK, and the Wellcome Trust.
- Greenhalgh, T., Raftery, J., Hanney, S., & Glover, M. (2016). Research impact: A narrative review. *BMC Medicine*, 14(1), 78. <https://doi.org/10.1186/s12916-016-0620-8>, PubMed: 27211576
- Guerrero-Bote, V. P., Chinchilla-Rodríguez, Z., Mendoza, A., & de Moya-Anegón, F. (2021). Comparative analysis of the bibliographic data sources Dimensions and Scopus: An approach at the country and institutional levels. *Frontiers in Research Metrics and Analytics*, 5, 593494. <https://doi.org/10.3389/frma.2020.593494>, PubMed: 33870055
- Guthrie, S., Cochrane, G., Deshpande, A., Macaluso, B., & Larivière, V. (2019). Understanding the contribution of UK public

- health research to clinical guidelines: A bibliometric analysis. *F1000Research*, 8, 1093. <https://doi.org/10.12688/f1000research.18757.1>, PubMed: 33552472
- Hanney, S., Packwood, T., & Buxton, M. (2000). Evaluating the benefits from health research and development centres: A categorization, a model and examples of application. *Evaluation*, 6(2), 137–160. <https://doi.org/10.1177/13563890022209181>
- Harzing, A.-W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804. <https://doi.org/10.1007/s11192-015-1798-9>
- Hicks, D. (2010). Overview of models of performance-based research funding systems. *Performance-based Funding for Public Research in Tertiary Education Institutions—Workshop Proceedings* (pp. 23–52). <https://doi.org/10.1787/9789264094611-4-en>
- Hicks, D., & Melkers, J. (2013). Bibliometrics as a tool for research evaluation. *Handbook on the theory and practice of program evaluation* (pp. 323–349). Edward Elgar Publishing. <https://doi.org/10.4337/9780857932402.00019>
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431. <https://doi.org/10.1038/520429a>, PubMed: 25903611
- Jiménez-Contreras, E., Anegón, F. d. M., & López-Cózar, E. D. (2003). The evolution of research activity in Spain: The impact of the National Commission for the Evaluation of Research Activity (CNEAI). *Research Policy*, 32(1), 123–142. [https://doi.org/10.1016/S0048-7333\(02\)00008-2](https://doi.org/10.1016/S0048-7333(02)00008-2)
- Karvonen, M., & Kässi, T. (2013). Patent citations as a tool for analysing the early stages of convergence. *Technological Forecasting and Social Change*, 80(6), 1094–1107. <https://doi.org/10.1016/j.techfore.2012.05.006>
- Kousha, K., & Thelwall, M. (2008). Assessing the impact of disciplinary research on teaching: An automatic analysis of online syllabuses. *Journal of the American Society for Information Science and Technology*, 59(13), 2060–2069. <https://doi.org/10.1002/asi.20920>
- Kousha, K., & Thelwall, M. (2017). Are Wikipedia citations important evidence of the impact of scholarly articles and books? *Journal of the Association for Information Science and Technology*, 68(3), 762–779. <https://doi.org/10.1002/asi.23694>
- Kryl, D., Allen, L., Dolby, K., Sherbon, B., & Viney, I. (2012). Tracking the impact of research on policy and practice: Investigating the feasibility of using citations in clinical guidelines for research evaluation. *BMJ Open*, 2(2), e000897. <https://doi.org/10.1136/bmjopen-2012-000897>, PubMed: 22466037
- Kuhn, J., Younge, K., & Marco, A. (2020). Patent citations reexamined. *The RAND Journal of Economics*, 51(1), 109–132. <https://doi.org/10.1111/1756-2171.12307>
- Lahtinen, E., Koskinen-Ollonqvist, P., Rouvinen-Wilenius, P., Tuominen, P., & Mittelmark, M. B. (2005). The development of quality criteria for research: A Finnish approach. *Health Promotion International*, 20(3), 306–315. <https://doi.org/10.1093/heapro/dai008>, PubMed: 15964888
- Larivière, V., & Gingras, Y. (2011). Averages of ratios vs. ratios of averages: An empirical analysis of four levels of aggregation. *Journal of Informetrics*, 5(3), 392–399. <https://doi.org/10.1016/j.joi.2011.02.001>
- Lopez Pineiro, C., & Hicks, D. (2015). Reception of Spanish sociology by domestic and foreign audiences differs and has consequences for evaluation. *Research Evaluation*, 24(1), 78–89. <https://doi.org/10.1093/reseval/rvu030>
- Márquez, M. C., & Porras, A. M. (2020). Science communication in multiple languages is critical to its effectiveness. *Frontiers in Communication*, 5, 31. <https://doi.org/10.3389/fcomm.2020.00031>
- Martin, B. R. (1996). The use of multiple indicators in the assessment of basic research. *Scientometrics*, 36(3), 343–362. <https://doi.org/10.1007/BF02129599>
- Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, 12(4), 1160–1177. <https://doi.org/10.1016/j.joi.2018.09.002>
- Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & Delgado López-Cózar, E. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics*, 126(1), 871–906. <https://doi.org/10.1007/s11192-020-03690-4>, PubMed: 32981987
- Mas-Bleda, A., & Thelwall, M. (2018). Estimación del valor educativo de los libros académicos que no están en inglés: El caso de España. *Revista Española de Documentación Científica*, 41(4), e222. <https://doi.org/10.3989/redc.2018.4.1568>
- May, R. M. (1997). The scientific wealth of nations. *Science*, 275(5301), 793–796. <https://doi.org/10.1126/science.275.5301.793>
- Meyer, M. (2000). Does science push technology? Patents citing scientific literature. *Research Policy*, 29(3), 409–434. [https://doi.org/10.1016/S0048-7333\(99\)00040-2](https://doi.org/10.1016/S0048-7333(99)00040-2)
- Moed, H. F. (2005). *Citation analysis in research evaluation*. Springer.
- Moed, H. F. (2010). CWTS crown indicator measures citation impact of a research group's publication oeuvre. *Journal of Informetrics*, 4(3), 436–438. <https://doi.org/10.1016/j.joi.2010.03.009>
- Moed, H. F., Burger, W. J. M., Frankfort, J. G., & Van Raan, A. F. J. (1985). A comparative study of bibliometric past performance analysis and peer judgement. *Scientometrics*, 8(3–4), 149–159. <https://doi.org/10.1007/BF02016933>
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- Morton, S. (2015). Progressing research impact assessment: A 'contributions' approach. *Research Evaluation*, 24(4), 405–419. <https://doi.org/10.1093/reseval/rvv016>
- Narin, F., & Hamilton, K. S. (1996). Bibliometric performance measures. *Scientometrics*, 36(3), 293–310. <https://doi.org/10.1007/BF02129596>
- Newson, R., Rychetnik, L., King, L., Milat, A., & Bauman, A. (2018). Does citation matter? Research citation in policy documents as an indicator of research impact—An Australian obesity policy case-study. *Health Research Policy and Systems*, 16(1), 55. <https://doi.org/10.1186/s12961-018-0326-9>, PubMed: 29950167
- NIH. (2008). (NOT-OD-09-025) Enhanced review criteria have been issued for the evaluation of research applications received for potential FY2010 funding and thereafter. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-09-025.html>
- Olensky, M., Schmidt, M., & van Eck, N. J. (2016). Evaluation of the citation matching algorithms of CWTS and iFQ in comparison to the Web of Science. *Journal of the Association for Information Science and Technology*, 67(10), 2550–2564. <https://doi.org/10.1002/asi.23590>
- Ophhof, T., & Leydesdorff, L. (2010). Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance. *Journal of Informetrics*, 4(3), 423–430. <https://doi.org/10.1016/j.joi.2010.02.003>
- Orduna-Malea, E., Thelwall, M., & Kousha, K. (2017). Web citations in patents: Evidence of technological impact? *Journal of*

- the Association for Information Science and Technology, 68(8), 1967–1974. <https://doi.org/10.1002/asi.23821>
- Overton. (2022). *How are scholarly references matched in policy documents?* Retrieved May 9, 2022, from <https://help.overton.io/article/how-are-scholarly-references-matched-in-policy-documents/>.
- Pallari, E., Eriksson, M., Billhult, A., Billhult, T., Aggarwal, A., ... Sullivan, R. (2021). Lung cancer research and its citation on clinical practice guidelines. *Lung Cancer*, 154, 44–50. <https://doi.org/10.1016/j.lungcan.2021.01.024>, PubMed: 33611225
- Pallari, E., & Lewison, G. (2020). The evidence base of international clinical practice guidelines on prostate cancer: A global framework for clinical research evaluation. In C. Daraio & W. Glänzel (Eds.), *Evaluative informetrics: The art of metrics-based research assessment* (pp. 193–212). Springer International Publishing. [https://doi.org/10.1007/978-3-030-47665-6\\_9](https://doi.org/10.1007/978-3-030-47665-6_9)
- Pendlebury, D. A. (2009). The use and misuse of journal metrics and other citation indicators. *Archivum Immunologiae et Therapiae Experimentalis*, 57(1), 1–11. <https://doi.org/10.1007/s00005-009-0008-y>, PubMed: 19219526
- Penfield, T., Baker, M. J., Scoble, R., & Wykes, M. C. (2014). Assessment, evaluations, and definitions of research impact: A review. *Research Evaluation*, 23(1), 21–32. <https://doi.org/10.1093/reseval/rvt021>
- Pinheiro, H., Vignola-Gagné, E., & Campbell, D. (2021). A large-scale validation of the relationship between cross-disciplinary research and its uptake in policy-related documents, using the novel Overton altmetrics database. *Quantitative Science Studies*, 2(2), 616–642. [https://doi.org/10.1162/qss\\_a\\_00137](https://doi.org/10.1162/qss_a_00137)
- Potter, R. W., Szomszor, M., & Adams, J. (2020). Interpreting CNCIs on a country-scale: The effect of domestic and international collaboration type. *Journal of Informetrics*, 14(4), 101075. <https://doi.org/10.1016/j.joi.2020.101075>
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). *Altmetrics: A manifesto* (Tech. Rep.). <https://altmetrics.org/manifesto>
- Rafols, I., Ciarli, T., & Chavarro, D. (2020). Under-reporting research relevant to local needs in the global south. Database biases in the representation of knowledge on rice. *SocArXiv*. <https://doi.org/10.31235/osf.io/3kf9d>
- Ravenscroft, J., Liakata, M., Clare, A., & Duma, D. (2017). Measuring scientific impact beyond academia: An assessment of existing impact metrics and proposed improvements. *PLOS ONE*, 12(3), e0173152. <https://doi.org/10.1371/journal.pone.0173152>, PubMed: 28278243
- REF2020. (2020). *Guidance on revisions to REF 2021* (Tech. Rep.). UKRI. <https://www.ref.ac.uk/media/1417/guidance-on-revisions-to-ref-2021-final.pdf>
- Reinhardt, A., & Milzow, K. (2012). *Evaluation in research and research funding organisations: European practices* (Tech. Rep.). European Science Foundation. <https://doi.org/10.22163/iteval.2012.97>
- Roach, M., & Cohen, W. M. (2013). Lens or prism? Patent citations as a measure of knowledge flows from public research. *Management Science*, 59(2), 504–525. <https://doi.org/10.1287/mnsc.1120.1644>, PubMed: 24470690
- Rogers, G., Szomszor, M., & Adams, J. (2020). Sample size in bibliometric analysis. *Scientometrics*, 125(1), 777–794. <https://doi.org/10.1007/s11192-020-03647-7>
- Salter, A. J., & Martin, B. R. (2001). The economic benefits of publicly funded basic research: A critical review. *Research Policy*, 30(3), 509–532. [https://doi.org/10.1016/S0048-7333\(00\)00091-3](https://doi.org/10.1016/S0048-7333(00)00091-3)
- Shema, H., Bar-Ilan, J., & Thelwall, M. (2015). How is research blogged? A content analysis approach. *Journal of the Association for Information Science and Technology*, 66(6), 1136–1149. <https://doi.org/10.1002/asi.23239>
- Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, 126(6), 5113–5142. <https://doi.org/10.1007/s11192-021-03948-5>
- Sivertsen, G. (2018). The Norwegian Model in Norway. *Journal of Data and Information Science*, 3(4), 3–19. <https://doi.org/10.2478/jdis-2018-0017>
- Szomszor, M., Adams, J., Fry, R., Gebert, C., Pendlebury, D. A., ... Rogers, G. (2021). Interpreting bibliometric data. *Frontiers in Research Metrics and Analytics*, 5, 628703. <https://doi.org/10.3389/frma.2020.628703>, PubMed: 33870066
- Tattersall, A., & Carroll, C. (2018). What can Altmetric.com tell us about policy citations of research? An analysis of Altmetric.com data for research articles from the University of Sheffield. *Frontiers in Research Metrics and Analytics*, 2, 9. <https://doi.org/10.3389/frma.2017.00009>
- Tennant, J. (2020). Web of Science and Scopus are not global databases of knowledge. *European Science Editing*, 46, e51987. <https://doi.org/10.3897/ese.2020.e51987>
- Thelwall, M. (2016). The discretised lognormal and hooked power law distributions for complete citation data: Best options for modelling and regression. *Journal of Informetrics*, 10(2), 336–346. <https://doi.org/10.1016/j.joi.2015.12.007>
- Thelwall, M. (2018). Dimensions: A competitor to Scopus and the Web of Science? *Journal of Informetrics*, 12(2), 430–435. <https://doi.org/10.1016/j.joi.2018.03.006>
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do Altmetrics work? Twitter and ten other social web services. *PLOS ONE*, 8(5), e64841. <https://doi.org/10.1371/journal.pone.0064841>, PubMed: 23724101
- Thomas, D. A., Nedeva, M., Tirado, M. M., & Jacob, M. (2020). Changing research on research evaluation: A critical literature review to revisit the agenda. *Research Evaluation*, 29(3), 275–288. <https://doi.org/10.1093/reseval/rvaa008>
- Tijssen, R. J. W., Yegros-Yegros, A., & Winnink, J. J. (2016). University–industry R&D linkage metrics: Validity and applicability in world university rankings. *Scientometrics*, 109(2), 677–696. <https://doi.org/10.1007/s11192-016-2098-8>, PubMed: 27795591
- Traag, V. A., & Waltman, L. (2019). Systematic analysis of agreement between metrics and peer review in the UK REF. *Palgrave Communications*, 5(1), 29. <https://doi.org/10.1057/s41599-019-0233-x>
- UKRI. (2018). *Gateway to research API 2* (Tech. Rep. Version 1.7.4). UKRI. <https://gtr.ukri.org/resources/GtR-2-API-v1.7.4.pdf>
- Valderrama-Zurián, J.-C., Aguilar-Moya, R., Melero-Fuentes, D., & Aleixandre-Benavent, R. (2015). A systematic analysis of duplicate records in Scopus. *Journal of Informetrics*, 9(3), 570–576. <https://doi.org/10.1016/j.joi.2015.05.002>
- van der Meulen, B., & Rip, A. (2000). Evaluation of societal quality of public sector research in the Netherlands. *Research Evaluation*, 9(1), 11–25. <https://doi.org/10.3152/147154400781777449>
- Van Eck, N. J. (2021). CWTS Leiden Ranking 2021. <https://doi.org/10.5281/zenodo.4889279>
- van Raan, A. F. (2017). Patent citations analysis and its value in research evaluation: A review and a new approach to map technology-relevant research. *Journal of Data and Information Science*, 2(1), 13–50. <https://doi.org/10.1515/jdis-2017-0002>
- Vera-Baceta, M.-A., Thelwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics*, 121(3), 1803–1813. <https://doi.org/10.1007/s11192-019-03264-z>
- Vinkler, P. (2012). The case of scientometricians with the “absolute relative” impact indicator. *Journal of Informetrics*, 6(2), 254–264. <https://doi.org/10.1016/j.joi.2011.12.004>



- Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1), 20–41. [https://doi.org/10.1162/qss\\_a\\_00112](https://doi.org/10.1162/qss_a_00112)
- Wallace, M. L., Larivière, V., & Gingras, Y. (2009). Modeling a century of citation distributions. *Journal of Informetrics*, 3(4), 296–303. <https://doi.org/10.1016/j.joi.2009.03.010>
- 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., & Costas, R. (2014). F1000 recommendations as a potential new data source for research evaluation: A comparison with citations. *Journal of the Association for Information Science and Technology*, 65(3), 433–445. <https://doi.org/10.1002/asi.23040>
- Waltman, L., & van Eck, N. J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, 9(4), 872–894. <https://doi.org/10.1016/j.joi.2015.08.001>
- Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, 5(1), 37–47. <https://doi.org/10.1016/j.joi.2010.08.001>
- Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics*, 10(2), 347–364. <https://doi.org/10.1016/j.joi.2016.02.003>
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., ... Johnson, B. (2015). The metric tide: Report of the independent review of the role of metrics in research assessment and management. <https://doi.org/10.13140/RG.2.1.4929.1363>
- Yin, Y., Gao, J., Jones, B. F., & Wang, D. (2021). Coevolution of policy and science during the pandemic. *Science*, 371(6525), 128–130. <https://doi.org/10.1126/science.abe3084>, PubMed: 33414211
- Yücel, A. G., & Demir, S. B. (2018). Academic incentive allowance: Scientific productivity, threats, expectations. *International Online Journal of Educational Sciences*. <https://doi.org/10.15345/iojes.2018.01.003>
- Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of ‘alternative metrics’ in scientific publications. *Scientometrics*, 101(2), 1491–1513. <https://doi.org/10.1007/s11192-014-1264-0>