

Position Paper

Measuring Research Outputs Through Bibliometrics

2026 Update *(First edition 2016)*

UNIVERSITY OF WATERLOO

https://doi.org/10.15353/10012/23385

{Intentionally left blank}

Position Paper

Measuring Research Outputs Through Bibliometrics

Updated by the Bibliometrics Working Group, University of Waterloo, 2026

Suggested citation:

University of Waterloo Bibliometrics Working Group, 2026. Position Paper on Bibliometrics: Measuring Research Outputs Through Bibliometrics: 2026 Update. Waterloo, Ontario: University of Waterloo

Update prepared by:

Laura Bredahl[1], Jana Carson, Wai Yin Winnie Chan, Nancy Collins[2], Wendy De Gomez[3], Annamaria Feltracco, Ana Ferrer[4], Jordan Hale[5], Jenny Hirst, Kumudinie Kariyapperuma[6], Antonio Marques[7], Ian Milligan[8], Martha Muszycka-Jones[9], Selena M. Santi, Peter Stirling[10], Jon Walgate, of the Task Forces for the University of Waterloo Bibliometrics Working Group.

First edition prepared by:

Lauren Byl, Jana Carson, Annamaria Feltracco, Susie Gooch, Shannon Gordon, Tim Kenyon, Bruce Muirhead, Daniela Seskar-Hencic, Kathy MacDonald, Tamer Özsu, Peter Stirling, of the University of Waterloo Bibliometrics Working Group



For further information regarding this report, please contact:

Bibliometrics Working Group
University of Waterloo
200 University Avenue West
Waterloo, ON Canada
N2L 3G1
Email: BibliometricsWG@uwaterloo.ca

[1] Laura Bredahl: https://orcid.org/0000-0002-9397-0764
[2] Nancy Collins: https://orcid.org/0009-0004-9717-9197
[3] Wendy De Gomez: https://orcid.org/0000-0002-7769-0893
[4] Ana Ferrer: https://orcid.org/0000-0002-7385-2381
[5] Jordan Hale: https://orcid.org/0000-0002-1534-2367
[6] Kumudinie Kariyapperuma : https://orcid.org/0000-0003-4119-932X
[7] Antonio Marques: https://orcid.org/0009-0005-9196-5530
[8] Ian Milligan: https://orcid.org/0000-0002-1470-7723
[9] Marth Muszycka-Jones: https://orcid.org/0000-0002-2672-4643
[10] Peter Stirling: https://orcid.org/0000-0002-8837-1660

TABLE OF CONTENTS

TABLE OF CONTENTS ..... **IV**

FOREWORD ..... **VI**

EXECUTIVE SUMMARY AND RECOMMENDED PRACTICES ..... **VIII**

Purpose VIII
Objectives of the Current Update VIII
Background VIII
Key Findings IX
Core Principles and Summary of Recommendations XII

1 MEASURING RESEARCH OUTPUTS THROUGH BIBLIOMETRICS ..... ERROR! BOOKMARK NOT DEFINED.

1.1 Purpose 1
1.2 What are Bibliometrics and Research Impact 1
1.3 Definitions 2
1.4 The Use of Bibliometrics in the Post-Secondary Sector 3
1.5 The Risks of Bibliometrics Analysis 5

2 CITATION-TRACKING DATABASES **6**

3 LIMITATIONS OF CITATION-TRACKING DATABASES **7**

3.1 Data Accuracy 8
3.2 Database Collection Scope 8
3.3 Reproducibility Challenges 10
3.4 Discipline Variations 10
3.5 Regionally Specific Research 12
3.6 Publication Type Variances 12
3.7 Language Representation 14
3.8 Attributing Authorship 14
3.9 Summary 15

4 BIBLIOMETRIC MEASURES AND THEIR LIMITATIONS **16**

4.1 Publication Counts 17
4.2 Citation Counts 18
4.3 H-index and Other Combined Measures 20
4.4 Collaboration 21
4.5 Journal Level Impact Rankings 22
4.6 Top Percentiles 23
4.7 Summary 23

5 EQUITY AND GENDER BIAS IN BIBLIOMETRICS **28**

5.1 General Benefits of More Diversity in Research 28
5.2 Bibliometric Methods for Determining Equity Dimensions 29
5.3 Measures of Equity at Waterloo 29
5.4 Bibliometrics and Equity Impacts 30
5.5 Recommendations For Equity and Gender Bias in Bibliometrics 34

6 MEASURING RESEARCH IMPACT IN THE HUMANITIES, ARTS, AND SOCIAL SCIENCES 36

6.1 Limitations of Bibliometrics Affecting HASS 36
6.2 Case Studies of Research Outputs in the Humanities, Arts, and Social Sciences (HASS) 40
6.3 Recommendations for Use of Bibliometrics in HASS 42

7 MEASURING SOCIETAL IMPACT WITH ALTMETRICS 43

7.1 What are Altmetrics 43
7.2 Differentiating Impacts 44
7.3 Case Study of Using Altmetrics To Understand Funding Impact 44
7.4 Limitations to Altmetrics 45
7.5 Case Study of Nuances in Citations from Policy 46
7.6 Recommendations For the Use of Altmetrics 47

8 MEASURING INTERDISCIPLINARITY WITH BIBLIOMETRICS 48

8.1 Defining Interdisciplinarity 48
8.2 Measuring Interdisciplinarity 49
8.3 Interdisciplinary Networks 51
8.4 Limitations of Measuring Interdisciplinarity 53
8.5 Recommendations for Measuring Interdisciplinarity with Bibliometrics 53

9 BIBLIOMETRICS AND COVID IMPACTS 54

9.1 Impact of COVID-19 on Academic Research 54
9.2 Recommendations on Monitoring the Effects of COVID-19 on Bibliometrics 55

10 USING BIBLIOMETRICS TO MEASURE AND INFLUENCE 56

10.1 Responsible Research Assessment 56
10.2 Using Bibliometric Indicators Responsibly 56
10.3 Quality Over Quantity Case Studies 57
10.4 Recommendations For Responsible Research Assessment 57

11 SUMMARY AND RECOMMENDATIONS 59

11.1 Limitations on the effective use of bibliometrics 59
11.2 Recommended Practices for Bibliometric Analysis 61

APPENDIX A: 2025 BIBLIOMETRICS WORKING GROUP MEMBERSHIP 68

APPENDIX B: TASK FORCE MEMBERSHIP FOR 2026 UPDATE 69

APPENDIX C: EVIDENCE-GATHERING PROCESS USED 70

APPENDIX D: POSITION PAPER CONSULTATION AND COMMUNICATION PLAN 72

APPENDIX E: LIST OF TABLES AND FIGURES 73

REFERENCES 74

FOREWORD

The University of Waterloo is committed to better understanding how to measure and reflect research output and impact using a range of measures. Important stakeholders, including funders, ranking organisations, and various accountability organisations, are increasingly using bibliometrics as one way to understand research outputs. Further, individual academics want to better understand bibliometrics and how they are used.

In 2012 the University of Waterloo formed the inaugural Bibliometrics Working Group (BWG) comprised of key partners across the University, including the Libraries, Office of Research, Institutional Analysis and Planning (IAP), and representatives from all faculties. During this original work the BWG was tasked with creating resources to better support institutional understanding of bibliometrics and their effective use, with the original paper as a key deliverable. This is an update to the 2016 "White Paper[11] Measuring Research Outputs Through Bibliometrics" (Byl et al., 2016). In 2022, the Bibliometrics Working Group (BWG) created a project team to update the paper with new evidence and emerging issues in bibliometrics.

Since first publishing the original paper, there have been consistent questions about; how to better serve the Arts, Humanities and Social Sciences in bibliometrics and research impact, the role of alternative metrics in analyses, the intersection of equity, diversity and inclusion and bibliometrics, collaboration and interdisciplinarity, and practical guidance on the responsible use of metrics. There was also interest in looking at the impact that COVID has had on research and related outcomes. This update incorporates reflections, lessons learned and the modern context into its existing form.

In this update, six task forces explored how bibliometrics intersects with:

- the humanities, arts and social sciences
- altmetrics
- equity, diversity, inclusion and accessibility (EDIA)
- interdisciplinarity
- impacts of COVID
- responsible, in-practice use

The six task forces, comprised of members who had a working knowledge of bibliometrics in their task force area, focused on updating evidence related to their discrete area, conducted a review of the literature and developed six draft working papers, one on each of the above topics, to inform this update. In addition to integrating this work into the paper and the creation of some new sections, the BWG updated content where new information was available, or the original information was outdated.

A draft of the updated paper was shared with the BWG for feedback, with equity offices, and then shared more widely on campus, and feedback received during these consultations was incorporated into this final paper. Appendix D provides more details on the groups included in the consultation and communication process.

This paper is a contribution to Waterloo's evolving understanding of how to use bibliometrics to measure research output and impact. It provides a high-level review of issues relevant to understanding this topic

[11] Since this update began, the White Paper is now referred to as the Position Paper or simply 'the paper' throughout this document and into the future.

and practical recommendations for how to improve our bibliometric practices. Further efforts have built on this paper and its predecessor and include the creation of an online guide which synthesizes content from this paper and provides recommended readings. Additional supplemental documentation may be developed as needed, for example, to further explore topics such as impacts of bibliometrics on equity in academia, and to capture feedback from the engagement efforts with the broader university community when seeking input and providing communication and education.

For more information, please contact:

Bibliometrics Working Group

BibliometricsWG@uwaterloo.ca

EXECUTIVE SUMMARY AND RECOMMENDED PRACTICES

PURPOSE

Bibliometrics is one family of measures that uses a variety of approaches for counting publications, citations, and authorship. This position paper represents the University of Waterloo's (the University) evolving stance on the responsible and effective use, and associated risks, of bibliometrics to measure research outputs in the pursuit of research assessment and analysis. It updates the original 2016 paper, which has served as a foundational guide for the University and has gained recognition beyond the institution. The 2026 update reflects the University's commitment to maintaining a "living document" that evolves with emerging research, institutional priorities, and societal shifts.

The purpose of this position paper is to incorporate current dialogue, literature, and institutional experience into the existing framework. It aims to expand the scope of bibliometric analysis to better serve all disciplines—particularly the Arts, Humanities, and Social Sciences (HASS)—and to address key themes such as equity, diversity, and inclusion (EDI), alternative metrics, interdisciplinarity, and the impact of COVID-19 on research. The paper does not seek to be prescriptive but rather to gather and synthesize the latest thinking, highlight caveats, provide practical recommendations and advance the conversation on responsible use of research output metrics.

Recommendations offered at the end of this paper provide a range of opportunities for how researchers and administrators at Waterloo and beyond can integrate bibliometric analysis into their practice. Additionally, Table 8 provides a summary of levels (i.e. individual, group/departmental/centres/institutes, and institutional/regional) and types of inquiry that were considered by the working group as appropriate and not appropriate in a variety of situations.

OBJECTIVES OF THE CURRENT UPDATE

The objectives of this update include:

- Integrating current research and institutional reflections into the original document.
- Expanding the scope to include a broader range of research outputs and practical examples from the Waterloo context.
- Providing guidance on the responsible and ethical use of bibliometrics, with attention to disciplinary diversity and equity considerations.
- Engaging university partners and the broader academic community in dialogue, education, and outreach.

BACKGROUND

Important university partners, including funders and ranking organisations, include bibliometrics as a part of their decision-making and accountability measures. As far back as 2011, the University of Waterloo, along with other Canadian universities, began to build institutional understanding and awareness of bibliometrics. In 2012, several key University units, including the Libraries, Office of Research, Institutional Analysis and Planning (IAP), and representatives from all Faculties formed a Bibliometrics Working Group (BWG). A full listing of current and past members of the BWG is provided in Appendix A. The purpose of the

BWG is to assist the University in understanding new realities of how bibliometrics are used, and to provide resources to support researchers and administrators in their effective, fair and responsible use, while highlighting limitations and risks. One of the working group's initial steps was to identify the need for resources to foster a common institutional understanding of bibliometrics and its role in capturing research performance. This resulted in the original "White Paper: Measuring Research Outputs with Bibliometrics", published in 2016 (Byl et al., 2016).

The current update to this original paper was initiated in 2022 and endorsed by the WBG. Six task forces (Appendix B), with broad representation across the University, were tasked with conducting a comprehensive literature review of peer-reviewed and grey literature published since the last paper was completed. Each task force identified, reviewed, and summarized key articles and drafted new sections for the update with support from a project manager, the project sponsors and a writer supported by the IAP. Further details of the evidence gathering process are outlined in Appendix C. The resulting document is a resource for leadership at all levels of the University, research groups, research centres and institutes, individual researchers, students, and other members of the campus community who are interested in better understanding bibliometrics and research impact metrics.

KEY FINDINGS

The key findings offered here include both the findings from the 2016 paper as well as those from this update. It was important to retain these core findings while adding these updates.

Bibliometrics offers a useful approach for measuring only some aspects of research output and impact; as it is subject to significant limitations on their effective, fair and responsible use.

Bibliometrics are Most Useful When:

a) used in combination with qualitative review, such as peer and other expert review, to assess the categorical or non-comparative impact of scholarly work.
b) used alongside other robust quantitative measures to help shape a more complete picture of research and its impacts.
c) used as indicators of quantity and not quality.

Bibliometrics Limitations

The **most significant limitations in bibliometrics** involve the **strong differences in disciplinary cultures** making sound comparisons between disparate disciplines very difficult, especially when comparing individuals and small groups. These differences mean that assigning a major role to bibliometric measures is strongly discouraged for hiring, merit review, tenure, and promotion decision-making, as well as using bibliometric measures alone as a measure for inter-departmental research activity comparisons. The scientific and scholarly content and quality of research outputs, understood by the disciplinary norms is more important than simple publication metrics for these purposes.

LIMITATIONS ON THE EFFECTIVE USE OF BIBLIOMETRICS INCLUDE THE FOLLOWING:

- **All bibliometric measures have limits:** Bibliometrics measures have specific uses and risks. Users need to understand the underlying calculations that go into metrics to properly interpret them. This is one reason why responsible use of research impact metrics declarations such as DORA have gained

significant support. Even normalized measures (such as the Field-weighted citation impact) can be skewed by outliers in small datasets.

- **Different methodologies:** Citation-tracking databases use different methodologies for collecting and reporting bibliometric measures, and their differential indexing of research publications from various fields of study can produce significant limitations to some disciplines.

- **Different Indexes:** Proprietary citation-tracking databases (such as Web of Science and Scopus) index different collections defined by the publications their commercial enterprises hold. Google Scholar, although not defined by proprietary collections, is significantly limited by search conventions that can include non-scholarly works and duplicates citations, and lacks precision and depends on user-created profiles that can be manipulated.

- **Data inaccuracies:** Citation-tracking databases are not infallible. They contain errors in data such as author name misspellings, disambiguation challenges, affiliation errors, or not capturing name changes. Care needs to be taken with smaller datasets, especially author-level analyses, where missed data can have larger proportional impacts on the analysis.

- **Data sovereignty:** Ethical considerations around informed consent and data sovereignty mean institutions must protect anonymity and limit sharing of socio-demographic data. While these safeguards are critical, they limit the granularity of author-level metadata often used for equity analyses. This is a necessary and generally accepted trade-off between privacy and comprehensive measurement. Instead, proxy measures or probabilistic methods of identification are often used in bibliometric methodologies, although they tend to introduce assumption biases in author demographic identification.

- **No comprehensive solution:** No citation-tracking database indexes every type of publication, and comprehensive coverage of research publications is not possible. This limited coverage is reflected in the bibliometrics tools (see Table 1) that draw on data from citation-tracking databases.

- **Disciplinary differences in research outputs:** Academic disciplines produce a range of research outputs, and not all of these are indexed equally well by citation-tracking databases. Therefore, research impact metrics should not be applied equally across disciplines and comparisons between disciplines should be avoided. Research outputs may include papers, patents, conference proceedings, application or systems, databases, data sets, hardware or software artifacts, policy papers, position papers, and government reports, books, exhibitions or creative works, etc.

- **Citation practices differ by discipline:** The practice of attributing citations, and collecting citation data, differs across disciplines and fields. In some fields citations accrue only many years after a work is published, in other fields citations accrue primarily within only a few years after publication. These differences carry over into every bibliometric measure that uses citations, including the h-index.

- **Poor coverage of non-English, interdisciplinary and regional research:** Citation-tracking databases do not have good coverage of research that is not published in English, interdisciplinary research or research of regional importance, and cannot provide field-specific context for research outputs like the extent and type of some research collaborations. This further supports the need for caution when making comparisons across disciplines.

- **Context is lost:** Retrospective analysis of research metrics cannot provide field-specific context for research outputs, such as the extent and type of some research collaborations, or how integrated the research is within the citing works.

- **HASS disciplines deserve special attention:** The humanities, arts and social sciences disciplines tend to be subject to many, if not all the limitations related to disciplinary differences. In particular, citations are typically lower and accrue slower, research is more often locally focused, include more non-English works, there are typically less authors per publication, and the publication types, such as books and creative works, are poorly indexed in citation-tracking databases.

- **Altmetrics aren't always the answer:** Although altmetrics, which expand the breadth of measures of research impact beyond academia, aid in our understanding of social and societal impact, they are also constrained by many of the same challenges as bibliometrics measures and sources. Altmetrics are not direct measures of impact but merely proxies of the attention that a particular study, researcher, or institution may be gaining from sources beyond academia.

- **Bibliometrics are dynamic:** Bibliometric measures taken at different times cannot always be meaningfully compared. First, citations accrue with time after publication. Second, the time required for understanding the impact of a paper using citations differs by discipline. And finally, citation databases themselves change their methodology and journal coverage over time, making longitudinal comparisons problematic.

- **Inequities represented in the data:** There is evidence of biases in citation and collaboration practices (such as self-citations and who gets cited) based on gender and race, and systemic barriers that manifest in slower career paths for women and racialized researchers. The COVID-19 pandemic also affected researchers unequally. These biases cause metrics to underestimate contributions made by women and racialized researchers, especially women of colour. These factors must be taken into consideration when conducting bibliometric analysis, especially if the metrics are being used for assessment activities such as promotion and tenure or grant application reviews.

- **Interdisciplinarity measures are not one-size fits all:** Combining cognitive and organisational approaches to analyze interdisciplinarity helps capture the breadth of ways that researchers collaborate across disciplines. The cognitive approach focuses on information flows and citation links between published literature, while the organisational approach looks at collaboration based on researcher affiliations and subject classifications. Simply looking at journal classification systems within citation-tracking databases may lose the nuance of interdisciplinary innovations.

- **Metric manipulation and behavioural distortions:** Bibliometric measures influence researcher behaviour in unintended ways. They may shape where and how researchers choose to publish to maximize their expected metrics or potentially steer what researchers choose to study. At worst, metrics create incentives for gaming or manipulation. Disciplinary differences in tool coverage and usability can be misinterpreted as differences in research activity or impact, rather than reflecting limitations of the tools themselves.

Tables 3 and 8 provide summaries of bibliometric measures, limitations, levels and types of inquiry and the appropriate uses.

CORE PRINCIPLES AND SUMMARY OF RECOMMENDATIONS

The position paper reaffirms core principles established in 2016:

- bibliometric measures should be used transparently
- bibliometric measures should be field-normalized
- bibliometric measures should never be used in isolation.

It then expands on these principles through a series of thematic sections, each offering updated recommendations (expanded within each section) that are summarized here:

- **Citation-Tracking Databases:** Emphasizes the need for multiple data sources, validation through persistent identifiers like ORCID, and awareness of disciplinary coverage gaps.
- **Bibliometric Indicators and Measures:** Advocates for normalized citation metrics, integration with peer review, and caution in cross-disciplinary comparisons.
- **Equity and Bias:** Calls for ethical data governance, inclusive evaluation practices, and institutional support for equity-deserving scholars.
- **HASS Disciplines:** Recommends combining bibliometrics with qualitative assessments and altmetrics to capture long-term and societal impacts.
- **Altmetrics and Societal Impact:** Encourages the use of altmetrics as complementary tools, with clear distinctions between social attention and societal change.
- **Interdisciplinarity:** Suggests using cognitive and organisational approaches, including network analysis, to assess interdisciplinary collaboration.
- **COVID-19 Impacts:** Highlights the need to monitor long-term effects on research productivity and equity, particularly for early-career and marginalized researchers.
- **Institutional Responsibility:** Warns against over-reliance on metrics and promotes a balanced, context-sensitive approach to research evaluation.

While these are important considerations the previous recommendations still hold true:

- Define a researcher's identity convention as an author early and use that convention systematically throughout their career.
- Approach the process of analysing research outputs in the same way that one would conduct good research.
- Consider bibliometrics as one measure among a set of others for understanding research output and impact.
- Understand and account for variations in how disciplines produce and use research publication.
- Involve those being evaluated in the process and provide them with interpretive information.
- Understand the distinctions among bibliometric measures.

This position paper is the product of a collaborative effort led by the Bibliometrics Working Group, supported by task forces focused on each thematic area. It is intended as a resource for researchers, administrators, and evaluators, offering practical guidance and reinforcing the university's commitment to ethical, inclusive, and evidence-informed research assessment.

1 MEASURING RESEARCH OUTPUTS THROUGH BIBLIOMETRICS

Bibliometrics serve as one tool, among many, used by universities, funders, ranking organisations, and others to measure research outputs.

Post-secondary institutions face increasing pressures from funding bodies, the public, and other institutions to measure and understand the amount, and impact of, research conducted by their institution. Bibliometrics are a series of measures used by universities, funders, ranking organisations, and others to assess research outputs. However, bibliometrics present both opportunities and challenges for accurate assessment. The process of creating understanding and interpreting bibliometric analysis can be significant. In organisations where resources are limited, staff and researcher time and funds spent on research metrics has both real and opportunity costs (Van Raan, 2007). A thorough understanding of bibliometrics, their limitations, and how they capture – or do not – academic impact in a variety of situations – can optimize the appropriate use of bibliometric measures and ensure that resources used to collect, analyse, and understand them, are well spent and aligned with priorities.

1.1 PURPOSE

The purpose of this Position Paper is to:

1. define and summarize evidence that addresses appropriate use of bibliometric analysis at the University of Waterloo and the intended goals of performing the analysis
2. develop recommendations about their use and understanding
3. explore how the institution can continue to support the appropriate use of these measures.

Understanding bibliometric measures and their limitations makes it possible to identify how, and under what circumstances, bibliometrics can be used effectively.

In 2026, the original White Paper was updated to include new research in vital areas and more deeply explore implications in priority areas of interest.

This update maintains the original content that describes commonly used citation-tracking databases and bibliometric measures, describes their proper uses, and provides an overview of the limitations of bibliometric analysis as both a method and within various disciplines. This update adds perspectives on the use of bibliometrics in the humanities, arts, and social sciences, including deeper analysis of altmetrics as short- and long-term social impact measures; looks at the use of bibliometrics for understanding interdisciplinarity; examines equity considerations; and the impact of the global COVID pandemic on research impact. Lastly, practical guidance is given on the responsible use of bibliometrics. This position paper concludes with recommendations that provide a thoughtful approach for how bibliometrics can be used effectively for research impact assessments.

1.2 WHAT ARE BIBLIOMETRICS AND RESEARCH IMPACT

Bibliometrics are one important tool among a basket of potential processes and related tools used to understand aspects of research output, such as authorship, documents, citation, affiliation and collaboration. Common activities that incorporate bibliometric measures can include: individual peer

Bibliometric analysis is one important tool among a basket of potential processes and related tools used to understand aspects of research output.

review of funding applications and institutional funding, granting of awards, university rankings, individual assessment for promotion and tenure, annual reporting or strategic planning at the group, unit or institutional level, and library collections assessment and planning.

In academia, the culture of 'publish or perish' has long been linked to determining the success of individual researchers. In recent years, there has been a shift from the understanding that publishing is an important result of research productivity, to a market competitions approach driven by the need to produce a variety of quantitative measures of research impact (Van Dalen & Henkens, 2012).

1.3 DEFINITIONS

1.3.1 Research Impact

Research impact is considered to be "...the social, economic, environmental and/or cultural benefit of research to end users in the wider community regionally, nationally, and/or internationally" (Bornmann & Marx, 2014, p. 212). Also, "the impact of a piece of research is the degree to which it has been useful to other researchers" (Bornmann et al., 2008, p.93). Assessing research impact in a fulsome way is not possible without a complex set of indicators, some of which may include bibliometrics, but others that do not. However, there is often a demand for simple quantitative measures because they are easier to use, can facilitate comparisons and are easier to decipher. In part, interest in trying to understand research impact gave rise to interest in measurement of research outputs, including bibliometrics.

In recent years, there has been a shift from the understanding that publishing is an important result of research productivity, to a market competitions approach driven by the need to produce a variety of quantitative measures of research impact.

1.3.2 Research Output

Research output is the measure of the research activity. It is considered a key element of a university's and an academic's achievements and is typically defined by the number and quality of research products a researcher, department, or institution has produced within a specific timeframe. Typical research outputs can include research published in refereed journals, conference proceedings, books, patents, policy reports, and other artifacts (e.g., exhibitions, developed systems, data sets, software, and hardware artifacts). Depending on the context, scholarly blogs and radio or television broadcasts may be categorized as research output.

1.3.3 Research Metrics

Research metrics are the quantitative measures that are used to quantify research output. These measures may include, but are not limited to, bibliometrics. Other commonly used research metrics include research funding, awards, publication acceptance rates, and the development of highly qualified personnel (HQP).

1.3.4 Bibliometrics Measures

Bibliometric measures are used to express an amount or degree of research or academic output and some associated measures of impact. Measures typically include the number of academic publications and

citations of a single researcher, group of researchers, or an institution. An important component of bibliometric measures are the citation-tracking databases that capture and report bibliometrics.

1.3.5 Citation-Tracking Databases

Citation-tracking databases track citations by counting how many times a particular work has been cited in other works included in the same database. As each of these databases are closed systems (each one indexing different content), publication and citation counts will naturally differ based on the data resource. Common citation-tracking databases include Web of Science, Scopus, and Google Scholar. Data collected and indexed by citation-tracking databases are often used as the basis of bibliometric measures.

1.3.6 Altmetrics

Altmetrics, or alternative metrics (not to be confused with the company Altmetrics) refers to a group of indicators that propose measuring research impact differently than traditional citation and research output measures. Altmetrics focus on broader social impact indicators (García-Villar, 2021) through tracking online mentions of research outputs. These metrics are leveraged primarily from online social media and blog sources (Shema et al., 2014) but include a growing number of sources such as traditional news outlets, policy documents, patent documents, usage/views data from reference managers (i.e. Mendeley), online scholarly contents such as datasets, syllabi, websites, and post-publication peer-review (Bredahl, 2022).

The main goal of altmetrics is to make tools available to the scientific community that allow for a more globalised and social vision of the impact of both publications and researchers, as an alternative or better yet a complement, to the strictly scholarly attention received through traditional peer-reviewed citation paths. Altmetrics are covered in greater detail in Section 7.

1.4 THE USE OF BIBLIOMETRICS IN THE POST-SECONDARY SECTOR

Key partners – governments, industry partners, and other funders – use bibliometric measures to assess research outputs.

In the last two decades, the use of bibliometrics has gained popularity in the post-secondary sector with a broad range of uses among partners. These include: governments funding decisions, accountability, and external assessments, university rankings, university internal assessments and strategic planning, university hiring, tenure and promotion, and university library collection analysis. These are explored in the following sections.

Although not exhaustive these uses cover:

1.4.1 Government Funding Decisions, Accountability and External Assessment

Governments are interested in ensuring public accountability in their investments in research institutions. In the past, bibliometrics have been widely used as an integral part of the evidence of the benefits of these investments. However, there have been several movements in recent years to deemphasize the focus on using quantitative measures to assess research for accountability purposes.(CoARA, 2022; Curry et al., 2022; DORA, 2012; Hicks et al., 2015; Wilsdon et al., 2015). The Tri-Agencies along with Canadian Foundation for Innovation and Genome Canada have signed DORA, reaffirming their commitment to meaningfully assess research quality through broader definitions of research outputs and impacts In some jurisdictions, the focus has begun to shift to more extensive consultation, peer-review, expert review and

autonomy across organisations in sharing their impact stories. Instead of a complete dismissal of bibliometrics, responsible research assessment encourages that bibliometrics be used ethically; ensuring that quantitative measures are indicators of quantitative outputs, not the quality of the research (CoARA, 2022), and that the methodologies used in assessments are transparent. Both the Ministry of Colleges, Universities, Research Excellence and Security (MCURES) in Ontario's Strategic Mandate Agreements (SMA) and the United Kingdom's (UK) Research Excellence Framework are governmental oversight bodies that have recently moved away from an "all-metrics" approach (Curry et al., 2022) to combine metrics with qualitative narrative submissions for accountability assessments. The Canadian federal government, among other funders, have used bibliometric measures and other means to inform decisions about funding opportunities in priority areas. For example, 2022 during the Canada First Research Excellence Fund competition, where international leadership in proposed interdisciplinary research areas was evidenced by appropriate and complementary bibliometrics, in line with the tri-agencies' commitment to DORA (Government of Canada, 2022).

1.4.2 University Rankings

Beyond governmental accountability programs, university ranking programs produce a rank-ordered list of post-secondary institutions based, in part, on bibliometric measures (Hazelkorn, 2013). The primary motivation for participating in these rankings is their influence in the national and international reputation of the institution. There is no universally accepted set of measures that fully and appropriately assesses university research outputs. Nonetheless, bibliometrics as a component of research output have been used, and continue to be used extensively, for this purpose (Van Vught & Ziegele, 2011). Currently, one standout ranking is the Leiden Ranking created by the Centre for Science and Technology Studies (CWTS) at Leiden University. They published "Ten principles for the responsible use of university rankings" in 2017, which focus on the design, use and interpretation of rankings, leading to some changes in the way they present their own rankings (Waltman et al., 2017). In 2024 they announced a move to an open edition of their rankings relying on open-source data provided by OpenAlex, instead of the proprietary data from Clarivate's Web of Science dataset. This shift to OpenAlex is aimed at increasing the transparency of the indicators and the underlying data. Currently, open source bibliometrics data has its own set of challenges, in particular its data integrity, which will need to be overcome for the Leiden's open edition ranking to be successful.

There is no universally accepted set of measures that fully and appropriately assesses research output.

1.4.3 University Internal Assessments and Strategic Planning

In addition to external assessment, bibliometric analysis has been used to understand and interpret research outputs internally within institutions. These analyses may take several forms including internal and external benchmarking, identifying research strengths or weakness, identifying emerging research trends, identifying potential or existing collaborations and formal partnerships, supporting strategic planning, or reporting on or celebrating research group performance (Gordon & Hitchens, 2020; Morphew & Swanson, 2011). For example, researchers, departments or a whole institution commonly track bibliometric measures of their own research performance at a single point, or over time, as a way to gauge their productivity against peers or those with whom they seek to collaborate.

1.4.4 University Hiring, Promotion and Tenure

Hiring and promotion and tenure committees sometimes use selected bibliometric measures for an individual researcher to assess their relative research. These are some of the most problematic uses of bibliometrics and have been the focus of many of the bibliometrics reforms that have occurred over the last decade. Some institutions, such as the University of Waterloo, explicitly caution against such use, encouraging that the "primary assessment of quality, originality and impact" is made based on the work itself, rather than secondary indicators. Researchers themselves may choose to submit metrics but it's suggested they do so autonomously, present relevant comparisons and only provide the data as supplemental evidence to their narratives. This is a stance that should be agnostic to the disciplinary field, meaning some disciplines likely have more reform work to do than others. These responsible research assessment reforms are discussed in further detail throughout this update.

Institutions, such as the University of Waterloo, may explicitly caution against use of bibliometrics by reviewers on promotion and tenure committees. The "primary assessment of quality, originality and impact" should be made based on the work itself, rather than secondary indicators.

1.4.5 University Library Collections Analysis

Although bibliometrics have a long history of use for library collection analysis and decisions, the interest in their use for these purposes has been renewed as open science initiatives take hold across the academic sector. With the growth of open access resources and the development of new publishing models there is interest in understanding the impact of open access on the publication behaviours of researchers, implications for access to research and compliance with open science policies such as the Tri-Agency Open Access Policy on Publications (Government of Canada, 2016) and the Tri-Agency Research Data Management Policy (Government of Canada, 2021).

Systemic issues with bibliometric measures and their use, however, do exist. These limitations must be carefully identified and considered before making any determinations or judgements using bibliometrics analysis.

1.5 THE RISKS OF BIBLIOMETRICS ANALYSIS

The very process of measuring research can lead to practices that challenge research integrity. Whenever productivity is measured, the opportunity to manipulate metrics (even to commit academic misconduct) exists (Furner, 2014; Gagolewski, 2013; Hazelkorn, 2013; Hicks et al., 2015). For example, researchers, funders, and institutions may elect to pursue research in areas that are more highly cited in the short term, rather than focusing on the researchers' areas of interest (Wouters, 2014). A focus on only highly cited research may draw attention away from research areas that may be just as worthwhile, but are less highly cited (Michels & Schmoch, 2014; Paiva et al., 2012; Van Vught, 2013).Indeed, it is well known that some disciplines, particularly the humanities, arts and social sciences (HASS), that are under-represented in bibliometrics, due to their disciplinary publishing norms, and not because of their inadequate citation or research excellence. Discipline variations are further explored in Section 3.

The very process of measuring research productivity can lead to practices that challenge research integrity, such as manipulation of metrics, or academic misconduct.

Post-doctoral fellows and new faculty are under particular pressure to publish in order to build their academic profiles and be considered for grants (Johann et al., 2024; Lawrence, 2008). Consequences of these pressures to publish range from researchers reading fewer papers, reduced interest in producing research in layperson's terms for policy papers and for the public, and lowered incentives to publish work in domestic journals (Van Dalen & Henkens, 2012). This may cause researchers to engage in extrinsically motivated behaviours that will produce impressive metrics for themselves, rather than intrinsically motivated behaviours that will benefit a larger group of people or are done out of interest or curiosity (Johann et al., 2024).

There are systemic issues associated with the use of bibliometrics and they must be carefully identified and considered before making judgements based on these measures. Section 2 and Section 3 further addresses these challenges through describing common citation-tracking databases, bibliometrics measures, and their limitations. As well, Section 4 further describes appropriate uses and recommended approaches for using bibliometric indicators and measures in any analysis.

2 CITATION-TRACKING DATABASES

Citation-tracking databases and common bibliometric measures are all used to report bibliometric data. Before interpreting the bibliometric outputs, it is critical to understand how citation-tracking databases capture and report data and measures, what they quantify, how the data are collected and when, and to use that information to select meaningful measures appropriate to the discipline and the context that is being measured.

Some post-secondary institutions and academic-industry collaborations have developed their own bibliometric measures using citation-tracking databases. For example, the Centre for Science and Technology Studies (CWTS) at Leiden University has developed a set of bibliometric measures to support the Leiden Ranking (https://www.leidenranking.com/), and a consortium of academic-industry partners have developed a "recipe book" called "Snowball Metrics" (https://snowballmetrics.com/), now primarily used by the SciVal system by Elsevier, to create global standards for institutional benchmarking. These tools are referenced in the examples below and use elements of the same citation-tracking databases and bibliometric measures described here.

Understanding how citation-tracking databases capture and report data and measures, what they quantify, how the data is collected and when, and using that information to select meaningful measures that are appropriate for the discipline and the context that is being measured, is a crucial task.

Citation-tracking databases are used extensively to collect and report a range of bibliometric measures. The most commonly used citation-tracking databases include proprietary databases that index citations along with the publications within their collections (see Table 1); key tools include Clarivate's Web of Science, Elsevier's Scopus, Digital Science's Dimensions, Lens.org and OurResearch's OpenAlex. Google Scholar is another tool that links publications to their citations but fails to provide robust data and access for bibliometric uses. Each citation-tracking database applies its unique methodological approach to determine how to collect data, including which publication and document types to index, such journal articles and conference proceedings. These differences, combined with differences in disciplines covered and the methodologies used by the database, directly impact the bibliometric measures derived from using each database.

In addition, these common Bibliometrics tools, listed in Table 1, sometimes also referred to as 'research impact analytics' or 'research intelligence' tools, are most often proprietary tools that generate bibliometric data based on, and therefore limited by, the underlying data of the citation-tracking database it uses for creating the bibliometrics, such as those mentioned above. These tools/products offer enhanced ways to explore research outputs which are not possible in the stand-alone citation-tracking database. They allow institutions to analyze their research output, benchmark data from a researcher level to an institutional level on a global scale and offer opportunities to observe research outputs across a period of time. In addition to the standard bibliographic data that these systems use they are constantly evolving their metrics and incorporating more data sources such as patents, news, policy, and grants. These systems are trying to link research with citations or mentions to allow institutions to analyse their impact on an ever-expanding ecosystem of scholarly impacts.

Each citation-tracking database applies its unique methodological approach to determine how to collect data, which journals and other works to index, as well as preferred document types. These differences, combined with differences in areas or disciplines covered by it and the methodologies used by the database, directly impact the bibliometric measures derived from using each database.

TABLE 1: COMMON CITATION-TRACKING DATABASES AND THEIR ASSOCIATED BIBLIOMETRIC TOOLS.

CITATION-TRACKING DATABASE	BIBLIOMETRIC TOOL	COMPANY
SCOPUS	SciVal	Elsevier (RELX Group)
WEB OF SCIENCE	InCites	Clarivate
DIMENSIONS	Dimensions Analytics	Digital Science
LENS.ORG	Lens Analytics	Cambia (non-profit)
OPENALEX	OpenAlex / API-based tools / Freemium services	OurResearch (non-profit)
CROSSREF	OpenCitations / Custom tools	Crossref (non-profit)
GOOGLE SCHOLAR	None (used in Publish-or-Perish)	Alphabet Inc.
MICROSOFT ACADEMIC (RETIRED 2021)	None (custom tools)	Microsoft

3 LIMITATIONS OF CITATION-TRACKING DATABASES

Bibliometrics tools are based on their underlying proprietary citation-tracking database as the data source. Thus, while they offer interesting opportunities for different types of analysis, they retain the flaws of the underlying citation-tracking database on which they are based. It is therefore strongly recommended that any analysis reports their data source for transparency, data validation, and

reproducibility, to uphold responsible use of research assessment practices (further discussed in Section 10)

Citation-tracking databases are susceptible to limitations based on two main factors, further outlined below [12]:

Citation-tracking databases are susceptible to limitations resulting from the data accuracy; parameters of their proprietary collections; the nature of disciplines, sub-disciplines, and related impacts; and how authorship is attributed.

1. Data accuracy
2. Database collection scope
3. Reproducibility challenges
4. Discipline variations
5. Authorship

3.1 DATA ACCURACY

The accuracy of the data reported through a database is dependent on the accuracy of how the data is initially entered. As an example, misspellings in author names and errors in institutional attribution are commonly found in these resources. A citation-tracking database is only as good as the data that it indexes. Persistent identifiers (PID) are a valuable element the research information infrastructure. They allow unique identifiers to be assigned to people, institutions, and objects such as documents or datasets. A commonly known PID is the DOI assigned to digital objects. An important approach researchers can use to ensure that their works are accurately attributed to them within digital systems is to create an author profile like Open Researcher and Contributor ID (ORCID) that allows the author to control how their name will appear in published form throughout their career. There are also persistent identifiers for institutions such as the ROR.

A fundamental limitation of citation-tracking databases is that the data reported through the database is dependent on the accuracy of how the data is initially entered. Garbage in, garbage out.

3.2 DATABASE COLLECTION SCOPE

3.2.1 DATABASES ARE NOT COMPREHENSIVE

No citation-tracking database is comprehensive. At best, each tool offers a large dataset of research outputs based on specific collection parameters. No single database indexes every type of publication, and no single citation-tracking database has the same coverage. Therefore, citation analysis does not provide a comprehensive indication of any given researcher's, or institution's, research output.

No single database indexes every type of publication, and no single citation-tracking database has the same coverage. Citation analysis does not provide a comprehensive indication of a given researcher's, or institution's, research output.

As an example, Waterloo used InCites, a research analytic tool to analyze research productivity within a sample of researchers affiliated with the Institute for Quantum Computing (IQC). This exercise compared IQC's Publications Database to the results of a search for publications produced by IQC researchers within InCites. While the results revealed comparable trend data between the two sources, there were important differences: InCites contained 94% of the publications

[12] Gender bias – this discussion has been significantly expanded and moved to a separate section. See Section 5 Equity and Gender Bias in Bibliometrics.

that were part of the IQC Publications Database. IQC used multiple data sources to gather citation data for these publications (Web of Science, Scopus, and Google Scholar). The use of InCites (which relies on Web of Science data alone) resulted in a 23% lower citation count for the publications in the IQC Publications Database. This exercise indicates that it may not be possible to reproduce bibliometric data generated from one citation-tracking database with that created by another and emphasizes the differences in indexing previously discussed.

Bibliometrics cannot offer a comprehensive data set.

3.2.2 Uncertainty and Sampling

Bibliometric measures cannot offer a comprehensive dataset. Consequently, anyone using bibliometric measures is advised to consider bibliometrics as reflecting a certain degree of uncertainty. Provided that one does not expect too much precision from the exercise, one may treat the range of such analyses as a large sample of data indicating trends over time within a specific context (Allen, 2010).

3.2.3 Changes in Database over Time

Even analysing data collected from the same citation-tracking database can be problematic. If one were to compare research outputs over a ten-year timeframe using a particular set of bibliometrics, the methodology used to gather data may have changed, and certainly the data sets themselves (authors, publications) are constantly evolving. This makes comparisons of annual reports a challenge as the data extracted in one year may not be under the same conditions when extracted in a subsequent year. Databases are continuously indexing more items, which increases their coverage, even for backdated items. This means that increased citations over time may be partly due to the fact that the underlying database simply includes more material. It is therefore recommended to recompute trend data in annual reporting, instead of relying on prior datasets.

Comparing research outputs over time can be problematic. Methodology used to gather data may have changed and the data sets themselves are constantly evolving.

3.2.4 Google Scholar's Inflated Counts

Google Scholar, a popular citation-tracking tool, has its own collection scope limitations. Google Scholar finds citations by searching the web, which means that citations to papers are not always from scholarly or peer-reviewed sources, causing confusion in the types of citations and inflating the citation counts received. For example, an undergraduate thesis or an acknowledgement from a paper might be counted as a citation. In addition to collection scope issues, Google Scholar offers limited precision for author names searches and lacks the capability for searching affiliations/institutions. This can result in problematic results for common author names, as it is difficult to do proper refinement. It also only provides bibliometric measures (h-indexes, among others) for researchers that have a Google Scholar Citations Profile, a service which requires researchers to set up a profile and validate suggested publications. This significantly limits its usefulness as a robust bibliometric tool beyond individual research level analyses.

Citations derived from Google Scholar are not always from scholarly or peer-reviewed sources.

3.3 REPRODUCIBILITY CHALLENGES

It is impossible to reproduce bibliometric data generated from one citation-tracking database with that created by another. Completely accurate and comprehensive citation counts are a myth.

It is impossible to reproduce bibliometric data generated from one citation-tracking database with that created by another for three main reasons:

Citation-tracking databases calculate their bibliometric data based on the items they index. For example, Web of Science does not index the journal *21st Century Music*. Therefore, if a researcher publishes an article in *21st Century Music*, neither that article, nor any citations it garners, will be captured by bibliometrics that use data from the Web of Science.

Coverage in citation-tracking databases differs by discipline or even sub-discipline. Therefore, being aware of acknowledging disciplinary effects on research outputs is essential.

1. Bibliometric indicators offered by one source may not be offered by another source. For example, InCites, which uses Web of Science data, offers the metric 'Category-normalized citation impact'. In contrast, SciVal is based on Scopus data offers the 'Field-weighted citation impact'. Although these metrics sound similar they are not calculated the same and therefore cannot be compared.
2. Validation of data is difficult within sources. To report the number of citations a paper receives citation tracking databases match the references from the end of each paper to other papers, some within their index and some not. When the references match up, it is counted as a cited reference in the bibliometrics. The problem arises when there are errors in the reference list and a paper is then not counted; even a simple error in the page numbers of a reference might mean that a citation is not counted. Completely accurate citation counts are a myth. [13]

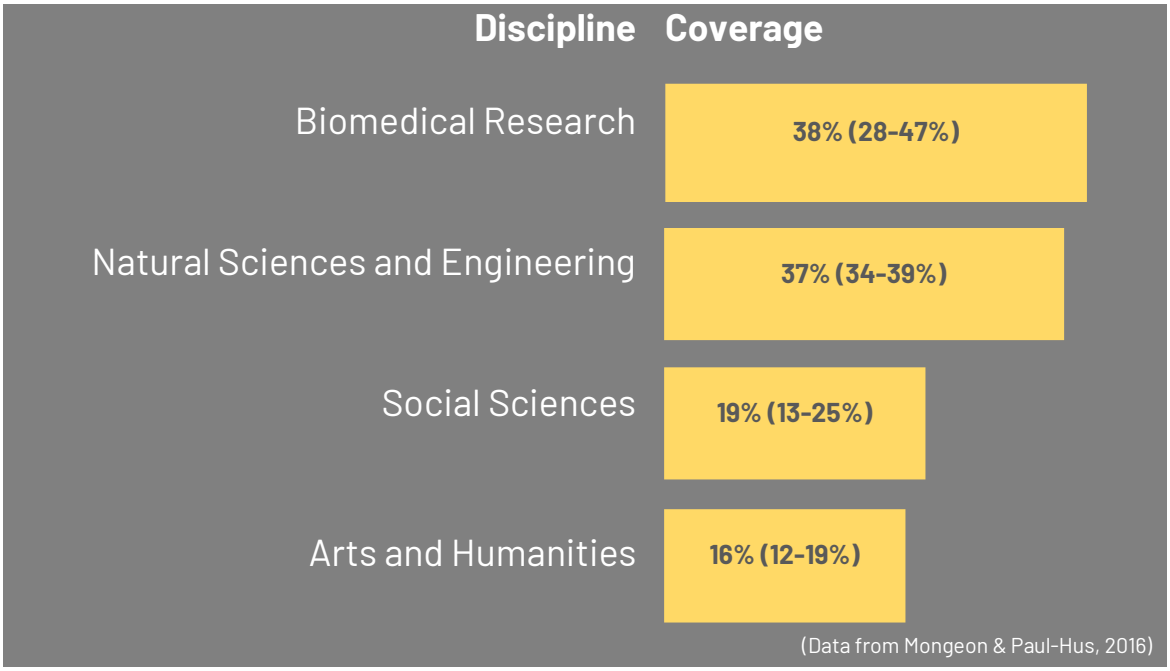
The lack of diversity in publication types in many databases also means that citation-tracking databases and bibliometric data capture only a subset of academically acknowledged research outputs and research. This is problematic where literature such as position papers or policy papers produced for governmental and other public organisations are important research outputs within a discipline. Section 6 discusses Humanities, Arts and Social Science (HASS) disciplines which are particularly affected by these omissions and offers some recommendations for how to better incorporate HASS into research impact assessments.

3.4 DISCIPLINE VARIATIONS

Proprietary citation tracking databases only include publications within their collection, but their coverage may also differ by discipline or sub-discipline. Each citation-tracking database offers different coverage of disciplines by nature of the publications in their collections. Therefore, individuals or groups interested in understanding bibliometrics within a specific discipline must acknowledge the discipline-specific effects of using that citation-tracking database on research outputs (Mryglod et al., 2013).

[13] The University of Waterloo offers a number of resources to support researchers and academics to build strong citation-tracking practices. One of these resources is the Libraries's Academic Footprint Guide. This resource was developed to give authors a comprehensive process that makes tracking citation counts and the h-index relatively self-sustaining over time.

FIGURE 1: DISCIPLINE VARIATIONS IN JOURNAL COVERAGE WITHIN CITATION-TRACKING DATABASES PERCENT OF TOTAL JOURNALS IN ULRICH'S PERIODICAL DIRECTORY.



Singh et al. (2021) reported discipline coverage proportions of each of WOS, Scopus and Dimensions using article and review article level analysis from 2010-18 among 20 countries. They found similar results as Mongeon & Paul-Hus (2016a) presented in Figure 1, with Life Science (43-45%) as the most well represented discipline in the databases and in descending order of representation by Physical Sciences (19-23%), Technology (19-24%), Social Sciences (7-14%) and Arts and Humanities (1-4%) (Singh et al., 2021).

Wainer et al. (2011) noted that even within a discipline, bibliometric database coverage can vary significantly. For example, the subdisciplines of artificial intelligence, architecture and systems within the computer science (CS) discipline were found to have proportionately less coverage in Web of Science than other CS subdisciplines. This demonstrates of the variability with which research output is captured within various disciplines. Dorta-González & Dorta-González (2013) suggest that "systemic differences in publication and citation behaviours across disciplines" (p. 670) are cultural variations likely only understood by individuals within a field.

3.4.1 INTERDISCIPLINARY RESEARCH

Interdisciplinary research has become a fundamental pillar of many universities strategic research agendas. In Canada this has been supported by federal research grant requirements that aim to strengthen research innovation and knowledge integration goals through interdisciplinary research. However, assessing interdisciplinary research through bibliometrics is a great challenge because interdisciplinarity can mean different things to different people or in different situations and is therefore difficult to measure consistently. Therefore, it is important to first understand the

Using citation-tracking databases to assess research output for individuals or institutions that publish regionally or in interdisciplinary journals will under-represent actual output.

goals of the interdisciplinary research and align assessments with those goals. Section 8 explores Measuring Interdisciplinarity with Bibliometrics in greater detail, including different approaches to interdisciplinarity, and outlines their limitations.

However, two main limitations can be highlighted here. First, there are research fields that are naturally more interdisciplinary than others making their publication data difficult to capture. For example Khan and Ho (2012) used Web of Science categories to find environmental science articles but found that many of the discipline's landmark articles were not published in the journals included in this subject area. Second, when looking at interdisciplinarity through co-authorships there are challenges in understanding the precise level and/or type of interactions occurring in the collaboration. For example, Vieira et al. (2024) studied co-authorships in the field of computer science and warned against generalizing their results to other fields due to differences in the nature of co-author relationships in other fields. See Section 8 for further discussion and some possible solutions to these limitations in using bibliometrics to explore interdisciplinarity.

3.5 REGIONALLY SPECIFIC RESEARCH

Researchers that focused on regional issues are at a disadvantage. As an example, major databases do not provide adequate coverage of regional journals. For example, a researcher who publishes about New Brunswick history, or another regionally specific topic, may produce excellent quality research in regional publications; however, those publications may not be covered by citation-tracking databases. Using these databases to assess research output for individuals or institutions that publish regionally or in interdisciplinary journals will under-represent actual output, and comparing them against researchers who publish in different regional, national, international, or discipline-specific journals is not appropriate (Moed, 2005; Van Raan, 2005).

3.6 PUBLICATION TYPE VARIANCES

Disciplines also vary by the type of research output that is produced. In some disciplines, aspects of output are well captured by journal publications and citations. However, in other disciplines, the number of journal article publications and times cited can be less of an indicator of impact than is the number of patents, papers in conference proceedings, produced systems developed and widely used, data sets, or hardware and software artifacts. In the social sciences and humanities, policy papers, position papers, reports produced for government and other public organisations, and books can provide more accurate understanding of research output. In the arts, works produced, exhibitions, and performances may be more important.

In some disciplines, aspects of output are well captured by journal publications and citations. In others, conference proceedings, books and book chapters constitute the major scholarly publication venues.

Researchers in disciplines such as computer science or engineering who predominantly publish via conference proceedings will have a different understanding of their research output depending on the database they use for analysis. For example, conference proceedings indexing in Web of Science and Scopus differs significantly. The Scopus database indexes 10.5 million conference papers, 1823 to present (Scopus search July 2025). In contrast, Web of Science indexes 10.1 million conference papers, 1900 to present (Web of Science search July 2025). As both data sources are unique, this means that computer

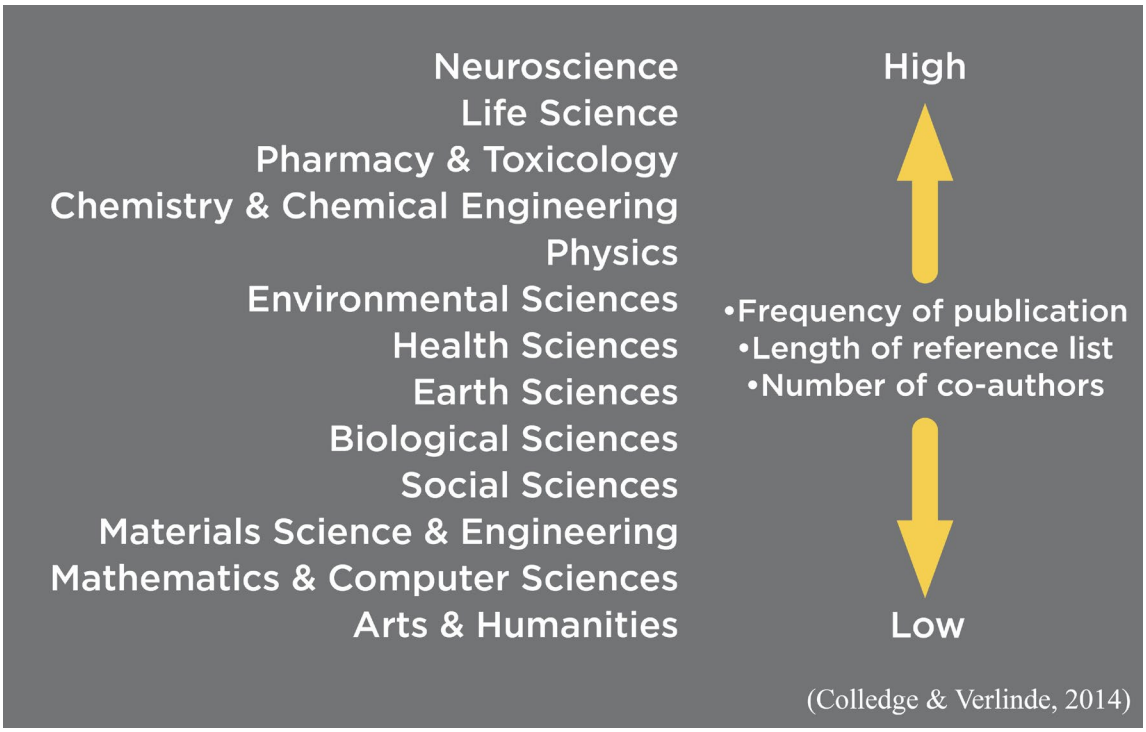
science or engineering researchers using Scopus will have a different understanding of their output vis à vis conference proceedings compared to those using Web of Science.

In some fields of the arts and humanities, books and book chapters – not journals – constitute the major scholarly publication venues (Federation for the Humanities and Social Sciences, 2014). These are notoriously greatly outnumbered by journal article records research databases. By contrast, in the medical sciences almost all research publications are made through serial publications (E. Archambault & Larivière, 2010; Chang, 2013) and are very well covered in the same databases. This means that the use of bibliometrics to assess research output would not be effective for disciplines such as language and linguistics, law, political science, sociology, and educational sciences (Van Raan, 2007). The Federation for the Humanities and Social Science Research recommends that bibliometrics should not be the only tool used to assess research productivity and outputs in the humanities and the social sciences (2014). How representative the bibliometric data are for different disciplines and their coverage of different research outputs make it integral to understand the strengths and weaknesses of a data source in order for the data to be meaningful. Measuring Research Impact in the Humanities, Arts, and Social Sciences is explored further in Section 6.

How representative the bibliometric data are for different disciplines, and their coverage of different research outputs is integral to understand the strengths and weaknesses of a data source, and to understand the meaningfulness of the data.

Citation-tracking databases do not consistently capture the various types of research productivity produced by different disciplines. As a result, there is frequently a disconnect between the amount and calibre of research produced by various disciplines and the research productivity and output data indexed by bibliometric services. Figure 2 illustrates, at a high level, publication behaviours across disciplines (*Research Intelligence: Research Metrics Guidebook*, 2019). It is necessary for users of bibliometric measures to develop a deeper understanding of how publication behaviours vary within disciplines and how that translates into coverage by citation-tracking databases. Moreover, recognition of the capacity of bibliometric measures to effectively represent discipline-level research publication and impact is also required.

FIGURE 2: PUBLICATION BEHAVIOUR ACROSS DISCIPLINES.



3.7 LANGUAGE REPRESENTATION

Another issue of concern regarding citation-tracking databases is their coverage of non-English publications. Citation-tracking databases, such as the Web of Science and Scopus, favour English-speaking countries and institutions (Mongeon & Paul-Hus, 2016b; Van Raan et al., 2011). Moreover, some disciplines publish primarily in English, while others do not. For example, Archambault et al. (2006) showed that more research in the social sciences and humanities is published in non-English languages than in the natural sciences and engineering. This language bias means that the social sciences and humanities are less well represented than other disciplines where publishing in English is the norm (É. Archambault et al., 2006).

In practice, awareness of language bias would influence how an individual uses a tool to make international comparisons. For example, a researcher publishing in another language should not be compared with those publishing in English. Van Raan et al. (2011) note that the language bias in bibliometrics is carried over into university ranking programs that use bibliometrics as a measure of institutional success.

Citation-tracking databases favour English publications; this language bias means that the social sciences and humanities are less well represented than other disciplines where publishing in English is the norm (Archambault, Vignola-Gagnè, Côtè, Larivière & Gingras 2006).

3.8 ATTRIBUTING AUTHORSHIP

Another limitation of citation-indexing databases stems from the different ways in which authorship of multi-authored publications can be attributed in citation analyses. Authorship can be

attributed to all of a publication's authors equally (full counting), or 'fractional counting' might be used, in which relative weights are given to authors in collaborative publications.[14] This means that when citation-tracking databases are used in ranking programs and full counting authorship is used, when an author from Institution A collaborates with an author from Institution B, both institutions get credit for this paper. When fractional counting is used, weights are provided on some basis (for example, first-listed author might be weighted at 1, second-listed author at 0.8, and third-listed author at 0.6, etc.).

Authorship can be attributed to all of a publication's authors equally (full counting), or relative weights can be given to authors in collaborative publications (fractional counting). Authorship is assigned differently across disciplines and citation databases lack the subtlety to differentiate between disciplines and how they attribute authorship.

However, authorship is assigned differently across disciplines, and sometimes even within a discipline (Abramo, D'Angelo, & Rosati, 2013; Retzer & Jurasinski, 2009). For example, in one discipline author names might be placed in alphabetical order, while in another, author names might be placed in order of contribution level. Therefore, since citation-tracking databases lack the subtlety to differentiate between disciplines and how they attribute authorship, treating all articles the same may introduce unintended biases in the data and analysis.

3.9 SUMMARY

Citation-tracking databases are widely used tools to collect and report research outputs using a series of bibliometric measures. Understanding how these databases work and their limitations supports more effective use and accurate reporting of bibliometric measures. Therefore, when reporting bibliometrics it is important to report the data sources used, acknowledge any potential limitations and describe mitigation strategies for these limitations.

Understanding how citation-tracking databases work and their limitations supports more effective use and accurate reporting of bibliometric measures.

[14] The Leiden Ranking methodology provides the following example, "For instance, if the address list of a publication contains five addresses and two of these addresses belong to a particular university, then the publication has a weight of 2/5 = 0.4 in the calculation of the indicators for this university" (Centre for Science and Technology Studies, Leiden University, 2015). InCites uses whole counting for authorship and credits all authors of a paper equally (http://researchanalytics.thomsonreuters.com/m/pdfs/indicators-handbook.pdf] and Leiden provides users the option to use full or fractional counting for authorship. It is unclear whether QS uses whole or fractional counting for faculty and citation attributions. Similarly, Webometrics uses an excellence rating provided by ScImago (10% of papers by citations) but it is unclear if it is fractional. In contrast, ARWU distinguishes the order of authorship using weightings to credit the institutions to which the author is affiliated. This means that a 100% weight is assigned for corresponding author affiliation with 50% for first author affiliation (or second author affiliation if the first author is the same as the corresponding author affiliation), 25% for the next author's affiliation, and 10% for subsequent authors (Shanghai Ranking Consultancy, 2014).

4 BIBLIOMETRIC MEASURES AND THEIR LIMITATIONS

In this section bibliometric measures that are captured by citation-tracking databases are described and their limitations assessed to further advance the understanding and reliability of bibliometric reporting. Bibliometric measures are one type of metric, within a basket of measures, used to assess research outputs both quantitatively and qualitatively. The most commonly known measure is peer-review. This scholarly review of a researcher's body of work by a group of peers, or experts in the same field, has long been considered the gold standard for understanding output (Abramo, D'Angelo, & Costa, 2011; Abramo, D'Angelo, & Di Costa, 2011; Haeffner-Cavaillon & Graillot-Gak, 2009; Lovegrove & Johnson, 2008; Lowry et al., 2013; Mryglod et al., 2013; Neufeld & von Ins, 2011; Rodríguez-Navarro, 2011; Taylor, 2011; Wainer & Vieira, 2013).

Other measures of research output include:

- the type and amount of intellectual property produced (e.g., patents, licenses, spin-offs, datasets, software, etc.)
- the type, amount and success rates of research awards received
- the nature and number of highly qualified personnel developed by the researcher or group
- the impact on public policy and social practice
- altmetrics [15] such as attention in social media or news media
- publication acceptance rates (the proportion of papers or conference presentations accepted compared to the number submitted). Acceptance rates are sometimes used as a proxy for scholarly quality.

Quantitative bibliometric measures can offer important contributions to understanding research output when used in conjunction with other measures (Pendlebury, 2009; Rodríguez-Navarro, 2011). As an example, the League of European Research Universities and the Australian Group of Eight Coalition of research universities each identify the use of bibliometrics as one tool, among a suite of tools, to assess research output (Phillips & Maes, 2012; Rymer, 2011). Moreover, there are cases where some papers considered in the field as 'the best' by experts in that field are not always the most highly cited (Coupé, 2013). Since it is widely acknowledged that the peer-review process is susceptible to bias (Butler & Mcallister, 2009; Van Raan, 1996), Van Raan recommends that blending of bibliometrics and peer-review could possibly mitigate prominent concerns with both methods.

Bibliometric measures can offer important contributions to understanding research output when used in conjunction with other measures (Pendlebury, 2009; Rodríguez-Navarro, 2011).

This section provides a description of six of the most common bibliometric measures used by post-secondary institutions, along with limitations of their use. In addition to the definitions and limitations outlined here, Table 3 provides a brief look at each measure, its intended function, and its appropriate use. Because data for bibliometric measures are collected and analysed using citation-tracking databases, the concerns outlined for using citation-tracking databases should all be considered in concert with these measures.

[15] See Section 7 Measuring Societal Impact with Altmetrics for more details.

4.1 PUBLICATION COUNTS

Absolute number of publications. An absolute count is the cumulative total number of publications produced by an entity (researcher, centre, institute, department, institution, etc.). Publication counts are a basic component of the formula used to calculate other measures, such as citation counts, h-index, normalized citation impact, international and industrial collaborations, among others.

A simple count of publications cannot always be considered a positive indicator of research impact or quality.

It is important to recognize that the terminology used to describe publications sometimes differs across resources. InCites, for example, uses the term 'Web of Science Documents' to capture total publications indexed by Web of Science. In contrast, Snowball Metrics, which was initially developed with SciVal, uses the term 'Scholarly Output' for the same purpose within the context of its data source (*Research Intelligence: Research Metrics Guidebook*, 2019). Regardless, the term *'publication count'* represents the total number of publications indexed by a particular database, within a specified area.

Bibliometric databases also use different methodologies for counting publications based on authorship. These different methods are discussed under Section 3.

4.1.1 LIMITATIONS

A simple count of publications cannot always be considered a positive indicator of research impact or quality. The significance of the publication to other researchers in the field, the type of publication in which the researcher's work is published, and the number of collaborators on the paper are other issues related to publication counts that must be considered.

Citation-tracking databases, which provide publication count data, are inherently limited by their proprietary collection. Publication counts are only provided for the articles found in the journals indexed by that particular database. This is explored in Section 3.

Research output also differs across disciplines. Academic disciplines – and sub-disciplines – each have their own traditions and practices for publishing and disseminating research outputs. For example, in a discipline like physics, one might expect a researcher to produce many publications with many collaborators. In the humanities, a single author may produce a book within a much longer timeframe. Analysing publication counts across disciplines is not advised.

Analysis of publication counts among small sample sizes can lead to challenges with outliers – one, or a couple, of researchers with heavily-cited papers. It is difficult to normalize for these outlier effects. This is particularly true where research units are small and in low-citation culture disciplines (Abramo, D'Angelo, & Viel, 2013; Hicks et al., 2015; E. S. Vieira & Gomes, 2010). At the institutional level, or in disciplines that do not publish frequently, publication counts are susceptible to outliers.

Normalizing a measure such as a publication count provides a more appropriate metric for comparing research productivity across research areas, document type, and period. For example, Institution X has a total of 2,000 published journal articles indexed in Web of Science. Normalized publication counts may weigh an institution's publication rate against the expected performance in the specific field or discipline (field normalized publications). While normalized measures are powerful, normalized measures with small

sample sizes may be susceptible to outliers. As a result, percentiles may be a more suitable approach depending on the context in such cases.

4.2 CITATION COUNTS

Citation counts are the absolute number of times that a given article is cited. For example, Article X has been cited 11 times, by documents indexed by Scopus. As with publication counts, citation counts may also be normalized or refined to reflect expected performance within a specific field or discipline. For example, mean normalized citation score normalizes citation counts by subject category and publication year (Waltman & Van Eck, 2013). The mean normalized citation score is used in the Leiden Ranking as one measure of impact.

Normalizing a measure such as a publication count provides a more appropriate metric for comparing research productivity across research areas, document type, and time period.

Measures based on citations are among the most frequently used bibliometric indicators, and they are used in a myriad of ways (Mercier & Remy-Kostenbauer, 2013). As an example, citation count based research productivity metrics used by InCites include:

- Total citations - the absolute number of citations to a specific work or group of publications.
- Documents cited - the proportion of total publications cited one or more times.
- Citation impact - average (mean) number of citations per paper.
- Normalized citation impact - citation impact (citations per paper) normalized for subject, year and document type.
- Highly cited papers - papers that rank in the top 1% by citations for field and year (Clarivate, 2021).

Citation-tracking databases use different types of counting to report citations. Understanding how these databases report citations is important, as they typically offer differing snapshots of research impact. For example, applying whole counting to citations for all co-authors on a given paper would typically result in higher citation totals, whereas applying fractional counting to citations for co-authored papers based on a protocol would most often present a slightly different snapshot of a researcher's impact. As an example, consider a single paper with 100-plus contributors. University ranking results have been directly skewed by incidences of 100-plus contributors in situations where citation data was not normalized (Holmes, 2013).

4.2.1 LIMITATIONS

Publications might be cited for reasons other than direct influence. The fact that an article has been cited does not necessarily indicate the influence or impact of a particular researcher (Johnson, Cohen, & Grudzinskas, 2012).

Bibliometric measures that use citations are fundamentally limited by the scope or coverage of the citation-tracking tools on which they rely (see Section 3). While citations offer many opportunities as a bibliometric measure, questions exist about how citation counts contribute to bibliometric analysis. Fundamentally, citation counts are based on the assumption that the greater the number of citations a publication receives, the more influential it is. Publications, however, might be cited for reasons other than direct influence: to support arguments, as an example of a flawed methodology or weak analysis, or to indicate the amount of research conducted on an issue. The fact that an article has been cited does not necessarily indicate the influence or impact of a particular researcher (Johnson et al., 2012).

Citations may also include self-citations by authors and may be perceived as inflating an individual researcher's citation count. However, there are contexts where self-citations are warranted and even essential to a work or progress in a field of research, such as when a researcher has published a seminal work and not citing that work would be ill-advised (Carley et al., 2013).

There also may be occasions where a discredited paper may receive many citations before it is retracted, and it may continue to receive citations post-retraction. In a well-publicized case, a study on genetics predicting longevity was retracted before publication in *Science* due to technical errors in the scientific data (Ledford, 2011). However, as of Sept 2024 the retracted article has been cited 56 times with the latest citation from as recent as 2022. This illustrates the enduring nature of retractions. Fortunately, this is a case where the authors corrected their mistakes and republished a version that has withstood review.

Citation counts can also be manipulated. In one example, a university actively recruited highly cited researchers to become "distinguished adjunct professors" at its institution. In exchange for a fee, airfare, and hotel stays to visit the institution, the researchers were asked to update their Thomson Reuters' (now Clarivate) highly cited researcher listing to include their affiliation to the university and occasionally attach the university's name to papers that they publish. The result of this "citations for sale" approach is that the university's rankings in certain areas were raised to levels that are generally considered unjustified (Messerly, 2014).

Another limitation is that citations are time-sensitive. A researcher's impact is understood to change over time. More established researchers will naturally have higher citation counts than researchers early in their career, regardless of the quality of their research or findings. Citations accrue over time, thus the number of citations a publication receives will differ based on how long the publication has been published. To address this issue, citation counts should be normalized over time. However, some authors suggest citations within one to two years of publication cannot be counted accurately, even with field normalization efforts (Wang, 2013).

Moreover, each discipline is unique in how new research is disseminated and integrated into a field. The time required for research impact to be understood in the field varies by discipline. Chang notes that researchers in arts and humanities tend to cite older literature (Chang, 2013). In anatomy it can take fifty years or longer for a publication's findings to be analyzed. Taxonomy papers, the branch of science that classifies organisms, are regularly cited decades, even a 100 years, after publication (Cameron, 2005). A three-to-five year window from publication time is recommended as the ideal choice for citations within the natural sciences and engineering (*Informing Research Choices*, 2012). Others have suggested a three-year citation window is necessary for effective citation analysis (Bornmann, 2013; Wang, 2013). Using citations to understand research impact must reflect the citation culture of the discipline(s) being assessed.

Articles of questionable research quality can receive numerous citations before they are retracted.

Despite these limitations, citation analysis remains one of the most commonly used bibliometric measures, as well as a component of other measures, including the h-index and its related iterations. Using citation rates as a bibliometric measure can be problematic because of limitations of citation-tracking databases, differing rationale for citing publications, potential for reporting citations on retracted works, inflated citation records, opportunities for manipulation, and the time-sensitive nature of citations. Solutions are being proposed to address some of the complexity of citation analysis to achieve greater objectivity (Retzer & Jurasinski, 2009).

4.3 H-INDEX AND OTHER COMBINED MEASURES

A researcher's h-index is x if the researcher has x papers each of which has received at least x citations (Hirsch, 2005), see Table 2. The h-index is one of a series of measures that captures output using both total number of publications and number of citations. It is a productivity measure that is only useful for a focused snapshot of an individual's research performance but is not useful for comparison between researchers. Therefore, it is recommended to recalculate it regularly. However, more recently the scientific community has recommended avoiding the use of the h-index altogether, as it has significant limitations, as discussed below.

TABLE 2: HOW TO DETERMINE AN H-INDEX

Articles (ordered by citation count)	**Citation counts**
1	65
2	30
3	20
4	15
5	7
6	6 = h-index
7	5
8	1

A further discussion of the h-index and how to create a custom h-index using an individual's comprehensive publication and citation history can be found at http://nfgwin.uni-duesseldorf.de/sites/default/files/Ireland.pdf (Ireland et al., 2012).

Other measures have been developed which are modifications of the h-index, such as the g-index ("h-index for an averaged citations count"), the i10-index (number of publications with at least ten citations), the m-index ("the m-index is a correction of H-index for time"), and the Py-index ("the mean number of published items per year") (Von Bohlen und Halbach, 2011). Measures like the h-index, m-index, and Py-index suggest trends and provide a snapshot of performance over the career of the researcher (Bornmann, 2013).

Any measure that uses citations must consider that citation measures require time to accumulate and are time dependent.

Total publication and citation counts can also be combined to create new measures of research output. The Snowball Metrics (Colledge, n.d.) initiative uses citations per output (average citations received by each output in a particular dataset) as a measure.

4.3.1 LIMITATIONS

Although using measures like the h-index that combine publication and citation outputs helps to mitigate some of the limitations of using citation and publication counts alone; there has been growing and significant critique of the h-index, culminating in the general recommendation that it be avoided, particularly "for individuals competing for the same resource" (Hirsch, 2005, p. 16572).

Limitations:

- The h-index **masks the full story** of an individual publication and citation history.
- The h-index is **dependent on the data source**, therefore always reference the citation-database sources.
- The h-index **ignores the context** of a researcher, such as their career length, their discipline or the authorship contribution patterns (Bi, 2023).
- The h-index **increases with career length** regardless of new publications (Hicks et al., 2015).
- The h-index **does not account for outstanding or highly cited works**.
- The h-index **favours publications that are published more quickly**, such as reviews or letters, potentially discounting contributions to innovation in science produced through experimental, primary or fundamental science and research (Akhtar, 2024).

4.4 COLLABORATION

Type and extent of research collaborations in publications. How collaboration is measured depends on the discipline being examined. Measuring publishing collaborations may refer to collaborations with researchers from different geographic regions, in different sectors (industry, government, etc.), or among defined groups of individuals (service providers, participants). Collaborations in the industry sector may identify opportunities for funding and partnerships. Spinoffs of industry collaborations can result in real world experiences in the form of co-op placements as well as job prospects for recent graduates. International or industry collaborations are sometimes given significance based on the citation impact of the paper produced; however, collaborations are by their nature context-dependent (Moed, 2005). As the proportion of international collaborations captures an institution's international research collaborations, and the proportion of industry collaborations captures an institution's research collaborations with industry, this information can provide useful data to inform strategic directions.

Ultimately, measures of collaboration only have meaning if they are considered within the discipline and relevant period in which the data were collected.

Systems like SciVal and InCites use both international collaborations (number of papers having one or more international co-authors, and the proportion of publications having international co-authors) and industry collaborations (proportion of publications having co-authors from industry). In these systems, a range of both impact and collaboration measures are presented for institutions, groups or individual researchers, including collaboration with international co-authors and collaboration with industry partners.

4.4.1 LIMITATIONS

Collaboration measures are fundamentally contextual in nature. They have limited application because their meaning is specific to a field of study and the measure itself is imprecise. For example, research collaborations with industry may be more common within certain disciplines, and international collaborations may be more common within specific nations or geographic regions. Industry collaborations may be governed by non-disclosure agreements regarding intellectual property policy, and so the research results of these collaborations may never be published in peer-reviewed venues. As well, a high proportion of collaborations with industry or international peers could reflect a few collaborations among many researchers, or a high proportion of collaborations could represent numerous collaborations

with a single researcher, making the measure ill-defined and requiring further investigation as to the nature of the relationships.

Researchers generally believe that the quality of an individual publication should be judged on its own merit. Individual article-based citation counts, rather than journal-based citation counts, are the preferred metric.

It is also true that collaboration practices change over time and, to a certain extent, are influenced by an institution's local culture and traditions. The internet age has directly influenced a researcher's ability to more easily explore collaborative opportunities. This could possibly complicate a retrospective analysis of collaborative measures data if the context for collaborative practices at different points of time, within the context of a specific discipline, is unknown. Ultimately, measures of collaboration only have meaning if they are considered within the discipline and relevant time-period in which the data were collected.

4.5 JOURNAL LEVEL IMPACT RANKINGS

The relative importance of the journal, not individual articles in the journal. This measure uses aggregate data of citations from articles published in the journal to determine the journal's impact ranking (Falagas & Alexiou, 2008; Krauskopf, 2013). The Journal Impact Factor (JIF) by Clarivate is a widely known example of this measure. There are similar indicators from Elsevier such as the SCImago Journal Rank and Source-Normalized Impact Paper (SNIP). The common rationale for continued use of these rankings is that an individual researcher who publishes their research in journals with a high impact ranking produce work of higher quality.

4.5.1 Limitations

Researchers generally believe that the quality of an individual publication should be judged on its own merit. Individual article-based citation counts, rather than journal-based citation counts, are the preferred metric. This critique is the hallmark of the Declaration on Research Assessment (DORA, 2012), which has gained international recognition and contributed to substantial reform in assessment processes such as the UK Research Excellence Framework. Harnad (2008) observes, "comparing authors in terms of their JIFs (journal impact factor) is like comparing university student applicants in terms of the average marks of the secondary schools from which they graduated, instead of comparing them in terms of their own individual marks" (p. 104). The key to understanding this metric is that the JIF is a journal-level metric, not an article-level measure.

The JIF is also problematic when considering the differences in citation culture between disciplines. Dorta-Gonzalez and Dorta-Gonzalez (2013) compared science and social science using the JIF. They noted that while there are, on average, 30% more references in social science publications than in science publications, 40% of the sources in social science are not indexed by Clarivate's Journal Citation Reports (JCR) and therefore do not have a journal impact factor, compared to only 20% of science references that are not indexed in the same. As a result, the aggregate impact factor in science is 58% higher than in social science (Dorta-González & Dorta-González, 2013, p.667-8).

Journal Impact Factor (JIF) calculations provide a quantitative value of the impact of a journal solely through citation count. They are susceptible to many confounding variables such as article type, self-citations, language of articles, research fields, which do not reflect researcher productivity (Ali, 2022; Priem et al., 2010). Finally, the citation distribution is skewed (Ali, 2022; Priem & Hemminger, 2010), as most of the papers in a journal are cited infrequently compared to a few highly-cited outlier authors. This

means the JIF is not a true representation of the journal's average citation rate (Glänzel & Moed, 2002; Kiesslich et al., 2021; Seglen, 1997).

This evidence does not mean that journal impact factor measures do not have a place in bibliometrics. What it does mean is that bibliometric measures need to be validated to ensure that metrics are being used for their intended purpose and capable of measuring what they were intended to measure. The fundamental point should be to encourage researchers to publish in venues where their publications are likely to have the highest impact on the field, as opposed to publications that only provide high impact factor scores.

4.6 TOP PERCENTILES

Bibliometric measures, when used appropriately and with a good understanding of the underlying limitations of citation-tracking databases used to collect them, may offer meaningful insight into research outputs.

The top percentile (for example 1% or 10%) is typically a measure of the most cited documents or citations in a subject area, document type, and year. For example, the top 10% most cited works in a specific discipline or among an institution's publication output. The Snowball Metrics Recipe Book includes several metrics that use impact factors as part of bibliometric measures. The Publications in Top Journal Percentiles measure establishes citation limits for the top 1%, 5%, 10%, and 25% of journals being used in each publication year, and it measures the absolute count, and the proportion of total counts, of outputs that fall within each of the identified limits.

4.6.1 Limitations

Top percentiles measures require medium to large datasets to be reliable.

4.7 SUMMARY

Bibliometric measures, when used appropriately and with a good understanding of the underlying limitations of citation-tracking databases used to collect them, may offer meaningful insight into research outputs. For example, when working with a large sample of works predominantly published in English which are well-represented in citation-tracking databases, it may be possible to conduct a bibliometric analysis via journal-level subject categories. In situations like this, close and careful collaboration with the researchers who produced the work is critical, and such collaboration can also support identifying peer institutions with which the final data can be compared against.

Researchers should be encouraged to publish in venues where their publications are likely to have the highest impact on the field, as opposed to publications that only provide high impact factor scores.

Table 3 provides a summative view of appropriate uses for bibliometric measures commonly used to assess research outcomes. Figure 3 illustrates the summary of recommended uses of bibliometrics by level of analysis adapted from Abbott et al. (2010), Gagolewski (2013) and (2012). These summative tools can guide researchers, administrators, and others in making decisions about which measures to use for research assessment activities, as well as in interpreting the relevance and accuracy of bibliometric measures already in use.

TABLE 3: BIBLIOMETRIC MEASURES, DEFINITIONS, USES AND LEVELS OF USE

<table>
<tr><td colspan="3">Publication Count: The total count of items identified as scholarly output. Journal articles, books, conference proceedings, etc.
Example: I have published 11 articles in my research field over the past 5 years.</td></tr>
<tr><th>USEFUL FOR</th><th>NOT USEFUL FOR</th><th>POSSIBLE LEVEL OF USE</th></tr>
<tr><td>▪ Assessing outputs of an individual, discipline, or institution</td><td>▪ Assessing quality of a work</td><td>▪ Individual
▪ Subject / Discipline
▪ Group / Institutional

▪ E.g., funding applications, CVs, strategy, collaborations, reporting</td></tr>
<tr><td colspan="3">Citation count: The total number of citations received to date by the publications.
Example: This research group has been cited 94 times over the past 5 years.</td></tr>
<tr><th>USEFUL FOR</th><th>NOT USEFUL FOR</th><th>POSSIBLE LEVEL OF USE</th></tr>
<tr><td>▪ Use of, or one element of impact or influence of a work or set of works
▪ Attention by other researchers</td><td>▪ Understanding context of the Impact (positive vs negative impacts)</td><td>▪ Individual
▪ Subject / Discipline
▪ Group / Institutional

▪ E.g., funding applications, CVs, strategy, collaborations, reporting</td></tr>
<tr><td colspan="3">Field-Normalized Citation Impact: Calculated using an entity's actual citations per paper compared to the expected citations per paper by similar papers within the same subject area, year, and publication type, globally. A value of 1.00 indicates that the work performs at the expected global average. A value >1.00 indicates that the publication exceeds the world average. This allows for normalization across entity sizes, research fields, and publication dates.

Examples metrics include:
• Field-weighted citation impact (FWCI) from Snowball Metrics is used by SciVal and OpenAlex
• Category-normalized citation impact (CNCI) from InCites
• Field Citation Ratio (FCR) from Dimensions

Example: Our institution's scholarly works received 95% more citations than expected with a FWCI of 1.95.</td></tr>
<tr><th>USEFUL FOR</th><th>NOT USEFUL FOR</th><th>POSSIBLE LEVEL OF USE</th></tr>
<tr><td>▪ Comparing between different subjects and sample.</td><td>▪ Small sets of publications, as a single highly cited paper can easily skew the calculation through inflation.</td><td>▪ Individual
▪ Subject / Discipline
▪ Group / Institutional

▪ E.g., funding applications, CVs, strategy, collaborations, reporting</td></tr>
</table>

H-Index: A researcher's h-index is x if the researcher has x papers each of which has received at least x citations (Hirsch, 2005).
Example: With an h-index of 22, I have a strong research record in comparison to researchers in my field and with a similar career length, whose average is about 15.

USEFUL FOR	NOT USEFUL FOR	POSSIBLE LEVEL OF USE
▪ Comparing researchers of similar career length ▪ Comparing researchers in similar field / subject / department and who publish in the same journal categories ▪ Focused snapshot of a researcher's performance	▪ Comparing ranked lists of researchers for merit or promotion ▪ Comparing researchers from different fields / disciplines /subjects ▪ Assessing fields / departments /subjects where research output is typically books or conference proceedings as they are not well represented by databases providing h-indices	▪ Individual ***Caution should be used when comparing researchers.*** See notes on when the h-index might be useful or not.

Proportion of International Collaborations: Proportion of publications having at least two different countries among the co-authors' affiliations.
Proportion of Industry Collaborations: Proportions of publications having the organisation type "corporate" for one or more co-author affiliation
Example: Our research institution's has high international collaboration, with 65% of our research outputs having international co-authors, in comparison to the institutional average of 52%.

USEFUL FOR	NOT USEFUL FOR	POSSIBLE LEVEL OF USE
▪ Proportion of work that is co-authored with international colleagues ▪ Compare groups of different size but not different disciplines ▪ Discovering research with industry that may be affected by non-disclosure agreements (NDAs), and hence not subject to publishing.	▪ Comparing different disciplines ▪ Comparing institutions in vastly different regions of the world. ▪ Small datasets, i.e. Small research groups, etc.	▪ Individual ▪ Subject / Discipline ▪ Group / Institution ▪ E.g., funding applications, CVs, strategy, institutional accountability processes/reporting.

Journal Impact Rankings: Measures of this type use aggregate citation data of articles published in a journal to capture the journal's relative citation performance over a standard period of time (often 3-5 years).
Examples of Rankings:

- Journal Impact Factor (JIF) by Clarivate
- SCImago Journal Rank (SJR) by Elsevier – Field-normalized
- Source-Normalized Impact Paper (SNIP) by Elsevier – Field-normalized
- CiteScore by Elsevier

USEFUL FOR	NOT USEFUL FOR	POSSIBLE LEVEL OF USE
▪ Identifying journal for manuscript submissions ▪ Identifying relative citation performance of journals within the database ▪ Comparing journals within the same discipline	▪ Comparing individuals ▪ Comparing disciplines if not field-normalized ▪ Open access journals as many are not captured in citation-databases ▪ Determining quality or importance of individual journal articles	▪ Journals

Top Citation Percentiles: Top percentile (for example, 1% or 10%) is typically a count of documents that are in the top percentage of cited documents in a subject area, document type, and year.
Example: The university has nearly 5000 works, or 17% of its total, in the top 10% most cited publications globally.

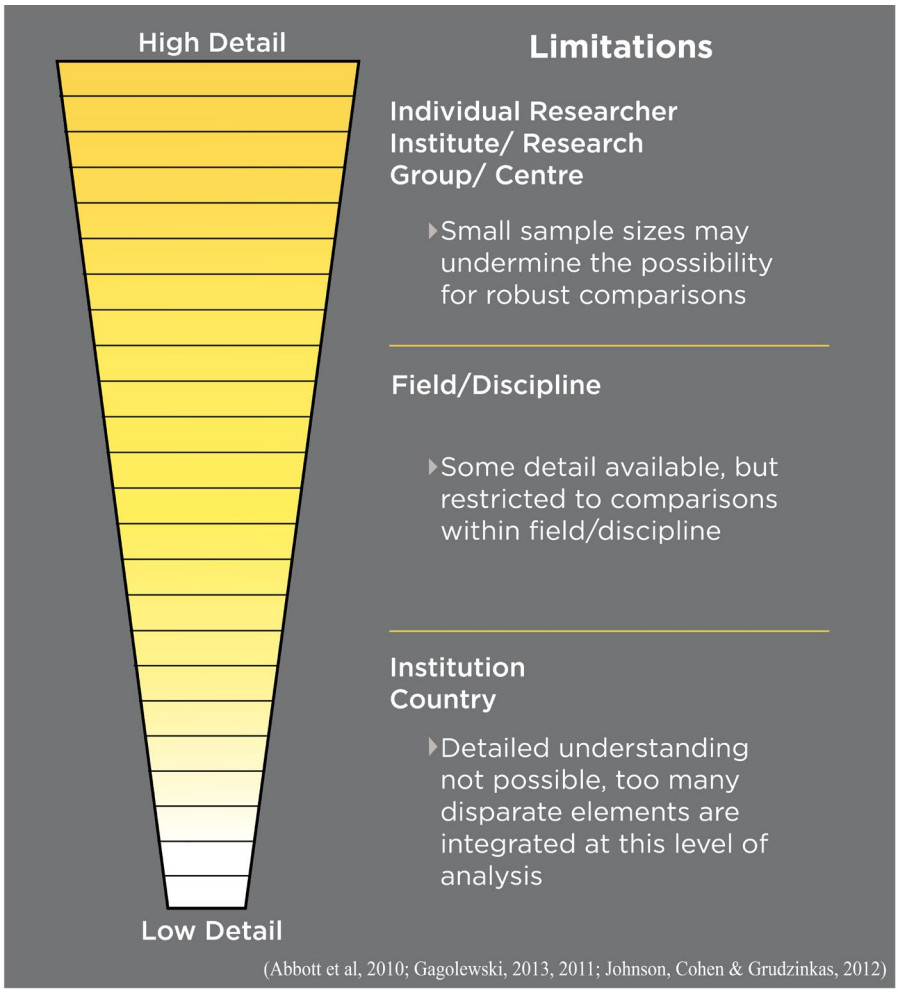
USEFUL FOR	NOT USEFUL FOR	POSSIBLE LEVEL OF USE
▪ Medium to large datasets ▪ Compare groups of different size (using %) but not different disciplines ▪	▪ Small datasets, unless accompanied by additional information, or gaps identified ▪ Comparing disciplines ▪ Early-career researchers	▪ Individual ▪ Subject / Discipline ▪ Group / Institutional ▪ E.g., funding application, peer review,

Highly Cited Researchers: A controversial list of highly cited Sciences and Social Sciences researchers created by Clarivate. Highlights researchers whose work represents the top 1% of researchers in a field for citations.

USEFUL FOR	NOT USEFUL FOR	POSSIBLE LEVEL OF USE
▪ Understanding an individual researcher's impact as it relates other papers in the subject matter in which they have published. ▪ Only relevant for researchers publishing in the Sciences and Social Sciences.	▪ Comparing researchers from different fields / subjects / departments. ▪ Not relevant for researchers outside the Sciences or Social Sciences. ▪ Early-career researchers	▪ Individual ▪ Institutional ▪ E.g., Celebrating

Altmetrics: Methods of evaluating and discovering scholarly work that focuses on the use of open data, social media sources, news media sources, policy and patent citation. Altmetrics diverge from traditional bibliomentric measures reliant on scholarly publication citations.		
USEFUL FOR	**NOT USEFUL FOR**	**POSSIBLE LEVEL OF USE**
▪ Online attention ▪ Social or societal interest by public, industry or governments.	▪ Contextualize or add perspective to traditional bibliometric measure ▪ Analysis of works published before ~2011. ▪ Comparing researchers from different fields / subjects / departments.	▪ Individual ▪ Subject/Discipline ▪ Group / Institutional ▪ E.g, peer review, funding applications, CVs, strategy, institutional accountability planning, research collaborations. ▪ Use to

FIGURE 3: RECOMMENDED USES OF BIBLIOMETRIC DATA BY ANALYSIS LEVEL



5 EQUITY AND GENDER BIAS IN BIBLIOMETRICS

Forms of discrimination found in other areas of society are also present within the scientific community, such as stereotypes about gender and race, anti-Black institutional policies, structural racism, and undermining of indigenous knowledge (Eaton et al., 2020; Malapane et al., 2022; Mustaffa, 2017; Odekunle, 2020). In 2020, Waterloo made a public commitment to address systemic racism across its campuses and formed the President's Anti-Racism Taskforce (PART). Seven of the 88 recommendations are relevant to the research enterprise and demonstrate an effort to stitch anti-racism and equity principles into the fabric of University life (PART Report UWaterloo, 2021). These recommendations (see Table 4) have the distinct potential to influence more equitable participation in research activities that will benefit researchers and be reflected in research impact data. They touch on creating guidelines, training, fair representation and membership, consultation, and recognition through awards.

TABLE 4: UNIVERSITY OF WATERLOO PART RECOMMENDATIONS RELEVANT TO THE RESEARCH ENTERPRISE.

- Develop guidelines for ethical data governance respecting principles for administrative data. (#10)
- Train faculty and departmental review committees on how to evaluate non-Western based (particularly, Indigenous) research, teaching, and service and should evaluate bodies of work with appropriate lenses. (#21)
- Ensure representation of different ways of knowing and being, equity, diversity, inclusion and anti-racism training, and best practices for the Research Ethics Board membership, so it has the capacity to support, assess, and approve ethics requests for research which impacts or involves racialized communities, with particular emphasis on Indigenous communities. (#25)
- Ensure diverse/broad membership and equity, diversity, inclusion and anti-racism training for dissertation and thesis committees to better evaluate non-Western research methodologies and projects (formally recognize the contributions of faculty, particularly Black, Indigenous, and other racialized faculty, who may be tasked with more expectations on their labour for serving on these committees. (#26)
- Undertake a consultation process to learn about and address research barriers and particular needs of Black, Indigenous, and other racialized researchers, including a review of recent research in this field.(#63)
- Create research awards and additional funding opportunities for Black and Indigenous faculty researchers (e.g., support for open-access publication fees; recruitment of Highly Qualified Personnel (HQP); mentorship or networking events; fellowships, research awards and other opportunities for research advancing culturally diverse ways of learning and knowing; and community-based research). (#68)
- Provide training for staff who support the administration of research, especially for projects with Black, Indigenous, and other racialized communities. (#69)

5.1 GENERAL BENEFITS OF MORE DIVERSITY IN RESEARCH

While striving for equity, diversity, inclusion and accessibility (EDIA) within academia is a worthy goal in and of itself, it is worth noting that increased diversity also produces concrete benefits for research. This includes increased productivity in industry (Herring, 2009) and in productivity more generally, as diverse groups have been found to outperform homogeneous ones in cognitive tasks (Freeman & Huang, 2015). In

the science fields, diverse scientific teams have been linked to higher citations and tied to gains in innovation (Lerback et al., 2020). While, unequal representation in science has been found to lead to under-investigation of particular topics (those often with racial dimensions) and may serve to suppress innovation (Kozlowski et al., 2022). In contrast, it has been found that racialized and gendered groups are more likely to contribute novel scientific contributions, although their work is often neglected by other scientists (Hofstra et al., 2020). Neglecting difference has real implications. As such, the NSERC Guide on Integrating EDI Considerations in Research and the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans: Research Involving the First Nations, Inuit, and Métis Peoples of Canada discuss numerous examples of harm that are directly correlated to a lack of diversity in research and research design. These include but are not limited to the misappropriation of sacred activities or artifacts, or the devaluation of Indigenous knowledge as primitive or superstitious.

5.2 BIBLIOMETRIC METHODS FOR DETERMINING EQUITY DIMENSIONS

In bibliometrics studies looking at inequity within academia a narrowly defined set of research articles and their author level metadata are most often used. Specifically, authors' given and family names. These studies most often only consider one of only two elements: racial identity or gender identity. Race and gender are commonly inferred based on the author's name, as self-identification is extremely time consuming and comes with its own set of challenges. In most cases, gender and race are predicted using an algorithm, examples include Caron et al. (2022), and Larivière et al. (2013), with the Gender-API.com becoming a common resource (Kwon, 2022; Wu, 2024). The algorithms used in these studies rely on census data and country-specific lists of men and women's names to best match an author's gender and/or race. Most often binary categories (man / woman; white / non-white) are used. In some cases, researchers assign a probability distribution for the gender categories or over a range of racial categories. This method of identity assignment can be seen as problematic; as at best these are probabilistic representations of perceived identities and at worst mis-assign demographic identities to individuals potentially skewing data and results. Fortunately, bibliometrics studies that use these types of methodologies often acknowledge these limitations and strive to make the data variability as transparent as possible. Currently, there are no strong alternatives to these methodologies for large-scale studies (Akbaritabar et al., 2024; Kozlowski et al., 2022; Zhao et al., 2023). However, more accurately identifying demographic information at the institutional level is discussed below.

5.3 MEASURES OF EQUITY AT WATERLOO

5.3.1 Data Privacy and Sovereignty Considerations

An important consideration for staff, students, and faculty is informed consent when providing self-identified data on socio-demographic information such as gender and race. Waterloo's equity survey in 2021 used principles of informed consent which included de-identifying data, role-based access to record-level data and pseudo-IDs during analysis as data privacy and security processes to ensure ethical data governance. Processes like this that ensure anonymity can help us understand the gaps and barriers to representation in research for equity deserving groups without impinging on their privacy. Ideally institutions should also offer guidance on how to provide these types of data to third party vendors at the manuscript submission and reporting stage, and the necessary privacy protections for that sharing process.

Although, like many universities, the University of Waterloo does not provide record-level (i.e. publication-level) institutional data to third parties explicitly for analysis purposes, the university does use platforms that may allow for this data collection. Research profiles may contain information from which inferences regarding gender and race (e.g., pronouns or country of birth or citizenship) could be made, and are subject to the same issues of mis-assignment as discussed above. Moreover, researcher Curriculum Vitae (CVs) and profiles are public and may be reviewed for identifying features which could be gathered by third party companies.

5.4 BIBLIOMETRICS AND EQUITY IMPACTS

There is no consistent citation or publication advantage linked to gender or race that is applicable across all research fields. However, gender and racial gaps in publication and citation still exist. There are disciplines and areas of research where these gaps are larger than others and merit additional equity considerations. This section specifically discusses recent literature on gender disparities related to productivity, citations rates, and career advancement, and citation of racialized groups, however, many of the equity considerations are intersectional, meaning that disparities seen in one group, such as women, are often tied to or enhanced when intersectional identities are considered. Therefore, although generalization should be cautioned it is not a broad leap to understand that there are considerable inequities experienced by groups who tend to experience inequities in other aspects of their lives. The literature in bibliometrics has largely focused on gender, therefore that where this report focuses.

5.4.1 Gender and Productivity

A global analysis of articles' authors gender by country from 2008-2012 Larivière et al. (2013) found that almost all countries have a majority of articles authored by men and presumes that the difference is due to gender differences in employment more than differences in career outputs. That is, men are more likely to experience fewer gaps or pauses in employment as science-producing researchers than women. These gaps or pauses are often due to parental leaves, elder care or temporary duties taking them away from their core research agenda. Countries with the most research output also had a citation advantage for men (Larivière et al., 2013). Women are less likely to be published in high-impact journals, and they are less likely to have their research cited by others (Burstin & Arora, 2021; Chatterjee & Werner, 2021).

5.4.2 Gender Disparity and Citation Rates

There has been largely three different approaches to studying gender disparity and citation rates; the article-level approach, the author-level approach, and the reference ratio approach (Wu, 2024). This matters because much of the constant debate in the literature around gender disparities in citation rates have stemmed from how the data is analysed. Therefore, Wu (2024) summarizes the different approaches and the preponderance of evidence, such that:

- The article-level approach focuses on analysing articles by separating them as either authored by men or women. Overall, it is found that at a per-article basis men and women, with some marginal differences in some studies, are cited at the same rates.
- The author-level approach focuses on authors and their career stage or total citations over a timespan. Overall, it is found that at a per-author basis men receive more citations than women, and that the gap tends to grow over their career progression.
- The reference-ratio approach takes the expected ratio of men and women based on known productivity rates and compares that to the ratio of men and women in the reference lists of articles.

Therefore, if there are 10:10 ratio of men and women authoring articles than the assumption is that you would expect that same 10:10 ratio in the reference lists. Overall, it is found an underrepresentation of women in reference lists. Additionally, although men tend to cite men more this homophily pattern is also found among women, with women tending to cite women more often.

What might be happening to explain these seemingly contradictory findings? Wu (2024) suggests that since the gender disparities are largely concentrated the author-level, meaning women tend to accumulate less citations over their career than do men, that the disparity may be caused by the lower overall productivity of women, rather that direct biases in citation behaviours. Instead, focus should be on the systemic barriers that women may face in academia like gendered role expectations within and outside of academia, such as, service responsibilities and caretaker roles (Wu, 2023, 2024).

Women tend to be less cited than men and even less so in field where there is more representation by women, contrary to assumption. (Chatterjee & Werner, 2021; Kozlowski et al., 2022; Wu, 2024). Globally, the number of lead women authors has increased significantly (Aguilar et al., 2019). While the scientific publishing rates of women have improved, the rate at which women authors have been cited have not increased. As the proportion of women per article increases, the number of citations tend to decrease (Kwiek & Roszka, 2021). Mayer and Rathman claim that "men cite men (at a rate double to women)" creating "cohesive networks of male authorship and citation" (2018, p.11). Men also self-cite more often than women and produce a greater proportion of the few extremely highly-cited articles (Baltussen & Kindler, 2004; Graham et al., 2019; Heng Wong et al., 2019). These citation patterns impact not only women but other marginalized researchers (Kozlowski et al., 2022; Wu, 2024).

5.4.3 Gender and Career Advancement

Bibliometrics has a multiplicative impact on career advancement. The "publish or perish" culture in contemporary academia means that citation and publication rates impact the ability of scholars to advance in their careers. At the same time, bibliometric measures are used, at least in part, to determine career advancement. As of 2021, women scientists are still reported to occupy more junior positions and to have lower salaries, be more often in non-tenure-track and teaching-only positions, to receive less grant money, be promoted more slowly, and are less likely to be listed as either the first or last author on a paper, as is highlighted by the past decade of research at all levels: from institutional, national and global.

Gender disparities in bibliometrics can have significant repercussions on the careers of women researchers, and these obstacles are only compounded for those whose identities are marginalized in intersectional ways. White et al. (2021), for example, found significant effects of both gender and race on productivity and impact for scholars in clinical psychology. They found that the impacts of race were present despite career stage, while the impact of gender increased the further along researchers are in in their careers. For example, men and white faculty in associate and full professor ranks had more publications per year and higher h-index than women and non-white faculty in the same field. Among associate professors, men were more likely to get tenure earlier, even, for example, when measures of scientific impact such as h-impact were the same (White et al., 2021).

Males tend to have more career citations, at least partly because they are less likely to work part-time or take career gaps for career responsibilities (Thelwall, 2020). A 2018 study indicated that measuring academic productivity with metrics outside of simply publishing journal articles led to a better understanding of productivity rates for women researchers. The proposed metrics controlled for: leaves of absence, including a wider range of publication types outside of journal articles, fractionalizing for co-

authorship, and de-emphasizing journal rankings (Nygaard & Bahgat, 2018). Therefore, productivity gaps between genders should be viewed skeptically, as these gaps may be highly influenced by how productivity is being defined, measured, and controlled for.

5.4.4 Citation of Individuals Who are Members of Racialized and Indigenous Groups

PART recommendations #21, 25 and 26 call for the University to acknowledge and respect non-Western ways of knowing. In so doing, it is important to recognize that bibliometrics as a tool and discipline captures primarily traditional, Western-centered focused knowledge. Kubota (2020) points out that traditional academia is biased by centering English publications, even in "international" journals, whereas the work of non-Anglophone scholars are rejected for not being "international enough," or that their work needs to be framed "in a more global context" (Kubota, 2020). More subtlety, "white Euro-American values, beliefs, and worldviews, including individualism, meritocracy, progress, and so on" (Kubota, 2020) are seen as universal, while any scholarly knowledge existing outside of this narrow lens is othered as "elective" or "alternative". These subtle biases play out in practice, when discrepancies arise in their attainment of academic recognition. For example, a recent study found that papers authored by Black and Hispanic scholars in the United States are cited less frequently than those authored by white researchers, even when the content is similar (F. Liu et al., 2023). These external pressures often force scholars of colour to conform to Euro-American ways of performing research, in order to avoid having their work taken less seriously.

Minoritized groups are over-represented in lowly cited topics and under-represented in highly cited topics and their work is less cited within and across topics, especially where they are under-represented (Kozlowski et al., 2022). Chakravartty et al. (2018) found that in general, non-White scholars are under-represented among the published first authors in communication journals, authoring only 746 out of 5,262 (14%) documents published from 1990 to 2016. Non-white first authors are cited significantly fewer times than their White counterparts. Articles authored by non-white scholars are cited significantly fewer times, on average, than white authors. Possible reasons behind this may be a lack of inclusivity and equity and related structural/cultural support that benefits team science. In addition, acceptance rates may be related to bias in peer-review, and citations may also reflect network effects (Lerback et al., 2020).

Williams (2022), examined citation practices of black women in anthropology, and argued that although Black women have contributed much to the field, they are credited for their work on methodological innovation or ethnographic detail, but much less so for theoretical insights. When scholars intend to demonstrate their expertise, they will reference the most historically cited works, which are almost always what exists within the white, male-dominated mainstream. Marginalization of Black women's work means that their work is never considered seminal, and it will never be cited as essential theoretical frameworks (Williams, 2022).

Bibliometrics favours researchers in the later stages of their careers, so if a field was historically dominated by white scholars, the metrics will serve this demographic better. To further complicate the issue, researchers with fewer citations are less likely to receive opportunities and resources from their organisations, meaning that the impacts of citational erasure are often self-reinforcing.

5.4.5 Citing Indigenous Knowledge (IK) within the academy

Indigenous Knowledge (IK) can be defined as consisting of "traditional local, non-scientific beliefs, customs, and methods that are normally described as informal forms of knowledge" (Malapane et al., 2022)

and passed down from generation to generation, often through oral histories. IK is largely supressed and marginalized within Western academia, and its methods interpreted as lacking scientific value. The legacy of colonialism has led to the disappearance of IK all around the globe. Studies have found that there is a lack of a standardized form of citing the contributions of IK Keepers or local experts within social science research (Malapane et al., 2022). Other studies have found that IK is overlooked despite its expertise on a topic, in favour of non-Indigenous experts, leading to IK being underprivileged "due to political, economic, and administrative reasons" (Malapane et al., 2022). Citation of IK has been seen as not as legitimate as white-male scholar, used as a form of tokenization and difficult to apply due to traditional standards of Western academic publications.

In Canada, there is now the TriCouncil guidance on research with Indigenous people: Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – TCPS 2 (2022) – Chapter 9: Research Involving the First Nations, Inuit, and Métis Peoples of Canada.

Contemporary Citation Practices

The academy has traditionally neglected to see the scientific value of Indigenous knowledge, as it has often existed outside of Western epistemology practices. While there are standard APA and MLA formats for acknowledging "personal communications" as an official reference, this would place a conversation with an Indigenous Knowledge Keeper on the same level as a phone call. Instead, librarian Lorisia MacLeod has created templates to use for receiving oral history from an Indigenous Knowledge Keeper, see Figure 4.

FIGURE 4: CITATION FORMATS FOR ORAL HISTORIES IN APA AND MLA.

APA

Last name, First initial. Nation/Community. Treaty Territory if applicable. Where they live if applicable. Topic/subject of communication if applicable. personal communication. Month Date, Year.

Example: Cardinal, D. Goodfish Lake Cree Nation. Treaty 6. Lives in Edmonton. Oral teaching. personal communication. April 4, 2004.

MLA

Last name, First name. Nation/Community. Treaty Territory if applicable. City/Community they live in if applicable. Topic/subject of communication if applicable. Date Month Year.

Example: Cardinal, Delores. Goodfish Lake Cree Nation. Treaty 6. Lives in Edmonton. Oral teaching. 4 April 2004.

MacLeod included the individual Indigenous Nation of the Knowledge Keeper in the citation in order to break down pan-Indigenous stereotypes and to acknowledge that there is variation within Indigenous communities. Another aspect to consider when citing Knowledge Keepers is presenting their names in non-Latin alphabets, as a way of respecting traditional languages.

Reconsidering how historical works are cited

Scholar Indu Ohri's advocacy brings awareness to the fact that many historical volumes which are currently only accredited to Western authors would not have existed without the contribution of Indigenous individuals who sourced, transcribed, and translated oral histories for British colonizers. Ohri advocates giving these Indigenous collaborators, who currently are not acknowledged in these works, co-authorship status. Historians typically refer to Indigenous contributors in these works as "native assistants," "research assistants," or othering them as "Indigenous experts," which do not appropriately express how integral their efforts were into producing these colonial-era publications. Ohri's reframing can encourage institutions to reconsider how historical works are currently being credited. In many instances, the work may have not existed without the efforts of Indigenous individuals or communities, and it may be appropriate to update their contributions to that of coauthors.

5.5 RECOMMENDATIONS FOR EQUITY AND GENDER BIAS IN BIBLIOMETRICS

Being aware of and developing approaches and strategies to mitigate and address equity issues in how research productivity is understood and captured is important. In this section we offer practical researcher-level and organisational-level practices that can help to offer some recourse to equity issues in bibliometrics.

At the individual researcher level, researchers can ask themselves the following questions:

- Whose words and ideas should I cite and draw on?
- Who do I cite in my writing? Why? What are the consequences?
- Do the cited scholars reflect the dominant voices historically privileged in this field or institution? How have scholars from underrepresented communities contributed to its development? Who else has shaped this field in ways that may not be widely recognized?
- If I value anti-racist and anti-sexist causes but have to compromise those values in order to be published, is this the right venue for my work?

Additionally, Kwon (2022), suggests several ways of diversifying citations including:

- Include citation diversity statements in manuscripts [16]
- Expand citable sources, for example:
    - Cite oral teachings from Indigenous communities as legitimate expertise not simply personal communications [17]. See Figure 4 for guidance.
- Seek out advocacy groups curating lists of underrepresented authors, for example [18]:
    - 500 Women Scientists
    - Cite Black Women collective

[16] See Dworkin, J., Zurn, P. & Bassett, D. S. 2020 as cited in (Kwon, 2022; Zurn et al., 2020)
[17] See MacLeod, L. 2021 as cited in (Kwon, 2022).
[18] As cited in (Kwon, 2022).

At the organisational level, it is recommended that:

- Establish inclusive research impact evaluations that include equity-aware metrics that account for:
    - Career gaps
    - Diversity in publications types
    - Known disparities in citation practices at the discipline level
- The university communicate the risks and benefits of submitting personal data to third party platforms.
- Use algorithmic prediction of gender or race for bibliometric study with full transparency about their limitations and provide probability and uncertainty measures
- Enhance and expand our institutional understanding of "other ways of knowing" through training, representation and memberships that influence our research enterprise.
- Allow more tailored approaches to bibliometrics and research impact assessments when necessary in order to create a structural and culture shift away from merely counting total research outputs to one that strives to acknowledge and celebrate differences from discipline to discipline.

6 MEASURING RESEARCH IMPACT IN THE HUMANITIES, ARTS, AND SOCIAL SCIENCES

Humanities, Arts and Social Science (HASS) disciplines seem to be the most heavily impacted by an increased concentration on the narrow set of scholarly outputs considered in bibliometric analysis. Despite this, in HASS disciplines, bibliometrics are used in a similar way to other disciplines, and increasingly used to measure research impact for grant-funding models activities (Anania & Caruso, 2013; Bonaccorsi et al., 2017; Tripathi et al., 2018). At the same time the HASS disciplines are also becoming increasingly interdisciplinary and multidisciplinary (Ochsner et al., 2016) allowing HASS to see some of the benefits from this heightened concentration on bibliometrics. While HASS scholars argue that peer-review alone may not be enough to assess research impact and quality, and that bibliometrics can offer a means to overcome subjective biases; their use has also been challenged based on concerns that intrinsic benefits of the arts and humanities will be overlooked if peer-reviewers value quantitative measures over subjective qualitative evidence (Broz & Stöckelová, 2018; Butler & Henadeera, 2007; Finkenstaedt, 1990; Ochsner et al., 2017; Tripathi et al., 2018).

Limitations of Bibliometrics in the HASS disciplines include:

1) Quantity is an ineffective marker of quality
2) Citation practices are discipline specific
3) Regionally focused research doesn't have less value
4) Trajectory of knowledge advancement builds over time
5) Publication age and citation rates are long and persistent
6) Single authorship results in less attention

6.1 LIMITATIONS OF BIBLIOMETRICS AFFECTING HASS

The use of bibliometrics in the HASS has been contested due to the limitations of the tools and approaches when evaluating HASS research outputs (Ochsner, 2021). HASS scholars argue that, while bibliometrics may be useful when applied to STEM fields, issues arise because of fundamental differences in HASS research and scholarly communication. Although these limitations have been discussed in other parts of this paper, it is important to recognize how they are particularly applicable to the HASS disciplines.

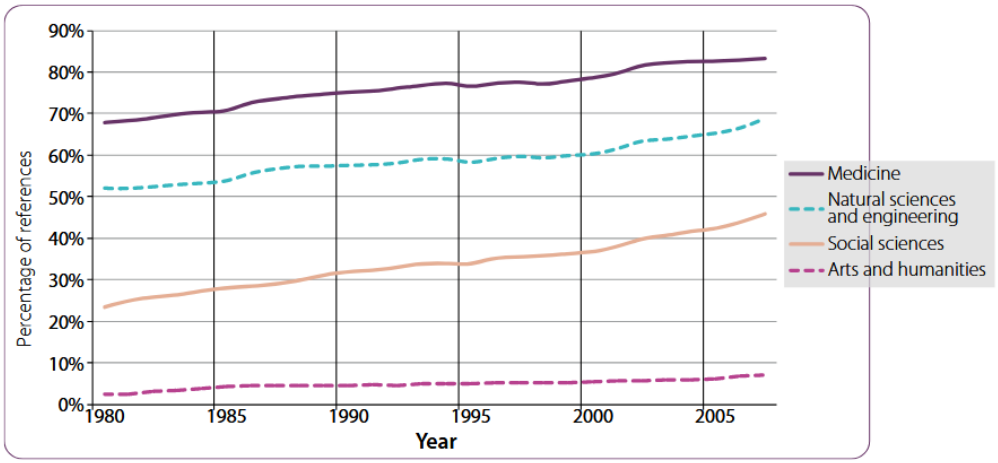
These limitations of bibliometrics in the HASS disciplines include:

- **Quantity is ineffective marker of quality:** outputs and impact of HASS research is often not quantifiable as they often include creative works such as exhibitions, performances, digital media, sound and images, as well as policies, data sets, media coverage, new processes, and more that are not well captured in bibliometric tools. Although there are advances in capturing some of these outputs in some tools, there are still significant gaps. Quantity should not be used as a direct indicator of quality whether counting publications or the citations to those publications (Broz & Stöckelová, 2018; Ochsner et al., 2017; Tripathi et al., 2018).
- **Citation practices are discipline specific:** citation practices in HASS differ greatly from other disciplines and typically result in fewer citations across the HASS area (Anania & Caruso, 2013; Colavizza et al., 2022; Federation for the Humanities and Social Sciences, 2021; Ochsner, 2021; Ochsner et al., 2017). Figure 5 shows the stark contrast between research fields and their share of

references in a database. It's clear that the HASS research field simply doesn't reference journals as frequently as other research fields (E. Archambault & Larivière, 2010).

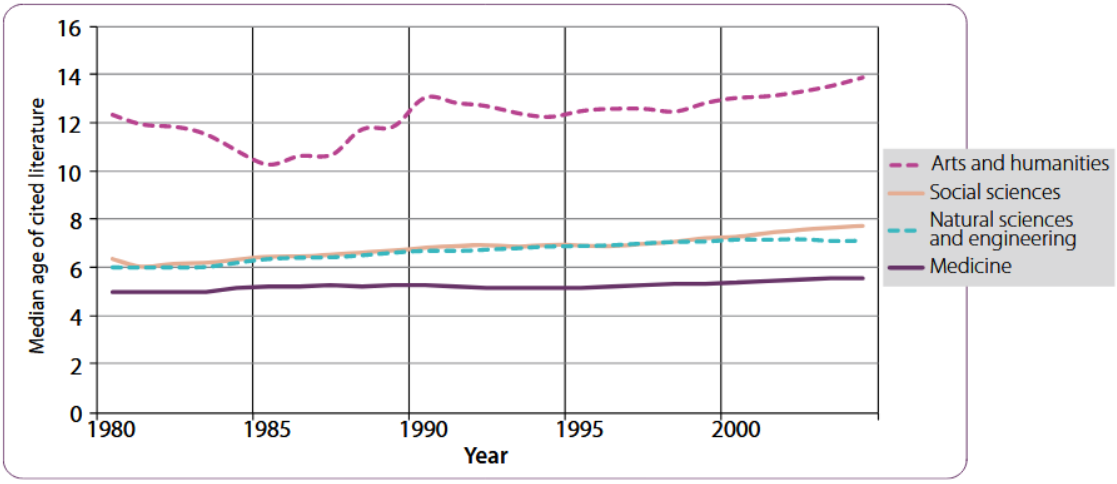
- **Regionally focused research does not have less value:** HASS research is more likely to be local, regional, and published in the local languages and resulting in less indexing by bibliometric databases (E. Archambault & Larivière, 2010; Colavizza et al., 2022; Doğan & Taşkın, 2020; Federation for the Humanities and Social Sciences, 2021; Franssen & Wouters, 2019; Ochsner, 2021; Yang & Zheng, 2019).
- **Trajectory of knowledge advancement builds over time**: advancement of knowledge in HASS is typically incremental and cumulative, building knowledge over time and reaches obsolescence slowly (Federation for the Humanities and Social Sciences, 2017). In other words, HASS knowledge is less linear than STEM, as HASS are based on knowledge that is complementary, segmented, or alternative to each other (Nederhof, 2006; Ochsner, 2021).
- **Publication age and citation rates are long and persistent**: These knowledge production practices result in indirect impacts that often take longer to realize (Federation for the Humanities and Social Sciences, 2017; Nederhof, 2006; Ochsner, 2021; Ochsner et al., 2017). This leads to the accrual of citations over an extended period of time (see Figure 6 and Figure 7) with older publications continuing to garner influence (Archambault & Larivière, 2010, p.252). This is contrary to STEM where the newest research on a topic tends to garner the most attention with citations quickly trailing off (Anania & Caruso, 2013; Bonaccorsi et al., 2017; Colavizza et al., 2022; Federation for the Humanities and Social Sciences, 2017; Nederhof, 2006; Ochsner, 2021).
- **Single authorship results in less attention:** HASS researchers publish single-authored publications more often, which tend receive fewer citations than multi-authored works (Ho & Ho, 2015). Fewer authors potentially results in less attention due to a known 'advertising effect' from multiple authors (Franceschet & Costantini, 2010). This results in the perception that collaborative research is more impactful (Tripathi et al., 2018).
- **Publication types:** HASS encompasses a wide range of disciplines and sub-disciplines, publish work (e.g., policy papers, monographs, book chapters, archival work, and others) in venues that are typically not factored into citation counts and impact factors as they are not indexed (Doğan & Taşkın, 2020; Federation for the Humanities and Social Sciences, 2017; Nederhof, 2006; Ochsner, 2021; Ochsner et al., 2017; Yang & Zheng, 2019), and monographs have a different distribution of citations than journal articles (Chi et al., 2015) and they have complex referencing practices (Colavizza et al., 2022), making it challenging to measure outputs, especially in terms of the impacts that a work has on the field.

FIGURE 5: SHARE OF REFERENCES MADE TO JOURNAL ARTICLES INDEXED IN THE WOS, BY FIELD, 1980–2007.



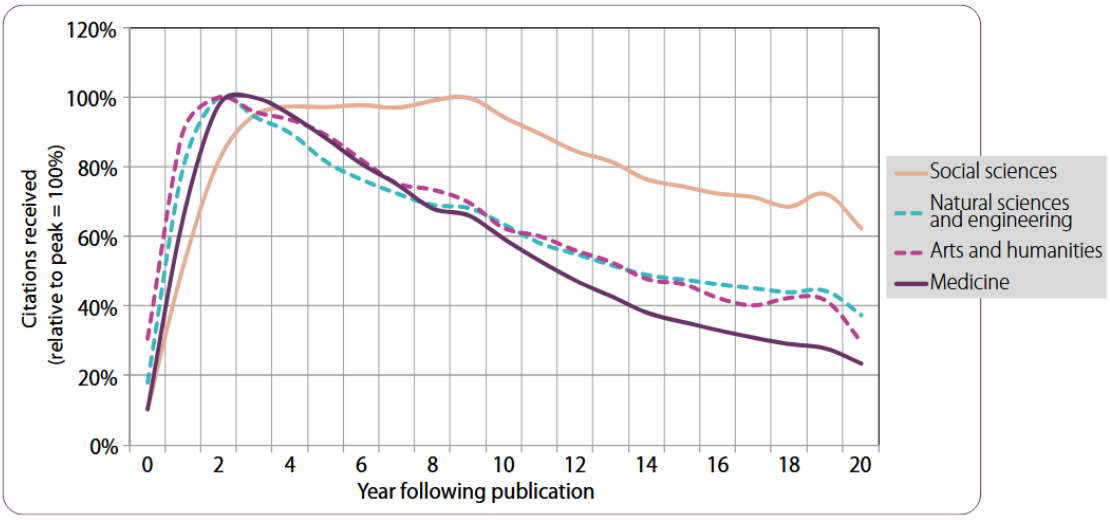
Source: (Archambault & Larivière, 2010, p.252) CC-BYSA UNESCO

FIGURE 6: MEDIAN AGE OF CITED LITERATURE BY FIELD (100-YEAR CITATION WINDOW), 1980-2005.



Source: (Archambault & Larivière, 2010, p.252) CC-BYSA UNESCO

FIGURE 7: CITATIONS OF PAPERS PER YEAR FOLLOWING PUBLICATION.



Source: (Archambault & Larivière, 2010, p.252) CC-BYSA UNESCO

While these limitations are important, they do not entirely preclude the use of bibliometrics in HASS where they can be used in conjunction with the more subjective peer review process, to contextualize and benchmark impacts situated in the HASS discipline or support the grant and evaluation review process. Moreover, as more HASS-specific journals and books are indexed, bibliometric measures will better reflect the overall impact of HASS scholars. The challenge, however, is that the problems with using bibliometrics for HASS go beyond database coverage, as is seen in the fundamental differences in knowledge production and advancement discussed here. Therefore, there are some strategies to more effectively reflect the HASS with bibliometrics. For example, bibliometric measures that may measure HASS

disciplines more effectively including the k-index, which takes into account all citations received for most-cited publications, or the w-index, which accounts for all citations received by all publications (Anania & Caruso, 2013).

6.2 CASE STUDIES OF RESEARCH OUTPUTS IN THE HUMANITIES, ARTS, AND SOCIAL SCIENCES (HASS)

In the Fall of 2022, interviews were conducted with social scientist Lai-Tze Fan, archeologist Robert Park, and historian Ian Milligan.

6.2.1 Themes

There were shared over-arching themes in how these HASS scholars' disciplines intersect with bibliometrics:

- Automating some types of legitimate research outputs within the HASS disciplines is a challenge, especially compared to the STEM fields. For example, a research output like an exhibit will not be indexed in any databases and therefore it would be more appropriate to use qualitative, rather than quantitative, measures to assess HASS research outputs.
- First, last, or single authorship may mean different things in different fields.
- There is considerable invisible labor that goes into their work, which may not be apparent due to their inabilities to be measured by bibliometrics. For example, knowledge translation activities may be a notable aspect of their fields, and that publishing pieces within non-academic sources such, as news sources, that prioritizes making timely research accessible to non-experts are outputs that are completely ignored by traditional bibliometrics, despite having positive social effects.
- Research may involve more community-building efforts, and the amount of time and labor that goes into this process may not be fully appreciated by those working outside of the HASS disciplines, as it is not a task that can be easily counted or measured. For example, researchers could struggle with explaining community-building efforts in quantifiable ways that do not simultaneously objectify those very communities that they are collaborating with.
- With collaborative research, there may be a misunderstanding that multiple authors does not necessarily mean less labor for each individual, as there is additional work that goes into effective collaboration. This may discourage crediting all contributors, which often overlooks traditionally women-intensive forms of labor such as task managing, which may not be properly credited due to the emphasis on finished, not process-based, work.

All three researchers emphasized the importance of understanding that there are vast differences even between the sub-fields within HASS, and as such, it is essential to not amalgamate the entirely of HASS into a few standardized bibliometric practices. For example, institutional performance review forms that ask researchers to describe all the work they have achieved over the last two years are often formatted in the exact same way across all the HASS faculties. Additionally, HASS department reviews may involve committee members with no experience within the candidate's field and may extend their research fields norms onto the review, impacting their ability to assess research outputs accurately. Outsiders may never be able to understand all the nuances of publication differences in fields, so involving community-members from the fields being evaluated are often essential.

6.2.2 Case Studies

Ian Milligan, historian describes that within his field, producing a book or monograph is the primary form of research production. However, as an interdisciplinary scholar, he also publishes in the computer science field. One form of invisible labor within his own discipline is recognition of technical peer-reviewed conference proceedings, which involve a much more rigorous review process than typical humanities conferences, sometimes including as many as 4-5 double blind reviews, compared to other HASS conferences which may only require an abstract being reviewed by a single committee. Additionally, producing code is a labor-intensive process that often is not recognized within the field. For historians who work in tandem with the communities they are researching, developing those relationships based on trust takes considerable time and effort, and while this is acknowledged within history, outsiders may not fully grasp this. This research process will look very different than a historian whose primary sources are books, articles, or archives. A unique aspect of being a historian is special attention is given to one's ability to bring a narrative element, or a point-of-view, to the topic of history you are focusing on. As such, extra credibility is given to single author works, compared to the STEM fields, where having a higher number of researchers attributed to an article tends to give it the appearance of more legitimacy. Edited collections of works (for example books where each chapter is the contribution of a different author), are going to be less discoverable and therefore less cited. Due to the standard being single authorship, the hierarchy placed on first or last authorship placement is foreign to most historians. It is an implicit standard in history that in order to achieve tenure, one has to publish a least one book, and two for promotion at most universities, and that one may even feel a pressure to produce a book initially, in order to get to the research they are really passionate about. If a researcher's work is not primarily book publishing, they are expected to explain why their outputs should be considered equivalent to producing a monograph.

Lai-Tze Fan, social scientist who focuses on design and technology, explains that publish or perish models apply much more to STEM fields than HASS ones, as HASS fields have a lot more variety in their research outputs. For example, her colleagues in STEM may produce eight papers in hopes that one will be accepted for a conference, whereas within her field, the standard practice would be to produce one in-depth work. Researchers in the arts may have to contend with how to define outputs that are not obviously quantifiable. For example, if the output of one's research is an art exhibition, a researcher may propose that impact could be determined by how many attendees the exhibition garnered, or how many views a YouTube video attracted. It may be inappropriate to use quantifiable metrics for some work, for example, if it involves working with vulnerable communities, certain standard practices such as peer-review may not be in spirit with the research or consented to by those communities. Interdisciplinary social scientists or artists may face a stigma that they are a "jack of all trades but a master of none."

Robert Park, anthropologist, and archeologist articulates that books are less coveted within anthropology compared to other HASS fields (though edited collections are much more common), and that most publishing occurs in traditional, peer-reviewed journals. One form of invisible labor for archeologists is that a lot of time and effort tends to go into producing reports for the provincial government that request access to sites for excavation, and these reports are written for a non-professional audience that are never indexed or published. Additional reports may also be required if you are working on sites that belong to a certain community, and many provinces have laws stating that reports justifying the research must be shared with those communities. For anthropologists who work in other countries, a great deal of time may be spent into meeting licensing requirements to do so. Public talks that emphasize knowledge translation on the findings of one's work are another common research output that are often not acknowledged.

6.3 RECOMMENDATIONS FOR USE OF BIBLIOMETRICS IN HASS

There are several other approaches that can be used to adapt or adjust bibliometrics to ensure a better representation of HASS disciplines-related research outputs in bibliometric measures. These are outlined in here:

- Include book reviews, when available, as a measure of merit and impact (Hammarfelt, 2016).
- Use a breadth of altmetrics as further outlined in Section 7 - Measuring Societal Impact with altmetrics.
- Contextualize bibliometric results by including descriptions of additional research impacts that are relevant to the discipline but potentially more difficult to measure (e.g., cultural, policy, and social implications).
- Include assessment criteria/indicators of societal impact that are outside of academia.
- Include digital object identifiers (DOIs) when publishing to increase the likelihood that your publications are included databases.
- Adjust the citation window to a minimum of five years when using standard bibliometric measures, to acknowledge differences in the longstanding and viable publication practices within the HASS disciplines that affect citation rates.
- Avoid using the JIF as a proxy for quality of research outputs especially for individuals, acknowledging that the JIF is a grossly inadequate measure of quality in the HASS disciplines.
- Consider reviewing the number of times that a work has been cited within the text, rather than how frequently an article is cited as a whole. This offers a better understanding of how much author(s) are engaging with a text and may be a more accurate way of assessing research outputs within HASS disciplines where authors are more likely to engage deeply with one or a few texts. It is worth considering whether a text has been cited merely in passing, or if it provides the entire framework for which a work is based on (Rosati & Mongeon, 2022).

7 MEASURING SOCIETAL IMPACT WITH ALTMETRICS

*"We will continue to be bold innovators and leaders in entrepreneurship, adding pillars in **social innovation and policy engagement** aligned with the Global Futures." - Waterloo at 100*

7.1 WHAT ARE ALTMETRICS

At the highest level, altmetrics [19] measure the real-time digital attention received by an article using multiple online sources. The aim of altmetrics is to measure the impact of science on society (García-Villar, 2021; Sugimoto & Larivière, 2017), broadening the sphere of impact beyond merely academic. This offers new ways of measuring science compared to traditional citation and journal impact measures to give a fuller picture of how research outputs influence conversation, thought, and behaviour (Piwowar, 2013). Altmetrics may include but are not limited to mentions of research outputs in blogs, news media, social media, websites, reddit posts, reference managers, and Wikipedia citations (Ortega, 2018). Altmetrics also incorporates data sources outside of the social and mainstream media ecosystem, such as patent and policy documents, that have potential to influence governments, law and commercial innovation (Haustein, 2016).

In this paper we make an important distinction between short-term ***social*** impact (in the form of news, Facebook, Tweets, etc.) and a longer-standing ***societal*** impact (as embodied in patents, policy and clinical practice guidelines). This gives a more nuanced view of the attention types that altmetrics measure, highlighting how the different sources might be tracked according to their relevant timelines (Pal & Rees, 2022).

- ***Social*** impact includes *attention that occurs shortly after publication* via news and social platforms and may be an *early predictor of the intermediate attention* that occurs in other online sources such as blogs and Wikipedia, which take more time to produce.
- ***Societal*** impact is quite rare and *takes significantly more time* to appear and are usually found in forms such as patents, clinical practice guidelines and policy documents. This societal impact bears little relation to earlier forms of attention.

Each source of attention is also associated with a characteristic time signature (Llewellyn et al., 2022) as outline in Table 5:

TABLE 5: CHARACTERISTIC TIME SIGNATURE OF SOCIAL AND SOCIETAL ATTENTION TO SCHOLARLY OUTPUT.

	← Timeline of mentions →		
	Early	**Intermediate**	**Late**
Social	News, Facebook, Tweets	Blogs, Wikipedia	
Societal			Patents, Policy, Clinical Practice Guidelines

[19] The altmetric measures discussed in this report are separate and distinct from Altmetrics (www.altmetric.com-) which is a specific proprietary altmetric provider (Altmetric website. https://www.altmetric.com/).

Proposed over a dozen years ago, the concept of measuring the interest in academic research as found in the broader sphere of social media (Priem et al., 2010) has since grown to maturity. Changes in the scholarly communications landscape since 2010 mean that using altmetrics to understand social and societal impact is now more relevant than ever. These changes include the proliferation of:

- ***Social media***: is now omnipresent, greatly facilitating how interest in research is shared beyond the walls of the academy.
- ***Open Science****:* has emerged as a model for the barrier-free exchange of research outputs in all forms, including datasets.
- ***Pre-prints****:* Once found only in Physics, the COVID-19 crisis intensified the acceptance of pre-prints as a legitimate form of publication in other fields, notably the Health Sciences.
- **Commercial tools:** There are now several commercial solutions for gauging the interest in research by the broader, non-academic public through altmetrics.

Speed, diversity, openness and the ability to measure the impact of a topic, all contribute to the advantages of altmetrics. Yet the practical applications of how alternative metrics help universities understand their social and societal impact is still poorly understood.

7.2 DIFFERENTIATING IMPACTS

Several studies have found a weak correlation between the commonly used Altmetric Attention Score and how often an article is more formally cited (Barakat et al., 2019; Kolahi et al., 2021; Viana Lora & Nel-lo Andreu, 2020). This highlights the disconnect between research that attracts a lot of attention and that which is highly cited; with tweets operating along a different dimension than downloads to reference managers (i.e. Mendeley) and citations. Accordingly, undifferentiated metrics like the Altmetric Attention Score, because they are so influenced by tweets, do not actually measure the quality of research in a meaningful way:

> *"The top articles based on Altmetric scores were not highly cited, suggesting that publications receiving the most media attention may not be the most scientifically rigorous, or that this audience places greater value on different subjects than the scientific community."* (O'Connor et al., 2017)

For example, a comparison of the various types of research impact in the field of shark conservation by Toupin *et al.* (2023) showed that the various sources of altmetric attention are not related to each other, finding no correlation between news mentions and either academic citations or policy documents. Even within the small and well-defined field of shark conservation, the research articles that attract many tweets are not the same ones that inform patents and policy.

7.3 CASE STUDY OF USING ALTMETRICS TO UNDERSTAND FUNDING IMPACT

A case study found in Dinsmore, Allen & Dolby (2014) documented the use of altmetric data for the UK based medical charity The Welcome Trust, who wanted to understand the impact of their funding. They were particularly interested in how altmetrics might be used to detect early engagement with research in the policy sphere. They examined the influence of a single research paper published in *PLOS Medicine* whose author alleged that, during a 2008 consultation with the Scottish government on alcohol policy, some alcohol corporations/companies misrepresented research to slant support towards policies which

favoured their interests. The PLOS Medicine paper remained uncited for three months following its publication, yet it was tweeted (on X) at about four times the average rate for *PLOS Medicine* articles published in 2013 and 70 times the average rate of all Welcome Trust associated articles published across the family of PLOS journals. Many of the tweets belonged to key influencers in the government, non-governmental organisations, and the World Bank. This suggests that although the research focus was very narrow, the ability for altmetrics to detect early engagement with research and its rapid influence on global policy was clear.

7.4 LIMITATIONS TO ALTMETRICS

The limitation to altmetrics span challenges with commercialisation, data quality and usage risks. These limitations do not preclude the use of altmetrics for the quantitative assessment or evaluation of research impacts. However, like traditional bibliometrics it is strongly recommended that these limitations are transparently reported and acknowledged in any analyses and are only used alongside more qualitative disciplinary situated reviews.

7.4.1 Commercialisation

Profit driven agendas: Social media companies profit from their users engaging as much as possible with one another, and therefore actively encourage the proliferation of communication on their platforms. This could artificially inflate the baseline use of such platforms, which would include use of altmetrics.

7.4.2 Data Quality of Altmetrics

There are several data quality issues related to the use of altmetrics including:

Bias: Not unlike traditional bibliometrics, there are language, geographical and cultural biases inherent in altmetrics. Social media platforms may be government regulated (i.e. Facebook in China, and TikTok in Canada). This can lead to unequal data distributions of altmetrics.

Target: There is sometimes no differentiation between what types of documents the altmetrics are coming from. This makes defining the actual type of impact difficult (i.e. policy documents | social media comment sites)(J. Liu & Adie, 2013).

Versioning: Often different versions of papers exist. This creates redundancy in data tracking if different versions of the same paper are included in analyses (J. Liu & Adie, 2013).

Meaning: Different types of altmetrics attention carry different meanings. A mention on X may carry a different contextual intention than a lengthy discussion. It is advisable to define sources and differentiate between source types(Neylon et al., 2014; M. Taylor, 2013).

Normalization: There is no mechanism to consistently standardize and compare altmetrics across periods of time. Altmetrics accumulate in shorter timeframes, however, mentions also increase over time and it is difficult to understand the meaning behind different rates of attention (Bornmann et al., 2013; Holmberg & Thelwall, 2014; M. Taylor, 2013; Thelwall et al., 2013; Vinkler, 2010).

Replication: Altmetric data is susceptible to the rapidly changing needs of the customer facing platforms on which they exist, making replication through methodological consistency difficult (Haustein, Peters, Bar-Ilan, et al., 2014; Haustein, Peters, Sugimoto, et al., 2014; Thelwall et al., 2013).

7.4.3 Usage Risks

Data manipulation risks in altmetrics: Altmetrics can be manipulated, similar to traditional bibliometrics (i.e. citation inflation rings or encouraging short editorials to increase citations). Indeed, altmetrics can be easily manipulated due the lack of quality control of social media platforms, inconsistent measurements, and susceptibility of social media platforms to fake accounts and bots (Darling et al., 2013; J. Liu & Adie, 2013).

Disciplinary differences: Like traditional bibliometrics, almetrics differ among disciplines. Research fields differ in the amount of people working in them and the attention they receive. A highly specialized research topic may have fewer people engaging with it on social media, but this may not reflect the importance or impact of the research output. The relative size and popularity of the fields should not impact the significance of the research output (J. Liu & Adie, 2013; Sugimoto & Larivière, 2017). A case study from the University of Waterloo is presented in Section 7.5 that makes the argument that connecting health with technological innovations at a university that lacks broad medical research funding can still potentially have large societal impacts.

Privacy: Similar to tracking other forms of activity on the internet, users cannot actively consent to their user data being used for altmetrics, or other purposes. Unless users explicitly deny that their data can be collected, interactions with articles such as likes or mentions etc. (along with the user's demographic data) can be are included in the aggregation of altmetrics (M. Taylor, 2013).

7.5 CASE STUDY OF NUANCES IN CITATIONS FROM POLICY

To better quantify policy citations at the University of Waterloo the percentage of the total output cited by policy was used to normalise the metric and allow comparisons between institutions of varying sizes. The SciVal database was used for this data. A five-year window (2020-2024) was used to compensate for the known time lag typical of policy citation accrual.

Benchmarked against its U15 peers Waterloo was to have the lowest proportional influence on policy documents of all U15 institutions, with 6.5% of outputs at Waterloo cited by policy compared to the U15 average of 7.9%. This is likely due to research in the Health Sciences being more relevant to the broad public interests of international organisations (e.g. the *World Health Organization is a common policy body that cites research*). Accordingly, universities with medical schools will tend to have a higher proportion of policy-cited outputs. Waterloo is the only U15 university without a medical school.

However, Waterloo ranks higher (about mid-range) when compared to 25 selected international comparator universities that reflect the technologically focused research mission of Waterloo and also do not have medical schools, with the 6.5% of papers at Waterloo cited by policy compared to the 5.9% for the international comparators. When comparing the averages Waterloo is much closer to its international comparators than to the U15 institutions. Overall, these comparisons indicate that Waterloo's lower policy-citation rate reflects disciplinary focus rather than research quality, aligning more closely with technologically oriented international peers than with medically oriented U15 institutions.

7.6 RECOMMENDATIONS FOR THE USE OF ALTMETRICS

In addition to the recommendations for the HASS disciplines, the recommendations for altmetrics below are complementary and aim to enhance and expand the traditional notion of research impact beyond the standard academic citation.

- Avoid making the direct link between citations from altmetric sources (social media, blog, policy etc.) and real-world impact. The mere use of research does not in itself mean there as been any true social or societal change (pers. comm. David Phipps, 2025).
- Make the distinction between altmetrics that look at short-term social attention and those that look at longer-term societal attention (i.e. Policy citations and patents).
- Report on the limitations of altmetrics as they pertain to commercialisation, data quality and usage risk described in this paper.
- Consider altmetrics as a legitimate alternative to traditional bibliometrics especially when considering disciplines that focus considerable time and effort to knowledge mobilization activities that engage with broader types of research outputs than the typical research article or conference paper. This consideration is particularly important for the HASS disciplines where bibliometrics grossly favour the STEM disciplines.

8 MEASURING INTERDISCIPLINARITY WITH BIBLIOMETRICS

For the last two decades, interdisciplinarity has received considerable attention from the academic community and beyond (Franklin, 1988). Molas and Salter (2002) described interdisciplinary research as a key mechanism to recreate and evolve a new system for academic investigation and can provide an intellectual bridge between disciplines to inform more effective practical solutions.

The growth of interdisciplinary fields of study is driven by the complexity of modern problems and supported by governments and scientific institutions. The University of Waterloo's current strategic plan Waterloo@100 (University of Waterloo, 2022) and its associated Global Futures all require interdisciplinary collaborations. While there is still ample space and need for disciplinary research, paying attention to new methods to identify, develop and measure interdisciplinarity deserves the attention of academic institutions.

8.1 DEFINING INTERDISCIPLINARITY

Academia has not yet had consensus on a definition of interdisciplinarity and therefore has not devised standardized measurements for interdisciplinary research. However, interdisciplinary research has been gaining traction since the 1980's (Van Noorden, 2015) and its value in stimulating world-leading research is clear, as there is longstanding evidence that interdisciplinary research can lead to overall academic development and is related to growth in research productivity and citation impact (Advisory Panel on Federal Support for Fundamental Science, 2017; Bikard et al., 2015; Franklin, 1988; Hall et al., 2012; Levitt & Thelwall, 2009; NeSmith et al., 2013; Wang et al., 2015).

Canada's response to interdisciplinary research

In Canada, federally funded grants that focus on interdisciplinary research are available through the Natural Science and Engineering Research Council (NSERC) and are defined by grant applications that require multidisciplinary review (NSERC, 2012). In 2017, the Tri-Agency has responded to key recommendations from the "Investing in Canada's Future" (the "Naylor Report") that encourage interdisciplinary research through strengthened adjudication processes, boundary spanning programs, expanding lines of inquiry, joint responsibility for researchers at the edges of one particular councils mandates, and mechanism to improve large-scale research (Advisory Panel on Federal Support for Fundamental Science, 2017).

It might be most helpful to think of interdisciplinary research as being the knowledge integration (KI) of scholarly work; connecting and unifying information, ideas, and methods from various sources or disciplines (Naylor et al., 2017), as it spans across academic disciplines and beyond into technological fields or even industrial sectors (Rafols & Meyer, 2010). The idea of knowledge integration focuses on two key aspects of interdisciplinary research, that of 'diversity' and the 'coherence'. Simply put, it is not enough for researchers to draw upon several diverse fields of knowledge but they also need to integrate that knowledge in some way. With this definition, interdisciplinarity can be measured not only by the diversity of the fields of research but by the "intensity of the relationship" (Wang et al., 2015, p. 241) between those fields.

Other related concepts are often used interchangeably with the term interdisciplinary, including multidisciplinary research and transdisciplinary research, neither of which requires both collaboration and integration as interdisciplinary research does (Glänzel & Debackere, 2021).

8.2 MEASURING INTERDISCIPLINARITY

Several scientometric and bibliometric methods have been developed to study various aspects of interdisciplinarity in science, which can in turn, be aggregated at various levels - discipline, organisational unit, research group, scientist, journal, article, keyword, or citation level (See details in Table 6 below) although none are universally accepted measure of interdisciplinarity (Rinia, 2007). A recent seminal study by Rafols and Meyers (2010) better captures the essential "knowledge integration" concept embedded in interdisciplinary research by looking at both diversity – how many different fields are contributing to the bibliographic references, and coherence of outputs – how integrated the different contributing disciplines are becoming (Rafols & Meyer, 2010).

Building on this work, Glänzel and Debackere (2021) define two methodological approaches to the analysis of interdisciplinarity:

1) **The cognitive approach** (See Figure 8) is based on information flows reflected by citation links. This may also include analysis of author's collaborations, or analysis of text similarity between documents, which is particularly promising in the context of Social Sciences and Humanities research.
2) **The organisational approach** analyses collaboration and co-authorship based on the researcher's affiliation and subject classification (Abramo et al., 2012).

The key questions to be settled in using either approach are:

a. What is the level of granularity chosen for the study, i.e., how small is the chosen field, discipline or research area?
b. How are subjects classified and research assigned to disciplines, and
c. How is the information to be quantified and made measurable? (Glänzel & Debackere, 2021).

Importantly, at low granularity levels, distinction between disciplines becomes vague and may result in non-unique assignments of references/authors to disciplines, while at a high granularity level some areas of multidisciplinary research might be missed.

Although there is no agreement over which measure of interdisciplinarity to use, these studies highlight that it is best to consider the subject to be assessed (either an individual, research area or research group) and the goal of the assessment (whether it is finding related literature from outside the field or justifying interdisciplinarity as an output for a group or institution) when choosing a measure, as different measures are best suited for different purposes.

TABLE 6: MEASUREMENTS AND GOALS OF INTERDISCIPLINARY RESEARCH ANALYSIS.

(Updated and adapted from Rinia, 2007, p. 30)				
	Measurement item	**Analysis Goal**	**Pros**	**Cons**
Discipline	Disciplinary origin of researchers	Border crossing; collaborative input; Research practice	includes social and cognitive aspects	not fully identical to actual research
	Collaboration by researchers with other disciplines			
Department / Faculty/ Institution	Collaboration with other departments in other disciplines			
	Disciplinary composition of department/groups			
	Disciplinary composition of students and staff			
Research Group	Collaboration with groups in other fields			
	Disciplinary composition of researchers			
	Disciplinary composition of staff			
Scientist/Author	Collaboration with researchers from other disciplines		includes social and cognitive aspects	data consistency and completeness
	Migration to other disciplines			
	Publishing in journals in other disciplines			
	Co-publishing with authors from other disciplines			
Journal	Publishing articles from different disciplines	Scope of research; Information/ content	less elaborate; feasible	rough measure; general level
	Publishing by authors/affiliations from difference disciplines			
	Classification to different disciplines		stable basis; systematic and broad schemes	limited reach; conservative
	Citations by, or references to, journals in other disciplines		multi level; actual usage	less applicable to most recent research; elaborate process
Article	Keywords from distinct disciplines		direct link to content	dynamic/restricted frame of reference
	Classification to different disciplines		stable basis; systematic and broad schemes	limited reach; conservative
	Authors/affiliations from different disciplines		includes social and cognitive aspects	not fully identical to actual research
Citation	Referencing from or citations to different disciplines	influence (borrowing); knowledge flows between disciplines	multi level; actual usage	less applicable to most recent research; elaborate process
	Co-citation of articles from distinct disciplines			

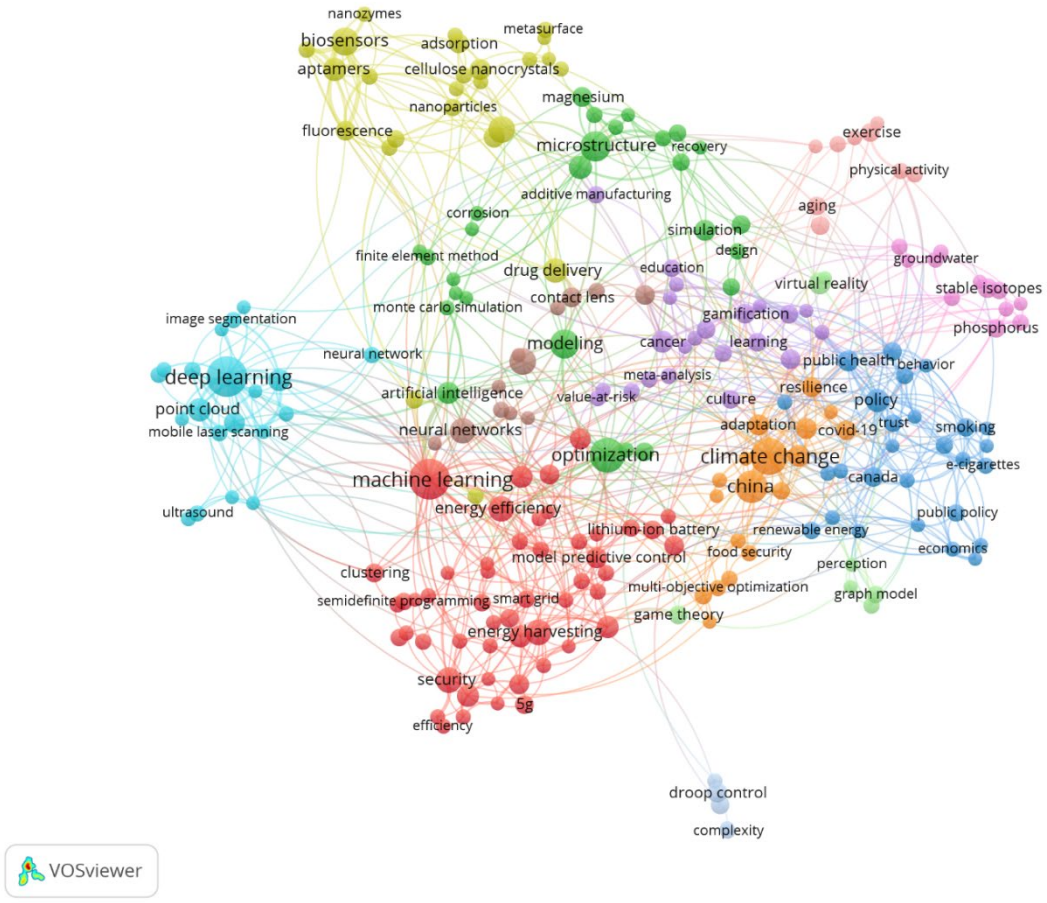
8.3 INTERDISCIPLINARY NETWORKS

Measuring interdisciplinary networks with bibliometrics requires a fundamentally different approach. Social Network Analysis (SNA) is used to map the relationships between objects, revealing underlying social structures. Networks are structurally/visually represented by nodes (objects or actors) and edges (relationships), see Figure 8 for an example a keyword co-occurrence visualization that uses the VosViewer software. In bibliometrics these objects can represent different disciplines or topics through the analysis of publications, journals, researchers, or keywords. In turn, the relationships between these objects are represented by co-citations (with authors or documents), keyword co-occurrence (between documents), bibliographic coupling (shared references between documents), direct citation or direct co-authorship. Termed "science mapping," these visualizations are a powerful way to understand the complex relationships in social networks in academia and can be a strong tool for exploring interdisciplinarity within these structures (C. Liu & Gui, 2016; van Eck & Waltman, 2014). Table 7 lists common tools used to perform social network analyses with bibliometrics data obtained from common citation-tracking database such as those already discussed in previous sections. VOSViewer is probably the most well known.

TABLE 7: COMMON TOOLS TO PERFORM SOCIAL NETWORK ANALYSES WITH BIBLIOMETRICS.

Tool	Description	Developers	Year
VOSViewer	Provides basic functionality for visualizing bibliometric networks using distance-based visualization.	van Eck & Waltman	2014
CitNetExplorer	Specialized tool for visualizing and analyzing citation networks.	van Eck & Waltman	2014
CiteSpace	Java-based software focused on dynamic visualizations of evolving bibliometric networks. [Website](http://cluster.cis.drexel.edu/~cchen/citespace/)	Chaomei Chen	2016
BibExcel	Supports generation of data files from Excel and visualization of tabbed data, especially in geodatabases. [Website](https://bibliometrie.univie.ac.at/bibexcel/)	Olle Persson	2009
ArcGIS (ArcMap)	Geographic information system for mapping and analyzing geographic data; more precise than BibExcel for discovering geographical distribution of authors and collaborative networks.	ESRI	1999

FIGURE 8: KEYWORD CO-OCCURRENCE SOCIAL NETWORK ANALYSIS, VISUALIZATION USING VOSVIEWER



8.3.1 CO-OCCURRENCE ANALYSIS

SNA co-occurrence analysis measures the co-occurrence of pairs of terms such as words, keywords or phrases, in a set of papers.

- Co-occurrence of highly cited papers using similar terms are used to explore research fronts (Morris et al., 2003)
- Keyword co-occurrence analysis can be used to identify research hotspots.

In keyword co-occurrence analysis the rates of the keywords chosen show the strength of the relationship between them. The assumption behind this analysis is that word pairs appearing together in many articles indicate some form of logical connection between the articles and can be mapped to represent common interests with or between given domains of research (C. Liu & Gui, 2016; Olmeda-Gómez et al., 2017; Trujillo & Long, 2018).

8.4 LIMITATIONS OF MEASURING INTERDISCIPLINARITY

In addition to the general limitations to bibliometrics discussed in Section 2 and Section 3, there are some specific limitations to bibliometrics analysis when looking at interdisciplinarity.

- Some critics are concerned that interdisciplinary research may result in research with little rigour because they miss discipline-based perspectives during the regular peer review and vetting (Graff, 2016) that is done during publishing.
- Bornmann and Daniel (2008) and Leydesdorff (2008) note that bibliometric analyses that rely on well-defined classification of disciplines of fields are counterintuitive to the innovative nature of interdisciplinarity research.
- Another concern is that publications by co-authors who research in distant fields, outside of their core field, often experience delayed recognition and are only highly cited over a longer term. Trying to capture bibliometrics over large time spans can be difficult (Wang et al., 2015; Zhang et al., 2024).
- Another concern is that interdisciplinary proposals can have lower funding success rates, at least in part because grant evaluation processes can be biased against interdisciplinary projects which may be assigned to a panel or reviewers who cannot evaluate all aspects of the project. In contrast, more narrowly-focused, less interdisciplinary, proposals are more likely to match assessors' expertise. It is important that interdisciplinary research proposals, and subsequent outputs, are assessed with careful consideration to the context / research area / collaboration in areas that are not usually considered for research (Bromham et al., 2016).

8.5 RECOMMENDATIONS FOR MEASURING INTERDISCIPLINARITY WITH BIBLIOMETRICS

- Combine cognitive and organisational approaches to analyze interdisciplinarity. The cognitive approach focuses on information flows and citation links between published literature, while the organisational approach looks at collaboration based on researcher affiliations and subject classifications.
- Consider using tools and platforms bibliometric tools like VOSviewer for social network analysis (SNA) to explore and identify cross-disciplinary collaborations and their research strength.
- Use bibliometric methods to measure interdisciplinarity by analyzing citation links, co-authorship, and keyword co-occurrence.
- Focus on both the diversity of fields contributing to research and the coherence (interconnectedness) of their knowledge integration. This can be done using citation network analyses.
- Be aware of the limitations of bibliometric analyses, such as delayed recognition of interdisciplinary work, biases in grant evaluation processes, and the context of interdisciplinary work across difference fields of research.

9 BIBLIOMETRICS AND COVID IMPACTS

The COVID-19 global pandemic began impacting the western world in March 2020. During the most intensive period of the pandemic, March 2020 and late spring 2021, there were worldwide lockdowns, travel bans, limitations on movement in public spaces, and most consequentially, waves of illness and loss of life. COVID-19 impacted the world unevenly. The virus itself impacted populations differently for a myriad of reasons, and countries and regions responded to the pandemic in a range of ways. Some countries and regions experienced more restrictive lockdown policies, and others did not. Individual personal circumstances including local geography, family statuses and situations, and academic discipline, meant that impacts were felt differently.

9.1 IMPACT OF COVID-19 ON ACADEMIC RESEARCH

In the academic sphere, human-focused research and fields requiring physical access to resources or where travel is integral, experienced more significant impacts than theoretical researchers. Following the second COVID-19 wave in 2021, academic research facilities began to reopen with a focus on laboratory work, while human-focused research followed later. Some work, such as those requiring access to national collections, may not have been fully possible until 2022. Importantly and throughout the period there was a pause in tenure and promotion processes or annual performance reviews.

One group that is particularly affected by these COVID-19 pandemic impacts are early career researchers. One of the key findings of the original 2016 Paper on "Measuring Research Outputs Through Bibliometrics" was that "assigning a major role to bibliometric measures for hiring, merit review, tenure, and promotion decision-making is strongly discouraged." Certainly, the complexities of COVID-19 reinforce the continued importance of that statement. As Stephanie Lechuga-Peña has argued, "[w]hile the pandemic has impacted faculty at all levels, the impact on pre-tenure junior faculty has different consequences. Pre-tenure faculty need to be engaged in ongoing research, maintain productivity in their scholarship and provide school service to ensure their indefinite academic appointments" (Lechuga-Peña, 2021). Research, teaching, and service were all dramatically impacted by COVID-19, and any uncritical reading of quantitative metrics would not accurately reflect this landscape. Given the long-term nature of research, and the impacts on promotion and tenure, these impacts are long -lasting.

There was also a gendered-impact in the COVID-19 pandemic. Research by Ribarovska et al. (2021) showed that lockdowns had a clear, and negative, impact on women's capacity to publish in their fields. While the reason behind this disadvantage has not been firmly established yet (Ribarovska et al., 2021) there are some indications of why this is so. It is suggested that women were disproportionately affected by changes to the work-life balance that occurred because of school and daycare closures. Academics were often working from home alongside their children, and women scientists tend to have fewer in-home supports than men, despite the increase in policies to even the playing field for men and women (Staniscuaski et al., 2021). Further research is needed to determine whether women scholars were also at a disadvantage in areas such as grant submissions, awards, networking, research, and teaching (Ribarovska et al., 2021). It is worth noting that skipping a single year of publication leads to the reduction of citation impact of any scientist.

Despite a considerable number of gains over the past few decades for women in academia, they remain substantially under-represented in faculty, particularly in tenured positions. Understanding effects on productivity in the pandemic is important. For example, highly educated women in the U.S. are more often becoming mothers than previously, which suggests that the childcare crisis brought on by the pandemic may have had widespread effects within academia. Before the pandemic, women were more likely to leave full-time STEM employment after the birth of a child. Gender diversity could increase scientific innovation and discovery if it becomes an institutional goal for organisations (King & Frederickson, 2021).

9.2 RECOMMENDATIONS ON MONITORING THE EFFECTS OF COVID-19 ON BIBLIOMETRICS

- **Monitor COVID-19 Impacts:** Pay close attention to how COVID-19 has affected bibliometric measures, particularly for women scientists and early career researchers, as the impacts vary by region and country.
- **Conduct Long-Term Research:** Engage in continued research over an extended period to better understand and quantify the effects of COVID-19 on research productivity, promotion, and tenure.
- **Address Work-Life Balance Disruptions:** Recognize the pandemic's potential to disrupt work-life balance and routines, which have been shaped by institutional demands for decades. Use these insights to support scholars' career development.
- **Identify and Mitigate Gender Biases:** Understand where and how gender biases occur in academia to create effective policies and programs that mitigate negative impacts on vulnerable gendered roles.
- **Address Omitted Groups:** Investigate the characteristics of groups omitted from gender analyses due to unreliable gender matching to gain a comprehensive understanding of the research landscape.

10 USING BIBLIOMETRICS TO MEASURE AND INFLUENCE

Institutional research impact is a complex and dynamic process that is ever-evolving. Keeping abreast of research, understanding emerging perspectives, and integrating new learnings into practices and processes, is important for developing responsible and appropriate measures for research productivity and impact.

10.1 RESPONSIBLE RESEARCH ASSESSMENT

There have been numerous efforts to understand how to better measure and assess research productivity and career progression. Four of the most developed and well-known of these are:

- Leiden Manifesto,
- San Francisco Declaration of Research Assessment (DORA),
- Framework for Research and Responsible Innovation (RRI), and
- Metric tide: review of metrics in research assessment.

Taken together, these efforts all conclude that there is a need for change in all areas or research assessment for a multitude of reasons, including assessment systems promoting competition between academics.

Some institutions and institutional networks have undertaken efforts to identify recommendations or best practices based on these, and other, metrics to understand research productivity and impact. In this section we provide a series of actions and supporting evidence, that can facilitate meaningful change in the use of bibliometric measures at the institutional and researcher levels. In these efforts, important institutional enablers include libraries, centralized university resource with bibliometric expertise, and committees, working groups, and task forces created to evaluate and drive the institutional change process. The long-standing concerns on the limitations and negative effects of current assessment practices have helped to lay the groundwork for more concrete changes; for example, some assessment systems promote more competition between academics, instead of strengthening collaboration to help drive interdisciplinarity.

10.2 USING BIBLIOMETRIC INDICATORS RESPONSIBLY

Indicators can be an important part of understanding research productivity, within parameters:

- Indicators should inform, but not over-ride expert judgement
- Indicators should consider differences between disciplines and career stages, equity and diversity factors. Research metrics that reflect or introduce bias shall not be used.
- Indicators should be transparent and understood by those being assessed, with methodologies and data available where possible, and the data itself should be reliable, statistically valid, multi-faceted, and any limitations identified and understood.
- Use of any indicators must follow research integrity standards and internal good research conduct
- Where research metrics are used, they will not rely on a single metric in isolation

Bibliometrics can contribute negatively to research culture through its emphasis on publications and citations; reinforcing "publish or perish," producing easily measured research, and a primary emphasis on research output over teaching and service contributions (Agate et al., 2020; Delgado-López-cózar et al., 2021).

Kowaltowski et al. succinctly describe the result of publish or perish culture as "Impatient Science" (2021). The increasing demands to produce more easily measured outputs within faster timeframes results in an avalanche of scientific information which researchers do not the time to read deeply, or the focus to fully comprehend all the new publications in their field. An unintended consequence of the overreliance of bibliometrics is that it supports "the unconstrained growth of the scientific editorial market," and the pressure to publish may encourage the amount of predatory journals and conferences to increase (Kowaltowski et al., 2021).

10.3 QUALITY OVER QUANTITY CASE STUDIES

The content and quality of research should always be the primary consideration over publication metrics. In considering better practices to evaluate research productivity, the University of Portsmouth outlined principles as part of an overall framework for promotion and tenure of academic staff, many of which are included in the recommendations below. They also developed a process to make the principles actionable, which stipulates that the university allows applicants to identify, in their own view, up to three of their best outputs, in terms of quality, and provide a short justification of their selection in terms of the research, significance, and contribution to knowledge and a description of their personal contribution to the output. This information is reviewed by the promotion and tenure panel when considering interviews. Their process also stipulates that staff promotion policies and the use of quantitative research metrics in other areas for informed decision-making must use metrics in a responsible manner as outlined in their principles.

Another framework was developed by the International Network of Research Management Societies (INORMS). This network brings together research management societies and associations from across the globe. They developed a five-step framework called SCOPE for operationalizing responsible evaluation building on three principles (evaluate only when necessary, include those being evaluated in the process, and build on evaluation expertise) (Inorms Research Evaluation Group, 2021).

10.4 RECOMMENDATIONS FOR RESPONSIBLE RESEARCH ASSESSMENT

- Move away from a limited set of assessment practices based on quantitative publication metrics. Aim to create a more holistic approach that improves academic culture by incentivizing and rewarding a broader range of academic activities.
- Recognize that alternative, non-bibliometric, evaluation of research productivity may require more time and effort. This needs to be recognized in evaluation of academic labour.
- Support and implement open, accurate, transparent, and responsible practices that also prioritize equity. Too little structure permits uneven comparisons that propagate inequities in the academic workforce, but being overly prescriptive can lead to "gaming" of the system. This should not primarily or even necessarily aim to add more indicators, but rather seek to find dynamic, context-sensitive, and

above all holistic approaches that allow researchers and universities the freedom to pursue / manage academic activities in any way they believe is most effective in service to society.

- Focus on raising awareness, community engagement, and building capacity. Bibliometric measures and approaches should foster personal accountability in staff and faculty.
- Aim for institutional initiatives backed by a concerted approach. Improving academic career assessment is a shared responsibility and requires a systems approach uniting the main actors.
- Ensure that research assessment practices and evaluation processes should have clear, strategic objectives.

11 SUMMARY AND RECOMMENDATIONS

This examination of peer-reviewed literature and selected grey literature offers an exploration of common bibliometric measures and the databases used to capture them. Bibliometrics offer a useful approach for measuring some aspects of research outputs. Table 3 provides a helpful summary of bibliometric measures, their limitations, and their appropriate uses. A summary of the levels and types of inquiry that are appropriate, and not, in a variety of situations is provided in Table 8.

Bibliometrics offer a useful approach for measuring some aspects of research outputs.

Bibliometric measures increase in utility when they are used in combination with other measures – most notably peer and other expert review – to assess research outputs.

Bibliometric measures increase in utility when they are used in combination with other measures – most notably peer and other expert review – to assess research outputs. As an example of this principle, Waterloo's Policy on Tenure and Promotion stipulates that "The primary assessment of quality, originality and impact is made by referees and DTPC [Department Tenure and Promotion Committee] members on the basis of examining examples of the candidate's work." Other less direct indicators include the rigour of the review processes for journals and conferences in which the candidate has published, the standards of publishing houses for books, and the extent to which other scholars have made reference to the work. Moreover, bibliometric measures cannot reliably be used as a single indicator to reflect the breadth and depth of an academic research programme. They should not be used to compare research outputs across disciplines, or across academic units or institutions of varying sizes or research orientations, without considering context.

11.1 LIMITATIONS ON THE EFFECTIVE USE OF BIBLIOMETRICS

- **All bibliometric measures have limits:** Bibliometrics measures have specific uses and risks. Users need to understand the underlying calculations that go into metrics to properly interpret them. This is one reason why responsible use of research impact metrics declarations such as DORA have gained significant support. Even normalized measures (such as the Field-weighted citation impact) can be skewed by outliers in small datasets.
- **Different methodologies:** Citation-tracking databases use different methodologies for collecting and reporting bibliometric measures, and their differential indexing of research publications from various fields of study can produce significant limitations to some disciplines.
- **Different Indexes:** Proprietary citation-tracking databases (such as Web of Science and Scopus) index different collections defined by the publications their commercial enterprises hold. Google Scholar, although not defined by proprietary collections, is significantly limited by search

Bibliometric measures cannot reliably be used as a single indicator to reflect the breadth and depth of an academic research programme. They should not be used to compare research outputs across disciplines, or across academic units or institutions of varying sizes or research orientations, without considering context.

conventions that can include-non-scholarly works, duplicates citations, lacks precision and depends on user-created profiles that can be manipulated.

- **Data inaccuracies:** Citation-tracking databases are not infallible. They contain errors in data such as author name misspellings, disambiguation challenges, affiliation errors, or not capturing name changes. Care needs to be taken with smaller datasets, especially author-level analyses, where missed data can have larger proportional impacts on the analysis.
- **Data sovereignty:** Ethical considerations around informed consent and data sovereignty mean institutions must protect anonymity and limit sharing of socio-demographic data. While these safeguards are critical, they limit the granularity of author-level metadata often used for equity analyses. This is a necessary and generally accepted trade-off between privacy and comprehensive measurement. Instead, proxy measures or probabilistic methods of identification are often used in bibliometric methodologies, although they tend to introduce assumption biases in author demographic identification.
- **No comprehensive solution:** No citation-tracking database indexes every type of publication, and comprehensive coverage of research publications is not possible. This limited coverage is reflected in the bibliometrics tools (see Table 1) that draw on data from citation-tracking databases.
- **Disciplinary differences in research outputs:** Academic disciplines produce a range of research outputs, and not all of these are indexed equally well by citation-tracking databases. Therefore, research impact metrics should not be applied equally across disciplines and comparisons between disciplines should be avoided. Research outputs may include papers, patents, conference proceedings, application or systems, databases, data sets, hardware or software artifacts, policy papers, position papers, and government reports, books, exhibitions or creative works, etc.
- **Citation practices differ by discipline:** The practice of attributing citations, and collecting citation data, differs across disciplines and fields. In some fields citations accrue only many years after a work is published, in other fields citations accrue primarily within only a few years after publication. These differences carry over into every bibliometric measure that uses citations, including the h-index.
- **Poor coverage of non-English, interdisciplinary and regional research:** Citation-tracking databases do not have good coverage of research that is not published in English, interdisciplinary research or research of regional importance, and cannot provide field-specific context for research outputs like the extent and type of some research collaborations. This further supports the need for caution when making comparisons across disciplines.
- **Context is lost:** Retrospective analysis of research metrics cannot provide field-specific context for research outputs, such as the extent and type of some research collaborations, or how integrated the research is within the citing works.
- **HASS disciplines deserve special attention:** The humanities, arts and social sciences disciplines tend to be subject to many, if not all the limitations related to disciplinary differences. In particular, citations are typically lower and accrue slower, research is more often locally focused, include more non-English works, there are typically less authors per publication, and the publication types, such as books and creative works, are poorly indexed in citation-tracking databases.

- **Altmetrics aren't always the answer:** Although altmetrics, which expand the breadth of measures of research impact beyond academia, aid in our understanding of social and societal impact, they are also constrained by many of the same challenges as bibliometrics measures and sources. Altmetrics are not direct measures of impact but merely proxies of the attention that a particular study, researcher, or institution may be gaining from sources beyond academia.
- **Bibliometrics are dynamic:** Bibliometric measures taken at different times cannot always be meaningfully compared. First, citations accrue with time after publication. Second, the time required for understanding the impact of a paper using citations differs by discipline. And finally, citation databases themselves change their methodology and journal coverage over time, making longitudinal comparisons problematic.
- **Inequities represented in the data:** There is evidence of biases in citation and collaboration practices (such as self-citations and who gets cited) based on gender and race, and systemic barriers that manifest in slower career paths for women and racialized researchers. The COVID-19 pandemic also affected researchers unequally. These biases cause metrics to underestimate contributions made by women and racialized researchers, especially women of colour. These factors must be taken into consideration when conducting bibliometric analysis, especially if the metrics are being used for assessment activities such as promotion and tenure or grant application reviews.
- **Interdisciplinarity measures are not one-size fits all:** Combining cognitive and organisational approaches to analyze interdisciplinarity helps capture the breadth of ways that researchers collaborate across disciplines. The cognitive approach focuses on information flows and citation links between published literature, while the organisational approach looks at collaboration based on researcher affiliations and subject classifications. Simply looking at journal classification systems within citation-tracking databases may lose the nuance of interdisciplinary innovations.
- **Metric manipulation and behavioural distortions:** Bibliometric measures influence researcher behaviour in unintended ways. They may shape where and how researchers choose to publish to maximize their expected metrics or potentially steer what researchers choose to study. At worst, metrics create incentives for gaming or manipulation. Disciplinary differences in tool coverage and usability can be misinterpreted as differences in research activity or impact, rather than reflecting limitations of the tools themselves.

Tables 3 and 8 provide summaries of bibliometric measures, limitations, levels and types of inquiry and the appropriate uses.

11.2 RECOMMENDED PRACTICES FOR BIBLIOMETRIC ANALYSIS

The position paper reaffirms core principles established in 2016:

- bibliometric measures should be used transparently
- bibliometric measures should be field-normalized
- bibliometric measures should never be used in isolation.

It then expands on these principles through a series of thematic sections, each offering updated recommendations (expanded within each section) that are summarized here:

- **Citation-Tracking Databases:** Emphasizes the need for multiple data sources, validation through persistent identifiers like ORCID, and awareness of disciplinary coverage gaps.
- **Bibliometric Indicators and Measures:** Advocates for normalized citation metrics, integration with peer review, and caution in cross-disciplinary comparisons.
- **Equity and Bias:** Calls for ethical data governance, inclusive evaluation practices, and institutional support for equity-deserving scholars.
- **HASS Disciplines:** Recommends combining bibliometrics with qualitative assessments and altmetrics to capture long-term and societal impacts.
- **Altmetrics and Societal Impact:** Encourages the use of altmetrics as complementary tools, with clear distinctions between social attention and societal change.
- **Interdisciplinarity:** Suggests using cognitive and organisational approaches, including network analysis, to assess interdisciplinary collaboration.
- **COVID-19 Impacts:** Highlights the need to monitor long-term effects on research productivity and equity, particularly for early-career and marginalized researchers.
- **Institutional Responsibility:** Warns against over-reliance on metrics and promotes a balanced, context-sensitive approach to research evaluation.

The use of bibliometrics, and bibliometric analysis, is a common approach for measuring research outputs. If used carefully, bibliometric measures can provide a data point, in conjunction with others, for evaluating research outputs. The following recommendations are suggested to researchers, administrators, and others interested in using bibliometrics or assessing the relevance of bibliometric results. These recommendations speak only to the methodological reliability of bibliometric measures, as indicated in the relevant literature. University policies (such as Waterloo's Policy 77 on Tenure and Promotion) may direct the use of these measures.

For Researchers:

- Define a researcher's identity convention as an author early and use that convention systematically throughout their career. Appropriate affiliation to the University of Waterloo is also important. As an example, researchers can increase the likelihood that their works will be accurately attributed to them within citation-tracking databases by proactively determining how their name will appear in published form throughout their career by creating an author profile such as an Open Researcher and Contributor ID (ORCID).

For All Users:

1) Approach the process of analysing research outputs in the same way that one would conduct good research:
   - develop a strong research question with the scope and clarity appropriate to the discipline and issue under consideration,
   - assess whether bibliometric measures can appropriately provide the information required to answer the research question; if not, it may be necessary to revise the research question or use other measures,
   - if bibliometric measures are indicated, select appropriate tools and measures to investigate the research question,
   - be explicit about other non-bibliometric data sources that should also be considered, and

- understand the research and comparison context, including discipline-specific effects and the implications of sample size.

2) Consider bibliometrics as one measure among a set of others for understanding research output and impact. Best practice is to work from a basket of measures. It is impossible for any bibliometric analysis to present a complete picture. Bibliometrics is optimally used to complement, not replace, other research assessment measures, such as peer review, keeping in mind that "both need to be used with wisdom, discretion and the rigorous application of human judgement" (Phillips & Maes, 2012, p. 3).

3) Understand and account for variations in how disciplines produce and use research publication. Avoid comparisons that the measurement tools and key concepts cannot support. The nature of research (and more generally, scholarly) output (e.g., journal articles, books and book chapters, conference proceedings, performances, social outputs, research artifacts) differs across disciplines, and thus the relevance and applicability of bibliometrics also differs across disciplines. It is important to use bibliometric measures relevant for each discipline and to recognize that meaningful comparisons across those measures may not be possible.

4) Involve those being evaluated in the process and provide them with interpretive information. Given the significant role and impact of context in the use of bibliometrics, researchers in the field or discipline in question may be best equipped to understand and explain the variability of how bibliometric measures capture and reflect research outputs in their field. This will help to ensure that using bibliometric measures incorporates a full understanding of their limitations, particularly at the discipline level.

5) Understand the distinctions among bibliometric measures. Be aware of the methodology, purpose, and limitations of bibliometric databases (such as Web of Science, Scopus, and Google Scholar) and of individual bibliometric measures (such as the Journal Impact Factor and h-index). As an example, it is important to recognize the value of normalized measures compared to whole/raw count while also recognizing that normalized measures can be vulnerable to outliers (e.g., a single highly cited paper can increase the average somewhat artificially). Regular review and updating of research methods and definitions will ensure a strong and current understanding of methodologies used.

6) Exercise caution when using journal impact rankings. Journal impact rankings such as JIF or SCImago Journal Rank (SJR) should not be broadly used as a proxy measure of the quality of individual research articles or an individual's overall performance. The quality of a researchers work deserve a more in-depth evaluation of individual publications, especially for high-stakes activities such as promotion and tenure, awards, grants and other forms of prestige attainment.

TABLE 8: BIBLIOMETRIC ANALYSIS COMPARISON PARAMETERS BY LEVELS AND TYPES OF INQUIRY.

LEVELS AND TYPES OF INQUIRY	COMPARISON PARAMETERS FOR BIBLIOMETRIC ANALYSIS	
	APPROPRIATE	**INAPPROPRIATE**
Individual Researchers: 1. Understand which of their articles have been cited, and how many times their articles have been cited within specific citation-tracking databases. 2. Gain insight into citing works and authors to assess possibilities for collaboration.	Types of analysis: ▪ Person-level analysis, to offer a personal understanding of research output and impact examples. ▪ Individual researchers may highlight highly cited documents in funding competitions, awards applications, or advancement packages. Possible measures: ▪ Total publications ▪ Number of publications per year ▪ Number of citations, total and per document ▪ Highly cited publications ▪ H-index or variant *Caveats and considerations:* ▪ *Comparisons must use same citation-tracking database for data collection (e.g., Web of Science or Scopus).*	▪ Measures in areas that are not robust or well captured in citation tracking databases. For example, regional and interdisciplinary disciplines, or fields where books or conference proceedings are the primary forms of research output. ▪ Performance measurement between researchers for personnel decisions, including hiring, merit review, and tenure. ▪ Comparisons of research output among individual researchers from different disciplines or fields, at different stages of career (early or established) or with different research foci (regional-focused research versus research focused at a broader level). ▪ Comparisons of women researchers' cited works with men (there is a known bias).
Unit Heads, Faculty Deans, Centre / Institute / School Directors and Planning Staff: 1. Understand their group's research activity in terms of production of publication outputs and associated citations. 2. Understand performance over time, or compared with peers. 1. Validate data that external agencies have published.	Types of analysis: ▪ Research publications and citations in fields or disciplines that are well covered in citation-tracking databases. ▪ Performance measurement and reporting analysis may appropriately examine total documents in a citation-tracking database, in selected journal classifications where faculty members are most active. As an example, Waterloo's Faculty of Engineering has used bibliometric measures reported using InCites along with other measures (research funding and research chairs, honours and awards) as part of a data package to support its annual strategic plan evaluation process. In 2014/15, Waterloo Engineering used: - Total Web of Science publications in three major journal classifications in which faculty members are most active, - Category-normalized citation impact of the above publications, and	▪ Measures that use the H-index of researchers as the only basis for comparison, individually or across fields or disciplines without appropriate normalizations. ▪ Measures in areas that are not robust or well captured in citation-tracking databases. For example, regional and interdisciplinary disciplines, fields where books or conference proceedings are the primary forms of research output. ▪ Performance measurement between different Faculties or institutes; for example, comparing the Waterloo Institute for Nanotechnology with the National Institute for Nanotechnology, a joint initiative of the Government of Alberta, University of Alberta and the National Research Council.

Levels and Types of Inquiry	Comparison Parameters for Bibliometric Analysis	
	Appropriate	**Inappropriate**
	- Percentage of documents in the top 10%, based on citations, of the selected publications (above). The Faculty has also used bibliometric measures internally to test the validity of data that other institutions or agencies have published about Waterloo Engineering or to provide input to internal exercises to better understand Waterloo Engineering relative to key peers. ▪ Internal reporting may examine the validity of data that other institutions or agencies have published (for example, examining Waterloo's performance in a specific subject area as captured by a journal subject category). Note: This can only be completed at the journal classification level, and will not map directly onto the work of the faculty members within the department, Faculty or centre under consideration. ▪ Internal reporting to provide an input to exercises aimed at better understanding the unit under consideration relative to key peers. ▪ Performance measurement relative to a similar Faculty, Department or Centre / Institute at another university only with acknowledgement and consideration of differing contexts that would impact comparison including different missions (research or teaching), program mixes or sub-disciplines within the overall unit, regional or international foci, age of the unit, dominant language for publishing, and administrative / funding environment . For example, a comparison of the Faculty of Engineering at Waterloo with that at the University of Toronto must consider the types and numbers of researchers in each sub-discipline, the age and stage of researchers, and the age of the Faculty. Possible measures: ▪ Number of documents cited (over time)	

Levels and Types of Inquiry	Comparison Parameters for Bibliometric Analysis	
	Appropriate	**Inappropriate**
	▪ Category normalized citation impact at the journal subject category level ▪ Percentage of documents in top 10%, or top 1% *Caveats and considerations:* ▪ *Analysis may also be possible with groups of authors if the data is properly cleaned and validated, and if it is considered as monitoring performance over time, rather than in comparison with others.* ▪ *Measures / analysis provided with contextual information, e.g., number of researchers, areas of specialty, career stage.* ▪ *Measures / analysis provided as part of a data package along with other measures, e.g., research funding, awards and honours.* ▪ *Measures provided with appropriate definitions (i.e., based on journal classification). For example, all research output from the University of Waterloo in journals classified as <area of interest> by Thomson Reuters.*	
Senior Administration and Planning Staff (institutional level): 1. Understand the institution's research activity in terms of production of publications and number of citations. 2. Understand performance over time, or compared with selected peers. 3. Validate data that external agencies have published.	Types of analysis: ▪ Research publications and citations in fields or disciplines that are well covered in citation-tracking databases. ▪ Performance measurement and reporting analysis at the organisational level may appropriately examine total documents in a citation-tracking database. ▪ Internal reporting to examine the validity of data that other institutions or agencies have published. ▪ Internal reporting to provide evidence for to exercises aimed at better understanding the institution relative to key peers. ▪ Performance measurement relative to another university / institution only with acknowledgement and consideration of differing contexts that would impact comparison, including different missions (research or teaching), program mixes, age of the institution, regional or international foci,	▪ Measures that use the h-index to produce a comparison of individual researchers. ▪ Measures that compare h-index across different fields or disciplines or the h-index of institutions without appropriate normalization. ▪ Publications and citations that have not been appropriately normalized. ▪ Performance measurement relative to other institutions where the universities or institutions are substantively different, for example, have different missions (liberal arts compared to a technical institute; regionally focused institution compared to a national-focused institution) or publish in predominantly different languages.

Levels and Types of Inquiry	Comparison Parameters for Bibliometric Analysis	
	Appropriate	**Inappropriate**
	dominant language for publishing and administrative / funding environment. For example, a comparison of a relatively new research-intensive institution with a liberal arts primarily undergraduate university must consider and acknowledge the institutional mission, Faculties and program mix, the number of researchers at the institution by discipline, and the age of the institution. ▪ Performance measurement can help an institution to understand and plan for strategic development of the institute. As an example, the Waterloo Institute of Nanotechnology's (WIN) International Science Advisory Board reviews the Institute's publication counts, top journals in which Institute researchers publish, citations and funding performance to provide advice and recommendations, including strategic directions and potential partnership opportunities with similar institutes. Possible measures: ▪ Total documents in Web of Science in selected journal classifications (where faculty members are most active) ▪ Category normalized citation impact ▪ Percentage of documents in top 10% or top 1% *Caveats and considerations:* ▪ *Analysis may also be possible with groups of authors if the data is properly cleaned and validated, and if it is analyzed over time only and not in comparison with others.* ▪ *Measures / analysis provided with contextual information, e.g., number of researchers, areas of specialty, career stage.* ▪ *Measures / analysis provided as part of a data package along with other measures, e.g., research funding, awards and honours.*	

APPENDIX A: 2025 BIBLIOMETRICS WORKING GROUP MEMBERSHIP

Advisory Group Members

Associate Provost, Institutional Data, Analysis and Planning: Allan Starr
University Librarian: Beth Sandore Namachchivaya
Vice President, University Research: Charmaine Dean

Working Group Members (2022-2026)

Chair: Ian Milligan, AVP, Research Oversight and Analysis (OVPRI)
Office of Research: Neil McKay, Antonio Marques, Wendy de Gomez, Swati Bodalia
Institutional Analysis & Planning: Daniela Seskar-Hencic, Jana Carson, Kerry Tolson
Libraries: Alison Hitchens, Laura Bredahl, Jenny Hirst, Mary Ochana
Information Services and Technology: Greg Smith
Health: Marina Mourtzakis
Arts: Ana Ferrer, Selena Santi
Science: Peter Stirling, Kim Arnold
Math: Sarindra Andrianasolo, Martha Foulds, Brittany Reiche
Engineering: David Clausi, Jon Walgate
Environment: Jeremy Pittman, Nancy Worth, Ian Rowland
Research Centres and Institutes: Sushanta Mitra (WIN), Lisa Pokrajac (WIN), Agata Antekiewicz (Games)

Past BWG Members

Past Advisory Group Members
University Librarian: Mark Haslett
Vice President, University Research: George Dixon

Past Working Group Members
Chair: Bruce Muirhead, Tamer Ozsu,
Office of Research: John Thompson, Brenda MacDonald
Libraries: Kathy MacDonald, Shannon Gordon, Pascal Calarco, Peter Stirling, MLIS co-op student, Lauren Byl, Susie Gooch
AHS: Brian Laird, John Mielke
Arts: Tim Kenyon, Angela Roorda, Jennifer Simpson
Science: Alain Francq, Bernie Duncker, Marc Gibson
Math: Tamer Özsu, Kim Tremblay
Engineering: Anwar Hasan, Martha Foulds
Environment: Maren Oelberman

APPENDIX B: TASK FORCE MEMBERSHIP FOR 2026 UPDATE

Task Force 1: How to better serve the Arts, Humanities and Social Sciences in bibliometrics and research impact
Task Force 2: The role of alternative metrics in analyses
Task Force 3: The intersection of equity, diversity, and inclusion and bibliometrics
Task Force 4: Exploring collaboration and interdisciplinary research with bibliometrics
Task Force 5: The impact of COVID on research
Task Force 6: Practical guidance on the responsible use of metrics

Member name	Member's Unit	Task Force Contributed
Ana Ferrer	Faculty of Arts	4
Antonio Marques	Office of Research	4
Ian Milligan	Office of the Vice President Research and International	5
Jana Carson	Institutional Analysis and Planning	3
Jenny Hirst	Libraries	2,6
Jon Walgate	Faculty of Engineering	3
Jordan Hale	Libraries	3
Kumudinie Kariyapperuma	Faculty of Math	4
Marian Davies	Libraries	2,6
Martha Foulds	Faculty of Math	6
Martha Muszycka-Jones	Libraries MLIS Coop Student	Project Manager - All
Nancy Collins	Libraries	1
Peter Stirling	Faculty of Science	5
Selena Santi	Faculty of Arts	1
Wai Yin Winnie Chan	Center for Teaching Excellence	4
Wendy de Gomez	Faculty of Engineering/Office of the Vice President Research and International	2, 6

APPENDIX C: EVIDENCE-GATHERING PROCESS USED

The goal of the 2026 update was to analyze how bibliometrics intersects with the following subjects:

1) the arts, humanities, and social sciences (HASS)
2) alternative metrics
3) equity, diversity, inclusion & accessibility (EDIA)
4) interdisciplinarity
5) COVID
6) responsible, in-practice use.

The six task forces began with a high-level review of known literature to inform the development keywords. Search strings were then developed for searching in Scopus and Web of Science over May to August 2022. Relevant literature was added to a Mendeley Library for review and selection. Using the task forces members working knowledge of the topics, search strings were iteratively developed as a combination of natural language keywords and the vocabulary found in initially scanned sources, see Table 9 for search string details. The search strings were refined by examining the indexed keywords and adding these keywords to the search strings. Keywords associated with incompatible results were filtered out, maximizing the relevance of the results. Task force 1 and 3 opted to do several more specified search strings, whereas the other groups used one all-encompassing search string per database. The results from the search strings were divided among team members to be scanned, with unrelated or redundant documents rejected. The remaining articles were then read, summarized, and their findings were compiled into drafts. Additional documents were found through snowball citation searching, often with the purpose of finding highly-cited foundational texts. Any knowledge gaps from the search string database results were filled with more targeted searches. Task forces 2 and 6 proceeded with supplemental searches in Factiva to review the grey literature on their areas of focus. By aggregating the findings from the literature review, each task force created a draft, which were later used for the update.

TABLE 9: SAMPLE SEARCH STRINGS USED BY EACH TASK FORCE.

Task Force	Scopus Search String	Web of Science Search String
1	("humanities and social sciences" OR HSS OR HASS) AND bibliometrics ("humanities and social sciences" OR HSS OR HASS OR humanities) AND bibliometrics "social sciences" and bibliometrics (limited to title, as adding "social sciences" generally to the search string ballooned the search results with a decline in relevance. Relevant articles all duplicates.	("humanities and social sciences" OR HSS OR HASS OR humanities) AND bibliometrics TS=("humanities and social sciences" OR HSS OR HASS OR humanities OR "social sciences") AND TS=(bibliometric* OR scientometric* OR metric* OR "research output" OR "research impact" OR "quantifying research" OR h-index OR citation* OR "scholarly assessment" OR "cumulative impact" OR humetrics OR DORA OR "declaration on research assessment" OR "valor* of research" OR "action research") AND TS= ("studio art*" OR "performance art" OR plays OR productions OR installation OR sculpture OR "oral history" OR "oral knowledge" OR music OR Indigen* OR "community knowledge" OR "community owned" OR exhibit OR dramaturgy OR "knowledge translation") AND PY=2017-2022)
2	TITLE-ABS-KEY("alt metric" OR "alt-metric" OR altmetric* OR "alternative metric*") AND ("crowdsource peer-review") OR ((plumx) OR ("plum print")) OR (("digital object identifier") OR (DOI) AND ("online attention") OR (("social media") OR ("twitter") OR ("facebook")) AND NOT health OR medic*) AND (LIMIT-TO (PUBYEAR,2022) OR LIMIT-TO (PUBYEAR,2021) OR	TS=(altmetric*) OR TS=(alt-metric*) OR TS=("alt metric*") OR TS=("alternative metric*") OR TS=(plumx) OR TS=(plumprint) AND ALL=(DOI) OR ALL=("digital object identifer") OR ALL=(mendeley) OR ALL=("online attention")

	LIMIT-TO (PUBYEAR,2020) OR LIMIT-TO (PUBYEAR,2019) OR LIMIT-TO (PUBYEAR,2018) OR LIMIT-TO (PUBYEAR,2017) OR LIMIT-TO (PUBYEAR,2016))	
3	TITLE-ABS-KEY((bibliometric* OR scientometric* OR "responsible metric*" OR "research evaluation" OR citation* OR "h-index" OR "reference list*") W/8 ("epistemic* injustice" OR inequality* OR inequity OR "gender bias*" OR "career advancement*" OR "gender disparity" OR "race*" OR racial* OR bipoc OR "ethnic" OR diversity)) AND PUBYEAR > 2016 TITLE-ABS-KEY(indigenous* AND citation*)	((TS=(bibliometric* OR scientometric* OR "responsible metric*" OR "research evaluation" OR citation* OR "h-index" OR "reference list"*) AND TS=("epistemic* injustice" OR inequity OR "gender bias*" OR "career advancement*" OR "gender disparity" OR "race*" OR racial* OR bipoc OR "ethnic" OR "sex disparity") NOT ALL=(biodiversity*) AND DOP=(2017/2022)) NOT WC=(medicine)) NOT WC=(Health Care Sciences & Services) and Highly Cited Papers
4	(TITLE-ABS-KEY (bibliometric OR scientometric OR "research impact")) AND (TITLE-ABS-KEY ("research output" OR (mapp* AND research) OR "research assess*" OR {evolution of research})) AND (TITLE-ABS-KEY (interdisciplin* OR "cross-disciplinary" OR crossdisciplinary OR multidisciplinary OR "multi-disciplinary" OR transdisciplinary OR "trans-disciplinary" OR "standard collaboration metric" OR "team work" OR "interdepartmental" OR "small team" OR "collaborative culture" OR ({diffusion score} OR {integration score}) OR "Co-author analysis" OR (("social network analysis" OR sna) AND (university OR academia OR "higher-ed" OR "higher-education")))) AND PUBYEAR > 2016	TS=(bibliometric OR scientometric OR "research impact") AND TS=("research output" OR (mapp* AND research) OR "research assess*" OR {evolution of research}) AND TS=(interdisciplin* OR "cross-disciplinary" OR crossdisciplinary OR multidisciplinary OR "multi-disciplinary" OR transdisciplinary OR "trans-disciplinary" OR "standard collaboration metric" OR "team work" OR "interdepartment*" OR "small team" OR "collaborative culture" OR ″diffusion score″ OR ″integration score″) AND PY= 2017-2022
5	TITLE((covid OR covid-19 OR pandemic) AND (publish* OR citing OR citation OR bibliometrics OR altmetrics OR tenure OR promotion OR hiring))	TITLE((covid OR covid-19 OR pandemic) AND (publish* OR citing OR citation OR bibliometrics OR altmetrics OR tenure OR promotion OR hiring))
6	(TITLE-ABS-KEY (bibliometric* OR scientometric* OR {publication metric*} OR {research impact} OR {research output} OR {scholarly impact}) AND TITLE-ABS-KEY ({Leiden Manifesto} OR dora OR {basket of measures} OR {basket of indicators} OR {complementary measure*} OR responsible OR {best practice*} OR {case stud*})) AND PUBYEAR > 2016	(((TS=(bibliometric* OR scientometric* OR publication metric OR research output OR scholarly impact)) AND ALL=(Leiden Manifesto OR DORA OR basket of measures OR {basket of indicators OR complementary measure*) AND PY=(2016-2022)))

APPENDIX D: POSITION PAPER CONSULTATION AND COMMUNICATION PLAN

A four-step consultation and communication process was used to consult with and gain input from campus partners (Table 10). These consultations resulted in important contributions, adding additional dimensions of analysis and understanding and enhancing clarity and relevance. Each consultation step informed relevant changes to the update or advised on future additional supplementary materials. The Bibliometrics Working Group endorsed the finalized document in Fall 2025 with additional endorsement by Waterloo Executive and Senior Institutional Leadership over the Winter and Spring 2026 terms. The finalized document was submitted to UWSpace and minted a DOI for robust sharing across online platforms and sharing with administrators and researchers at other universities.

TABLE 10: PHASES, TIMELINE AND AUDIENCES FOR THE CONSULTATION PHASE OF THE UPDATE.

Phase	Timeline	Stakeholders
1. Bibliometrics Leadership and Equity Review **Goal:** In-depth consultation and feedback.	2024 throughout	▪ Bibliometrics Working Group ▪ Research Equity and Office of EDI-R ▪ Advisory Group on Bibliometrics ▪ Provost
2.Broad Campus **Goal:** Email/news communications, review and minor feedback/changes.	2026 Winter/Spring	▪ Campus at large (promotions through Daily Bulletin) ▪ Associate Deans, Research ▪ Faculty Association of the University of Waterloo ▪ Libraries staff ▪ Institutional Analysis and Planning
3.Operational Groups **Goal:** Endorsement and final approvals. UWSpace deposit.	2026 Spring	▪ Research Operations Committee ▪ Advisory Group on Bibliometrics
4.Executive & Senior Institutional Leadership **Goal:** Information, endorsement and awareness, celebration.	2026 Spring	▪ Senate Graduate Council ▪ Senate Research Council ▪ Senate

APPENDIX E: LIST OF TABLES AND FIGURES

Table 1: Common Citation-Tracking Databases and Their Associated Bibliometric Tools.....7
Table 2: How To Determine an H-index 20
Table 3: Bibliometric Measures, Definitions, Uses And Levels Of Use..... 24
Table 4: University Of Waterloo Part Recommendations Relevant To The Research Enterprise. 28
Table 5: Characteristic Time Signature Of Social And Societal Attention To Scholarly Output. 43
Table 6: Measurements And Goals Of Interdisciplinary Research Analysis. 50
Table 7: Common Tools To Perform Social Network Analyses With Bibliometrics. 51
Table 8: Bibliometric Analysis Comparison Parameters By Levels And Types Of Inquiry..... 64
Table 9: Sample Search Strings Used by Each Task Force. 70
Table 10: Phases, Timeline and Audiences for the Consultation Phase of the Update. 72

Figure 1: Discipline Variations in Journal Coverage within Citation-Tracking Databases percent of Total Journals in Ulrich's Periodical Directory.....11
Figure 2: Publication Behaviour Across Disciplines. 14
Figure 3: Recommended Uses of Bibliometric Data by Analysis Level 27
Figure 4: Citation Formats for Oral Histories in APA and MLA. 33
Figure 5: Share Of References Made To Journal Articles Indexed In The Wos, By Field, 1980-2007. 38
Figure 6: Median Age Of Cited Literature By Field (100-Year Citation Window), 1980-2005. 39
Figure 7: Citations Of Papers Per Year Following Publication. 39
Figure 8: Keyword Co-occurrence Social Network Analysis, Visualization using VOSViewer 52

REFERENCES

Abbott, A., Cyranoski, D., Jones, N., Maher, B., Schiermeier, Q., & Van Noorden, R. (2010). Metrics: Do metrics matter? *Nature*, *465*(7300), 860–862. https://doi.org/10.1038/465860a

Abramo, G., D'Angelo, C. A., & Costa, F. D. (2011). National research assessment exercises: A comparison of peer review and bibliometrics rankings. *Scientometrics*, *89*(3), 929–941. https://doi.org/10.1007/s11192-011-0459-x

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2011). A national-scale cross-time analysis of university research performance. *Scientometrics*, *87*(2), 399–413. https://doi.org/10.1007/s11192-010-0319-0

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2012). Identifying interdisciplinarity through the disciplinary classification of coauthors of scientific publications. *Journal of the American Society for Information Science and Technology*, *63*(11), 2206–2222. https://doi.org/10.1002/asi.22647

Abramo, G., D'Angelo, C. A., & Rosati, F. (2013). Measuring institutional research productivity for the life sciences: The importance of accounting for the order of authors in the byline. *Scientometrics*, *97*(3), 779–795. https://doi.org/10.1007/s11192-013-1013-9

Abramo, G., D'Angelo, C. A., & Viel, F. (2013). The suitability of h and g indexes for measuring the research performance of institutions. *Scientometrics*, *97*(3), 555–570. https://doi.org/10.1007/s11192-013-1026-4

Advisory Panel on Federal Support for Fundamental Science. (2017). *Investing in Canada's Future – Strengthening the Foundations of Canadian Research*. Canada's Fundamental Science Review. https://ised-isde.canada.ca/site/canada-fundamental-science-review/sites/default/files/attachments/2022/ScienceReview_April2017-rv.pdf

Agate, N., Kennison, R., Konkiel, S., Long, C. P., Rhody, J., Sacchi, S., & Weber, P. (2020). The transformative power of values-enacted scholarship. *Humanities and Social Sciences Communications*, *7*(1). https://doi.org/10.1057/s41599-020-00647-z

Aguilar, I. N., Ganesh, V., Mannfeld, R., Gorden, R., Hatch, J. M., Lunsford, S., Whipple, E. C., Loder, R. T., & Kacena, M. A. (2019). Authorship Trends Over the Past 30-Years in the Annals of Biomedical Engineering. *Annals of Biomedical Engineering*, *47*(5), 1171–1180. https://doi.org/10.1007/s10439-019-02222-3

Akbaritabar, A., Castro Torres, A. F., & Larivière, V. (2024). A global perspective on social stratification in science. *Humanities and Social Sciences Communications*, *11*(1), 914. https://doi.org/10.1057/s41599-024-03402-w

Akhtar, M. K. (2024). The H-index is an unreliable research metric for evaluating the publication impact of experimental scientists. *Frontiers in Research Metrics and Analytics*, *9*. https://doi.org/10.3389/frma.2024.1385080

Ali, M. J. (2022). Questioning the Impact of the Impact Factor. A Brief Review and Future Directions. *Seminars in Ophthalmology*, *37*(1), 91–96. https://doi.org/10.1080/08820538.2021.1922713

Allen, M. A. (2010). On the current obsession with publication statistics. *ScienceAsia*, *36*(1), 1–5. https://doi.org/10.2306/scienceasia1513-1874.2010.36.001

Anania, G., & Caruso, A. (2013). Two simple new bibliometric indexes to better evaluate research in disciplines where publications typically receive less citations. *Scientometrics*, *96*(2), 617–631. https://doi.org/10.1007/s11192-013-0951-6

Archambault, E., & Larivière, V. (2010). The Limits of bibliometrics for the analysis of the social sciences and humanities literature. In *World social science report, 2010: Knowledge divides* (pp. 251–254). UNESCO.

Archambault, É., Vignola-Gagné, É., Côté, G., Larivière, V., & Gingrasb, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, *68*(3), 329–342. https://doi.org/10.1007/s11192-006-0115-z

Baltussen, A., & Kindler, C. H. (2004). Citation classics in anesthetic journals. *Anesthesia and Analgesia*, *98*(2), 443–451. https://doi.org/10.1213/01.ANE.0000096185.13474.0A

Barakat, A. F., Nimri, N., Shokr, M., Mahtta, D., Mansoor, H., Masri, A., & Elgendy, I. Y. (2019). Correlation of Altmetric Attention Score and Citations for High-Impact General Medicine Journals: A Cross-sectional Study. *Journal of General Internal Medicine*, *34*(6), 825–827. https://doi.org/10.1007/s11606-019-04838-6

Bi, H. H. (2023). Four problems of the h-index for assessing the research productivity and impact of individual authors. *Scientometrics*, *128*(5), 2677–2691. https://doi.org/10.1007/s11192-022-04323-8

Bikard, M., Murray, F., & Gans, J. S. (2015). Exploring Trade-offs in the Organization of Scientific Work: Collaboration and Scientific Reward. *Management Science*, *61*(7), 1473–1495. https://doi.org/10.1287/mnsc.2014.2052

Bonaccorsi, A., Daraio, C., Fantoni, S., Folli, V., Leonetti, M., & Ruocco, G. (2017). Do social sciences and humanities behave like life and hard sciences? *Scientometrics*, *112*(1), 607–653. https://doi.org/10.1007/s11192-017-2384-0

Bornmann, L. (2013). The problem of citation impact assessments for recent publication years in institutional evaluations. *Journal of Informetrics*, *7*(3), 722–729. https://doi.org/10.1016/j.joi.2013.05.002

Bornmann, L., & Daniel, H. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80. https://doi.org/10.1108/00220410810844150

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

Bornmann, L., & Marx, W. (2014). How should the societal impact of research be generated and measured? A proposal for a simple and practicable approach to allow interdisciplinary comparisons. *Scientometrics*, *98*(1), 211–219. https://doi.org/10.1007/s11192-013-1020-x

Bornmann, L., Mutz, R., Neuhaus, C., & Daniel, H.-D. (2008). Citation counts for research evaluation: Standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics*, *8*(1), 93–102. https://doi.org/10.3354/esep00084

Bredahl, L. (2022). The Current and Evolving Landscape of Bibliometric Tools and Technologies. *Library Technology Reports*, *58*(8). https://doi.org/10.5860/ltr.58n8

Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature*, *534*(7609), 684–687. https://doi.org/10.1038/nature18315

Broz, L., & Stöckelová, T. (2018). The culture of orphaned texts: Academic books in a performance-based evaluation system. *Aslib Journal of Information Management*, *70*(6), 623–642. https://doi.org/10.1108/AJIM-03-2018-0063

Burstin, H. R., & Arora, V. M. (2021). Gender Disparities in Journal Citations—Another Metric of Inequity in Academia. *JAMA Network Open*, *4*(7). https://doi.org/10.1001/jamanetworkopen.2021.14787

Butler, L., & Henadeera, K. (2007). *Is there a role for novel citation measures for the social sciences and humanities in a national research assessment exercise?* 170–178.

Butler, L., & Mcallister, I. (2009). *Metrics or Peer Review? Evaluating the 2001 UK Research Assessment Exercise in Political Science* (POLITICAL STUDIES REVIEW, Vol. 7, pp. 3–17).

Byl, L., Carson, J., Feltracco, A., Gooch, S., Gordon, S., Kenyon, T., Muirhead, B., Seskar-Hencic, D., MacDonald, K., Ozsu, T., & Stirling, P. (2016). White Paper: Measuring Research Outputs Through Bibliometrics. In *University of Waterloo*.

Cameron, B. D. (2005). Trends in the usage of ISI bibliometric data: Uses, abuses, and implications. *Portal*, *5*(1), 105–125. https://doi.org/10.1353/pla.2005.0003

Carley, S., Porter, A. L., & Youtie, J. (2013). Toward a more precise definition of self-citation. *Scientometrics*, *94*(2), 777–780. https://doi.org/10.1007/s11192-012-0745-2

Caron, R. M., Jamshed, S. Q., Goodman, M. S., & Kang, S. (2022). Editorial: Women in science: Public health education and promotion 2021. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.1011133

Chakravartty, P., Kuo, R., Grubbs, V., & McIlwain, C. (2018). #CommunicationSoWhite. *Journal of Communication*, *68*(2), 254–266. https://doi.org/10.1093/JOC/JQY003

Chang, Y. W. (2013). A comparison of citation contexts between natural sciences and social sciences and humanities. *Scientometrics*, *96*(2), 535–553. https://doi.org/10.1007/s11192-013-0956-1

Chatterjee, P., & Werner, R. M. (2021). Gender Disparity in Citations in High-Impact Journal Articles. *JAMA Network Open*, *4*(7). https://doi.org/10.1001/JAMANETWORKOPEN.2021.14509

Chi, P.-S., Jeuris, W., Thijs, B., & Glänzel, W. (2015). *Book bibliometrics-A new perspective and challenge in indicator building based on the book citation index*. 1161–1169.

Clarivate. (2021). *Indicators Handbook* (3.3). https://incites.help.clarivate.com/Content/Indicators-Handbook/ih-about.htm

CoARA. (2022). *AGREEMENT ON REFORMING RESEARCH ASSESSMENT*.

Colavizza, G., Peroni, S., & Romanello, M. (2022). The case for the Humanities Citation Index (HuCI): A citation index by the humanities, for the humanities. *International Journal on Digital Libraries*. https://doi.org/10.1007/s00799-022-00327-0

Colledge, L. (n.d.). *Snowball Metrics Recipe Book*. Elsevier.

Coupé, T. (2013). Peer review versus citations—An analysis of best paper prizes. *Research Policy*, *42*(1), 295–301. https://doi.org/10.1016/j.respol.2012.05.004

Curry, S., Gadd, E., & Wilsdon, J. (2022). *Harnessing the Metric Tide: Indicators, infrastructures & priorities for UK responsible research assessment*. (p. 11014215 Bytes). Research on Research Institute. https://doi.org/10.6084/M9.FIGSHARE.21701624

Darling, E. S., Shiffman, D., Côté, I. M., & Drew, J. A. (2013). The role of twitter in the life cycle of a scientific publication. *Ideas in Ecology and Evolution*, *6*. https://doi.org/10.48550/arxiv.1305.0435

Delgado-López-cózar, E., Ràfols, I., & Abadal, E. (2021). Letter: A call for a radical change in research evaluation in spain. *Profesional de La Informacion*, *30*(3). https://doi.org/10.3145/epi.2021.may.09

Dinsmore, A., Allen, L., & Dolby, K. (2014). Alternative Perspectives on Impact: The Potential of ALMs and Altmetrics to Inform Funders about Research Impact. *PLOS Biology*, *12*(11), e1002003–e1002003. https://doi.org/10.1371/JOURNAL.PBIO.1002003

Doğan, G., & Taşkın, Z. (2020). Humanities: The outlier of research assessments. *Information (Switzerland)*, *11*(11), 1–11. https://doi.org/10.3390/info11110540

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

Dorta-González, P., & Dorta-González, M. I. (2013). Comparing journals from different fields of science and social science through a JCR subject categories normalized impact factor. *Scientometrics*, *95*(2), 645–672. https://doi.org/10.1007/s11192-012-0929-9

Eaton, A. A., Saunders, J. F., Jacobson, R. K., & West, K. (2020). How Gender and Race Stereotypes Impact the Advancement of Scholars in STEM: Professors' Biased Evaluations of Physics and Biology Post-Doctoral Candidates. *Sex Roles*, *82*(3–4), 127–141. https://doi.org/10.1007/s11199-019-01052-w

Elsevier. (2023). *Scopus Content Coverage Guide*. Elsevier.

Falagas, M. E., & Alexiou, V. G. (2008). The top-ten in journal impact factor manipulation. *Archivum Immunologiae et Therapiae Experimentalis*, *56*(4), 223–226. https://doi.org/10.1007/s00005-008-0024-5

Federation for the Humanities and Social Sciences. (2014). *The Impacts of Humanities and Social Science Research: Working Paper*. https://www.federationhss.ca/sites/default/files/2014-10-03-impact-project-draft-report-english-version-final2.pdf

Federation for the Humanities and Social Sciences. (2017). *Approaches to Assessing Impacts in the Humanities and Social Sciences*.

https://www.federationhss.ca/sites/default/files/sites/default/uploads/policy/2017/impact_report_en_final.pdf

Federation for the Humanities and Social Sciences. (2021). *Think Big—How the social sciences and humanities are building a better Canada*. https://www.federationhss.ca/en/programs-policy/advocacy/RoleofHSS

Finkenstaedt, T. (1990). Measuring research performance in the humanities. *Scientometrics*, *19*(5–6), 409–417. https://doi.org/10.1007/BF02020703

Franceschet, M., & Costantini, A. (2010). The effect of scholar collaboration on impact and quality of academic papers. *Journal of Informetrics*, *4*(4), 540–553. https://doi.org/10.1016/J.JOI.2010.06.003

Franklin, M. N. (1988). *The community of science in Europe*. ommission of the European Communities, Directorate-General for Science, Research and Development, Brussels.

Franssen, T., & Wouters, P. (2019). Science and its significant other: Representing the humanities in bibliometric scholarship. *Journal of the Association for Information Science and Technology*, *70*(10), 1124–1137. https://doi.org/10.1002/asi.24206

Freeman, R. B., & Huang, W. (2015). Collaborating with People Like Me: Ethnic Coauthorship within the United States. *Journal of Labor Economics*, *33*(S1), S289–S318. https://doi.org/10.1086/678973

Furner, J. (2014). The Ethics of Evaluative Bibliometrics. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (p. 0). The MIT Press. https://doi.org/10.7551/mitpress/9445.003.0008

Gagolewski, M. (2013). Scientific impact assessment cannot be fair. *Journal of Informetrics*, *7*(4), 792–802. https://doi.org/10.1016/j.joi.2013.07.001

García-Villar, C. (2021). A critical review on altmetrics: Can we measure the social impact factor? *Insights into Imaging*, *12*(1), 1–10. https://doi.org/10.1186/S13244-021-01033-2/TABLES/4

Glänzel, W., & Debackere, K. (2021). Various aspects of interdisciplinarity in research and how to quantify and measure those. *Scientometrics*. https://doi.org/10.1007/s11192-021-04133-4

Glänzel, W., & Moed, H. F. (2002). Journal impact measures in bibliometric research. *Scientometrics*, *53*(2), 171–193. https://doi.org/10.1023/A:1014848323806/METRICS

Gordon, S., & Hitchens, A. (2020). *Library Impact Practice Brief: Supporting Bibliometric Data Needs at Academic Institutions*. https://doi.org/10.29242/brief.waterloo2020

Government of Canada. (2016, December 20). *Tri-Agency Open Access Policy on Publications (2015)*. https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/open-access/tri-agency-open-access-policy-publications-2015

Government of Canada. (2021, March 14). *Tri-Agency Research Data Management Policy*. https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/research-data-management

Government of Canada. (2022, March 17). *Canada First Research Excellence Fund – 2022 Competition Letter of Intent Instructions*. Canada First Research Excellence Fund. https://www.cfref-apogee.gc.ca/program-programme/competitions-concours/2022/loi-instructions-eng.aspx

Graff, H. J. (2016). The "Problem" of Interdisciplinarity in Theory, Practice, and History. *Social Science History*, *40*(4), 775–803. https://doi.org/10.1017/ssh.2016.31

Graham, A., Pratt, T. C., Lee, H., & Cullen, F. T. (2019). Contemporary Classics? The Early Onset of Influence of Articles Published in Criminology and Criminal Justice Journals, 2010–2015. *Journal of Criminal Justice Education*, *30*(3), 348–375. https://doi.org/10.1080/10511253.2018.1510018

Haeffner-Cavaillon, N., & Graillot-Gak, C. (2009). The use of bibliometric indicators to help peer-review assessment. *Archivum Immunologiae et Therapiae Experimentalis*, *57*(1), 33–38. https://doi.org/10.1007/s00005-009-0004-2

Hall, K. L., Stokols, D., Stipelman, B. A., Vogel, A. L., Feng, A., Masimore, B., Morgan, G., Moser, R. P., Marcus, S. E., & Berrigan, D. (2012). Assessing the Value of Team Science. *American Journal of Preventive Medicine*, *42*(2), 157–163. https://doi.org/10.1016/j.amepre.2011.10.011

Hammarfelt, B. (2016). Beyond Coverage: Toward a Bibliometrics for the Humanities. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research Assessment in the Humanities* (pp. 115–131). Springer International Publishing. https://doi.org/10.1007/978-3-319-29016-4_10

Harnad, S. (2008). Validating research performance metrics against peer rankings. *Ethics in Science and Environmental Politics*, *8*(1), 103–107. https://doi.org/10.3354/esep00088

Haustein, S. (2016). Grand challenges in altmetrics: Heterogeneity, data quality and dependencies. *Scientometrics*, *108*(1), 413–423. https://doi.org/10.1007/S11192-016-1910-9/METRICS

Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics*, *101*(2), 1145–1163. https://doi.org/10.1007/S11192-013-1221-3/METRICS

Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology*, *65*(4), 656–669. https://doi.org/10.1002/ASI.23101

Hazelkorn, E. (2013). World-class universities or world-class systems? Rankings and higher education policy choices. In P. T. M. Marope, P. J. Wells, & E. Hazelkorn (Eds.), *Rankings and Accountability in Higher Education: Uses and Misuses* (pp. 71–94). UNESCO Publishing.

Heng Wong, M. Y., Tan, N. Y. Q., & Sabanayagam, C. (2019). Time trends, disease patterns and gender imbalance in the top 100 most cited articles in ophthalmology. *British Journal of Ophthalmology*, *103*(1), 18–25. https://doi.org/10.1136/bjophthalmol-2018-312388

Herring, C. (2009). Does Diversity Pay?: Race, Gender, and the Business Case for Diversity. *American Sociological Review*, *74*(2), 208–224. https://doi.org/10.1177/000312240907400203

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

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, *102*(46), 16569–16572. https://doi.org/10.1073/pnas.0507655102

Ho, H.-C., & Ho, Y.-S. (2015). Publications in dance field in Arts & Humanities Citation Index: A bibliometric analysis. *Scientometrics*, *105*(2), 1031–1040. https://doi.org/10.1007/s11192-015-1716-1

Hofstra, B., Kulkarni, V. V., Munoz-Najar Galvez, S., He, B., Jurafsky, D., & McFarland, D. A. (2020). The diversity-innovation paradox in science. *Proceedings of the National Academy of Sciences - PNAS*, *117*(17), 9284–9291. https://doi.org/10.1073/pnas.1915378117

Holmberg, K., & Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics*, *101*(2), 1027–1042. https://doi.org/10.1007/S11192-014-1229-3/METRICS

Holmes, R. (2013, June 29). *Competition and controversy in global rankings*. University World News. https://www.universityworldnews.com/post.php?story=20130626160718267

Informing research choices: Indicators and judgment. (2012). Council of Canadian Academies.

Inorms Research Evaluation Group. (2021). *The Scope Framework*. *Emerald Publishing*.

Ireland, T., MacDonald, K., & Stirling, P. (2012). The H-index: What is it, how do we determine it, and how can we keep up with it? In A. Tokar, M. Beurskens, S. Keuneke, M. Mahrt, I. Peters, C. Puschmann, T. van Treeck, & K. Weller (Eds.), *Science and the Internet* (pp. 237–247). D/U/P, Düsseldorf University Press.

Johann, D., Neufeld, J., Thomas, K., Rathmann, J., & Rauhut, H. (2024). The impact of researchers' perceived pressure on their publication strategies. *Research Evaluation*, rvae011. https://doi.org/10.1093/reseval/rvae011

Johnson, M. H., Cohen, J., & Grudzinskas, G. (2012). The uses and abuses of bibliometrics. *Reproductive BioMedicine Online*, *24*(5), 485–486. https://doi.org/10.1016/j.rbmo.2012.03.007

Khan, M. A., & Ho, Y. S. (2012). Top-cited articles in environmental sciences: Merits and demerits of citation analysis. *Science of the Total Environment*, *431*, 122–127. https://doi.org/10.1016/j.scitotenv.2012.05.035

Kiesslich, T., Beyreis, M., Zimmermann, G., & Traweger, A. (2021). Citation inequality and the Journal Impact Factor: Median, mean, (does it) matter? *Scientometrics*, *126*(2), 1249–1269. https://doi.org/10.1007/S11192-020-03812-Y/FIGURES/5

King, M. M., & Frederickson, M. E. (2021). The Pandemic Penalty: The Gendered Effects of COVID-19 on Scientific Productivity. *Https://Doi.Org/10.1177/23780231211006977*, *7*. https://doi.org/10.1177/23780231211006977

Kolahi, J., Khazaei, S., Iranmanesh, P., Kim, J., Bang, H., & Khademi, A. (2021). Meta-Analysis of Correlations between Altmetric Attention Score and Citations in Health Sciences. *BioMed Research International*, *2021*(1), 6680764. https://doi.org/10.1155/2021/6680764

Kowaltowski, A. J., Silber, A. M., & Oliveira, M. F. (2021). Responsible science assessment: Downplaying indexes, boosting quality. *Anais Da Academia Brasileira de Ciencias*, *93*(1), 1–9. https://doi.org/10.1590/0001-3765202120191513

Kozlowski, D., Lariviere, V., Sugimoto, C. R., & Monroe-White, T. (2022). Intersectional inequalities in science. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(2). https://doi.org/10.1073/PNAS.2113067119

Krauskopf, E. (2013). Deceiving the research community through manipulation of the impact factor. *Journal of the American Society for Information Science and Technology*, *64*(11), 2403–2403. https://doi.org/10.1002/asi.22905

Kubota, R. (2020). Confronting Epistemological Racism, Decolonizing Scholarly Knowledge: Race and Gender in Applied Linguistics. *Applied Linguistics*, *41*(5), 712–732. https://doi.org/10.1093/APPLIN/AMZ033

Kwiek, M., & Roszka, W. (2021). GENDER DISPARITIES IN INTERNATIONAL RESEARCH COLLABORATION: A STUDY OF 25,000 UNIVERSITY PROFESSORS. *Journal of Economic Surveys*, *35*(5), 1344–1380. https://doi.org/10.1111/joes.12395

Kwon, D. (2022). The rise of citational justice: How scholars are making references fairer. *Nature*, *603*(7902), 568–571. https://doi.org/10.1038/D41586-022-00793-1

Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Global gender disparities in science. *Nature*, *504*(7479), 211–213. https://doi.org/10.1038/504211a

Lawrence, P. A. (2008). Lost in publication: How measurement harms science. *Ethics in Science and Environmental Politics*, *8*(1), 9–11. https://doi.org/10.3354/esep00079

Lechuga-Peña, S. (2021). Navigating pre-Tenure and COVID-19: A Testimonio of a BIPOC Junior Faculty Mother. *Https://Doi.Org/10.1177/08861099211048432*, *37*(1), 13–19. https://doi.org/10.1177/08861099211048432

Ledford, H. (2011). Paper on genetics of longevity retracted. *Nature*. https://doi.org/10.1038/news.2011.429

Lerback, J. C., Hanson, B., & Wooden, P. (2020). *Association Between Author Diversity and Acceptance Rates and Citations in Peer-Reviewed Earth Science Manuscripts*. https://doi.org/10.1029/2019EA000946

Levitt, J. M., & Thelwall, M. (2009). The most highly cited Library and Information Science articles: Interdisciplinarity, first authors and citation patterns. *Scientometrics*, *78*(1), 45–67. https://doi.org/10.1007/s11192-007-1927-1

Leydesdorff, L. (2008). Caveats for the use of citation indicators in research and journal evaluations. *Journal of the American Society for Information Science and Technology*, *59*(2), 278–287. https://doi.org/10.1002/asi.20743

Liu, C., & Gui, Q. (2016). Mapping intellectual structures and dynamics of transport geography research: A scientometric overview from 1982 to 2014. *Scientometrics*, *109*(1), 159–184. https://doi.org/10.1007/s11192-016-2045-8

Liu, F., Rahwan, T., & AlShebli, B. (2023). Non-White scientists appear on fewer editorial boards, spend more time under review, and receive fewer citations. *Proceedings of the National Academy of Sciences*, *120*(13). https://doi.org/10.1073/pnas.2215324120

Liu, J., & Adie, E. (2013). Five challenges in altmetrics: A toolmaker's perspective. *Bulletin of the American Society for Information Science and Technology*, *39*(4), 31–34. https://doi.org/10.1002/BULT.2013.1720390410

Llewellyn, N. M., Weber, A. A., Fitzpatrick, A. M., & Nehl, E. J. (2022). Big splashes & ripple effects: A narrative review of the short- & long-term impact of publications supported by an NIH CTSA pediatrics program. *Translational Pediatrics*, *11*(3), 411–422. https://doi.org/10.21037/tp-21-506

Lovegrove, B. G., & Johnson, S. D. (2008). Assessment of research performance in biology: How well do peer review and bibliometry correlate? *BioScience*, *58*(2), 160–164. https://doi.org/10.1641/B580210

Lowry, P. B., Moody, G. D., Gaskin, J., Galletta, D. F., Humpherys, S. L., Barlow, J. B., & Wilson, D. W. (2013). Evaluating journal quality and the association for information systems senior scholars' journal basket via bibliometric measures: Do expert journal assessments add value? *MIS Quarterly: Management Information Systems*, *37*(4), 993–1012. https://doi.org/10.25300/MISQ/2013/37.4.01

Malapane, O. L., Musakwa, W., Chanza, N., & Radinger-Peer, V. (2022). Bibliometric Analysis and Systematic Review of Indigenous Knowledge from a Comparative African Perspective: 1990–2020. *Land 2022, Vol. 11, Page 1167*, *11*(8), 1167–1167. https://doi.org/10.3390/LAND11081167

Mayer, S. J., & Rathmann, J. M. K. (2018). How does research productivity relate to gender? Analyzing gender differences for multiple publication dimensions. *Scientometrics*, *117*, 1663–1693.

Mercier, P., & Remy-Kostenbauer, P. (2013). *Citations Impact at the University of Ottawa* (pp. 1–31).

Messerly, M. (2014, December 5). *Citations for sale*. Www.Dailycal.Org. https://www.dailycal.org/archives/citations-for-sale/article_50580ac9-06b6-558b-9cf7-66d35636c0b3.html

Michels, C., & Schmoch, U. (2014). Impact of bibliometric studies on the publication behaviour of authors. *Scientometrics*, *98*(1), 369–385. https://doi.org/10.1007/s11192-013-1015-7

Moed, H. F. (2005). *Citation analysis in research evaluation*. Springer.

Molas-Gallart, J., & Salter, A. (2002). Diversity and excellence: Considerations on research policy. *IPTS Report*, *66*(5), 5–13.

Mongeon, P., & Paul-Hus, A. (2016a). 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

Mongeon, P., & Paul-Hus, A. (2016b). 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

Morphew, C. C., & Swanson, C. (2011). On the Efficacy of Raising Your University's Rankings. In *University Rankings* (pp. 185–199). Springer Netherlands. https://doi.org/10.1007/978-94-007-1116-7_10

Morris, S. A., Yen, G., Wu, Z., & Asnake, B. (2003). Time line visualization of research fronts. *Journal of the American Society for Information Science and Technology*, *54*(5), 413–422. https://doi.org/10.1002/asi.10227

Mryglod, O., Kenna, R., Holovatch, Y., & Berche, B. (2013). Comparison of a citation-based indicator and peer review for absolute and specific measures of research-group excellence. *Scientometrics*, *97*(3), 767–777. https://doi.org/10.1007/s11192-013-1058-9

Mustaffa, J. B. (2017). Mapping violence, naming life: A history of anti-Black oppression in the higher education system. *International Journal of Qualitative Studies in Education*, *30*(8), 711–727. https://doi.org/10.1080/09518398.2017.1350299

Naylor, C. D., Crago, M., Lazaridis, M., Malacrida, C., McDondal, A. B., Piper, M. C., Quirion, R., Wilson, A., & Advisory Panel for the Review of Federal Support for Fundamental Science. (2017). *Investing in Canada's Future – Strengthening the Foundations of Canadian Research*. https://ised-isde.canada.ca/site/canada-fundamental-science-review/sites/default/files/attachments/2022/ScienceReview_April2017.pdf

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, *66*(1), 81–100. https://doi.org/10.1007/s11192-006-0007-2

NeSmith, E. G., Medeiros, R. S., Ferdinand, C. H. B., Hawkins, M. L., Holsten, S. B., Zhu, H., & Dong, Y. (2013). "It takes a village" to raise research productivity. *Journal of Trauma and Acute Care Surgery*, *75*(1), 173–178. https://doi.org/10.1097/TA.0b013e31829383c4

Neufeld, J., & von Ins, M. (2011). Informed peer review and uninformed bibliometrics? *Research Evaluation*, *20*(1), 31–46. https://doi.org/10.3152/095820211X12941371876382

Neylon, C., Willmers, M., & King, T. (2014). *Rethinking impact: Applying altmetrics to Southern African research*. Scholarly Communication in Africa Programme. http://hdl.handle.net/10625/53461

Nygaard, L. P., & Bahgat, K. (2018). What's in a number? How (and why) measuring research productivity in different ways changes the gender gap. *Journal of English for Academic Purposes*, *32*, 67–79. https://doi.org/10.1016/j.jeap.2018.03.009

Ochsner, M. (2021). Bibliometrics in the humanities, arts and social sciences. *Handbook Bibliometrics*, 117–124. https://doi.org/10.1515/9783110646610-013/MACHINEREADABLECITATION/RIS

Ochsner, M., Hug, S. E., & Editors, H. D. (2016). *Research Assessment in the Humanities*. https://doi.org/10.1007/978-3-319-29016-4

Ochsner, M., Hug, S., & Galleron, I. (2017). The future of research assessment in the humanities: Bottom-up assessment procedures. *Palgrave Communications*, *3*(1). https://doi.org/10.1057/palcomms.2017.20

O'Connor, E. M., Nason, G. J., O'Kelly, F., Manecksha, R. P., & Loeb, S. (2017). Newsworthiness vs scientific impact: Are the most highly cited urology papers the most widely disseminated in the media? *BJU International*, *120*(3), 441–454. https://doi.org/10.1111/bju.13881

Odekunle, E. A. (2020). Dismantling systemic racism in science. *Science*, *369*(6505), 780–781. https://doi.org/10.1126/science.abd7531

Olmeda-Gómez, C., Ovalle-Perandones, M.-A., & Perianes-Rodríguez, A. (2017). Co-word analysis and thematic landscapes in Spanish information science literature, 1985–2014. *Scientometrics*, *113*(1), 195–217. https://doi.org/10.1007/s11192-017-2486-8

Ortega, J. L. (2018). Disciplinary differences of the impact of altmetric. *FEMS Microbiology Letters*, *365*(7). https://doi.org/10.1093/femsle/fny049

Paiva, C. E., Lima, J. P. da S. N., & Paiva, B. S. R. (2012). Articles with short titles describing the results are cited more often. *Clinics*, *67*(5), 509–513. https://doi.org/10.6061/clinics/2012(05)17

Pal, A., & Rees, T. J. (2022). Introducing the EMPIRE Index: A novel, value-based metric framework to measure the impact of medical publications. *PLOS ONE*, *17*(4), e0265381. https://doi.org/10.1371/journal.pone.0265381

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

Phillips, M., & Maes, K. (2012). *Research universities and research assessment*. League of European Research Universities. https://www.leru.org/files/Research-Universities-and-Research-Assessment-Full-paper.pdf

Piwowar, H. (2013). Value all research products. *Nature 2013 493:7431*, *493*(7431), 159–159. https://doi.org/10.1038/493159a

Priem, J., & Hemminger, B. M. (2010). Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. *First Monday*, *15*(7). https://doi.org/10.5210/FM.V15I7.2874

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010, October 26). altmetrics: A manifesto. *Altmetrics*. http://altmetrics.org/manifesto/

Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, *82*(2), 263–287. https://doi.org/10.1007/s11192-009-0041-y

Research Intelligence: Research Metrics Guidebook. (2019, March). Elsevier.

Retzer, V., & Jurasinski, G. (2009). Towards objectivity in research evaluation using bibliometric indicators—A protocol for incorporating complexity. *Basic and Applied Ecology*, *10*(5), 393–400. https://doi.org/10.1016/j.baae.2008.09.001

Ribarovska, A. K., Hutchinson, M. R., Pittman, Q. J., Pariante, C., & Spencer, S. J. (2021). Gender inequality in publishing during the COVID-19 pandemic. *Brain, Behavior, and Immunity*, *91*, 1–3. https://doi.org/10.1016/J.BBI.2020.11.022

Rinia, E. J. (2007). *Measurement and evaluation of interdisciplinary research and knowledge transfer*.

Rodríguez‐Navarro, A. (2011). Measuring research excellence: Number of Nobel Prize achievements versus conventional bibliometric indicators. *Journal of Documentation*, *67*(4), 582–600. https://doi.org/10.1108/00220411111145007

Rosati, D., & Mongeon, P. (2022). *The impact of in-text citations on journal rankings*.

Rymer, L. (2011). *Measuring the impact of research: The context for metric development*. The Group of Eight. https://files.eric.ed.gov/fulltext/ED536201.pdf

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ (Clinical Research Ed.)*, *314*(7079), 498–502. https://doi.org/10.1136/BMJ.314.7079.497

Shema, H., Bar-Ilan, J., & Thelwall, M. (2014). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. *Journal of the Association for Information Science and Technology*, *65*(5), 1018–1027. https://doi.org/10.1002/asi.23037

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

Staniscuaski, F., Kmetzsch, L., Soletti, R. C., Reichert, F., Zandonà, E., Ludwig, Z. M. C., Lima, E. F., Neumann, A., Schwartz, I. V. D., Mello-Carpes, P. B., Tamajusuku, A. S. K., Werneck, F. P., Ricachenevsky, F. K., Infanger, C., Seixas, A., Staats, C. C., & de Oliveira, L. (2021). Gender, Race and Parenthood Impact Academic Productivity During the COVID-19 Pandemic: From Survey to Action. *Frontiers in Psychology*, *12*, 1640–1640. https://doi.org/10.3389/FPSYG.2021.663252/BIBTEX

Sugimoto, C. R., & Larivière, V. (2017). Altmetrics: Broadening Impact or Amplifying Voices? *ACS Central Science*, *3*(7), 674–676. https://doi.org/10.1021/acscentsci.7b00249

Taylor, J. (2011). The Assessment of Research Quality in UK Universities: Peer Review or Metrics? *British Journal of Management*, *22*(2), 202–217. https://doi.org/10.1111/j.1467-8551.2010.00722.x

Taylor, M. (2013). Towards a common model of citation: Some thoughts on merging altmetrics and bibliometrics. *Research Trends*, *1*(35).

Thelwall, M. (2020). Author gender differences in psychology citation impact 1996–2018. *International Journal of Psychology*, *55*(4), 684–694. https://doi.org/10.1002/IJOP.12633

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–e64841. https://doi.org/10.1371/JOURNAL.PONE.0064841

Toupin, R., Melnick, K., Moharana, T., Curry, B., Gracey, C., Richards, N., MacDonald, B. H., & Mongeon, P. (2023). Measuring engagement with shark research in science, news and social media, and policy: A bibliometric analysis. *Marine Policy*, *157*, 105829. https://doi.org/10.1016/j.marpol.2023.105829

Tripathi, M., Kumar, S., & Babbar, P. (2018). Bibliometrics of social science and humanities research in India. *Current Science*, *114*(11), 2240–2247. https://doi.org/10.18520/cs/v114/i11/2240-2247

Trujillo, C. M., & Long, T. M. (2018). Document co-citation analysis to enhance transdisciplinary research. *Science Advances*, *4*(1). https://doi.org/10.1126/sciadv.1701130

University of Waterloo. (2022). *Waterloo at 100: University of Waterloo Strategic Vision* (No. FP2245). University of Waterloo. https://uwaterloo.ca/strategic-vision/sites/default/files/uploads/documents/fp2245_waterlooat100.pdf

Van Dalen, H. P., & Henkens, K. (2012). Intended and unintended consequences of a publish-or-perish culture: A worldwide survey. *Journal of the American Society for Information Science and Technology*, *63*(7), 1282–1293. https://doi.org/10.1002/asi.22636

van Eck, N. J., & Waltman, L. (2014). Visualizing Bibliometric Networks. In *Measuring Scholarly Impact* (pp. 285–320). Springer International Publishing. https://doi.org/10.1007/978-3-319-10377-8_13

Van Noorden, R. (2015). Interdisciplinary research by the numbers. *Nature*, *525*(7569), 306–307. https://doi.org/10.1038/525306a

Van Raan, A. F. J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics*, *36*(3), 397–420. https://doi.org/10.1007/BF02129602

Van Raan, A. F. J. (2005). *Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods Introduction: Crucial problems in bibliometric analysis* (Budapest Scientometrics, and Springer, Vol. 62, Issue 1, pp. 133–143).

Van Raan, A. F. J. (2007). Challenges in Ranking of Universities. *The World-Class University and Ranking: Aiming beyond Status, Higher Education for a Knowledge Society*, 87–121.

Van Raan, A. F. J., Van Leeuwen, T. N., & Visser, M. S. (2011). Severe language effect in university rankings: Particularly Germany and France are wronged in citation-based rankings. *Scientometrics*, *88*(2), 495–498. https://doi.org/10.1007/s11192-011-0382-1

Van Vught, F. (2013). U-Multirank: A user-driven and multi-dimensional ranking tool in global higher education and research. In P. T. M. Marope, P. J. Wells, & E. Hazelkorn (Eds.), *Rankings and Accountability in Higher Education: Uses and Misuses* (pp. 257–280). UNESCO Publishing.

Van Vught, F., & Ziegele, F. (2011). *Design and Testing the Feasibility of a Multidimensional Global University Ranking Final Report*.

Viana Lora, A., & Nel-lo Andreu, M. G. (2020). Alternative Metrics for Assessing the Social Impact of Tourism Research. *Sustainability*, *12*(10), 4299. https://doi.org/10.3390/su12104299

Vieira, E. S., & Gomes, J. A. N. F. (2010). A research impact indicator for institutions. *Journal of Informetrics*, *4*(4), 581–590. https://doi.org/10.1016/j.joi.2010.06.006

Vieira, V. D. F., Ferreira, C. H. G., Almeida, J. M., Moreira, E., Laender, A. H. F., Meira, W., & Gonçalves, M. A. (2024). A network-driven study of hyperprolific authors in computer science. *Scientometrics*, *129*(4), 2255–2283. https://doi.org/10.1007/s11192-024-04940-5

Vinkler, P. (2010). *The Evaluation of Research by Scientometric Indicators* (p. 313). Elsevier Ltd. https://doi.org/10.1533/9781780630250

Von Bohlen und Halbach, O. (2011). How to judge a book by its cover? How useful are bibliometric indices for the evaluation of " scientific quality" or " scientific productivity"? *Annals of Anatomy*, *193*(3), 191–196. https://doi.org/10.1016/j.aanat.2011.03.011

Wainer, J., Billa, C., & Goldenstein, S. (2011). Invisible work in standard bibliometric evaluation of computer science. *Communications of the ACM*, *54*(5), 141–148. https://doi.org/10.1145/1941487.1941517

Wainer, J., & Vieira, P. (2013). Correlations between bibliometrics and peer evaluation for all disciplines: The evaluation of Brazilian scientists. *Scientometrics*, *96*(2), 395–410. https://doi.org/10.1007/s11192-013-0969-9

Waltman, L., & Van Eck, N. J. (2013). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics*, *96*(3), 699–716. https://doi.org/10.1007/s11192-012-0913-4

Waltman, L., Wouters, P., & van Eck, N. J. (2017, May 17). *Ten principles for the responsible use of university rankings*. CWTS Meaningful Metrics Blog. https://www.cwts.nl/blog?article=n-r2q274

Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, *94*(3), 851–872. https://doi.org/10.1007/s11192-012-0775-9

Wang, J., Thijs, B., & Glänzel, W. (2015). Interdisciplinarity and Impact: Distinct Effects of Variety, Balance, and Disparity. *PLOS ONE*, *10*(5), e0127298–e0127298. https://doi.org/10.1371/journal.pone.0127298

White, S. W., Xia, M., & Edwards, G. (2021). Race, gender, and scholarly impact: Disparities for women and faculty of color in clinical psychology. *Journal of Clinical Psychology*, *77*(1), 78–89. https://doi.org/10.1002/JCLP.23029

Williams, B. (2022). Black Feminist Citational Praxis and Disciplinary Belonging. *Cultural Anthropology*, *37*(2), Article 2. https://doi.org/10.14506/ca37.2.04

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & 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

Wouters, P. (2014). The citation: From culture to infrastructure. In B. Cronin & C. R. Sugimoto, *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact* (pp. 47–66). MIT press.

Wu, C. (2023). The gender citation gap: Why and how it matters. *Canadian Review of Sociology/Revue Canadienne de Sociologie*, *60*(2), 188–211. https://doi.org/10.1111/cars.12428

Wu, C. (2024). The gender citation gap: Approaches, explanations, and implications. *Sociology Compass*, *18*(2), e13189. https://doi.org/10.1111/soc4.13189

Yang, S., & Zheng, M. (2019). Performance of citations and altmetrics in the social sciences and humanities. *Proceedings of the Association for Information Science and Technology*, *56*(1), 326–335. https://doi.org/10.1002/pra2.69

Zhang, Y., Wang, Y., Du, H., & Havlin, S. (2024). Delayed citation impact of interdisciplinary research. *Journal of Informetrics*, *18*(1), 101468. https://doi.org/10.1016/j.joi.2023.101468

Zhao, X., Akbaritabar, A., Kashyap, R., & Zagheni, E. (2023). A gender perspective on the global migration of scholars. *Proceedings of the National Academy of Sciences*, *120*(10), e2214664120. https://doi.org/10.1073/pnas.2214664120

Zurn, P., Bassett, D. S., & Rust, N. C. (2020). The Citation Diversity Statement: A Practice of Transparency, A Way of Life. *Trends in Cognitive Sciences*, *24*(9), 669–672. https://doi.org/10.1016/j.tics.2020.06.009