Lemke, Steffen

Doctoral Thesis — Published Version

An Assessment of Impact Metrics' Potential as Research Indicators Based on Their Perception, Usage, and Dependencies from External Science Communication

An Assessment of Impact Metrics' Potential as Research Indicators Based on Their Perception, Usage, and Dependencies from External Science Communication

Dissertation

zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften
(Dr.-Ing.)

an der Technischen Fakultät
der Christian-Albrechts-Universität zu Kiel

Steffen Lemke

Kiel, 2022

Erste Gutachterin: Prof. Dr. Isabella Peters
Zweiter Gutachter: Prof. Dr. Klaus Tochtermann
Datum der mündlichen Prüfung: 13.12.2022

Acknowledgements

First, I want to express my gratitude to my supervisors and doctoral committee for their guidance, their advice, and their time.

A heartfelt thank-you is due to Prof. Dr. Isabella Peters - I am deeply grateful for the outstanding support I received from her, for the innumerable amount of insights I learned from her about (information) science and about what it means to be a researcher, and for creating (and making me a part of) a working environment in which I always felt surrounded by such cooperative, enriching, appreciative, and friendly colleagues.

Furthermore, I wish to thank Prof. Dr. Klaus Tochtermann for his support regarding this thesis - as well as for allowing me to also dabble in teaching during my time at ZBW, which proved to be an illuminating and highly satisfying experience for me.

I also wish to thank my other colleagues at the ZBW Leibniz Information Centre for Economics - I feel truly privileged to have been able to undertake this endeavor at a place where I always felt recognized and supported and - as mentioned above - surrounded by such pleasant coworkers. Special thanks go out to Dr. Athanasios Mazarakis, Max Brede, Paula Bräuer, Dr. Kaltrina Nuredini, Dr. Nicholas Fraser, Anke Koslowski, Maryam Mehrazar, Dr. Hadas Shema, Oliver Hahn, Julian Sakmann, Anne Witthake; all of whom I had the pleasure of working more closely with over the years; as well as to Fidan, Till, Doreen, Ahmed, Lukas, Lennart, Maria, Kristin, Guido, Atif, Mohammad, Arben, Iacopo, and all the other colleagues at ZBW who I might not have worked *that* directly with, but who all contributed to making the office such a welcoming place.

I also want to thank my colleagues from the **metrics* project, the *MeWiKo* project, and the *QuaMedFo* project, as well as the other colleagues I had the joy of collaborating or co-authoring publications with.

Also, I thank my dear friends from around Kiel, Hamburg, Lübeck, and Hannover - thanks to our gatherings, trips, and shenanigans I never even had to worry about ever being mentally overwhelmed by the more stressful aspects of working towards a PhD.

And finally, I thank my family - more specifically my two brothers and my dear parents, whose support, love, and presence always provided the foundation for me from which to accomplish anything at all.

Abstract

The demand for practicable methods for quantitative assessments of scientific products' relevance has risen considerably over the past decades. Reasons for this are, among others, the rising diversity of research fields that are getting more and more specialized, continuously growing publication rates, as well as various stakeholders' increasing information needs regarding the added value of research for society. Especially this latter aspect recently has been cause for research to - next to traditional, citation-based indicators of scientific relevance - increasingly also consider alternative metrics (so called *altmetrics*) as impact indicators. This highly heterogeneous family of indicators is based on the principle of measuring interactions with scientific publications that are observable online, and covers for instance mentions of publications in social and journalistic media, in literature management software, or in policy documents. Continuous discussions on the theoretical validity of academic citations as indicators for scientific relevance have led to the development of (still frequently debated) theories of citing; altmetrics, on the other hand, almost completely lack such theoretical foundation. Questions regarding what it is that different metrics measure or express therefore in many facets remain unanswered.

This thesis makes two central contributions towards answering these questions. Its first part systematically assesses the status quo of various metrics' perception and usage by researchers, which constitute the user group most directly concerned by the use of research metrics. This assessment serves to determine the significance of metrics in academic daily routines, as well as to identify relevant perceived problems concerning their usage. The individual challenges identified this way will in later sections of the thesis be opposed with concrete measures that should be taken during the development of future research metrics and their infrastructure to effectively solve common criticisms that exist regarding current metrics and their use.

Proceeding from the first part's user studies, this thesis' second part examines the relationship between research metrics and external science communication. It this way addresses a wide research gap with considerable potential implications for metrics' validity as indicators for quality - the question to which degree these metrics are merely the result of promotion, which respective research publications receive. More particularly, the thesis investigates on the associations between various metrics' manifestations and promotion of research publications within two formats that for long have been important tools of external science communication: press releases and embargo e-mails.

To achieve its goals, the thesis applies a mixed-methods approach utilizing a combination of qualitative interviews, large-scale online surveys, interactive online-experiments analyzed via conjoint analysis, bibliometric and altmetric analysis of extensive datasets of publications, embargo e-mails and press releases, as well as confirmatory path analysis.

The user studies on researchers' perception and usage of research metrics reveal an ambivalent relationship between the two. On the one hand, in particular citation-based indicators are shown to frequently play a part in many researchers' daily routines, for instance in use cases of literature research or for the identification of suitable publication venues. On the other hand do the studies' participants voice a broad array of concerns regarding the used metrics' validity. Especially a limited understanding of various metrics' methodologies as well as a perceived general opacity of indicators stand out as

common problem areas. In the comparison of different metric families with each other, altmetrics perform considerably worse than bibliometrics regarding their perception and usage by researchers. The assessment of the status quo unveils two primary challenges concerning the future endeavor of improving the supply as well as the utilization of research metrics: the necessary realization of more far-reaching offerings of education about and information on research metrics, especially for early-career researchers, as well as the implementation of open and transparent technical infrastructures for the elicitation of indicators.

The analyses of the associations between research articles' promotion within embargo e-mails and press releases demonstrate substantial correlations between the two instruments of external science communication and the citations as well as altmetrics of respective promoted articles. Concerning the representation of research within these two formats, a strong focus on life sciences and multidisciplinary high-impact journals becomes apparent. While quantifications of direct effects of external promotion on research metrics have to be interpreted with great care due to the complexity of the system of science communication, a variety of findings does suggest that the publicity generated by the two examined promotional formats may substantially influence respective research metrics. These results underline the significant need of further research as well as critical reflection on the interplay between the medial and the academic sphere in view of research metrics and their adequate interpretation.

Through its elaborations on the two research gaps addressed within this thesis, it contributes to a more differentiated picture of the meaning and maturity of research metrics and especially altmetrics. This way, the thesis provides empirical substance that is urgently needed in theory building, which itself is a requirement for a reasonable and responsible use of metrics for the assessment of research. The thesis' findings inform several stakeholder groups, among them developers of software for the collection, processing and provision of research metrics; suppliers of scientific publications like libraries or scholarly publishers, which want to provide their users with informative research metrics; research administrations as users of metrics in evaluative contexts; as well as the scientometric research community, which seeks a profound understanding of research metrics including their potentials and limitations.

Zusammenfassung

Die Nachfrage nach praktikablen Methoden zur quantitativen Bestimmung der Relevanz wissenschaftlicher Produkte ist in den vergangenen Jahrzehnten erheblich gestiegen. Gründe dafür liegen unter anderem in der gestiegenen Vielfalt sich immer weiter ausdifferenzierender Fachgebiete, dem stetig wachsenden Publikationsaufkommen und einem vermehrten Informationsbedürfnis verschiedener Interessengruppen bezüglich der durch wissenschaftliche Anstrengungen erzielten Gegenwerte für die Gesellschaft. Insbesondere letzterer Aspekt hat einen erheblichen Anteil daran, dass neben den traditionell verwendeten, auf akademischen Zitationen basierenden Indikatoren wissenschaftlicher Relevanz in jüngerer Vergangenheit auch alternative Metriken (sogenannte *Altmetrics*) vermehrt in den Blickpunkt der Forschung gerieten. Diese hoch heterogene Familie von Indikatoren basiert auf dem Prinzip online erfassbare Interaktionen mit wissenschaftlichen Produkten zu messen und deckt beispielsweise Erwähnungen von Publikationen in sozialen oder journalistischen Medien, in Literaturverwaltungsprogrammen, oder in Policy-Dokumenten ab. Kontinuierliche Diskussionen über die theoretische Validität akademischer Zitationen als Indikatoren wissenschaftlicher Relevanz führten zur Ausbildung von (weiterhin kritisch diskutierten) Theorien des Zitierens; Altmetrics dagegen mangelt es an einer derartigen theoretischen Grundlage nahezu gänzlich. Fragen danach, was verschiedene Metriken messen und auszudrücken vermögen, bleiben in vielen Facetten unbeantwortet.

Diese Arbeit leistet zwei zentrale Beiträge dazu, die Beantwortung dieser Fragen voranzubringen. Der erste Teil dieser Arbeit gilt der systematischen Erhebung des Status Quo der Wahrnehmung und Nutzung verschiedener Metriken durch Forschende, welche die am unmittelbarsten durch Forschungsmetriken betroffene Gruppe darstellen. Diese Erhebung dient der Bestimmung der Bedeutung von Metriken zur Relevanzbestimmung im akademischen Alltag, sowie der Identifikation wesentlicher in Bezug auf ihre Nutzung wahrgenommener Probleme. Die einzelnen so identifizierten Herausforderungen werden im späteren Teil der Arbeit konkreten Maßnahmen gegenübergestellt, die bei der Entwicklung zukünftiger Indikatoren und ihrer Infrastruktur ergriffen werden sollten, um derzeitige Metriken und ihre Nutzung betreffende Kritikpunkte effektiv auszuräumen.

Ausgehend von den Nutzungsstudien des ersten Teils beleuchtet der zweite Teil dieser Arbeit die Beziehung zwischen Forschungsmetriken und externer Wissenschaftskommunikation. Er nimmt sich damit einer weiten Forschungslücke mit erheblichen potenziellen Implikationen für die Validität der Metriken als Qualitätsindikatoren an - der Frage, inwiefern diese Metriken lediglich ein Ergebnis der Promotion sind, welche die betreffenden Publikationen erfahren. Im Fokus steht dabei die Ermittlung der Zusammenhänge zwischen den Ausprägungen verschiedener Metriken und der Promotion wissenschaftlicher Publikationen in zwei Formaten, die seit langem bedeutende Instrumente externer Wissenschaftskommunikation darstellen: Pressemitteilungen sowie Embargo E-Mails.

Zur Erreichung ihrer Ziele stützt sich diese Arbeit auf einen Mixed-Methods-Ansatz unter Verwendung qualitativer Interviews, großangelegter Online-Umfragen, mittels Conjoint-Analyse ausgewerteter interaktiver Online-Experimente, biblio- und altmetrischer Analyse umfangreicher Publikations-, Embargo E-Mails- und Pressemitteilungsdatensätze, sowie konfirmatorischer Pfadanalyse.

Die Nutzerstudien der Wahrnehmung und Nutzung von Forschungsmetriken durch Forschende offenbaren ein zwiespältiges Verhältnis letzterer zu den Metriken. So sind insbesondere zitationsbasierte Indikatoren zwar regelmäßiger Bestandteil des Forschungsalltags vieler Befragter, welche in diversen Nutzungsszenarien, wie beispielsweise bei der Literaturrecherche oder der Identifikation geeigneter Publikationsorte, von diesen Gebrauch machen. Auf der anderen Seite äußern die Teilnehmenden ein breites Spektrum an Bedenken gegenüber der Validität der genutzten Metriken. Insbesondere stechen Unkenntnis bezüglich ihrer Methodiken als auch eine empfundene allgemeine Intransparenz der Metriken als Problemfelder heraus. Im Vergleich verschiedener Familien von Metriken schneiden Altmetrics gegenüber zitationsbasierten Indikatoren erheblich schlechter hinsichtlich ihrer Wahrnehmung und Nutzung ab. Als primäre Herausforderungen im Bestreben, das Angebot von und den Umgang Forschender mit zukünftigen Forschungsmetriken zu verbessern, ergeben sich aus dem erhobenen Status Quo die notwendige Realisierung weitreichender Lehr- und Informationsangebote zum Themenkomplex Metriken insbesondere für Nachwuchswissenschaftler:innen, sowie das Schaffen offener und transparenter technischer Infrastrukturen zur Erhebung von Indikatoren.

Die Analyse der Zusammenhänge zwischen Promotion wissenschaftlicher Artikel in Embargo E-Mails und Pressemitteilungen demonstriert substanzielle Korrelationen zwischen beiden Maßnahmen externer Wissenschaftskommunikation und den Zitationen wie auch Altmetrics der beworbenen Artikel, während sich hinsichtlich der in diesen Formaten repräsentierten Forschung ein starker Schwerpunkt auf Lebenswissenschaften und multidisziplinären High-Impact-Journalen zeigt. Während Quantifizierungen direkter Effekte externer Promotionmaßnahmen auf Forschungsmetriken aufgrund der Komplexität des Systems wissenschaftlicher Kommunikation mit großer Vorsicht interpretiert werden sollten, deutet eine Vielzahl von Befunden darauf hin, dass sich die aus den untersuchten Formaten resultierende Sichtbarkeit erheblich in den Forschungsmetriken niederschlägt. Diese Ergebnisse unterstreichen den erheblichen bestehenden Forschungs- und Diskussionsbedarf hinsichtlich der Wechselwirkungen zwischen medialer und akademischer Sphäre in Hinblick auf Forschungsmetriken und deren adäquate Interpretation.

Durch die Adressierung der beiden in dieser Arbeit bearbeiteten Forschungslücken trägt sie zu einem differenzierten Bild der Bedeutung und des Reifegrades insbesondere neuartiger Forschungsmetriken bei. Sie liefert so empirische Substanz, die bei der für einen nachvollziehbaren und verantwortungsbewussten Umgang mit Metriken erforderlichen Theoriebildung und Diskussion dringend benötigt wird. Verschiedene Interessenten ziehen aus den Erkenntnissen dieser Arbeit Nutzen, darunter Entwickler von Software zur Erhebung, Verarbeitung oder Darstellung von Forschungsmetriken; Anbieter wissenschaftlicher Publikationen wie Bibliotheken oder Verlage, welche ihren Nutzer:innen informative Forschungsmetriken zur Verfügung stellen wollen; wissenschaftliche Administrationen als Anwender:innen von Forschungsmetriken in evaluativen Kontexten; sowie die szientometrische Forschungscommunity, die ein fundiertes Verständnis von Forschungsmetriken inklusive ihrer Potenziale und Limitationen anstrebt.

Contents

List of Figures

List of Tables

Chapter 1: Introduction

The need for scalable methods for the assessment of individual research publications' relevance and value is no novel demand. Among the first to explicitly acknowledge this need were Gross & Gross (1927), who proposed to assess academic journals' usefulness by counting the number of references to them, which should result in a metric meant to help libraries decide which periodicals to provide to their clients. This now almost one hundred years old suggestion foreshadowed what would become one of the core principles of bibliometrics, i.e. citation analysis. Substantially propelled by the development and publication of large citation indices in the 1960s (Garfield & Sher, 1963; Garfield, 1972), tracking citations still is the foundation of many common forms of quantitative research assessment up to this day. This assumption, that the number of times a research publication has been cited provides an indicator for its importance, nowadays constitutes the rationale behind much-used instruments for evaluating the scientific relevance not just of individual articles, but also of academic journals (e.g., in the form of *Clarivate's* journal impact factor) or individuals (e.g., in the form of the h-index; Hirsch, 2005).

The caveats, limitations, and undesired ramifications of focusing research evaluation on citation indicators are a permanent issue of discussion in the field of scientometrics and beyond. Many biases of such indicators have been identified, their methodological weaknesses have been debated, and problematic implications for the reward system of science have been discussed (Rijcke, Wouters, Rushforth, Franssen, & Hammarfelt, 2016). More recently, such discussions and analyses culminated in the publication of several high-profile manifestos advising on how to prevent misuse of respective indicators, e.g., the Leiden Manifesto for research metrics (Hicks, Wouters, Waltman, de Rijcke, & Rafols, 2015), the San Francisco Declaration on Research Assessment (Cagan, 2013), or the Metric Tide report (Wilsdon et al., 2015).

The awareness of the flaws of a research evaluation culture that places such strong emphasis on a single indicator motivates a constant search for additional scalable possibilities of quantitative research assessment. Endeavors to come up with new metrics are driven by the prospect that these might circumvent certain known biases of citation counts (Wouters & Costas, 2012), as well as by the general intention to enable the academic reward system to recognize a more diverse set of facets of scientific work.

A multitude of new opportunities to measure the use of individual research publications arose with the steady increase of their electronic reproduction since the 1980s. With the establishment of the first online document repositories in the academic sector it became possible to count new types of interactions with research publications, e.g., numbers of downloads or page views. Another immense catalyst for the emergence of new potential metrics was the rise of the Social Web starting at the beginning of this millennium. The advent of numerous new participative online platforms - and their adoption by researchers for their own work-related purposes (Kramer & Bosman, 2016) - brought with it a sheer incalculable and ever growing number of opportunities to measure how academic publications are received and interacted with. References to and usages of individual articles could now for instance be counted in blogs, online news outlets, social networks, forums, literature management platforms, and many, many more. The general idea of utilizing traces on the web as indicators of scientific productivity had already been discussed for some time - Cronin, Snyder, Rosenbaum, Martinson, & Callahan (1998) for instance had proposed the concept of "polymorphous mentions" to describe such metrics. However, research in the field gained new momentum in 2010 when Priem et al. coined (the loosely defined) 'altmetrics' as an umbrella term for the highly heterogeneous variety of traces of online interactions with research (Priem, Taraborelli, Groth, & Neylon, 2010; Sugimoto, Work, Larivière, & Haustein, 2017).

Although since then research on altmetrics has grown to be an ever-present facet in scientometric literature and on respective conferences, the field still suffers from a severe lack of common definitions, clear terminology and conceptual frameworks (Glänzel & Gorraiz, 2015; Haustein, 2016). Moreover, while visualizations of research articles' altmetric impact have become a common sight on many publishers' article landing pages (a particularly prominent example of such visualizations being the Altmetric badge or 'donut'[1], which is provided by *Digital Science's* Altmetric.com), little is known about how users perceive, understand, interpret, and use such information, and how altmetrics compare to traditional bibliometric indicators in these regards. Learning about these perceptions and identifying common concerns regarding the underlying data and methodologies is an important prerequisite for the development and implementation of impact indicators that meet and counter justified points of criticism, and that thus have the potential to be both informative and widely accepted.

This thesis aims to make two primary contributions to the assessment of different metrics' potential as research indicators: in its first part, the thesis provides a thorough overview over the current status quo of the perception and usage of metrics (Chapter 2). Although there is a variety of stakeholder groups that may interact regularly with research metrics (e.g., researchers, research administration, funding agencies, librarians, etc.), the focus of this thesis is on the researchers, who constitute the user group that is arguably most directly affected by the use of impact indicators. The investigation of the state of usage and perception of metrics from the viewpoint of researchers shall help to identify particularly pressing issues that future impact metrics need to resolve to be able to meet researchers' demands; also, it shall inform an assessment of the maturity of the less established types of metrics. The perceptions and concerns detected in Chapter 2 will be consolidated with further challenges for metrics that have been determined by the scientometric literature, alongside recommendations of how to mitigate them.

The second part of the thesis, Chapter 3, contrasts these inquiries about the ways metrics are perceived with what scientometric research can actually tell us about how research metrics are affected by or

[1] https://www.altmetric.com/about-our-data/the-donut-and-score/

associated with various aspects of external science communication. Proceeding from common concerns and criticisms regarding the use of research metrics encountered in Chapter 2, Chapter 3 of the thesis systematically investigates the empirical evidence on one frequent but under-analyzed point of criticism regularly brought up against the adaptation of such indicators - namely, the suspicion that they to a large extent merely reflect effects of ‘marketing efforts’, i.e., are to a substantial degree the result of various means of external science communication. The concluding Chapter 4 shall then integrate the findings presented in the previous chapters, with the goals of providing concrete and precise recommendations for how to improve future implementations of research metrics and highlight open questions that are yet to be solved, to facilitate the identification of particularly pressing demands for action regarding metrics research and development.

With regard to different types of metrics, this thesis places a particular focus on altmetrics, as these represent a family of novel and in several ways (which will be discussed in detail below) promising complements to traditional bibliometric indicators that have not nearly received the amount of attention and scrutiny in research that citation-based indicators have. Nevertheless, the studies and arguments discussed in this thesis will for the most part also consider citations and their derivatives, as these constitute the type of metrics that is used most commonly in research assessment exercises and thus embodies the baseline altmetrics will have to be compared to.

Within chapters 2 and 3, four distinct studies are presented. Every study is preceded by a foreword to briefly describe the context of its original publication. The studies themselves, although being integrated into the logical structure of the thesis, are self-contained and thus can also be read independently from the overarching narrative of this manuscript.

From a methodological perspective, the studies presented within the thesis utilize a variety of methods stemming from diverse fields of science, ranging from the social sciences over economics, statistics, to computer science. While the first two studies contain substantial qualitative elements, over the course of the thesis quantitative perspectives move more and more to the center. Study A relies on user studies in the form of focus group interviews as well as online surveys to inquire about researchers’ perceptions of impact metrics and the platforms they are derived from. Study B expands on Study A’s findings and utilizes a self-developed online experiment, the results of which are analyzed with methods of conjoint analysis. Study C features the combination and statistical analyses of data from journalistic, altmetric, and bibliometric sources to investigate relationships between promotion of research in external science communication and respective research’s impact. Finally, Study D expands on the concepts from Study C while analyzing a larger sample, which covers a more diverse set of data sources, with means of path analysis. For the field of scientometrics, in which all four studies are located, the application of both conjoint analysis as well as path analysis represented methodical novelties (at least at the respective studies’ times of publication).

The insights presented in this thesis address a topic of major relevance for libraries and other information infrastructures, who as institutions that gather, administer, and distribute research in many forms naturally occupy a key role in the conceptualization and realization of research assessment exercises. Further stakeholder groups addressed by this thesis for instance come from the domains of metrics-related software development and -supply (e.g., retailers of aggregated metrics databases or other software based on research metrics), of research administration (e.g., university administrations responsible for the conduction of assessment exercises that might make use of metrics), or of policy making (e.g., institutional bodies issuing guidelines that determine the role of metrics in research evaluations of public interest). All of these stakeholders should base their decisions regarding the handling of research metrics on the empirical foundation provided by scientometric research, which is the foundation that this thesis contributes to.

While the thesis therefore has its scientometric roots within the domain of library and information sciences, it both methodologically and regarding the research interests it addresses lies in an intersection of information science, computer science and social science. The complex melange of agents, environments, and dependencies that the system of scholarly and science communication is composed of makes such mixed-methods approaches as taken in this thesis sensible.

Especially the subtopic of altmetrics does to an even greater extent than traditional bibliometrics allude to several aspects from the domain of computer science. Consulting for instance the *Association for Computing Machinery's Computing Classification System (ACM CSS)* as a framework defining the manifold areas of computer science, we find multiple concepts the matters addressed in this thesis touch upon. The perhaps most obvious example would be “Social media”, which as one of the major types of sources for altmetric data play a crucial role in most case studies involving altmetrics, as will also be seen in the later chapters. Other ACM CSS concepts with connections to the themes of this thesis include “Social content sharing”, “Collaborative content creation”, “Social networking sites”, “Social network analysis”, “Collaborative and social computing systems and tools”, “Web-based interaction”, “User studies”, “Link and co-citation analysis”, “Digital libraries and archives”, or “Publishing”. Alternatively, without sticking to the ACM CSS's vocabulary, one can locate this thesis' contributions within the emerging subfield of *Big Scholarly Data* (Xia, Wang, Bekele, & Liu, 2017). Big scholarly data (science) is a primarily data-driven field that is in need of a more robust theoretical foundation - a foundation that this thesis contributes to by providing empirical results. From the perspective of business informatics, on the other hand, in its endeavor of examining how certain technical systems (i.e., systems that collect, process, and provide research metrics) and socio-organizational systems (i.e., the systems of internal and external science communication) affect each other, this thesis connects to the subfield of *Computer-supported cooperative work.*[2]

To summarize its primary mission in short, the thesis' findings and conclusions shall support the development of improved tools for research evaluation, which make better use of the vast opportunities for impact measurement that today's landscape of scholarly communication offers, while they at the

[2] See also https://fg-cscw.gi.de/fileadmin/FG/CSCW/dokumente/Flyer_CSCW.pdf (link in German).

same time have to consider the concerns and problems related to metrics-based impact assessments that the studies incorporated in this manuscript - alongside the existing body of scientometric research - have revealed. Although the assessment of research objects is the metrics use case that this thesis focuses on, its findings are also relevant for other areas in which metrics or citation analysis come to use, e.g., in certain cases of recommender systems, information retrieval, or predictive analytics (see also National Information Standards Organization (NISO) [2016] for an overview over metrics use cases).

The remainder of this chapter will introduce and clarify core concepts of which a degree of understanding is necessary (or at least helpful) to be able to follow the thesis' later chapters. After a brief discussion of the varying existing definitions and taxonomies of altmetrics (Section 1.1), prevalent drivers and motives behind their adaptation (Section 1.2), use cases (Section 1.2.1), as well as known challenges and criticisms (Section 1.3) will be presented. Section 1.4 will then close this chapter by offering a brief explanation of the concepts of *science communication* and *scholarly communication,* as they will play an important role from Chapter 3 onwards.

1.1 Altmetrics - Terminologies and Taxonomies

The challenges yet to solve for altmetrics research start with their lack of a common definition (Haustein, 2016). The original altmetrics manifesto just introduces altmetrics as “filters” enabled by the growth of new online scholarly tools, whose usage would “reflect and transmit scholarly impact” (Priem et al., 2010). Although concrete platforms like Twitter, Mendeley, CiteULike, or Zotero are mentioned as exemplary sources to derive altmetrics from, within the manifesto its authors do not make an attempt to suggest a truly comprehensive definition. In his later work, Priem (2014, p. 266) defines altmetrics as the “study and use of scholarly impact measures based on activity in online tools and environments”. On a similarly abstract level, Moed (2016, p. 362) describes altmetrics as “traces of the computerization of the research process”. Another definition is proposed by the *National Information Standards Organisation (NISO)*, which defines altmetrics as “indicators [...] derived from activity and engagement among diverse stakeholders and scholarly outputs in the research ecosystem” (National Information Standards Organization (NISO), 2016, p. 1). Haustein (2016, p. 416) proposes a slightly less abstract view that explains altmetrics as a heterogeneous subset of *scholarly metrics* (Haustein, Sugimoto, & Larivière, 2015), which themselves are defined as “indicators based on recorded events of acts (e.g., viewing, reading, saving, diffusing, mentioning, citing, reusing, modifying) related to scholarly documents (e.g., papers, books, blog posts, datasets, code) or scholarly agents (e.g., researchers, universities, funders, journals)”. This way, Haustein (2016) places altmetrics within the taxonomy of metrics as a subset of scholarly metrics (Haustein, Sugimoto, et al., 2015), webometrics (Almind & Ingwersen, 1997), informetrics (Nacke, 1979), and scientometrics (see also Figure 1).

While all these definitions of altmetrics remain broad, they can incorporate the full diversity of data captured by altmetrics providers like *Altmetric.com*, *Plum Analytics* or *ImpactStory*, and account for the fact that the landscape of potential sources for altmetrics is ever-changing. In this thesis we will adapt the open definition from Haustein (2016) explained above, i.e., we use the term *altmetrics* to denote any type of indicator derived from recorded events of acts related to scholarly documents or agents that can be tracked online and are not based on academic citations (we specifically exclude the latter from our definition, as indicators based on academic citations in most contexts constitute the default that altmetrics are meant to complement). This includes any traceable kinds of interactions with scholarly documents in social media, but also other events that altmetric data providers harvest online, e.g., mentions of research in online news outlets, policy documents, syllabi, etc.. Thus, the sometimes-used term *social media metrics* (Haustein, Larivière, Thelwall, Amyot, & Peters, 2014) should not be considered a synonym, but merely a subset of altmetrics as the term is used in this work.

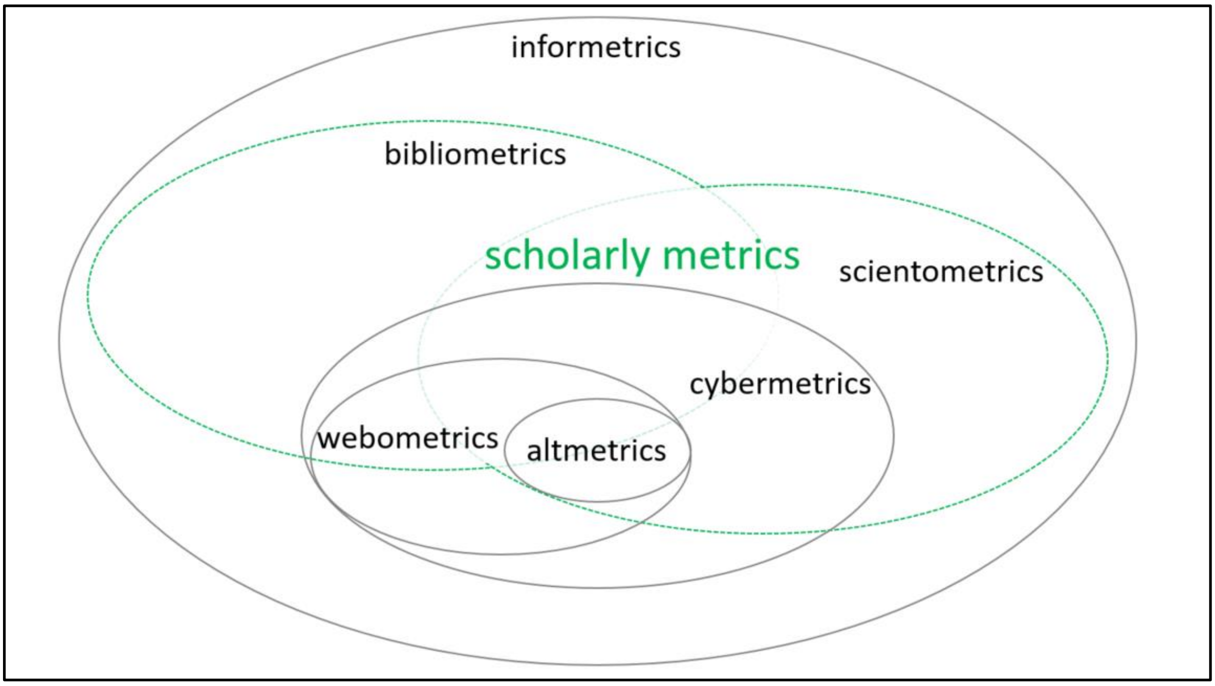


Figure 1: Positioning of altmetrics within similar concepts; sizes of ellipses are not representative of field sizes; adapted from Haustein (2016).

To conclude our brief terminological distinction, the juxtaposition of altmetrics and *usage metrics* deserves additional attention. The concept of usage metrics - in the context of scientometrics - has been around for a much longer time than altmetrics, early examples being libraries' loan statistics or user surveys conducted to track usage of physical monographs and journal issues (Glänzel & Gorraiz, 2015). Presently, the term usage metrics is most commonly associated with downloads and page views (although one could argue - just like with altmetrics - that these so-called 'usage metrics' reflect merely an *intention* of using a respective resource instead of its actual usage; Gorraiz, Gumpenberger, & Schlögl, 2014). While our aforementioned open definition of altmetrics would allow to consider downloads and page views as their subset (see also Haustein, Bowman, & Costas, 2016), we will still use the term *usage metrics* to specifically distinct downloads and page views - this differentiation between altmetrics and usage metrics is congruent with large parts of the literature, as well as with the altmetrics manifesto and the ontology of metrics applied by *PLOS Article Level Metrics* (Glänzel & Gorraiz, 2015; Lemke, Nuredini, & Peters, 2020; Lin & Fenner, 2013).

It should be noted that the term *altmetrics* itself - a portmanteau from *alternative* and *metrics* originally proposed in a tweet by Jason Priem in 2010 (Figure 2) - has also caused criticism, mainly due to the facts that (1) the status of something being *alternative* is very likely to change over time (Rousseau & Ye, 2013) and (2) many studies of altmetrics' worth and their correlations with academic citations conclude that altmetrics should be considered a complement and not an alternative to more traditional indicators (Cress, 2014; Haustein, Costas, & Larivière, 2015; Lemke, Zagovora, et al., 2020; Tahamtan & Bornmann, 2020). Nevertheless, 'altmetrics' still is the prevalent term in scientific literature and at conferences to denote new web-based measures of scholarly impact, which to this day no suggested alternative term succeeded to replace.

Figure 2: Priem's tweet that introduced the term 'altmetrics'.

The heterogeneity and multidimensionality of altmetrics has caused several attempts to classify existing types of altmetrics. In context of the development of the metrics aggregator service *ImpactStory*, Piwowar & Priem (2012) suggest an ontology that groups types of article-level metrics (and among them altmetrics) into five classes that correspond to different levels of engagement with respective research articles, the levels being *viewed*, *discussed*, *saved*, *recommended*, and *cited*, for which individual metrics are then differentiated between the audiences *scholars* and *public* (see Table 1). Lin & Fenner (2013) later refined and adapted this ontology to structure PLOS Article Level Metrics (see Figure 3). Drawing from these engagement-related views on different altmetrics, Haustein, Bowman, & Costas (2016) propose a framework to group the various *acts* that lead to events on which the metrics are based (Figure 4). The framework proposes three categories, which again correspond to varying degrees of engagement with a research object: *access*, *appraise*, and *apply*.

Table 1: Ontology of article-level metrics applied by ImpactStory, source: Piwowar & Priem (2012)

	scholars	**public**
recommended	citations by editorials, f1000	press article
cited	citations, full-text mentions	wikipedia mentions
saved	citeulike, mendeley	delicious
discussed	science blogs, journal comments	blogs, twitter, facebook, etc.
viewed	pdf downloads	html downloads

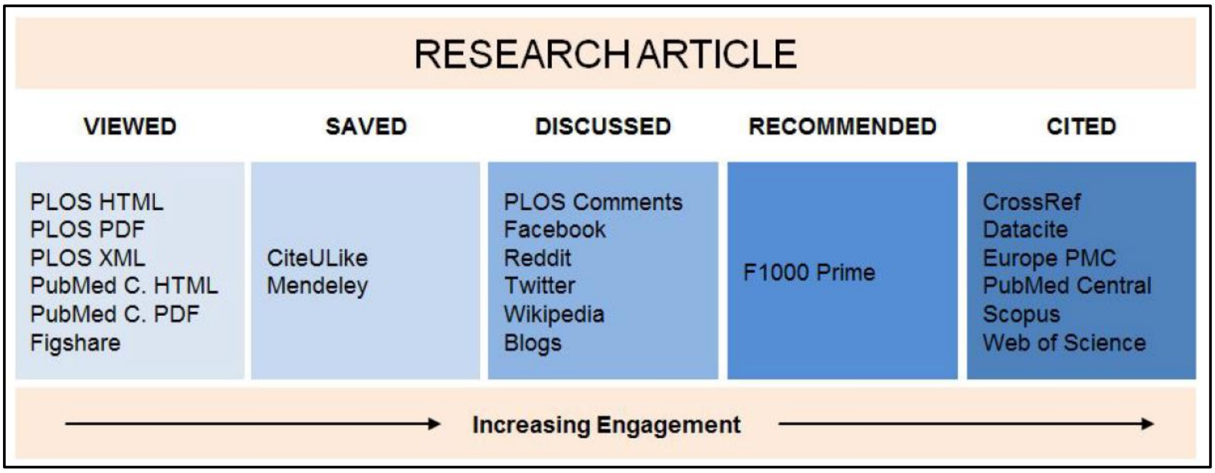


Figure 3: Taxonomy of PLOS ALM based on engagement with respective research articles; source: Lemke, Nuredini, & Peters (2020), based on Lin & Fenner (2013).

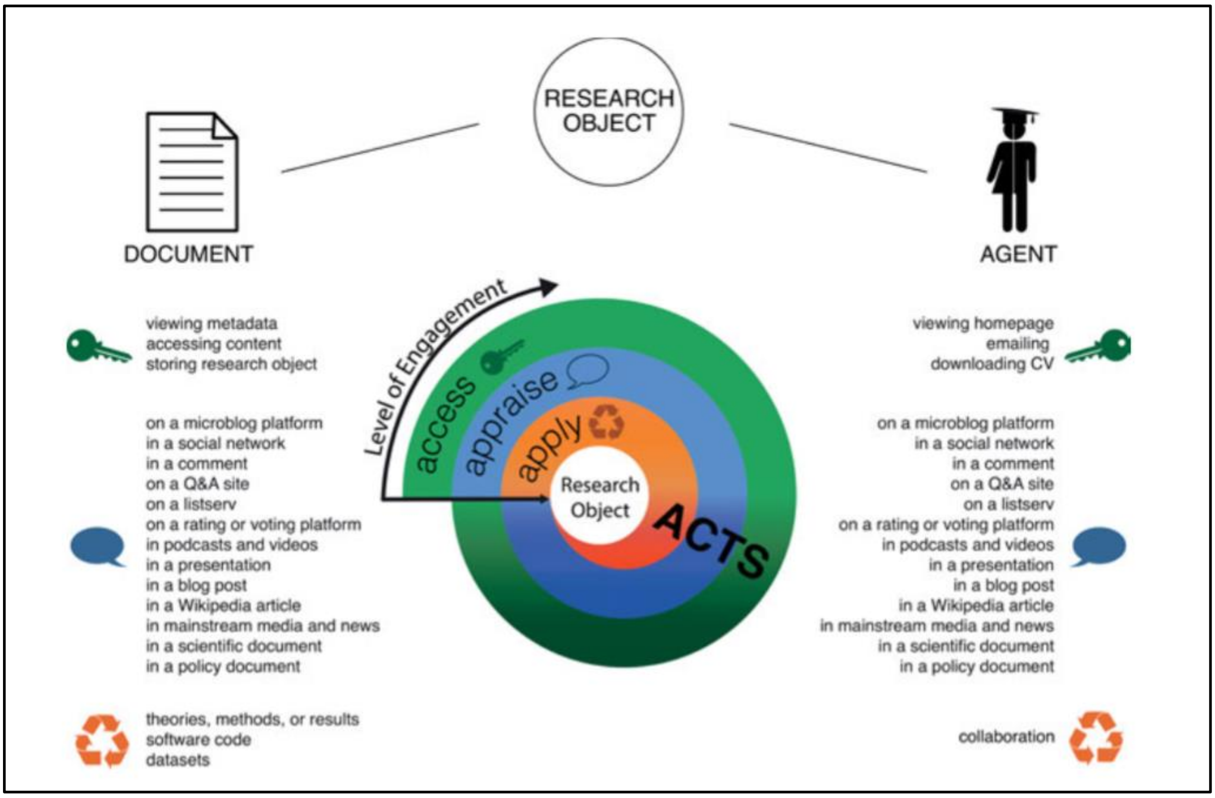


Figure 4: Framework of categories and types of acts referring to research objects (scholarly documents and agents); source: Haustein, Bowman, & Costas (2016).

These examples of classification systems for altmetrics and their limited congruency reflect how the field of altmetric research is still in search of a common, accepted definition of the concept. At the same time, the top-down nature that virtually all these approaches towards a classification of altmetrics take reflects a widely shared understanding that 'altmetrics' denote an ever-changing and continuously growing set of metrics, a fact which any attempt of a definition must account for via openness, extensibility and flexibility.

1.2 Altmetrics - Motives and Drivers

The accelerating interest in altmetrics as a research subject is driven by globally increasing expectations in researchers to demonstrate the value their research has for a broader audience (Moed, 2017; Nicholas et al., 2015; Wilsdon et al., 2015). Funding agencies, research administrations, and governments increasingly demand from scientists to document the societal impact of their research and how it aids in solving real world problems (Tahamtan & Bornmann, 2020), and novel opportunities of capturing research impact electronically might provide new means to do so (Bornmann, 2016). While citations are widely used as indicators for *scientific* impact - the influence of research within the academic sphere, that is - many recent research endeavors imply the hope that altmetrics might be capable of also expressing impact within different spheres, e.g., public, social, environmental, health-related, or economic spheres (Bar-Ilan et al., 2012; Bornmann & Haunschild, 2016a, 2017; Haustein, Larivière, et al., 2014; Konkiel, Madjarevic, & Rees, 2016; Zhang & Wang, 2021). Applying a different terminology, Hicks, Stahmer, & Smith (2018) distinguish between *inward*- and *outward-facing goals of scholarship.* The fulfillment of inward-facing goals - for instance the goal of producing research insights that are useful for the own scientific community - might be evaluable through traditional metrics such as citation counts. For the adequate assessment of outward-facing goals (i.e., the application of research to solve societal needs), however, developing new metrics might be necessary. To which degree altmetrics actually succeed in reflecting such diverse kinds of impact remains an unsolved and much-debated question (Tahamtan & Bornmann, 2020).

Other key arguments behind the interest in altmetrics are based on properties that are easier to prove. One of them is their availability for a more diverse set of types of scholarly outputs. Where academic citations are usually restricted to 'traditional' publication types indexed by the relevant bibliometric databases, altmetrics can also easily be collected for other research products like datasets, software, presentation slides, podcasts, videos, blog posts, etc.. This might also make altmetrics a particularly interesting prospect for disciplines in which research is not typically published in journals, but in books and which are therefore historically underrepresented in bibliometric databases, e.g., many fields from the humanities (Cronin & Sugimoto, 2015; Nederhof, 2006; Thelwall, 2017b). In this light, Kousha & Thelwall (2016) for instance assessed the utility of reviews on Amazon.com for the evaluation of books, while Zuccala, Verleysen, Cornacchia, & Engels (2015) analyzed the value of Goodreads reader ratings for impact evaluations in the humanities.

Moreover, the wider adoption of altmetrics for alternative publication types might in some cases help to encourage a more value-adding use of such types (Konkiel, 2013); e.g., in a scientific reward system that appreciates citations of datasets, researchers would be encouraged to spend more effort and care on making their data truly reusable, which would also help to boost other beneficial initiatives like the FAIR data movement (Wilkinson et al., 2016).

Another characteristic of many altmetrics that sets them apart from academic citations is their accumulation speed (Priem et al., 2010; Wouters & Costas, 2012). While the lion's share of an article's citations typically takes years to accumulate, most of its social media buzz tends to happen close to its publication (Zhang & Wang, 2021). This temporal difference also makes altmetrics that tend to correlate substantially with citations, like for instance Mendeley readerships, interesting tools for early predictions of a publication's estimated citation impact (Thelwall, 2017a, 2018b).
The comparative speed of online assessments of research is also one major driver behind the development of post-peer-review sites like F1000Prime or thirdreviewer.com, which enable a type of micro-level research assessment similar to traditional peer review, but with less overhead and structural limitations than the peer review at a journal or conference would usually bring with it (see also Mandavilli, 2011). For publication assessment, post-peer-review platforms might therefore represent an interesting middle-ground between the robustness of traditional peer-review and the flexibility and speed of altmetrics.
A further argument frequently voiced in favor of altmetrics is the openness of the respective raw data, compared to citations. While citation analysis typically relies on commercial databases like Web of Science or Scopus, which need to be licensed for considerable costs to be accessed, many altmetrics are based on publicly available APIs of online platforms. Principally this would make altmetrics more transparent and easier to reproduce than citations, although it should be noted that recent initiatives for the construction of open access publication graphs like *I4OC*[3] or *OpenAlex*[4] might soon diminish this advantage.

1.2.1 Exemplary Use Cases for Altmetrics

Research on altmetrics identified a number of different use cases for altmetrics that are more differentiated than the rather vague and abstract purpose of 'research assessment'. In their recommended practice on alternative assessment metrics, the NISO lists 38 individual use cases from the perspectives of 8 stakeholder personas (National Information Standards Organization (NISO), 2016). These use cases apply to the stakeholder groups of librarians, research administrators, hiring committees, funding agencies, academics/researchers, publishers/editors, media officers/journalists, and content platform providers and are grouped into three main themes: *showcase achievements, research evaluation*, and *discovery*. The theme *showcase achievements* refers to use cases that indicate a stakeholder's interest in highlighting positive achievements garnered by one or more scholarly outputs - e.g., from the perspective of librarians, altmetrics might be a way to add value to an existing institutional repository by enabling the creation of reports that showcase the frequencies of views and downloads of deposited works, which might in turn encourage researchers to deposit their works. An

[3] https://i4oc.org/
[4] https://openalex.org/

example for a use case from the theme *research evaluation* comes from the perspective of research administrators, for whom altmetrics might constitute another tool for benchmarking the performance and achievements of departments within a respective institution. A straightforward example for the theme *discovery* on the other hand refers to the persona of a hiring committee, which might make use of altmetrics to identify new talent whom it may want to recruit.

Zahedi (2018, pp. 11-14) arrives at a different categorization of use cases for altmetrics, which comprises four categories or "main venues of application":

1. *As indicators of presence and reception of research*; e.g., to compare the research produced by various units of analysis like universities, institutions, research groups, or countries. In such a scenario, altmetric information could for instance inform research managers about the visibility of research of their units across different platforms.
2. *As indicators of thematic interest, and local or global reach;* e.g., to "unveil communities of attention around scholarly documents and scientific topics" (Zahedi, 2018, p. 12). In such use cases altmetric information could for instance solve the purpose of informing about hot topics and trends, or of tracking the reach of scholarly topics across different audiences.
3. *As a form of capturing societal relevance or impact.* Aware of the difficulties of measuring or even defining *societal impact* of research in a concise way, Zahedi (2018) notes that there still lies promise in certain altmetrics to reflect a broader impact of scholarly research beyond the scientific sphere.
4. *As an early impact indicator or predictor of future citation impact.* In this case, the comparable speed with which the majority of many altmetrics for an article accumulate could make them helpful for early predictions of said article's expected academic impact.

These examples of use case categorizations for altmetrics illustrate that the question about altmetrics' potential as research indicators is a multifaceted one, as different use cases impose different demands on metrics' validity. While this thesis puts a focus on the notion of using metrics in evaluative scenarios (so on use cases of *research evaluation* in the NISO's categorization, or of venues 1 and/or 3 in Zahedi's (2018) schema), the diversity of potential use cases is a circumstance that will be taken up repeatedly over the course of the thesis.

1.3 Altmetrics - Criticism and Challenges

Needless to say that since the very beginning of discussions about web-based metrics for purposes of research assessment a plethora of challenges and criticisms of the concept has been identified. Such criticism starts on a conceptual level: different from citations, where theories to explain citation behavior were laid out as early as in the 1970s (most prominently in the form of Merton's *normative theory* and Gilbert's *social constructivist theory* of citing; Bornmann & Daniel, 2008), for altmetrics no cohesive theory of citing has been established so far (Glänzel & Gorraiz, 2015; Haustein, 2016). This lack of an underlying theoretical framework or "interpretative lens" (Sugimoto et al., 2017, p. 2046), which helps in assigning meaning to altmetrics, leads to difficulties in systematically assessing their construct validity, i.e., the degree to which altmetrics are able to measure what they claim or purport to measure (Rowlands, 2018; Tahamtan & Bornmann, 2020).

Thus, a considerable share of past and current altmetric research approaches the question of which phenomena or qualities are actually measured by different altmetrics. While the hope that altmetrics might capture societal impact of research is a frequently invoked argument for their potential value in research evaluation (see also Section 1.2), many researchers resort to more humble claims, suggesting that altmetrics would merely express attention (Konkiel et al., 2016), public engagement (Khazragui & Hudson, 2015), or popularity (Xia et al., 2016) of research. In their review of research on altmetrics' potential to reflect societal impact, Tahamtan & Bornmann (2020, p. 17) conclude that "altmetrics measure public (online) discussions, but not the societal impact of research" and suggest to complement them with other indicators such as patent citations or qualitative case studies to arrive at valid indications of research's societal impact.

In addition to such more theoretical approaches, also several empirical studies tackled the question of what altmetrics capture and encountered further challenges concerning their practical use. In an endeavor of coming to conclusions about altmetrics' ability to reflect scientific quality, some studies analyzed the relationship between different metrics and assessments by peers (Bornmann & Haunschild, 2018; Bornmann, Haunschild, & Adams, 2019). Bornmann & Haunschild (2018) found associations between various altmetrics (research mentions on Twitter, Wikipedia, Facebook, blogs, in news, and in policy documents) and peer assessments, but found the relationship between citations and peer assessments to be about two to three times stronger. Bornmann et al. (2019) studied correlations between case study data from the UK Research Excellence Framework (REF) and altmetrics, finding close to zero or negative correlations. The authors conclude that altmetrics might capture different aspects of societal impact than what REF reviewers see, who are primarily interested in causal links between research and action in society.

Analyzing about 8,000 tweets on dentistry research papers, Robinson-Garcia, Costas, Isett, Melkers, & Hicks (2017) found that a share of 74% tweets had apparently been tweeted automatically, either by bots or humans who behaved like bots, i.e., they did not contain any noteworthy original thought. The concerns of metrics being shaped significantly by automated accounts is also shared by Haustein, Bowman, Holmberg, et al. (2016), who found automated Twitter accounts to account for 9% of tweets to 2012 arXiv submissions subsequently published in journals indexed in Web of Science. Such an influence of bots might constitute a challenge concerning a multitude of altmetrics that are based on interactions in openly accessible online fora (e.g., research mentions on social networks like Facebook, on wikis like Wikipedia, or on scholarly literature platforms like Mendeley), while altmetric indicators relying on interactions happening in more restricted environments (e.g., research mentions in journalistic media or in policy documents) will be less prone to such distortions.

Closely connected to the question of what altmetrics measure is the question of *whose* interactions are reflected by them. Most thoroughly this aspect has been investigated for research mentions on Twitter. While some researchers suggest that users from outside of academia are responsible for a high share of Twitter altmetrics (Mohammadi, Thelwall, Kwasny, & Holmes, 2018; Yu, 2017), which might support the notion of tweet mentions being a valid indicator for the reach of research among society, other empirical case studies indicate that most tweets to scientific papers are likely to come from researchers (Birkholz, Seeber, & Holmberg, 2015; Tsou, Bowman, Ghazinejad, & Sugimoto, 2015). In their review of altmetrics literature, Sugimoto et al. (2017, p. 2052) conclude regarding this matter that "initial studies suggest that social media has rather opened a new channel for informal discussions among researchers, rather than a bridge between the research community and society at large".

Other challenges for the use of altmetrics are of technical nature. Just like many bibliometric analyses to this day rely on a small group of proprietary bibliographic databases (most prominently on Clarivate's Web of Science or Elsevier's Scopus), most large-scale altmetric studies obtain their data from a similarly small set of data providers, e.g., Altmetric.com, Plum Analytics, PLOS Article-Level Metrics, or Crossref Event Data. A variety of studies reviewed and compared these providers to investigate individual strengths and disadvantages - the review by Sugimoto et al. (2017) lists 21 such studies alone. Various comparisons have shown that altmetrics obtained for the same articles can differ considerably between different providers (Jobmann et al., 2014; Zahedi, Fenner, & Costas, 2014). At least partly these differences can be explained with the variations regarding retrieval strategies. For instance, while some aggregators calculate Facebook mentions solely based on public posts, other aggregators also consider private posts, shares, and likes (Zahedi et al., 2014). The partial lack of transparency regarding individual aggregators' strategies and processes for the collection of data can for many indicators be circumvented by mining data directly from the source platforms' own APIs (see also Section 1.2). The existence of this additional option, however, adds another reason as to why different altmetric datasets are often difficult to compare, even if they in principle aim to describe the same or very similar entities.

A further limitation regarding the reproducibility of altmetric analyses results from the dynamic nature of the interactions that are counted as altmetrics (Haustein, 2016). Different from citations, which under regular circumstances will not decrease for a respective research publication, many altmetrics can and frequently will, e.g., due to deletions of postings or withdrawals of certain online actions.

Finally, another substantial challenge for altmetrics related to aspects of data quality is their reliance on correctly linked metadata in the research mentions that are meant to be tracked. Different from academic citations, which mostly follow established codifications in the form of citation norms and -styles and therefore typically appear in fairly well-structured formats, mentions of research on altmetric platforms can appear in a myriad of forms. Most altmetric implementations and aggregators therefore rely on persistent identifiers - most commonly DOIs - to identify research mentions. While the availability of DOIs for research publications overall increases, this method introduces its own problems: metrics that rely on DOIs will for instance always result in an underestimation of the actual amount of references to research on a respective platform (which might be severe; see Lemke, Bräuer, & Peters, 2021). Furthermore, the restriction of data collection to publications with DOIs might increase biases resulting from discipline-specific publication behaviors, particularly to the disadvantage of disciplines that rely more commonly on other document formats than journal articles or with overall lower presence of DOIs, e.g., social sciences or humanities (Haustein, Costas, et al., 2015).

One factor of uncertainty regarding altmetrics, which received very little research attention so far, is their perception among stakeholders. When thinking ahead about future practical implementations of (alt-)metric data (be it within larger scale research assessment exercises or in smaller scale information offerings, e.g., as informative tools on the online portals of libraries, publishers, or other providers of scholarly works), detailed knowledge about said perception becomes crucial, as it is a requirement for both the development of useful and accepted indicators as well as for the prevention of unwitting indicator misuse on the users' side. This is the research gap that studies A and B within Chapter 2 of this thesis contribute to.

Within this section it has been made clear that altmetrics research is in many regards still in its infancy, many practical challenges have yet to be overcome, and many research gaps remain to be filled. A particularly glaring need of research, however, appears to concern the question of what altmetric signals actually measure. While this question has at least as many facets to it as there are altmetric indicators (times the number of factors with potential effects on these indicators), it is also the second major research gap that this thesis aims to contribute to. More specifically, Chapter 3 presents two studies (C & D) that analyze a factor of significant meaning for virtually all fields of research, whose interactions with (alt-)metrics still have only scarcely been researched: external science communication. In doing so, the thesis scrutinizes whether a suspicion holds true that Phillips, Kanter, Bednarczyk, & Tastad (1991) raised already over 30 years ago for the case of citations, but which has barely been investigated since - namely, that impact indicators might to a substantial degree be the result of the marketing- and dissemination efforts a research publication receives.

1.4 Scholarly Communication and External Science Communication

Science communication summarizes all communicative activities with the goal of disseminating, informing about, or raising awareness of science-related topics or discoveries. Depending on the stakeholders involved, we further differentiate between *scholarly communication* and *external science communication*.

The *Association of College & Research Libraries* defines *scholarly communication* as “the system through which research and other scholarly writings are created, evaluated for quality, disseminated to the scholarly community, and preserved for future use”[5]. As such, scholarly communication denotes the system through which scientific knowledge is distributed *within* the academic sphere, e.g., through publications in academic journals, presentations at scientific conferences, or grant proposals. This includes all formats for disseminating research findings that are primarily aimed at audiences with academic backgrounds themselves and thus constitute a form of expert-to-expert communication. For this reason, the term *scholarly communication* is sometimes used interchangeably with *science inreach, inward-facing*- or *internal science communication* (Illingworth & Allen, 2020).

Endeavors of informing about or distributing science-related content that involve agents without backgrounds in research are referred to as *external science communication.* Examples are research publications that target non-scientific audiences, e.g., books, podcasts, or videos of popular scientific style, or research lectures aimed at non-academic stakeholders like the general public. However, in many cases of external science communication, not only the recipients but also the emitters have no direct background in science: for instance, cases of science journalists reporting on research findings in newspapers or of PR officers informing via press releases about new publications from their institutions, publishers, or journals can all also be considered activities of external science communication. In such cases, the non-scientific intermediaries not only distribute the research, but oftentimes also perform the tasks of translating, condensing, and explaining it while transforming it into new media formats.

External and internal science communication are not clearly distinct, as it is not always possible to precisely define the target groups of respective emissions. Also, with the many new opportunities for researchers to directly engage with diverse audiences through the participatory web, overlaps between both forms of science communication become even more common, as online platforms increasingly provide researchers with straightforward means to broadly distribute their research insights without any involvement of intermediaries. Examples for this might range from individual Twitter accounts used to inform only a few dozen followers about news from the lab, to scientific Youtube channels with millions of subscribers.

[5] https://www.ala.org/acrl/publications/whitepapers/principlesstrategies

Figure 5 presents a schematic overview of the field of science communication as a whole and its central agents (based on Könneker, 2017). Könneker (2017) identifies scientists, science journalists and media-/PR officers as the three central professional groups in science communication. Of these, scientists and media-/PR officers usually appear as institutionally bound communicators, i.e., in science communication they explicitly represent certain institutions or are at least perceivable as such institutional representatives.

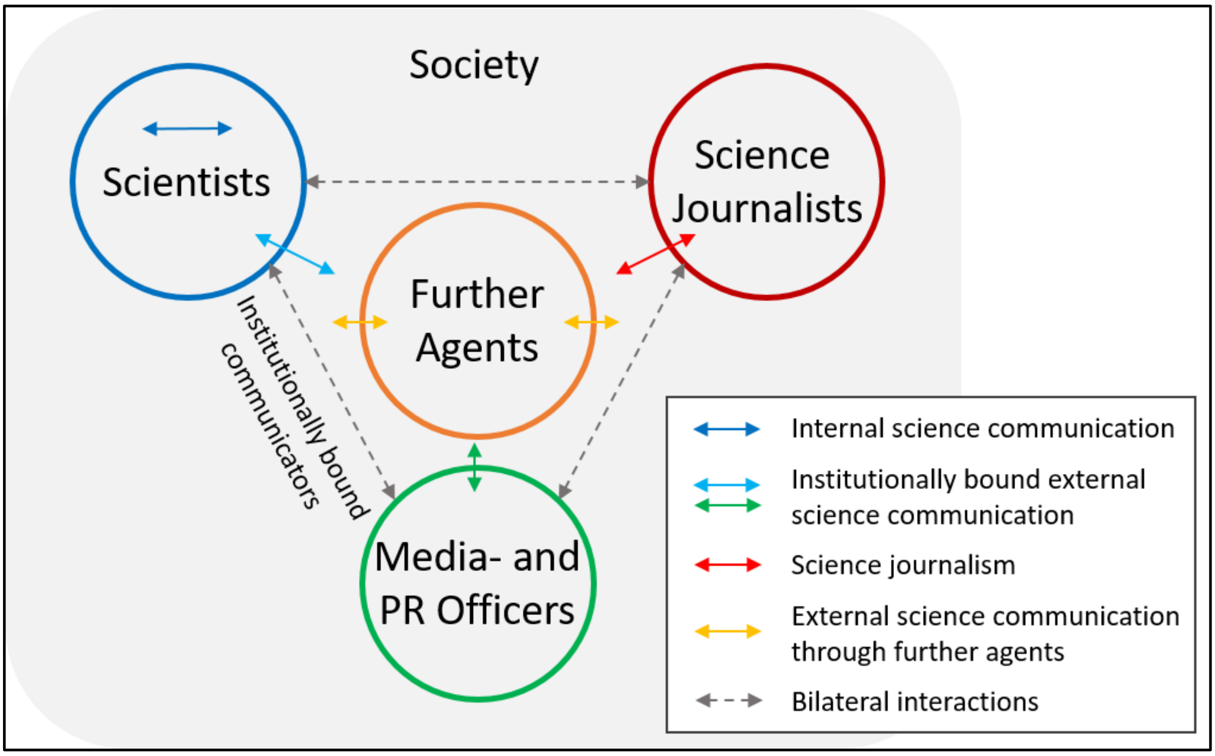


Figure 5: Schematic overview of the field and agents of science communication; based on Könneker (2017).

Bibliometrics developed as a set of measures and methods for analyzing the structure of and the processes within scholarly communication (Borgman & Furner, 2002). The analysis of citations as connections between documents enables us to model the landscape of scientific publications as a networked system, in which underlying interdependencies between entities - be them publications, journals, authors, institutions, disciplines, topics - can be made explicit; quantitative indicators like citation counts, impact factors, or h-indexes were developed to provide estimations of entities' influence within the system of scholarly communication. Given this close connectedness between the core methods of bibliometrics and the system of scholarly communication, it is no surprise that scientometric research over the decades identified and described numerous ways in which various properties of the entities that make up the scholarly communication system affect related indicators' expected values.

In their extensive literature review of 198 articles that investigate factors influencing citation frequencies, Tahamtan, Afshar & Ahamdzadeh (2016) identified 28 different factors. Besides attesting what a considerable amount of research has already been carried out to get a better understanding of what shapes citation counts, the review by Tahamtan et al. (2016) also demonstrates that the overwhelming majority of such research put entities from within the system of scholarly communication and their properties at the center of attention. This is already indicated by the three categories Tahamtan et al. (2016) suggest to group the 28 identified factors, which are *paper-related*, *journal-related*, and *author-related* factors. Similarly, Sugimoto et al. (2017) reviewed literature examining factors associated with higher altmetric performance, finding coverage and density of certain altmetrics to for instance correlate with article-related factors such as their publication dates, disciplines, geographic origins, study types and lengths, or publishing journals. Jin, Duan, Lu, Ni, & Guo (2021) on the other hand found the Altmetric Attention Score to significantly correlate with the readability of respective articles' abstracts.

Compared to the amount of research that went into studying how the properties of entities within scholarly communication affect research metrics, so far less attention was spent by the scientometric community on interactions between activities of external science communication and research assessment measures. An overview over such research, as well as our own studies of interactions between promotion of research within external science communication and respective metrics, shall be reviewed in Chapter 3.

Chapter 2: Researchers' Perception and Usage of Metrics

In recent years, the scholarly debate on the use of quantitative metrics for research assessment has gained traction, as is evidenced by several much-noticed publications on the matter. In December 2012, a group of publishers and editors of academic journals met at the Annual Meeting of the American Society for Cell Biology (ASCB) in San Francisco to discuss potential improvements for the ways in which academic institutions, funding agencies, and other stakeholders evaluate outputs of scientific research (Cagan, 2013). The resulting set of 18 recommendations was subsequently published as the *San Francisco Declaration on Research Assessment (DORA)*, which since then has been signed by more than 2,600 organizations and over 19,000 individuals.[6] DORA particularly focuses on practices related to assessments of peer-reviewed journal articles, and first and foremost warns of using journal-based indicators, such as Journal Impact Factors, as surrogate measures of individual research articles' quality.

About two years later, a multidisciplinary steering group of experts from the domains of scientometrics, publishing, research funding, research policy, and university management chaired by James Wilsdon came together in an 'Independent Review of the Role of Metrics in Research Assessment and Management', the results of which should later culminate in a monograph titled *The Metric Tide* (Wilsdon et al., 2015). In it, Wilsdon et al. (2015) discuss both strengths and weaknesses of assessments based on quantitative indicators, and highlight the importance of qualitative peer review, open and interoperable data infrastructures, as well as ongoing investments in the study of research systems. Moreover, they suggest a framework under the notion of *responsible metrics* that describes five dimensions of importance regarding metrics use: *robustness, humility, transparency, diversity,* and *reflexivity*. To offer *robustness*, metrics should be based "on the best possible data in terms of accuracy and scope"; to maintain *humility*, users should recognize "that quantitative evaluation should support - but not supplant - qualitative, expert assessment"; to achieve *transparency*, "data collection and analytical processes [should be kept] open and transparent, so that those being evaluated can test and verify the results"; to consider *diversity*, users should account "for variation by field, and [use] a range of indicators to reflect and support a plurality of research and researcher career paths across the system"; and to show *reflexivity,* users should recognize and anticipate "systemic and potential effects of indicators, and [update] them in response" (Wilsdon et al., 2015, pp. 134-135).

Another high-profile publication on the theme of responsible metrics use arrived in the same year in the form of the *Leiden Manifesto for research metrics* (Hicks et al., 2015). The Leiden Manifesto concisely proposes ten principles for a responsible use of bibliometrics, exposing several significant biases and

[6] https://sfdora.org/signers/

weaknesses of citation-based indicators along the way. Although the Leiden Manifesto specifically focused on bibliometrics, it was soon pointed out by Bornmann & Haunschild (2016b) that substantial parts of it would apply to altmetrics as well.

In addition to sharing the consensus that the use of impact metrics for assessments of individual publications and scholars is problematic and should be limited, all of the aforementioned wake-up calls share a focus on administrative use cases of metrics: the decisions criticized within these papers mostly refer to cases of the funding, promotion or hiring of researchers. This focus is understandable of course - after all, these types of decisions have immediate impact on both the careers and lives of researchers as well as on the structure of the research system as a whole, and it is thus easy to imagine how grave the negative consequences of an inappropriate use of metrics within such contexts could be. However, with such large parts of the debate about responsible use of impact metrics focusing on administrative scenarios, critical discussions of other usage scenarios fell behind. For instance, comparatively few studies so far investigated the status quo of metrics usage in individual day-to-day scenarios - little is known on questions of how impact metrics affect everyday decisions by individual scholars. It is fair to assume that researchers will be aware of the incentive system imposed to them through metrics and react to it.

Examining researchers' stance towards and usage of metrics in micro-decisions, e.g., during literature selection processes or manuscript submissions, is relevant for several stakeholders. First, the ways researchers interact with metrics on such basic levels are a reflection of a scientific culture that (at least in part) also determines metrics' role within the 'grander picture', i.e. the acceptance and prevalence of metrics within macro-decisions related to for instance funding, hiring, or promotion (Lemke, 2022). Awareness of researchers' conceptions of metrics in daily contexts therefore also informs the debate about their appropriate place within overarching research evaluation systems, e.g., by revealing what researchers actually find useful about certain metrics and which properties suitable alternatives thus would need to offer; or by indicating prevalent and potentially harmful misconceptions about metrics' applicability that should be combated with information. Knowing about such misconceptions - and about what researchers will likely already know and think about different research indicators - would also be important for individuals or institutions taking on the role of advisors on the appropriate use of metrics; a role that information infrastructures like libraries commonly will fulfill. Finally, knowing about common concerns regarding metrics is also relevant for content providers that within their catalogs aim to provide helpful data to their clients, e.g., publishers, aggregating services, or again libraries.

In this chapter, two published studies will be presented. The first one, Study A, consists of a combination of interviews and surveys with participating researchers, primarily coming from the Social Sciences. In the context of this thesis, Study A provides a detailed overview over the status quo of researchers' perceptions and concerns regarding impact metrics in general and altmetrics in particular. We will see that the reservations researchers have against the use of impact metrics are diverse, ranging from general to metric-specific.

In a survey that aims to capture the participants' perceptions of altmetrics (among other impact metrics), it makes sense to also incorporate segments that investigate the same participants' usage and perceptions of the online platforms that altmetrics are most typically based upon, i.e. social media. There are two reasons for this: first, stakeholders' perceptions of altmetrics as research indicators will likely be intertwined with those same stakeholders' perceptions of the platforms the indicators are derived from. If an individual for instance perceives a certain online platform as obscure or unfamiliar, or has certain preconceptions about its active community, then these experiences will very likely also affect that individual's perception of indicators derived from said platform. Second, to learn more about altmetrics and their meanings themselves, a better understanding of social media's uses and users in the context of scholarly communication remains a challenge of central importance (Haustein et al., 2016; Sugimoto et al., 2017; Tsou, Bowman, Ghazinejad, & Sugimoto, 2015) - who for instance are the communities that engage with scholarly content in such online fora? The upcoming Section 2.1 presents a study that therefore sets out to gather researchers' perceptions of the use of both social media as well as research metrics in scientific contexts.

Study B can be seen as a continuation of Study A and more closely investigates how researchers' perceptions of metrics actually find expression in the researchers' own behavior within the prototypical use case of literature selection. While Study A provided us with a broad impression of the present state of perceptions and concerns researchers have regarding impact metrics and their utility, Study B shall go deeper regarding the function several prominent metrics fulfill for researchers within a concrete everyday scenario they are faced with on a regular basis, the process of literature selection. The study is meant to provide us with insights on the role of metrics in such publication assessment situations relative to other potentially relevant (mostly qualitative) selection criteria, as well as on how different types of metrics compare to each other regarding their perceived utilities.

Together, the two studies illustrate the actual state of the use of impact metrics on the micro-level and shed light on some of the major hindrances that researchers see regarding their utility. The chapter will be closed by interim conclusions to summarize some of the lessons learned from the two reported studies and to clarify how these findings motivate the examination of factors influencing impact metrics that will follow in Chapter 3.

As indicated above, researchers are not the only stakeholder group in research metrics, as for instance funding agencies or research administrations have their own interests in this matter. Although the literature reviewed in the following studies will at times also touch on those perspectives, a clear focus on the stakeholder group of researchers within our user studies is justified, as they are both the group that is most directly affected by impact metric use, and - in the field of scientometrics - the group most actively involved in the suggestion and development of new indicators.

2.1 Study A - "When You Use Social Media You Are Not Working": Barriers for the Use of Metrics in Social Sciences

2.1.1 Foreword

The study presented here was originally published in 2019 in *Frontiers in Research Metrics and Analytics* as part of a special issue on *Increasing the Visibility of Research in the Social Sciences and Humanities (SSH)*[7]. Thus, it focuses on these specific disciplines, which tend to struggle particularly with the utilization of citation-based metrics - for reasons which are discussed in more detail within the article (see also Hicks, 2005; Nederhof, 2006; Thelwall, 2017). Nevertheless, we will see that many of the concerns and problems expressed in this study do not seem to be specific to the social sciences and humanities, but echo common conceptions of researchers about metrics that also came up in other studies with different discipline-wise foci.

The study was co-authored with Maryam Mehrazar, Athanasios Mazarakis, and Isabella Peters as part of the DFG-funded project *metrics (grant number 314727790). Supplementary material (i.e. the questionnaires used in interviews and surveys) can be found at https://www.frontiersin.org/articles/10.3389/frma.2018.00039/full. In addition to my co-authors, I wish to thank the original manuscript's reviewers, Erjia Yan and Raf Guns, as well as the special issue's editor Andreas Ferus for their valuable comments.

[7] Full reference: Lemke, S., Mehrazar, M., Mazarakis, A., & Peters, I. (2019). "When you use social media, you are not working": Barriers for the Use of Metrics in Social Sciences. *Frontiers in Research Metrics and Analytics, 3*(39), 1-18. https://doi.org/10.3389/frma.2018.00039

2.1.2 Abstract

The Social Sciences have long been struggling with quantitative forms of research assessment - insufficient coverage in prominent citation indices and overall lower citation counts than in STM subject areas have led to a widespread weariness regarding bibliometric evaluations among social scientists. Fueled by the rise of the social web, new hope is often placed on alternative metrics that measure the attention scholarly publications receive online, in particular on social media. But almost a decade after the coining of the term *altmetrics* for this new group of indicators, the uptake of the concept in the Social Sciences still seems to be low. Just like with traditional bibliometric indicators, one central problem hindering the applicability of altmetrics for the Social Sciences is the low coverage of social science publications on the respective data sources - which in the case of altmetrics are the various social media platforms on which interactions with scientific outputs can be measured. Another reason is that social scientists have strong opinions about the usefulness of metrics for research evaluation, which may hinder broad acceptance of altmetrics too. We conducted qualitative interviews and online surveys with researchers to identify the concerns which inhibit the use of social media and the utilization of metrics for research evaluation in the Social Sciences. By analyzing the response data from the interviews in conjunction with the response data from the surveys, we identify the key concerns that inhibit social scientists from (1) applying social media for professional purposes and (2) making use of the wide array of metrics available. Our findings show that aspects of time consumption, privacy, dealing with information overload, and prevalent styles of communication are predominant concerns inhibiting Social Science researchers from using social media platforms for their work. Regarding indicators for research impact we identify a widespread lack of knowledge about existing metrics, their methodologies and meanings as a major hindrance for their uptake through social scientists. The results have implications for future developments of scholarly online tools and show that researchers could benefit considerably from additional formal training regarding the correct application and interpretation of metrics.

2.1.3 Introduction

The first to introduce the idea of evaluating the importance of scientific work based on quantitative metrics - more specifically citation counts - were Gross and Gross in 1927 (Bornmann & Daniel, 2008). Since then, the assessment of research, which had historically been based on the qualitative practice of peer review, has incorporated a multitude of quantitative methods and indicators (Desrochers et al., 2018). Among these quantitative methods the most commonly used techniques are still bibliometric, i.e., based on output and citation analysis, well-known examples being the Journal Impact Factor or the h-index (Hirsch, 2005). Other developments in quantitative research evaluation, like the "Norwegian model" (Sivertsen, 2016) or the book-oriented "libcitation count" (White et al., 2009), try to solve known problems of an evaluation system predominantly focusing on citations. Moreover, recently numerous promising alternatives and complements to citations as indicators for research impact have been enabled by the proceeding digitalization of the scientific community.

Because scientific publications are to an increasing extent accessed as electronic documents online, the providers of publication outlets hosting those documents can without difficulty record and display the attention that individual publications receive as usage metrics, i.e., as download- or page view counts. Another prevalent family of web-based metrics is called altmetrics, a term coined by Priem et al. (2010) to comprise various signals of the buzz scholarly products receive on social media. The concept of altmetrics includes a heterogeneous multitude of indicators, ranging from counts of postings referring to a publication on social networks like Twitter, over numbers of bookmarks pointing to that publication on the literature management system Mendeley, to the amount of online news outlets and blogs citing the respective publication. Altmetrics have been shown to circumvent several weaknesses of citations as indicators for scientific attention (Wouters & Costas, 2012): they can be collected for a large variety of scientific products, e.g., for software, presentation slides, posters, individual book chapters, et cetera; altmetrics are available much faster than citation counts as the major part of altmetric resonance toward a publication happens very shortly after its publication (see also Thelwall, Haustein, Larivière, & Sugimoto, 2013); they show a broader spectrum of scientific impact than citations, as they are able to also reflect resonance among non-scientific audiences; most altmetrics are based on publicly available APIs which are open and free to use, unlike the commercial databases commonly used for citation analyses.

Still, the scientometrics community is widely concordant that altmetrics are by no means meant to be used as self-sufficient, flawless indicators for scientific relevance, but merely valuable complements to existing research impact measures (see e.g., Hicks et al., 2015). Just like bibliometrics, altmetrics come with their own shortcomings and yet unsolved challenges. Haustein (2016) identified issues of data quality, heterogeneity and technological dependencies as three “grand challenges” of altmetrics. Another frequently stated problem of altmetrics is their susceptibility to gaming (Bellis et al., 2014). And altmetrics are - just like the Journal Impact Factor - not fit for cross-discipline comparisons: for example, STM subject areas and Life Sciences tend to be significantly better represented on various altmetric data sources than the Social Sciences, Arts, and Humanities (Jobmann et al., 2014; Peters et al., 2014).

For a metric's applicability to a discipline, that discipline's degree of coverage in the metric's data base is a crucial factor. The smaller the share of a discipline's output that is represented in such a data base, the less truthful and comprehensive measurements based on it will be. In other words, low degrees of coverage diminish the validity of both macro- (e.g., institutional-level) and micro-level (e.g., author- or article-level) assessments of research performance in respective disciplines. In the context of alt- and usage metrics for the Social Sciences this means: as long as only few Social Science publications are made visible on the web, web-based metrics' applicability to the discipline is substantially restricted. For the case of Social Sciences their current low coverage online seems especially deplorable, as due to their also non-satisfying representation in prevalent citation indices (Archambault, Vignola-Gagné, Côté, Larivière, & Gingrasb, 2006; Sivertsen & Larsen, 2012) and their compared to “hard” sciences usually lower volume of citations (Glänzel, 1996; Nederhof, 2006) they could benefit particularly from alternatives to citation-based indicators for quantitative research evaluation.

Only if researchers perceive the work-related usage of social media as genuinely beneficial, they will spend time and effort to disseminate and discuss their research on the platforms that can be used to derive web-based metrics for research evaluation. Hence, identifying the barriers that keep social scientists from utilizing social media for work and thus inhibit an increase of Social Science publications' coverage on social media is a necessary part of the endeavor to make altmetrics useful for the Social Science-related fields of research. Such barriers might lie in a wide array of researchers' general concerns regarding the usage of individual social media platforms, which are the main data sources for altmetric data. These concerns might range from concerns regarding technical aspects (e.g., concerns regarding the security of data uploaded to a certain platform) to user- or content-related concerns (e.g., concerns regarding the target groups assumed to be represented on a certain platform). With its goal of identifying such reservations inhibiting researchers from utilizing social media for their work, this study follows previous studies: Nicholas & Rowlands (2011) examined researchers' utilization of social media in the research workflow in a large-scale survey study, which also inquired about barriers inhibiting such usage. Analyzing about 2,000 responses, they found *Lack of time*, *Problems of authority and trust* and *Unclear benefits* to be the most prevalent reasons for researchers not to use social media. In another survey study specifically targeting researchers that already use social media, Collins, Shiffman, & Rock (2016) asked their participants to state suspected reasons why many of their colleagues would refrain from using Twitter. The most commonly given responses were *Fear of the unknown* and *Lack of time*. In reference to ResearchGate's success, Van Noorden (2014) suggests further possible reasons that might demotivate scientists to use social media professionally: researchers might for instance be wary to openly share data and papers, or they might be repelled by high volumes of emails automatically sent from the platforms.

To better understand and to develop strategies to overcome such concerns we conducted qualitative interviews and a subsequent online survey, primarily addressing researchers from the Social Sciences. By studying the response data we aim to answer the following research question:

RQ1: Which concerns inhibit researchers in their work-related usage of social media?

While identifying social scientists' concerns regarding researchers' professional usage of social media might reveal what would need to be done to increase Social Science publications' coverage on social media, altmetrics' (and usage metrics') usefulness for the discipline is limited by at least one other major factor: their acceptance among stakeholders - most of which will be researchers. Hammarfelt & Haddow (2018, p. 924) analyzed the attitudes of Australian and Swedish researchers from the Social Sciences and Humanities toward bibliometric indicators, finding that “scholar's attitudes regarding bibliometrics are mixed; many are critical of these measures, while at the same time feeling pressured to use them”. Also they found the shares of researchers that had already used bibliometrics to vary significantly between the two countries. Rousseau & Rousseau (2017) surveyed economists about their knowledge about several citation-based indicators, identifying the Journal Impact Factor as the most well-known indicator, followed by the h-index. Overall they found the bibliometric knowledge of their respondents to be fairly heterogeneous. They propose the concept of “metric-wiseness” to describe a researcher's capacities to appropriately use scientometric indicators. And, among other things, Rousseau & Rousseau (2017) provide arguments why such metric-wiseness might be of particular importance for social scientists, as many researchers might for instance not be aware of the fact that Google Scholar also records citation counts and indices for non-English publications and working papers.

Biblio-, alt-, and usage metrics serve several purposes besides research evaluation (National Information Standards Organization (NISO), 2016), e.g., increasing scholarly outputs' discoverability or enabling researchers to showcase their achievements. The acceptance of metrics probably varies with the area of application - in this study however we focus on the most sensitive area, i.e., research evaluation. We therefore also aim to answer the following research question by using interview- and survey data:

RQ2: Which concerns do researchers have regarding various metrics used for research evaluation?

To get to a more accurate picture of whether researchers' stated concerns toward metrics for research evaluation affect their acceptance of certain types of metrics more strongly than others, we also aim to answer the research question RQ3 by drawing from the interview- and survey responses:

RQ3: Which metrics used for research evaluation do researchers consider as useful?

2.1.4 Materials and Methods

To learn about researchers' thoughts and concerns related to metrics as well as social media usage in professional contexts, we conducted usage studies following a two-step approach. As the first step, we interviewed 9 researchers face-to-face in groups about their work-related usage of social media and their notions on metrics used for research evaluation. Although these exploratory interviews allowed us to inquire about individual researchers' usage- and perceptual patterns in great detail, because of their low sample size we cannot assume their findings to be universally valid for whole disciplines. As the second step, we therefore conducted online surveys among the broader population of researchers, which more extensively investigated on the researchers' concerns we learned about during the interviews. This quantitative section of the study is our primary source from which we aim to derive insights that apply to the Social Sciences as a whole.

Methods Used for Qualitative Interviews

Interviews: Design

For the semi-structured group interviews we designed a questionnaire with three sections as a guideline: the first two sections consisted of questions about the interviewees' experiences and perceptions regarding the use of online tools and social media in their field of research, the third section contained questions about the interviewees' notions on various metrics for measuring research impact. We tested and adjusted the questionnaire over the course of four iterations during which various acquainted scientists (without direct relation to our research project) took the roles of the interviewees. After these test runs the final questionnaire contained a total of 25 questions, which in a group with two to three interviewees should in total take between 90 and 120 minutes to discuss. The interview questionnaire is part of this article's Supplementary Material.

Interviews: Sampling

To recruit researchers as interviewees, we resorted to a subset of the participants of the *metrics project's[8] first international survey on social media usage from spring 2017 (see also Lemke, Mehrazar, Mazarakis, & Peters, 2018; Mehrazar, Kling, Lemke, Mazarakis, & Peters, 2018). Like in this study, the 2017 survey‘s prioritized target groups during dissemination had been researchers from Economics, Social Sciences and respective sub-disciplines, which subsequently accounted for 83% of the survey’s 3,427 respondents. At the end of the survey, participants had been given the option to provide an email address in case they would be interested in taking part in other studies related to the *metrics project. From the list compiled this way, we extracted 22 mail addresses from research institutes situated in Northern Germany to allow for easy traveling to face-to-face interviews. We invited the respective researchers to take part in our interviews, offering them a reimbursement of 50€ for their participation. Seven researchers were recruited this way (more information on participants in section Interviews: Demographics).

Along with the responses given by said seven researchers in this article we will also report on the responses given during our fourth internal test run. The interviewees in that final test run were two computer scientists acquainted with the authors of this article. For this test run we used the same questionnaire as during the later, “real” interviews, while our test candidates had been given analogous preparatory information on the interviews' content and purpose as our later, “real” interviewees. Thus,

[8] https://metrics-project.net/

the conditions of that final test run and the real interviews were similar enough to justify the former's inclusion in the results.

While the later surveys should focus exclusively on social scientists, we decided to also allow researchers from other disciplines to the interviews. First, many more general concerns regarding the usage of social media and metrics will be applicable to all disciplines, meaning there would very likely be a lot to learn from the experiences of researchers from other disciplines that also holds true for many social scientists. Second, possible discipline-specific patterns should become easier to detect with interview data for different disciplines at hand.

Written informed consent was obtained from all interview participants before the start of the interviews (the consent form is available upon request).

Interviews: Conduction and Analysis

We conducted the interviews in groups with two to three interviewees, two interviewers and an assistant responsible for the data's later transcription. The role of the interviewers was filled by two of this study's authors (MM & SL). Interviewees were - as far as their availability allowed us to - grouped according to their field of research and their academic rank (see section Interviews: Demographics, Table A1). By interviewing multiple researchers from similar disciplines and career stages at the same time we hoped to allow for more extensive digressions regarding discipline- or role-specific phenomena. All interviews were conducted in English to minimize later translation requirements, regardless of the participants' mother tongues.

The transcribed interviews were analyzed with qualitative methods based on grounded theory (Burnard, 1991). A first topic-related coding of interview contributions was applied during the interviews' transcription. The transcribing assistant and the authors reviewed and discussed the preliminary coding in two iterations; the resulting adjusted coding scheme was subsequently applied to the full transcripts. The coding scheme was used to tag transcript sections in which interviewees stated their concerns and usage purposes, both regarding researchers' social media- and metrics usage. While the coding was used as guidance during the review of the interview data, in this article's results section slightly different, more self-explanatory category names are used to structure the interview results.

Methods Used for Online Surveys

Online Surveys: Design

As our second major step, we designed an online survey to check to which degree observations made during the analysis of the interview responses apply to the broader population of social scientists.

A crucial part of our survey design was the set of social media platforms to include in platform-related questions. As the landscape of social media platforms with potential relevance for researchers is multifaceted and vastly growing (see also Kramer & Bosman [2015] for a crowd-sourcing-based catalog of online tools used by researchers - in July 2018 it had about 680 entries, many of which could qualify as “social media”), we were forced to select a set of only particularly relevant or interesting platforms, so we would not overwhelm our survey participants with too many sub-questions. The basis for this set of platforms was the previous *metrics survey from 2017, in which we collected extensive data about the online tools used by social scientists, allowing us to derive a ranking of the most-used platforms among these researchers. Starting from the top of the ranking we added every platform to our set that could be classified as social media according to a definition from Kaplan & Haenlein (2010, p. 61), i.e., “Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content.” In the next step we removed every platform used by < 100 respondents of the previous *metrics survey from the set. Finally we added Xing and Quora to the set, both of which had not been included in 2017's survey but had been mentioned frequently there as “other platform” by respondents in a respective free text-field. This left us with a set of 18 social media platforms to include in our survey: Academia.edu, Facebook, GitHub, Google+, LinkedIn, Mendeley, Quora, ResearchGate, Scholarly blogs (e.g., WordPress), SlideShare, SourceForge, StackExchange (including its communities, e.g., StackOverflow), Twitter, Vimeo, Wikipedia, Xing, Youtube, and Zotero.

To avoid having an overly time-consuming and exhausting questionnaire, while still being able to ask all the questions we would need to solve our research questions, we decided to split the questionnaire into two separate surveys. The first survey - from here on referred to as survey A - contained a total of 14 questions, the pivotal 6 of them about the researchers' thoughts and concerns regarding their professional usage of various social media platforms. Survey B contained a total of 15 questions, the central 7 of them being about the participants' conceptions regarding metrics that are used to evaluate research impact. The remaining 8 questions were identical across both surveys and aimed at obtaining context information, namely about the participants' demographics and their previous knowledge regarding social media. Both questionnaires are included in the Supplementary Material for this article.

Online Surveys: Design - RQ1

To examine to which degree concerns regarding social media usage that came up in the interviews apply to the broader population of social scientists, we incorporated those previously identified concerns into a question for survey A. To account for the survey format we transformed the concern groups found during the interviews into more self-explaining individual concerns, sometimes splitting a group into multiple concerns when this seemed to increase their comprehensibility. This left us with twelve individual concerns to ask for in the survey, e.g., *Concerns about privacy, Concerns about data security,* et cetera (for the full list of concerns see section RQ1: Survey Results).

Each participant was asked to report which of these concerns demotivate them to use which specific social media platforms in the form of a matrix checkbox question. This matrix question listed on its vertical axis all social media platforms the participant had marked as “used at least once” in a previous question - this should prevent participants from being asked to make statements about platforms whose features they do not know. On its horizontal axis the question contained the 12 concerns mentioned above.

To allow our survey participants to provide further explanations or to add other individual concerns regarding their use of social media for work purposes, the survey question described above was followed by a free text question asking *"Are there any other concerns you have using the mentioned services? If so, please tell us."*.

Online Surveys: Design - RQ2

Our second research question - about the concerns researchers have regarding metrics used for the evaluation of research - was included into survey B in the form of a free text question: *"Do you have any thoughts or concerns about using metrics like these to evaluate research? If so, please tell us."*. Beforehand, the survey participants had already been asked to assess various types of metrics regarding their perceived usefulness for determining a scientific product's relevance (see section Online Surveys: Design - RQ3 below), so at this stage they would already have seen several examples for the types of metrics we are investigating.

Online Surveys: Design - RQ3

This study's third research question was represented by a matrix question in survey B: *"The following list contains several types of metrics that can be used to evaluate the impact of a scientific output (e.g., a publication, a video, …) and/or its author. Would you find these individual metrics useful to decide whether to consume (read/watch/...) a respective scientific output?".* The "list" mentioned in the question's text referred to the vertical axis of the matrix which listed 14 types of metrics, e.g., *Citation number of the scientific output, Number of downloads of the scientific output*, et cetera (for the full list of metrics we included see section RQ3: Survey Results). To every type of metric depicted on the vertical axis each participant had to assign one of the five options *Very useful, Useful, Hard to use, Useless, No answer/Don't know*.

Online Surveys: Sampling

Both questionnaires were implemented and distributed using the online survey tool *LimeSurvey*[9]. The sampling process followed the approach of the 2017 survey described in Lemke et al. (2018): a mailing list administered by the *ZBW Leibniz Information Centre for Economics* was used to contact about 12,000 researchers working in economic institutions from German-speaking parts of Europe; further invitations were sent to about 42,000 email addresses of authors of Economics- or Social Science-related papers mined from *RePEc* and *Web of Science*. As we had divided our questions into two surveys

[9] https://www.limesurvey.org/

as described above, we also divided these 54,000 mail addresses randomly into two lists of 27,000 addresses, each group receiving an invitation to one of our two surveys. As an incentive for participating, we gave participants the option to enter a drawing of 25 10€-*Amazon.com* vouchers at the end of the surveys.

Before their submission of responses, participants were asked to give their informed consent about their participation in the survey. On the first page of the survey (see also “Questionnaire for Survey A/B” in the Supplementary Material), participants were provided with respective information about it[10] along with the note that at the end of the survey they would be asked to confirm their consent to submit their answers under these terms. Accordingly, on the survey's last page, participants were asked to indicate that they had read all the given information and voluntarily agreed to participate by clicking on a submit-button.

Online Surveys: Conduction and Analysis

The initial dissemination of both surveys took place over the course of 20 days from June 25th to July 14th 2018. A wave of reminders was sent to those who had not yet responded to (or not yet opted out of) their first invitation during the second week of August. Afterwards, the survey was kept running till August 27th 2018.

2.1.5 Results

The following sections provide results from the interviews and online surveys regarding our three research questions.

Interviews: Demographics

Table A1 shows demographic information about the participants of the qualitative interviews. The recordings of the four group interviews added up to 375 min of interview material, which were subsequently transcribed and coded. The allocated time was almost equally distributed among the four groups.

---

[10] This included information about the purpose of the survey and the project behind it, about why the respective participant had been asked to participate, about which data will be stored for which purposes and in which locations, that stored answers will be anonymized, that their participation is entirely voluntary and that they can withdraw from the survey at any time, as well as our contact information in case of questions.

Table A1: Demographic details of interview participants

ID	*Group*	*Gender*	*Nationality*	*Academic Rank*	*Discipline*
P1	1	M	German	PhD Student	Marine Biogeochemistry
P2	1	F	German	PhD Student	Marine Biogeochemistry
P3	1	M	Ghanaian	PhD Student	Economics
P4	2	F	German	PhD Student	Economics
P5	2	M	German	PhD Student	Economics
P6	3	M	German	Postdoc	Economics
P7	3	M	German	Postdoc	Economics
P8	4	M	German	PhD Student	Computer Science
P9	4	M	Macedonian	PhD Student	Computer Science

Online Surveys: Demographics

Till the day on which we closed the two surveys, 1,065 participants had responded to survey A, 1,018 participants to survey B, meaning a rate of response of ~4% for both surveys. For our study we are primarily interested in the perceptions of all kinds of social scientists, therefore we considered only responses from researchers identifying themselves in the survey as primarily working in either *Social Sciences*, *Political Sciences*, *Sociology*, *Psychology*, *Demography*, *Human Geography*, *Economics* or *Business Studies*. This leaves us with 872 respondents for survey A and 948 respondents for survey B. Table A2 shows the demographic properties of those respondents.

The median time participants spent in the survey was 7 minutes 51 seconds for survey A and 6 minutes 54 seconds for survey B.

RQ1: Which Concerns Inhibit Researchers in Their Work-Related Usage of Social Media?

To provide answers to our first research question we will first review the segments of our qualitative interviews in which participants expressed concerns and reasons that might inhibit them or their colleagues from using social media platforms as part of their work life. Afterwards we will report data from our online surveys, in which we asked a larger sample of social scientists about the concerns we had compiled through the interviews and subsequent discussions.

Table A2: Surveys - demographics

	Survey A	*Survey B*
	N = 872	N = 948
Gender		
- Female	31.1%	31.6%
- Male	68.7%	68.3%
- Other	0.1%	0.1%
Academic Rank		
- Assistant professor	12.3%	12.6%
- Associate professor	16.8%	16.6%
- Other	11.3%	7.1%
- PhD student/Research assistant	16.6%	14.7%
- PostDoc/Senior researcher	15.2%	19.8%
- Professor	27.9%	29.1%
Country of affiliation (Top 5 + Other)		
- Germany	27.7%	33.4%
- USA	14.0%	14.7%
- United Kingdom of Great Britain	5.8%	5.9%
- Italy	5.8%	5.3%
- France	3.7%	3.5%
- Other (includes 65 countries)	43.0%	37.3%

RQ1: Interview Results

In this section we describe statements from the interviewees relevant to RQ1. We start with the concerns that were brought up more frequently and then move to the less often expressed issues.

Represented target groups/style of communication

A concern inhibiting social media usage for professional purposes mentioned in every single group interview was the suspicion, research-related communication on social media would often remain shallow due to the target groups represented and reachable on the respective platforms.

P2, P3, and P8 specifically mentioned Facebook as an example for a service that is usually more associated with non-professional, casual communication, which is why they would not expect researchers to share a lot of professionally relevant information or articles there. Independently confirming this, P7 stated: *"I wouldn't post a paper I published on Facebook, because I have so many friends who are not into research; who are not really interested in that."*

Also, P2 added that she would distinguish between media and platforms suitable for communicating with researchers and others suitable for the communication with policy makers or the broader public.

P1 mentioned scientific communication on social media sometimes being restricted by the need to address too many target groups at once: *"I get suspicious if it gets so superficial. I mean, if you communicate something that addresses many target groups – policy makers and economists and the broader public – then sure, Twitter can be used – for information that is not so into detail."* Related to that argument, P2 stated a concern regarding how Twitter's technical details restrict professional communication in a similar fashion: *"If you have only [140] signs[…], that's just too short. And my problem with that is that I would never know what I can put there while still being precise and basing on the facts. […] I could imagine that this applies to many researchers, so that's why they don't have an account there, because they simply don't really know how to use that and still be a researcher."*

Information overload/spamming

Another frequently stated concern referred to the problem of dealing with information overload or spamming during platform usage, either previously experienced or just expected by the participants.

While P1, P2 and P8 named Twitter as a service often sending overwhelming amounts of notifications, P1 and P7 mentioned similar problems with ResearchGate.

Moreover, referring to ResearchGate, P7 mentioned another related concern that might prevent researchers with teaching duties from using platforms that enable one-sided follower-relations: *"Also students use it. And then you will get like 20 students per semester which want to – or which will follow you. I mean, it's not 'want to follow', they follow you. […] And that's a bit annoying. So that's why I don't really use it."*

Another reported facet of information overload experienced when using social media for professional purposes lies in the difficulty of distinguishing important from unimportant information, as explained by P1 and P8. This leads P1 to believe that services like Twitter are more appropriate for achieving an overview than for investigating about *"actual research."*

Related to this, P3 explained that especially on ResearchGate he is sometimes missing compliance with quality standards: *"I know a lot of others who are using ResearchGate as a playground. Any little thing they do, they put it there. They developed a small proposal that has not seen any proofreading – they put it there."* Similar impressions were stated by P4: *"On ResearchGate I would always think that there is 'quantity over quality' for most of the people. Because they put all their work there and then, of course, I know that not all of the listed articles are of high quality."*

P2 explained a problem of being bothered with redundant information resulting from connecting to the same persons or institutions on several platforms in parallel: *"What I find irritating is when they share something on Facebook, Instagram and WhatsApp and I get the same message three times in a row."*

A very specific aspect related to the concern of spamming was described by P1, who referred to the experiences of a publicly well-known researcher working on emotionally charged topics: *"There is one [colleague] that says, he doesn't for example use Twitter or Facebook, because he would be spammed with emails or requests, because the research he undertakes is kind of a very emotional hot topic. […] So at some point I guess you have to refuse to use these kinds of media, because you otherwise would get spammed."*

Time consumption

Similarly consistently, interview participants mentioned the concern that social media usage could easily consume much time (P1, P2, P3, P4). Closely related to the problem of information overload, participants frequently brought up the assumption that making sense of the volume of information incurring in social media would cost time that could likely be better spent for other aspects of work. As P1 put it: *"And I think one thing stays constant the whole time and that's the time that people have during a day. I mean, there is more and more popping up, more and more to do, but everyone just has the same amount of time, so something has to fall over the table."*

While P4 acknowledged that a researcher's high degree of activity on social media could hint at that researcher being a good networker, she still expressed doubts about whether using the time for networking on social media is really well spent: *"I would also say the good networkers are those who are using [Social Media] more frequently, but one could also say that they could use the time they spend on social media, promoting and working on their profiles, they could rather use it to do research or something like that."* also adding that *"I also know that you can get lost and can waste a lot of time on those platforms. As I said, when I don't have anything to do then sometimes I go on ResearchGate […]."*

Separation of private and professional matters/Privacy

Often the aforementioned concern of time consumption seemed to be related to another concern: the question whether time spent on social media can actually qualify as "work" and whether it is therefore appropriate for researchers to spend the time necessary for social media's utilization during work time. Very clearly stated was this issue by P8, who explained: *"I think social media and Computer Science has always a little bit of... bad flavor? Kind of, if you use social media then you're not working (laughs)."*

P4 even reported that *"if my professor walks into my office and I have Facebook and Twitter open, I always close it (laughs), even though I might be on [professionally relevant sites]."*

Another problem related to the separation of private and professional matters on social media and mentioned during most interviews are the difficulties that arise from using private social media accounts for professional communication.

P8 stated that while he finds it easy to follow a quite strict policy of using Facebook for private and LinkedIn for professional communication, he misses this kind of clarity on Twitter, making it more difficult for him there to determine which information is important for him and which is not.

As was previously reported regarding the concern *Represented target groups*, P7 also would not use Facebook for professional postings as this would lead to private contacts being addressed that probably would not be interested in the respective postings. P7 even stated that he would be aware of functions helping in this case, but would find it too bothersome to use them. Additionally, he stated an analogous concern regarding his professional contacts; adding them on Facebook would enable those contacts in undesirable ways: *"I mean they could see some comments I did, I don't know, ten years ago on some picture and maybe, I don't know, at a late time at night. Yeah, I mean this kind of stuff…"*

P6 phrased his disinterest in using his private social media accounts for professional purposes differently: *"I [just don't think that] someone who I'm in a direct professional relationship with needs to see my interests, or needs to know me that close."*

Data security

P1 mentioned that concerns about the security of data uploaded to social media platforms led to restrictions regarding which platforms researchers at his institute are allowed to use for professional purposes, explaining that there would be *"clouds that are set up specifically for [storing] data,"* that these researchers had to use instead.

Paywalls

P3 stated a concern specifically regarding the usage of the academic social network Academia.edu, stating that the service would require users to make fee-based subscriptions to increase the visibility of their articles - an approach that according to P3 would have led to a lot of his colleagues moving to ResearchGate instead.

Apart from these concerns regarding the professional usage of social media, P4 and P9 also brought up the concern how the availability of social media might lead to disadvantages for researchers not using those new opportunities for their work. P9: *"When I went to a conference I saw that everybody was using Twitter, to share the program or to just follow each other, and then it was a must. […] Practically it was a requirement to fully participate and follow the event (laughs)."*

P4 speculated that the rise of social media might especially benefit researchers with certain skillsets: *"I don't know if it's a disadvantage for people who are not so good networkers, because they now are even less visible than they have been before."*

RQ1: Survey Results

As explained in section Online Surveys: Design - RQ2, transforming the concern groups we encountered in the interviews into individual concerns to ask for in the survey left us with 12 such concerns:

• Concerns about data security
• Concerns about privacy
• Concerns about the respective user base (e.g., in terms of expected reactions/addressable target groups)
• Costs too much time
• Lack of interest
• Overwhelming amount of data/news
• Spamming
• There are better alternatives
• Too few/restricted functions
• Too many emails sent from the platform
• Too many functions
• Usage feels inconvenient

Figure A1 shows how often the individual concerns were marked in survey A for all platforms aggregated. The percentage values express the shares that the occurrences of a specific concern have among all concerns reported in total. It can be seen that, if no distinction between platforms is made, *costs too much time* (887 occurrences) is the most frequently reported concern demotivating our survey participants from using social media for their work. Also relatively high rank a general *lack of interest* (727) as well as *concerns about privacy* (718). The lowest ranked concerns are *too few/restricted functions* (410), *spamming* (362), and *too many functions* (160).

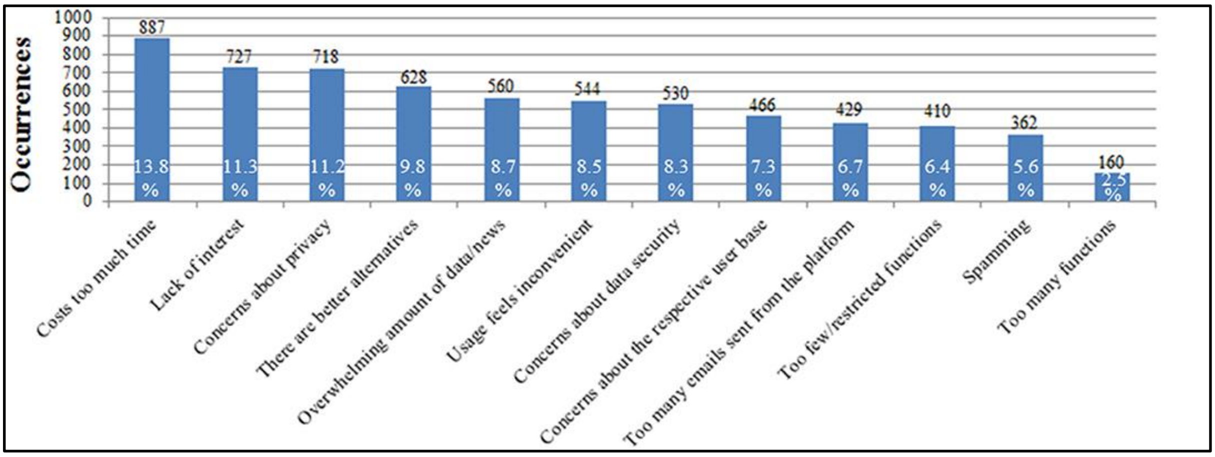


Figure A1: Social scientists' concerns regarding social media usage for work purposes (non-service-specific).

Figure A2 shows the data from Figure A1 broken down by gender, with positions on the y-axis in this case indicating the shares that occurrences of a specific concern have among all concerns reported by respondents of the respective gender in total.

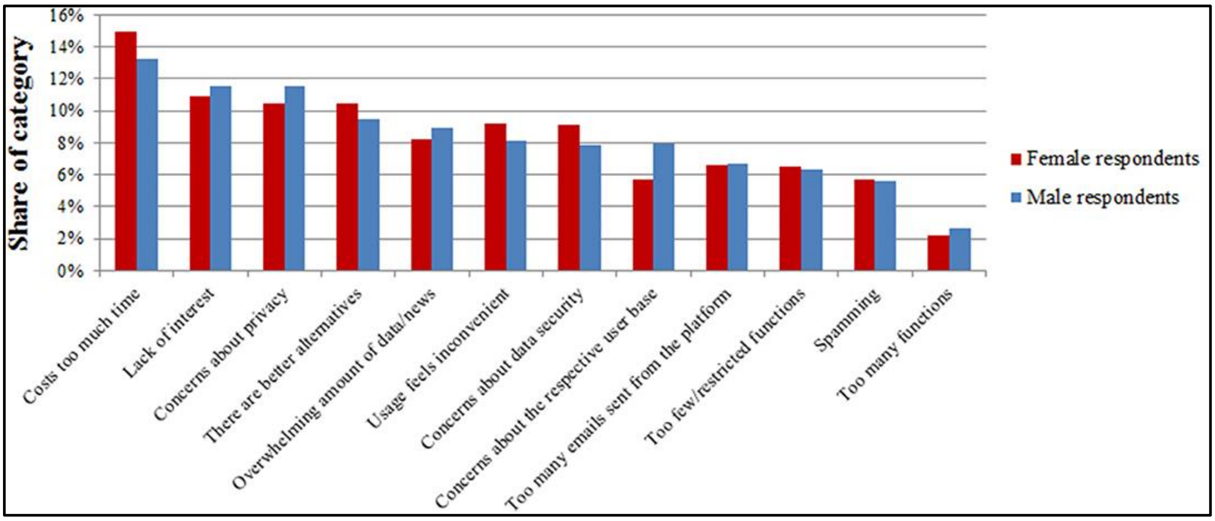


Figure A2: Concerns regarding social media usage for work purposes (non-service-specific) by gender.

Analogically, Figure A3 shows concerns broken down by respondents' research roles. The group of "Professors" in this case includes respondents identifying as associate-, assistant-, as well as full professors; "PostDocs" include postdocs and senior researchers; "Ph.D. students/Research assistants" include respondents identifying as either Ph.D. students, research assistants or a combination of the two.

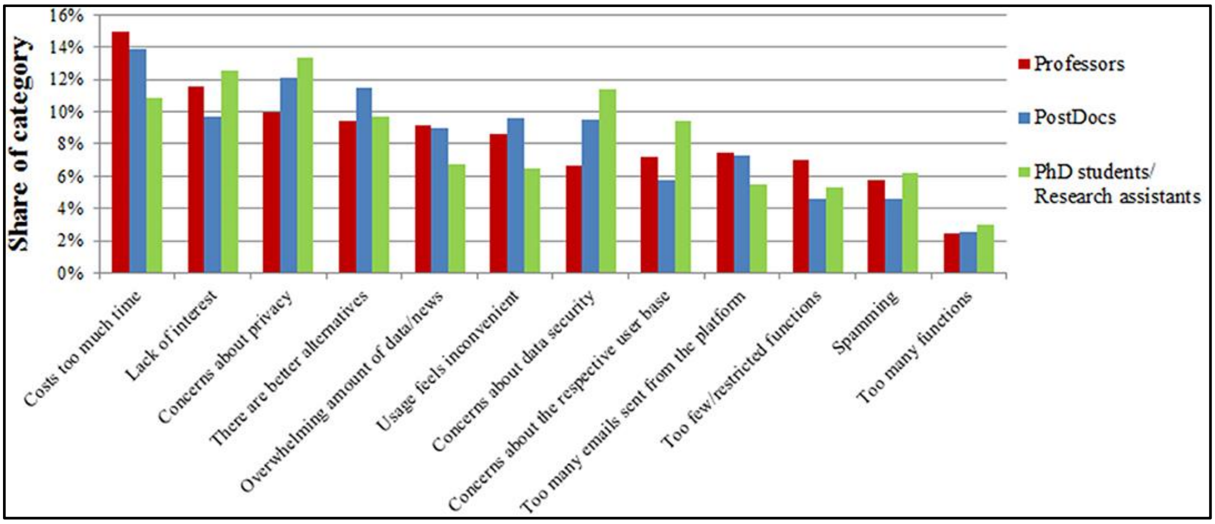


Figure A3: Concerns regarding social media usage for work purposes (non-service-specific) by role.

Drawing from the response data from the same survey question, Figure A4 shows concerns related to individual platforms as a heat map. The data from every cell of the map was normalized by the number of survey participants who previously had stated that they would have used that platform at least once

- this number equals the amount of participants who had been asked to voice their concerns regarding the respective platform. This way the heat map shows whether there are certain concerns that particularly large proportions of the users of a specific platform share. Darker cells represent concerns more commonly expressed in conjunction with the respective platform, brighter cells less frequent concerns. Our way of normalizing data means that the presented color coding is insensitive to the variation of usage degrees between the platforms - information on the percentage of survey respondents who reported to have used a respective platform for work at least once is therefore given in column UD, on the right side of Figure A4.

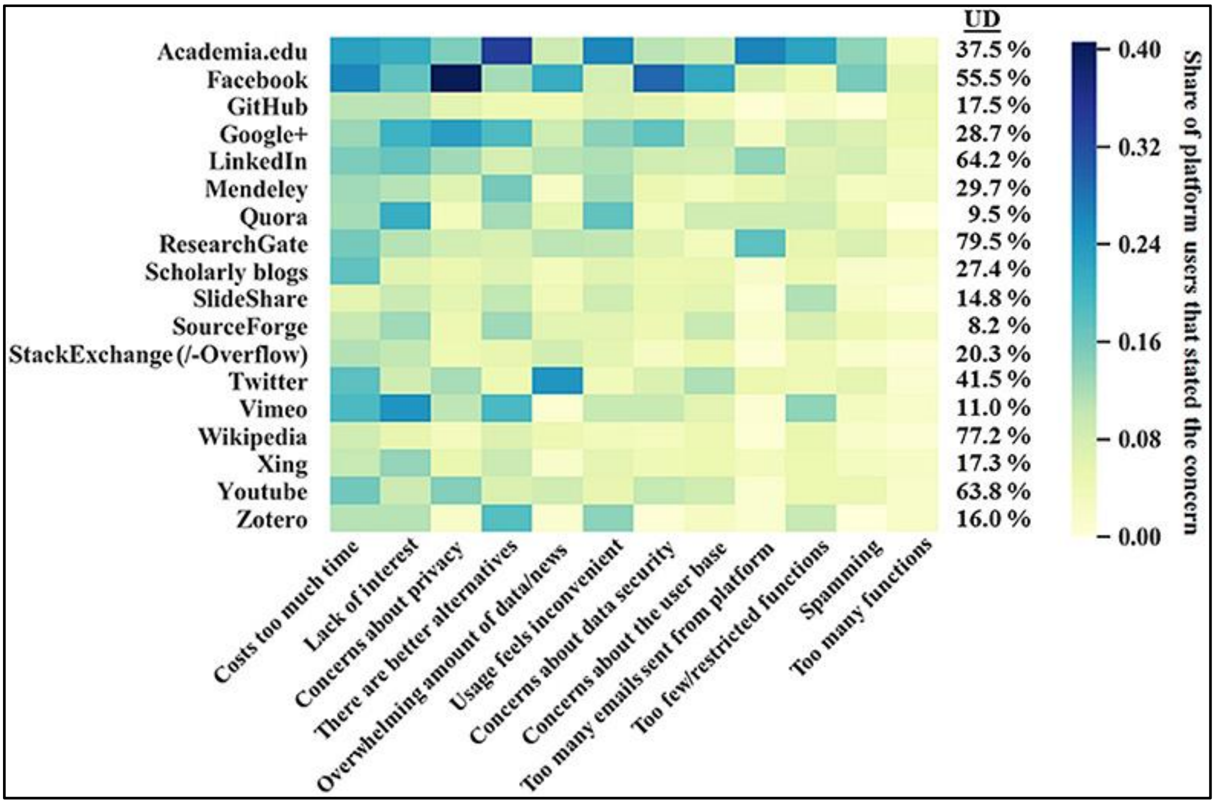


Figure A4: Social scientists' concerns regarding social media usage for work purposes (service-specific).

Comparisons of individual rows of the heat map show how platforms perform regarding user concerns: the darker a row, the more concerns were voiced regarding that platform's usage. It can for instance be seen that researchers have few complaints regarding GitHub, StackExchange or Wikipedia in general. Academia.edu, Facebook or Google+ on the other hand exhibit wider varieties of perceived deficiencies. Looking at the darkest cell of a given row reveals the most widespread concern related to a respective platform - this way we can for example see that Academia.edu, Zotero, and Mendeley are often considered to be suboptimal choices as there are better alternatives available; Facebook and Google+ prevalently arouse concerns regarding privacy; ResearchGate tends to annoy its users with too many emails; Quora and Vimeo simply do not catch many researchers' interest; and on Twitter the amount of news/data displayed is found to be overwhelming. Going through the heat map column by column leads to a view similar to the one presented by Figure A1, as particularly bright columns correspond to overall rarer types of concerns and vice versa.

Answers to free text question

In addition to marking their service-related concerns in the matrix question, a total of 125 participants also entered a response to the accompanying optional free text question *“Are there any other concerns you have using the mentioned services? If so, please tell us.”.* Cleansing this data from non-topical answers like *“no”, “don't want to answer”,* et cetera left us with 72 answer texts, which were subsequently coded according to the concern categories used before. This revealed that many answer texts once more confirmed the concerns asked about in the previous survey question - most frequently answer texts repeated that social media usage would cost too much time (13 times), platforms' target groups would not match the researcher's (8 times), and the usage of (often specific) platforms would feel inconvenient (5 times). Apart from such answers repeating previously identified concerns, three additional kinds of answers occurred repeatedly: a total of 8 respondents mentioned financial costs or *“paywalls”* as disincentives, often specifically referring to Academia.edu (Note: although this concern had also been brought up during the qualitative interviews, we had not included it as a predefined answer in the survey as we had deemed it to be applicable to too few of the platforms we planned to ask about). Another five respondents stated that they perceive the lack of quality assurance on social media as a problem, while two respondents mentioned the difficulty of sharing contents from within one platform with people outside of that platform as a disincentive - a result of the *“‘walled garden’ model”* of Academia.edu and ResearchGate, as one respondent called it.
Beyond that, some more specific concerns occurred in the responses only once each. These included gender bias and a missing openness to new Economics-related entries on Wikipedia, the fact that nothing one puts on the internet can really be deleted, and the fear that *“it can be seen as unprofessional to use social media as a researcher”.* Another researcher reported a very specific reason for frustration coming out of using social media: *“One of the platforms really annoyed me because there was a high access to one of my papers, but I could not retrieve any citation”.* Finally, one respondent just stated that *“some services are not meant to be used academically”* without further explanation.

RQ2: Which Concerns Do Researchers Have Regarding Various Metrics Used for Research Evaluation?

After having looked at the concerns that inhibit social scientists from using social media for their work, we in this section will report on the thoughts and concerns that researchers participating in our user studies stated regarding the usage of various research impact metrics, several of which draw from the previously examined online platforms. First we will review respective responses from our qualitative interviews, then we will inspect our survey data to examine how the perceptions of our interviewees compare to those of a larger sample of social scientists.

RQ2: Interview Results

Researchers' prior knowledge regarding metrics for research evaluation

An observation made repeatedly during all interviews was that when asked which indicators would come to the interviewees' minds upon hearing the terms “metrics for scientific impact” or “metrics for research evaluation”, in every single interview the first indicators to be mentioned were citation-based. P1, P2, and P3 started with mentioning the h-index, citations, and the Journal Impact Factor (in that order), P4 and P5 mentioned the Journal Impact Factor and citations, P6 and P7 mentioned the Journal Impact Factor and the h-index and P8 and P9 mentioned citations and then researchers' numbers of publications. After a bit of discussion P4 also suggested the *ResearchGate score*, P5 stated that he had also heard of the h-index before, P7 mentioned the *Handelsblatt-Ranking*, P6 the *GEWISOLA-Ranking* as well as university rankings. Beyond that, none of the interview participants seemed to have an idea of the concepts of altmetrics or web-based usage metrics for the evaluation of scientific impact yet. Also, even when interviewees named more intricate metrics like h-index or Journal Impact Factor, they were barely able to explain correctly how these indicators are calculated.
After these initial questions we provided the interview participants with a handout featuring a list of various existing indicators for research impact, explaining individual indicators where necessary. This list included citations, h-index, Almetric.com score, ResearchGate score, download counts, science rankings, Journal Impact Factor, Eigenfactor, and nine types of web citations as they can be collected with existing altmetric software such as Altmetric.com or Webometric Analyst (http://lexiurl.wlv.ac.uk/), e.g., Mendeley reader counts, Wikipedia citations, Google book citations, et cetera. After the interviewees had read that list, we asked them what they would think about metrics used for research evaluation in general and about the ways these metrics are used right now.

Lack of familiarity and transparency

In line with the previously identified interviewees' limited prior knowledge regarding metrics, a frequently stated concern regarding their usage involved a perceived lack of familiarity with and knowledge about them. As a result, many types of metrics remain non-transparent to our interviewees, which limits their abilities to trust in the metrics' validity. As P8 stated with regards to web-based metrics: *“Most of the [types of] web citations I just don't know, I have to admit. […]This is always a problem, when I don't understand their metric, what does it tell me? And if I then need to invest a lot of time to understand the metric or if it's not even publicly available, then I can just not use it”.* But, according to P8, classical bibliometric citations have similar problems, stating that *“nobody knows where [providers of citation data] get their numbers from, and how they aggregate them and in which* *intervals. So okay, what does it tell you now, that their Google Scholar says 1,700 citations? Nobody knows".* A similar point - regarding metrics in general - was made by P1, who stated: *"The point is – is [metric data] really transparent? So, is everyone in the same knowledge what it means? And the more [metrics] there are out there, the more – at least as an early career scientist – the more you resign. The more you kind of give up to really look through all this".* The problem of insecurity about how to interpret metrics was confirmed by P2, who mentioned that she had no idea about how to inquire truthful citation counts for a given article.

Reliability

In several cases interviewees went one step further than voicing concerns about metrics' missing transparency by questioning whether the metrics reliably captured what they might claim to capture at all.

Referring to citations, P2, P3, and P8 all mentioned that they would not believe them to be reliable proxies for scientific quality, but emphasized the necessity to check an article's content to be able to evaluate it truthfully. P8 illustrated citations' shortcomings as proxies for scientific quality with an anecdote from the field of Computer Science: *"You see with these neural networks, most of the publications I think were from the eighties and nineties. Nobody really cared about them – now everyone seems to care about them. The publications from 20, 30 years ago get really high citation counts, but although this means only now they have an impact, the quality was good 20, 30 years ago, when they had no impact. [...] And there I see a little bit of difficulty, because a low level, low quality paper can have a huge impact when it's just of popular interest, and the other way around".*

P6 showed awareness for the Matthew effect of citations, stating that *"once you are above a certain threshold of citations, you probably receive lots of more citations, even though [the article] is not that relevant".*

Another mentioned drawback of citation counts was the concern that their validity as proxies for relevance could easily be distorted by self-citations (P3).

Moving on to web-based metrics, P5 and P6 both stated that they would perceive social media-based metrics merely as *"network measures",* which indicate how well connected an author is and not necessarily the relevance or quality of a respective publication.

P4 mentioned that she would not likely trust in download counts as indicators for scientific relevance due to how easily they could be gamed. A similar mindset regarding the potential value of download counts was expressed by P6, who stated that it would be *"easy to download an article and throw it in the virtual trash".*

Regarding the differentiation between scientific quality and relevance, P6 and P7 shared thoughts on which types of metrics might better reflect which of these two properties. P7 said that his *"greatest concern"* regarding social media-based metrics would be that while they might successfully capture what a broad audience or the media perceives as relevant, highly theoretical or foundational research might have considerable disadvantages there, even though it might be of high quality, highly useful for its specific community and therefore often cited by it. P6 seconded this by adding an example from the field of food security: *"So, when it's about understanding when prices spike or why prices have a certain movement or behaviour, this is usually of high policy relevance and everyone wants to know about it. But the methods to understand price behaviour or to identify the drivers – these papers are more important but would never be on the media, because no one will be interested in understanding the estimator and the standard error or whatever. […] In our area […] works are based on a model which you need to calibrate, which is the high quality research that is in the shadows somewhere, because it [is neither empirical nor does it have any policy implications]. But it's the base for all the applied work".*

Restricted comparability

Another set of stated concerns referred to perceived limitations regarding the validity of cross-discipline or cross-community comparisons based on metrics.

Such concerns were particularly often related to the Journal Impact Factor. P4 explained this by mentioning how top journals from Natural Sciences would typically exhibit much higher impact factor scores than top journals from Economics. P6 described similar conditions regarding comparisons of different sub-fields from Economics.

P3, P6, and P7 reported of known cases in their disciplines in which impact factor comparisons would not reflect the relative prestige certain journals have among their research communities, with P7 assuming that the degree of interdisciplinarity of a journal would influence its Journal Impact Factor: *"For instance, in agricultural economics the journal with the highest impact factor is maybe ranked third or fourth if you consider the prestige of the journal. But it has the highest Impact Factor because it's a bit more interdisciplinary – it has a broader audience and higher citations. But everyone in the field knows that another journal is the number one in the field, although it has a lower impact factor".*

P9 mentioned the necessity to keep in mind that metrics need time to accumulate, something that should be considered especially when evaluating the impact of younger publications.

Increasing publication pressure

While the above mentioned concerns mostly deal with the correct application or interpretation of metrics, another set of concerns expressed in the interviews was linked to possible negative effects on the scientific system caused by the usage of quantitative impact metrics in general.

P1, P2, P3, and P5 voiced assumptions that the expanding quantitative evaluation of research outputs would increase the already existing pressure for researchers to publish. P5 hypothesized that the wide usage of quantitative impact metrics might lead to a gratification of quantity over quality, making publishing a higher priority than carrying out truly valuable research: *"You have to deliver. […] If you're this far away from the solution but you don't get it published, that doesn't make you a bad* *researcher, but it will get you very low scores. […] The pressure to publish something at some point is definitely something that is not really pushing quality".*

RQ2: Survey Results

Similar to the question we asked in the interviews, we also in the survey inquired about the participants' concerns regarding the usage of metrics by asking *"Do you have any thoughts or concerns about using metrics like these to evaluate research?"* as a free text question. In total 241 responses were collected this way, of which 215 answer texts remained after the removal of non-topical answers. These 215 answers were coded manually for the themes and concerns regarding the usage of metrics they addressed, one theme per answer text. The high topical variation between the answers led to a large number of themes identified this way - nevertheless, certain themes reoccurred especially frequently in the answer texts. Table A3 shows the ten themes that occurred more than five times along with examples taken from the response data.

Most of the concerns we encountered during the qualitative interviews reappeared in some form in the survey responses, although with different intensities. The most frequent kinds of concern stated in the survey refer to the notion that metrics might be misused as direct indicators for scientific quality, although they are perceived to primarily be indicators for popularity or the degree of dissemination efforts undertaken. Many researchers also suspect specific types of metrics to be inherently biased, be it toward certain fields, certain forms of publications, more *"mainstream"* research, *"fashionable topics",* English publications, and against *"hard science"* and small fields of research. Other widespread concerns relate to metrics susceptibility to gaming and their shortcomings during comparisons. Slightly more optimistic groups of answers stressed that metrics do have value, albeit use cases have to be selected carefully, they may not replace the consideration of a publication's content entirely, and instead of using isolated metrics they should be used in conjunction. A few answers described particular negative effects the reliance on metrics could have for science in its entirety, e.g., by leading to researchers spending less attention to the underlying research of publications, by incentivizing a *"click bait behavior"* among researchers, or by leading to *"decreased submits to lower ranked journals in specials"* and thus ultimately to *"more generic journal design".*

When we compare interview- with survey responses, one major difference becomes apparent: while for the interviewees their lack of familiarity with many metrics and their perceived non-transparency was a very present concern, in the survey only few participants reported similar issues (only 3 occurrences). We propose two explanations for this: (1) participants of the survey might just not have felt asked to explain their state of knowledge in this question, while in the interviews we purposefully led the conversation to this aspect, and (2) our interviewees' overall lower average academic experience might also explain comparatively lesser knowledge about impact metrics and their methodologies, making perceived lack of familiarity and non-transparency more apparent issues.

Table A3: Survey responses - concerns regarding the usage of impact metrics

Number of occurrences	*Theme*	*Example*
37	Measure popularity, not quality	*"Many of these metrics rely on some measure of "Popularity", which is usually a poor indicator of quality for new scholarly work."*
25	Low reliability / Inherent biases	*"Metrics are affected by superficial characteristics such as the use of terms in the title or abstract of a publication that connect with currently fashionable topics."*
25	Useful only for certain use cases	*"These metrics are sometimes helpful to find important research, but very limited in evaluating researchers."*
23	Manipulation / Gaming	*"They could be faked by bots, especially likes or posts and retweets."*
11	Metrics are (almost) useless	*"Yes, most metrics just serve non-useful or even non-desirable goals, they are a useless invention."*
9	Restricted comparability	*"These metrics are useful for comparison within sub-fields, but not across sub-fields and especially not across disciplines."*
9	Must not replace content analysis	*"None of the metrics substitute reading the paper."*
7	Need to be used in conjunction	*"They ought to be used very carefully and in a complex manner (combination of a few metrics)."*
6	Could have negative effects on science	*"Metrics may lead to decreased submits to lower ranked journals in specials and lead to more generic journal design."*
6	Measure dissemination efforts	*"Social media metrics are primarily a measure of time and effort the author puts in disseminating research."*
…	…	…

RQ3: Which Metrics Used for Research Evaluation Do Researchers Consider as Useful?

To obtain a precise picture of how the previously examined concerns affect researchers' perceptions of individual metrics in comparison, we will now review our interviewees' statements about how they utilize these metrics themselves, before again consulting the survey data on this matter.

RQ3: Interview Results

As seen in section RQ2: Interview Results, the interviewees' preconceptions regarding research metrics were mostly restricted to bibliometric indicators, in particular citations and Journal Impact Factor.

Accordingly, when asked if and how they would make use of such metrics themselves, most responses revolved around these indicators.

Journal Impact Factor

Although the Journal Impact Factor was, along with citations, the most frequently brought up metric during the interviews, notions about its usefulness seemed to vary a lot between the researchers. P3 explained that the Journal Impact Factor would play a major role for him during literature research because of a particular previous experience: *"I remember I was once using a paper to argue at one of my presentations and the professor mentioned "What is the source?" I mentioned the article and then the journal. Then he turned to the postdoc asking "Has it got an impact factor?" and the postdoc said "No, I don't think so"*.

P1 and P4 also mentioned that the Journal Impact Factor would sometimes help them as a filter mechanism, although they would not solely rely on it. P2, P6, and P7 on the other hand stated that they usually would not pay attention to the Journal Impact Factor due to their concerns regarding its comparability (see section Interview Results). Nevertheless, P6 said that looking into highly ranked journals according to Journal Impact Factor can be a good way to get informed about *"the newest kind of research",* as the most progressive research will more likely be found in highly ranked journals.

Citations

Despite the various concerns interviewees expressed toward citations' shortcomings as indicators for quality or relevance, many of our participants stated that citation counts would be helpful to quickly identify the most important publications in a certain field of research (P4, P5, P6, P8, P9).

According to P4 and P5, citation counts get more meaningful the higher they are - so although in many common cases they might not be reliable indicators for an article's relevance, if an article reaches an unusual high amount of citations one can fairly reliably assume that article to be of particular relevance for its field.

Furthermore, P5 and P6 stated that sometimes a particularly high citation count might indicate a *"mandatory"* citation in its field of research, *"[an article that] you have to cite to be taken seriously in the field"* (P5) or rather *"a citation [that] must not be disregarded when it comes to your own research"* (P6).

Hypothetical metrics

At some points during the interviews, interviewees described possible metrics they would find interesting, although they did not ever use anything like them until now. For example, P7 and P8 expressed interest in a (hypothetical) metric that would capture citations along with context information about the citing works. P8 suggested to somehow capture the shares of criticizing, negative citations, while P7 would like to see citation counts that only include citations from peer-reviewed sources, not knowing that Web of Science provides such features. P9 on the other hand said that he would imagine an article-level metric informing about the number of researchers currently considering that piece of work as a reference to be helpful - so effectively a metric reflecting the expected future citations of an article.

RQ3: Survey Results

Figure A5 shows the survey participants' responses to the question *"The following list contains several types of metrics that can be used to evaluate the impact of a scientific output (e.g., a publication, a video, ...) and/or its author. Would you find these individual metrics useful to decide whether to consume (read/watch/...) a respective scientific output?"*.

It can be seen that regarding the shares of participants describing a metric as very useful the bibliometric indicators clearly lead the field, although with considerable differences between each other: the highest acceptance receive citation counts, followed by the Journal Impact Factor, while the h-index ranks on the third position regarding participants judging that metric to be very useful. Looking at web-based metrics, only download numbers are considered to be either useful or very useful by a comparably large share of participants. The various altmetric indicators all perform drastically worse regarding their perceived usefulness - for all of them the shares of participants considering them to be useful or very useful are lower than the shares finding them either hard to use or useless. Another tendency indicated by Figure A5 is that the metrics that are perceived as less useful on average also seem to be unknown to larger shares of respondents. Regarding awareness levels, the most noteworthy case is the Altmetric attention score, for which the group of participants not knowing said metric (38%) is larger than any of the other four groups.

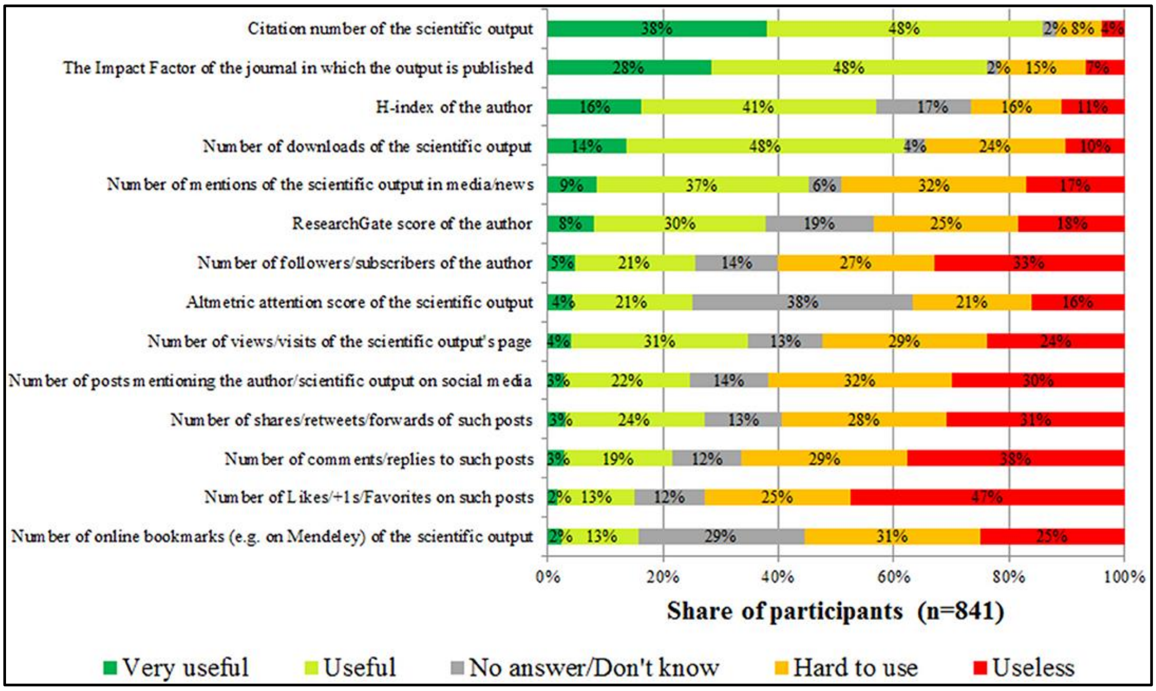


Figure A5: Perceived usefulness of different types of research impact metrics for social scientists.

2.1.6 Discussion

RQ1: Which Concerns Inhibit Researchers in Their Work-Related Usage of Social Media?

Our interviews revealed several distinctive types of concerns that demotivate social scientists to use social media in work-related contexts, the most frequent being: the platforms' target groups and prevailing styles of communication are often felt to be unsuitable for academic discourse; social media usage seems to cost much time; on several platforms separating personal from professional matters is bothersome; and the utilization of more platforms increases the efforts necessary to handle information overload. The wider prevalence of these concerns was also confirmed by the responses to our survey, where especially the aspect of time-consumption stood out as an often held concern - a finding in line with previous studies (Collins et al., 2016; Nicholas & Rowlands, 2011). Researchers' reluctance to use social media due to the platforms being perceived as unsuitable for scientific discussions on the other hand confirms a finding by Collins et al. (2016), who report similar concerns from researchers for the specific cases of Facebook and Twitter. Moreover, our survey data showed that complaints about the platforms' technological affordances, e.g., complaints about the amount of functionalities provided, play comparatively minor roles for the social scientists that participated.

Several of the identified concerns often seem to be intertwined: the impression that utilizing social media channels might cost so much time could well be a result of having to cope with an overload of available information there, a problem which bothers many respondents. Similarly, although an effective separation of professional from private matters could on most platforms be realized by consistently maintaining separate profiles for both scopes, this would often be inconvenient and time-consuming. In the interviews we learned that even when platforms already provide customizable filters to reduce the incoming amount of information or functionalities to manage different groups of contacts, respondents find their usage bothersome and thus ultimately not worth the effort. We think that in these aspects lies a lot of potential for technological improvements of the existing social media platforms which researchers could particularly benefit from, for instance in the form of easier-to-use and more transparent information filters, or in form of tools that assist in the creation and maintenance of multiple clearly divided profiles on the same platform.

The fact that academics hesitate to use certain social media platforms for scientific discussions because they perceive the style of communication on those platforms to be non-academic could be a self-fulfilling prophecy (Merton, 1948): similar to how rumors about a bank's insolvency - no matter whether true or false - can lead to that bank actually becoming insolvent as a result of customer actions motivated by the rumors, mere suppositions of a platform not being suitable for academic communication could lead to academics abandoning that platform, subsequently further depleting it of academic contents. Such developments would severely reduce social media platforms' value for the scientific community, both as tools for scholarly communication and as sources for altmetrics data. Overcoming this problem seems to be a particularly difficult task - what could help to showcase individual platforms' potentials as tools for scholarly communication would be to bring subject matter experts together in highly focused and easily discoverable discussion groups, similar to how mailing list services like *JiscMail*[11] manage email-based discussion lists for clearly defined interest groups. Although various social media platforms offer group- or list-functionalities that can be exploited for such discussion groups, browsing already existing groups can be difficult due to a lack of structured directories cataloging them. Another way of encouraging more social scientists to go online and increase the amount of scientific discussions on social media would be to explicitly integrate some online dissemination efforts into institutions' codified publication workflows. Such measures could also counter the researchers' feeling of not doing work when being on social media, as it was expressed in our interviews. This feeling is also fed by the fact that online visibility still barely counts in the reputation system of science, leading to a lack of external incentives to actively use social media platforms as a researcher.

In spite of all the barriers discussed in this article which inhibit social scientists in their work-related use of social media, interviewees also occasionally expressed the belief that not using social media could nowadays lead to noticeable disadvantages career-wise. Thus, right now many researchers might feel pressure to engage in professional social media activities, although various concerns make it a cumbersome or uncomfortable experience for them. Hence, developing tools to overcome these concerns is not only a necessary step to increase Social Science publications' visibility online and thus allowing the discipline to benefit from web-based impact measurement, but also a way of addressing everyday needs social scientists will most likely continue to face in the times to come.

RQ2 and RQ3: Which Concerns Do Researchers Have Regarding Various Metrics Used for Research Evaluation? Which Metrics Do They Consider as Useful?

Considering our second and our third research question, the first finding of the interviews was that the researchers' knowledge about metrics was mostly restricted to bibliometric measures, and even there some often-used concepts like the Journal Impact Factor or the h-index were in many cases not understood in detail. Nevertheless, it could be seen that the researchers do use bibliometric indicators, mainly for filtering purposes during literature research or to assess journals when considering where to publish their own research. Also, many interviewees and survey respondents showed awareness of some of the indicators' specific shortcomings, e.g., their restricted applicability in several kinds of comparisons.

As could also be seen in the interviews, non-bibliometric alternatives like usage- and altmetrics many researchers are not even aware of. Accordingly, the survey-based comparison of the perceived usefulness of various types of metrics revealed that bibliometric indicators are perceived as useful by

[11] https://www.jiscmail.ac.uk/

much larger shares of the community of social scientists than altmetrics, with usage metrics mostly lying in between. Two differing explanations for perceiving a metric as useless are possible: first, a metric can be seen as inherently flawed and thus not suitable for measuring what is meant to be measured; second, missing expertise about how to apply and interpret said metric might make its utilization so difficult it becomes effectively useless, although the metric in principle might have the desired properties to measure what is meant to be measured. While several free text responses to the survey indicated that some researchers reject certain metrics due to suspected inherent flaws, our interview and survey results strongly suggest that also the second reason might apply to many social scientists.

Hence, the lack of familiarity with existing metrics substantially constrains their usefulness for individual researchers. Many survey respondents voiced their concerns of metrics being misused - be it unintentionally or on purpose. Researchers' limited knowledge about the indicators' properties increases the risk of such unintentional misapplication and -interpretation. Thus, for social scientists the sheer lack of knowledge seems to be a decisive hindrance to making better use of metrics for research impact. This indicates that Social Science researchers could benefit from better formal training in the correct application and interpretation of metrics. Such training should ideally already be provided in "scientific working" courses at universities and be explicitly supported by thesis advisors, but also libraries can play an important role here by informing about the whole range of indicators available, their individual fields of application, strengths, and - especially - their limitations. Content-wise, the recommendations provided by the San Francisco Declaration on Research Assessment (Cagan, 2013) and the Leiden manifesto (Hicks et al., 2015) provide foundations for guidelines that researchers could be provided with. Moreover, various online resources exist that can be helpful for informing about metrics' peculiarities in more detail, e.g., the *Metrics Toolkit*[12], the Parthenos project's modules on research impact[13], or EC3metrics' periodic table of scientometric indicators[14]. A better familiarity with metrics among social scientists would address researchers' frequently brought up concern that metrics appear non-transparent in their methodologies, decrease the risk of unintentional misapplication, and could also dispel a commonly stated reason for frustration by clarifying which kinds of comparisons on basis of certain metrics are valid and which are not. Also, as Rousseau & Rousseau (2017, p. 483) argue, *"a basic knowledge of informetrics, including knowledge of scientometrics indicators and data sources, should be part of any doctoral education"* so that *"assessment processes […] would potentially be less distorted and the advantage of more knowledgeable researchers would be reduced".* Thus, a better and more comprehensive education about metrics could also lead to more fairness in research assessments by at least slightly leveling the field regarding researchers' knowledge about how to optimize metrics for their own research outputs.

[12] http://www.metrics-toolkit.org/
[13] http://training.parthenos-project.eu/sample-page/intro-to-ri/research-impact/
[14] https://ec3metrics.com/wp-content/uploads/2018/06/tablaper3.pdf

Beyond that, interviewees and survey respondents voiced a multitude of suspected negative effects that an excessive focus on metrics might have on science in general, e.g., higher publication pressure for individuals, increased concentration on more conservative and therefore “safe” research endeavors, and overall more generic journal design, to name a few (see also Rijcke, Wouters, Rushforth, Franssen, & Hammarfelt [2016] for a review of literature examining the potential effects of increased indicator use on science). Most of these undesirable scenarios follow the premise of metrics gaining a disproportionate amount of influence in hiring- and funding decisions in academia. While assuring that such decisions are not inappropriately based on impact metrics ultimately is a matter that governments and administrations have to administer to (Wilsdon et al., 2017), we believe that achieving a widespread awareness among researchers about what metrics can and what they cannot do is a major step toward preventing those scenarios from happening.

Limitations of the Study

A limitation of this study lies in the sample of researchers which participated in the qualitative interviews. First, the majority of our interviewees were fairly young researchers, which might be an explanation for their altogether restricted knowledge about indicators used for measuring research impact. Some use cases for such indicators these young researchers just might not have encountered yet, e.g., hiring decisions, promotion-, or grant applications. This hypothesis of younger researchers having less experience with metrics usage is also backed up by Hammarfelt & Haddow (2018), who found researchers with < 5 years of academic experience to be the group with the lowest usage of impact metrics in their sample.

Also, all our interviewees were from German institutions, so the interview results might be shaped by region-specific effects we are not aware of. Furthermore, our reliance on convenience samples for both the interviews and the surveys leads to potential self-selection biases; this could lead to overrepresentations of participants with a comparatively strong opinion on the subjects of social media and/or metrics. And finally, all interviews were conducted in English although our interviewers and participants were non-native speakers, which might have reduced our abilities to express and recognize linguistic nuances.

A more specific limitation results of the way we asked our survey question about concerns related to individual social media platforms. As explained in section Online Surveys: Design - RQ1, we only asked researchers to state their concerns regarding platforms if they had selected those platforms as “used at least once” in a previous question. There might of course be cases of researchers that know a platform well enough to have precise reservations against it despite never having used it before - these are cases that would not be covered by our survey.

Furthermore, there are potential barriers inhibiting an uptake of web-based metrics specific to the Social Sciences that are not explained by the researcher perspective which we examined in this article. Many metrics' reliance on specific persistent identifiers (like DOIs) for instance peculiarly penalizes publication patterns prevalent in Social Sciences, where monographs often play a more important role than in hard sciences (Glänzel, 1996; Nederhof, 2006). Another disadvantage for certain Social Sciences in web-based as well as bibliometric impact measurements results from their often comparatively high dependencies on local contexts. While such aspects' implications have been examined for bibliographic citation analysis (see for instance Hicks, 2005), regarding web-based metrics further work is necessary.

2.1.7 Conclusions and Future Work

In this study we examined researchers' concerns inhibiting the use of various metrics for research evaluation in the Social Sciences by analyzing the results of face-to-face interviews and two associated online surveys. More specifically, we identified problems that prevent Social Science publications from gaining visibility online and thus impede web-based metrics' usefulness for the discipline. Alongside these findings we provide recommendations for how to tackle these problems. Our findings and recommendations have implications for various stakeholders, among them platform providers and -developers, community managers, research instructors, libraries, policy makers, research administrations, and of course the researchers themselves.

This study will be followed by experiments in which we will examine how researchers from the Social Sciences make decisions based on various indicators for research impact. The experimental data in combination with this article's findings will provide us with a foundation to more clearly describe the state of how social scientists use and interpret the multitude of impact metrics and should thus help to point out misconceptions that might be prevalent in the discipline.

To obtain a more comprehensive picture of social scientists' notions on professional social media usage and metrics, further work should go into analyzing the relationships between various researcher demographics and specific concerns. Also, the conduction of similar studies like this one for other disciplines could highlight discipline-specific peculiarities that did not become apparent as such in this study. Another interesting subject for future research was exposed by our interviewees P6's and P7's hypothesis that certain types of metrics might give an advantage to certain types of publications. While work has been done to identify document properties that lead to increasing altmetric resonance (see e.g., Haustein, Costas, & Larivière, 2015; Zagovora, Weller, Janosov, Wagner, & Peters, 2018), the properties examined usually refer to document type, -structure, or -meta data, and not to the publications' contents.

2.1.8 Author Contributions

SL performed the statistical analysis, coded the survey free text answers according to their themes and wrote the first draft of the manuscript; MM and SL conducted the interviews, implemented and supervised the online survey and provided the figures used in the manuscript; IP acquired funding for

the research project. All authors contributed conception and design of the study, contributed to manuscript revision, and read and approved the submitted version.

2.1.9 Acknowledgments

We thank Felix Heute for transcribing the qualitative interviews as well as for his assistance during the phases of interview conduction and coding. Also, we wish to express our gratitude to all the researchers who helped us by participating in our interviews or surveys.

2.2 Study B - Conjoint Analysis of Researchers' Hidden Preferences for Bibliometrics, Altmetrics, and Usage Metrics

2.2.1 Foreword

Originally, the following study was published in 2021 in the *Journal of the Association for Information Science and Technology (JASIST)*[15]. It was co-authored with Athanasios Mazarakis and Isabella Peters and - just like Study A - a result of the DFG-funded project *metrics (grant number 314727790). The original article's supplementary information can be found on https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24445, the data collected within the study is openly available on https://zenodo.org/record/3560886. The web application within which the experiment took place, which was developed specifically for the presented study, is available for reuse on https://github.com/stlemke/metrics-conjoint/.

I wish to thank my co-authors, the three (anonymous) reviewers for their helpful feedback as well as the journal's editors for their work on improving it.

[15] Full reference: Lemke, S., Mazarakis, A., & Peters, I. (2021). Conjoint Analysis of Researchers' Hidden Preferences for Bibliometrics, Altmetrics and Usage Metrics. *Journal of the Association for Information Science and Technology (JASIS&T), 72*, 777-792. https://doi.org/10.1002/asi.24445

2.2.2 Abstract

The amount of annually published scholarly articles is growing steadily, as is the number of indicators through which impact of publications is measured. Little is known about how the increasing variety of available metrics affects researchers' processes of selecting literature to read. We conducted ranking experiments embedded into an online survey with 247 participating researchers, most from social sciences. Participants completed series of tasks in which they were asked to rank fictitious publications regarding their expected relevance, based on their scores regarding six prototypical metrics. Through applying logistic regression, cluster analysis, and manual coding of survey answers, we obtained detailed data on how prominent metrics for research impact influence our participants in decisions about which scientific articles to read. Survey answers revealed a combination of qualitative and quantitative characteristics that researchers consult when selecting literature, while regression analysis showed that among quantitative metrics, citation counts tend to be of highest concern, followed by Journal Impact Factors. Our results suggest a comparatively favorable view of many researchers on bibliometrics and widespread skepticism toward altmetrics. The findings underline the importance of equipping researchers with solid knowledge about specific metrics' limitations, as they seem to play significant roles in researchers' everyday relevance assessments.

2.2.3 Introduction

In an age of exponentially growing publication rates (Bornmann & Mutz, 2015; Pautasso, 2012; Tian, Wen, & Hong, 2008) and with an abundance of publication-related usage data readily available (Moed, 2018), basing assessment of research outputs on quantitative metrics becomes ever more tempting. No matter whether on publication-, author-, journal-, or institution-level, calculating indicators based on citations is most often faster than engaging in thorough qualitative peer review. And with reliance on online platforms within researchers' workflows increasing (Kramer & Bosman, 2016), indicators from more sources arise to complement citations - often summarized by umbrella terms like “altmetrics” (e.g., an article's mentions in news outlets, social media, policy documents, gray literature, academic syllabi, or clinical guidelines) or “usage metrics” (e.g., an article's download counts or online view counts). Linked to these alternative indicators is the hope that they might provide means to assess wider forms of research impact, especially for domains like social sciences and humanities, where the applicability of bibliometric assessments is limited due to idiosyncratic publication- and citation norms (Hicks, 2005; Sivertsen & Larsen, 2012) and which lie outside the major citation databases' focus on STEM fields.

Background and Related Work

A large body of work from numerous fields of research has studied the ramifications of using quantitative indicators in academic evaluation and assessment systems (see Rijcke, Wouters, Rushforth, Franssen, & Hammarfelt [2016] for an overview). Frequently such studies conclude with warnings of inappropriate applications of the indicators, pointing to undesirable effects an extensive reliance on them could have on science. Common concerns for instance include that the use of indicators as auditing instruments could lead to researchers and institutions becoming more market-oriented, and as a consequence unduly focusing on research areas or publication types that have been shown to attain high metrics (Butler, 2005; Willmott, 2011). Such a climate could hinder certain kinds of research proposals, for example, particularly unusual or innovative and therefore risky projects (Butler, 2007). Similarly, a rising importance of indicators could lead to increasing publication pressure for researchers and put fields with lower overall metrics at a disadvantage for funding. Particularly fierce criticism against the use of citation counts for evaluative purposes is voiced by MacRoberts & MacRoberts (2018), who argue that the practice of citation analysis would be based on fundamentally false assumptions. They point to citation analysis' political implications by referencing Sosteric (1999, p. 19), who states that "citation analysis tends to support a particular one-sided, reified, and elitist view of scientific contributions that ignores […] certain groups of scholars and by doing so justifies the highly stratified nature of the academy where certain groups are privileged over others." It should be noted that the much-debated aspect of evaluation is only one of research metrics' major use cases. Looking at eight types of typical metrics users, the NISO Alternative Assessment Metrics Project identified three overarching themes of use cases for metrics: showcase achievements, research evaluation, and discovery (National Information Standards Organization (NISO), 2016). Albeit the project's focus was placed specifically on altmetrics, all three themes apply to bibliometrics and usage metrics as well.

Individuals' Perceptions and Usage of Metrics

Several surveys investigated how indicators are used and perceived by individuals, most of them focusing on researchers. Hammarfelt & Haddow (2018) surveyed humanities scholars from Australia and Sweden about their publication practice as well as about their knowledge and use of evaluation tools. They found a considerable share of Australian (62%) and a significantly lower share of Swedish (14%) respondents to actively use indicators. In a related study, Haddow & Hammarfelt (2019) found slightly larger but overall similar proportions of indicators users among Swedish and Australian social scientists. Both studies identified citations as the most frequently used indicator, common use cases being CVs, promotion, or grant applications, which in the NISO's use case scheme would fall into *Showcase achievements*. In their survey of Norwegian scientists, Aksnes & Rip (2009) found oftentimes "somewhat cynical" but ambivalent stances on the subject of citation counts. Ma & Ladisch (2016) conducted semi-structured interviews with four researchers about how they are affected by metrics usage. They identified three main themes of metrics usage: self-monitoring, collaboration, and choice of journals and research topics. In their follow-up study, Ma & Ladisch (2019) revealed a discrepancy between interviewed researchers' attitudes toward metrics in practice and in principle: while in principle, most researchers would not trust metrics as objective indicators for quality, in practice they do actively use them for personal or administrative purposes, thereby relying on them as supposedly objective measures. A similar discrepancy was found in a recent qualitative survey across faculty, instructors, and researchers at the University of Minnesota about their attitudes toward research metrics (Bakker, Cooper, Langham-Putrow, & McBurney, 2020). The participants stated several use cases metrics would fulfill for them, for example, in information-seeking activities, as a means of self-assessment, or in the assessment of other individuals. They did however also voice severe concerns linked to indicator use, for instance the feeling that metrics on their own could not serve as robust representations of the participants' own impact.

In an international study combining a long-term interview stage with a large-scale online survey, Nicholas, Herman, et al. (2020) specifically inquired about early career researchers' (ECRs) attitudes and practices toward citation-based metrics and altmetrics. Their analysis revealed citation-based indicators to fulfill several purposes for many ECRs, while altmetrics were regarded less favorably - criticism included that altmetrics might primarily reflect curiosity instead of impact, and that they would be vulnerable to being gamed. For the field of bibliometric research on the other hand, Haustein, Peters, Bar-Ilan, et al. (2014) attested altmetrics potential as a valuable source of impact data, as their analysis had shown large shares of bibliometric literature to be represented on online reference managers and substantial parts of their sample of bibliometricians to be affected by social media tools in their professional lives. Also concentrating on altmetrics, Aung, Erdt, & Theng (2017) surveyed members of academia regarding their awareness and usage of their different types, differentiating between non-faculty staff and faculty staff among the participants. They found a tendency for non-faculty staff to be more aware of altmetrics than faculty staff. Moreover, mentions and shares on social networks were found to be the most used altmetrics, and usage of most altmetrics was shown to correlate with usage of social media. In the second phase of their study, Aung et al. (2019) investigated more broadly on scholars' familiarity with and usage of both traditional indicators and altmetrics. They found only few indicators to be widely known among scholars and the familiarity with and usage of altmetrics to be particularly low. As potential reasons for altmetrics' low popularity, Aung et al. (2019) mention academics' insecurities regarding altmetrics' added value, missing encouragement of altmetrics- and social media usage from institutions, and privacy concerns, among others. The reoccurring finding of a comparatively low familiarity with altmetrics is in line with multiple other surveys across academic librarians (Miles, Konkiel, & Sutton, 2018) and faculty (Bakker et al., 2019; DeSanto & Nichols, 2017).

Other research about individuals' use and perception of indicators placed its focus on stakeholders from academic administration. Abbott et al. (2010) combined a poll among researchers with interviews of academic administrators to investigate the former's perception and the latter's use of metrics. They find that researchers tend to overestimate the actual importance of metrics when it comes to how research administration measures their achievements, as most administrators reported not to use metrics to a large extent. Interestingly, when asked what researchers think should be the criteria for evaluating their careers, metrics were chosen very frequently. Regarding the use of indicators by administrations, McKiernan et al. (2019) arrived at different conclusions by analyzing 864 documents used in review, promotion, and tenure processes of North American research institutions. Their data suggests that research-intensive institutions in particular make heavy use of impact factors for evaluative purposes, noting that cautious statements about indicator usage were very rare in the analyzed documents.

In an own previous effort of inquiring about researchers' perceptions, knowledge, and use of various indicators, we conducted a series of group interviews and online surveys with social scientists (Lemke, Mehrazar, Mazarakis, & Peters, 2019). Our study's results revealed that even though many researchers are justifiably wary regarding metrics' reliability as indicators of scientific relevance or quality, they still are inclined to use some of them - particularly citation-based ones - quite regularly. Frequently metrics serve as filters, for example, during literature research, when deciding which sources to cite, or when planning where to publish own works. Also, while most specific concerns about using metrics for research evaluation voiced by the respondents could apply to all types of metrics (i.e., bibliometrics, altmetrics, usage metrics) in equal measure, altmetrics performed considerably worse than bibliometrics considering perceived usefulness and trustworthiness.

Previous Research on Researchers' Literature Selection Processes

Researchers' information-seeking behavior and the criteria with which they decide on literature to read have also been examined in previous studies, although not necessarily with a focus on properties of the documents that are available, like for instance their metrics, but more often on the readers themselves (Tenopir, King, Spencer, & Wu, 2009). Through building regression models on survey data from 2,063 researchers from natural sciences, engineering, and medical sciences, Niu & Hemminger (2012) found numerous factors to affect scientists' information-seeking behavior, proposing a framework that includes demographic, psychological, role-related, and environmental factors. The most important determinants of information-seeking behavior found included the academic position, gender, and discipline. Looking at both demographic and contextual factors, Tenopir et al. (2009) used an online survey to analyze reading patterns of academic staff of universities from the US and Australia. They found subject disciplines and work responsibilities to be important characteristics determining reading behavior - for instance, while medical faculty tend to read more, specifically for current awareness, engineering faculty tend to spend more time per article and more often read for research than others. In a more recent large-scale survey, Tenopir et al. (2016) examined which activities the over 3,600 participating researchers would find most important to determine an article's trustworthiness, finding "checking if the arguments and logic presented in the content are sound," "checking to see if the data used in the research are credible," and "reading the abstract" being ranked most highly. Overall, they found participants to quite consistently value content properties more than respective articles' meta-information, for example, author or publisher names. On the other hand, regarding criteria for judging reading trustworthiness, the most highly rated statement was "Peer-reviewed journals are the most trustworthy information source," indicating the type of publication venue to be an important aspect in this regard. In another international survey aimed at early career researchers, Nicholas, Jamali, et al. (2020) found a journal's prestige, rank, and impact factor as well as ease of access to be influential factors for participants when deciding what to read.

Research Problem

While several previous studies examined researchers' overall stance and practices toward different types of indicators for research assessment, little empirical research investigated indicators' role in the concrete practical decisions researchers take virtually every day when selecting literature to read. Furthermore, previous studies in this field are mostly restricted to surveys and interviews, which share some methodological weaknesses, for example, vulnerability to certain response biases. In this study, we approach the question of how researchers make use of indicators for evaluation purposes from a new angle by introducing conjoint analysis to the context of scientometrics. We designed an interactive online experiment in which invited researchers were asked to rank fictitious publications against each other regarding their expected scientific relevance - three publications at a time. Participants had to base their judgments on a set of preselected bibliometrics, altmetrics, and usage metrics, of which individual manifestations were presented for each single article. Through applying regression analysis we could later determine how different types of indicators overall affected the rankings made by participants of our experiment. The experiment was enclosed by a questionnaire that should help to put the results from our conjoint analysis into context.

With this study's results, we try to expand on insights from Lemke et al. (2019) on how indicators influence researchers' decisions during literature selection. By examining whether the researchers' ranking behavior in a fictitious practice situation complies with statements made in previous interviews and surveys, we aim to achieve a more truthful picture of researchers' perceptions of indicators than surveys and interviews alone could possibly provide us with. Moreover, we aim to detect and describe distinguishable types of indicator users by clustering respondents based on similarities in their ranking behavior. This way, the data obtained in our conjoint experiment should allow us to make detailed observations on the comparative values of individual indicators for certain users, which would be extremely difficult to obtain in such granularity in a regular survey. We primarily focus on researchers from social sciences, as our previous experiences with this target group should help us to better put our findings into perspective. Also, due to the Social Sciences being subject to discipline-inherent limitations regarding the applicability of citation-based assessments (see for instance Hicks, 2005), it should be particularly interesting to gain further insights into how open researchers from this domain are toward using alternative indicators, which might circumvent some of citations' shortcomings (Wouters & Costas, 2012), in relevance assessments.

We chose the specific use case of utilizing indicators for reading prioritization during literature research as our experiment's core scenario for two reasons: first, we can assume most researchers to regularly be in the situation of encountering individual articles' metrics as possible filter criteria during online literature research and thus being familiar with it. This was also evidenced by responses during our previous interviews with social scientists (Lemke et al., 2019), which showed that even researchers in very early career stages are aware of article-level metrics and sometimes use them as filters in this particular scenario. Second, in much previous literature about researchers' perceptions of indicators, an emphasis often lies on use cases surrounding their own evaluation or their choice of publication channels (see for example Abbott et al., 2010; Haddow & Hammarfelt, 2019; Hammarfelt & Haddow, 2018; Ma & Ladisch, 2016). How researchers perceive and use metrics when evaluating others' work is in our eyes an equally interesting question that seems to have received less attention from the scientific community so far.

Conjoint Analysis

The term conjoint analysis encompasses a variety of decompositional multivariate techniques with the goal of estimating consumers' preference structures (Green & Srinivasan, 1978). In most implementations of conjoint analysis, a sample of the respective target population is asked to evaluate multiple differing alternatives of a product in question in an experimental setting (a so-called choice- or ranking experiment). The presented alternatives feature different manifestations of attributes, which are hypothesized to be relevant for the participants' decisions. After the collection of data, statistical techniques (e.g., regression models) are used to model participants' preferences and allow for conclusions about individual attributes' effects on participants' overall choices. One main benefit of conjoint analysis is that it comparatively truthfully emulates participants' real decision-making situations, as participants evaluate a respective product's attributes implicitly while evaluating whole products - as they would most likely have to in the real world.

Originally stemming from the fields of mathematical psychology and psychometrics (Green & Wind, 1975), conjoint analysis has been used prevalently in marketing where it is typically utilized to estimate consumer behavior. However, conjoint analysis can in principle be applied to innumerable scenarios from other domains in which human subjects choose between multiple alternatives with the goal of maximizing fulfillment of their personal preferences. For an example for an application of conjoint analysis in the field of Computer Science see Kirchhoff, Capurro, & Turner (2014), who used it to assess users' “preferences” for different types of errors made by machine translation engines. Tenopir et al. (2011) even utilized conjoint analysis in context of a research question from the sphere of scholarly communication similar to ours: controlling for seven different characteristics of research articles, they found article topic, online accessibility, and peer review status to be the most important factors for researchers when deciding which articles to read.

To the best of our knowledge, conjoint analysis has not been used for assessing preferences of users of research indicators yet. In this study, we combine methods of conjoint analysis with an online survey to investigate factors that influence researchers' decisions about which scientific publications to consume with priority. In particular, we inquire about the role quantitative indicators play in this - how do different types of indicators compare regarding their perceived utility as selection-criteria during literature research? Also, the majority of previous studies on researchers' perceptions and use of indicators focused on bibliometrics. We address this gap by, in addition to bibliometric indicators, also incorporating a selection of altmetrics and usage metrics in our study.

2.2.4 Methods

The following sections provide information on the planning, implementation, and dissemination of our ranking experiment as well as on the methods used for analysis of the collected data.

Definition of Attributes and Levels

A key decision when planning a conjoint analysis concerns the attributes that should be incorporated in the experiment - meaning the features of the products in question that participants are meant to base their preference judgments upon. As we want to examine how different quantitative indicators influence researchers' choices when deciding which literature to read first, the *products* in our experiment will be fictitious scientific articles that showed up as results of our participants' hypothetical literature search, while their *attributes* will be a selection of indicators, for example, their individual citation counts, or numbers of mentions on Twitter. Different articles sport different *levels* of those attributes - one article could for example have five citations and five mentions on Twitter, while another article has zero citations but 250 mentions on Twitter.

The decision about how many attributes and how many levels per attribute to include in a conjoint analysis is closely connected to the chosen method for data collection. Two fundamentally different approaches for data collection exist (Green & Srinivasan, 1978; McCullough, 2002): partial-profile and full-profile approaches. In full-profile designs, participants are asked to rank products for which individual data on all of their attributes is visible; in partial-profile approaches, participants can only see a fraction of the attributes for all products during each ranking task. While partial-profile approaches make sense in scenarios with extremely high numbers of relevant attributes (such designs can include up to 50 or more attributes; McCullough, 2002), we decided for a full-profile design, as it provides more realistic and comprehensible tasks if the number of included attributes is fairly low.

To prevent our participants from information overload, we decided to follow the common recommendation for full-profile designs to include a maximum of six attributes (Green & Srinivasan, 1978; McCullough, 2002). An initial list of potential indicators to include as attributes was created through a combination of literature review, inspection of the data sources covered by prominent altmetrics providers, and brainstorming. Due to their extremely diverse sources, especially the different types of altmetrics quickly led this initial list grow to more than 20 entries. When we felt a saturation regarding the incorporation of further relevant indicators to be reached, the list was then iteratively reduced to six indicators we deemed to have particularly high presence and relevance in scientometric literature and practice. We balanced this criterion with the aims of incorporating at least one prototypical indicator from each of several different areas of metrics and picking indicators that as many participants as possible should already feel familiar with:

- the article's *citations* (e.g., on *Google Scholar*) as an article-level bibliometric indicator;
- the publishing journal's *Journal Impact Factor* as a prominent and much-debated journal-level indicator;
- the first author's *h-index* as a widely known author-level indicator;
- the article's number of *downloads* as an article-level usage indicator;
- the article's number of mentions in *tweets* as an altmetric drawn from a prominent social media platform targeted at a general audience;
- the article's number of readers on *Mendeley* as a comparatively well-examined altmetric drawn from a social media platform targeted at scholars.

Ideally, for each attribute the same number of levels should exist in the experiment to prevent attributes' numbers of levels from having an effect on individual attributes' estimated importance (McCullough, 2002). Moreover, different levels of the same attribute should be easily perceivable as distinct; the ranges covered by them may be slightly larger than in reality, but not so large as to be unbelievable (Green & Srinivasan, 1978). We decided to include three levels per attribute: level 1 should intuitively translate to *no occurrences of this indicator*, level 2 to *few occurrences of this indicator* and level 3 to *a high number of occurrences of this indicator*. Starting from this, the authors discussed and agreed upon values they deemed to be plausible for the examined domain while being in accordance with the advice from Green & Srinivasan (1978). Table B1 shows the resulting three levels for our six attributes. An example for how a publication profile based on this selection of attributes and levels finally looked like in our experiment is shown in Figure B1. It should be noted that in the actual experiment the order in which attributes were listed was randomized between participants to reduce a possible influence of their order of appearance on individual attributes' measured effect. The image of an article's front page on the left was depicted for illustration only and did not provide any bibliographic information, as participants should base their judgments solely on our six pre-selected attributes.

Table B1: Attributes and their levels in the experiment

	Level 1	*Level 2*	*Level 3*
Citations (e.g., on Google Scholar)	0	5	250
Journal Impact Factor	0	5.0	30.0
h-Index	0	5	30
Downloads	0	100	5,000
Tweets	0	10	500
Mendeley readers	0	10	500

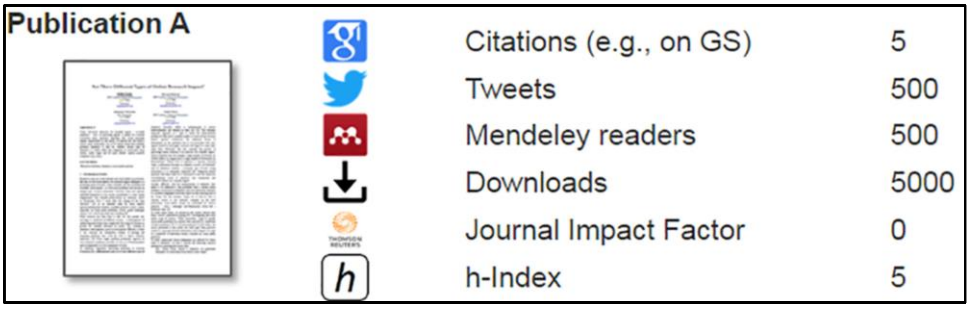


Figure B1: Random example publication based on our choice of attributes and attribute levels.

Design of Tasks

After the included attributes and their levels were defined, decisions had still to be made about how many alternative publications a participant should have to compare during a single task and about how many tasks each participant should be asked to complete during their run of the experiment. Based on pretests with a first prototype of the software used for our experiment (see below), we decided to let participants assess three fictitious articles at a time - this should keep individual tasks short and comprehensible. We planned for every participant to complete a total of 20 tasks, an amount that should be solvable in approximately 15 minutes without degradation of data quality due to participants' fatigue (McCullough, 2002). Following advice by McCullough (2002), we decided to regard the first two tasks completed by every participant as warm-up tasks which would not be considered during analysis. This should account for the fact that it can take a little while for participants' behavior to stabilize, as they might only get a feeling for the experiment's concept and scale after having completed one or two tasks. For the analysis, this would leave us with 18 tasks to evaluate for every participant who completed the full experiment.

To come up with a definite set of 20 tasks to be used in the experiment, we largely followed the guideline for the design of choice experiments using R by Aizaki & Nishimura (2008). This included the creation of a full factorial design based on our predefined attributes and levels and the subsequent generation of a fractional factorial design matching the number of tasks in our experiment. The approach makes use of Federov's exchange algorithm (as implemented in Bob Wheeler's R package *AlgDesign*[16]) that, given our restrictions regarding number of attributes, levels, and tasks per participant, creates a combination of tasks to include in our design to make model estimation as efficient as possible.

Another fundamental question during the planning of conjoint analysis experiments concerns the way in which participants are asked to express their judgments. Two widely used paradigms exist (Louviere, Flynn, & Carson, 2010): the “traditional,” rankings-based conjoint analysis and “discrete choice experiments” (DCEs), in which participants have to choose exactly one option out of the given alternatives in each task. In our experiment we want to emulate the situation of researchers prioritizing between different articles found during literature research - a scenario in which the respective researcher will usually intend to read several of the articles a search has led her to, not only the single most appealing one. We therefore believe a ranking of articles to more realistically represent the kind of decisions in question than a discrete choice. Moreover, literature has shown that rank order data can be expanded into sets of implied discrete choices (R. G. Chapman & Staelin, 1982; Louviere et al., 2010; Vermeulen, Goos, & Vandebroek, 2011). For us this means that relying on a ranking-based method of data collection increases the amount of information obtained per task without leading to a significant loss of flexibility regarding data analysis.

Implementation

We implemented our experiment as a web application, so that invited participants could access it via a web browser. Our software is open-source and can be obtained from GitHub.[17]

Figure B2 schematically summarizes the experiment's course a participant would follow after clicking on an invitation link. To collect demographic data as well as free text responses on our participants' thoughts regarding their literature research practices, the use of metrics, and our experiment in general, the experiment both began and ended with questionnaire segments (pages B, C, G, and H) surrounding the actual 20 tasks a participant was asked to complete (pages F). Pages D and E gave detailed information on the indicators referred to in the experiment and explained how to navigate through the tasks. A file with full screenshots of all pages of the experiment is available in this article's supplementary material.

[16] https://CRAN.R-project.org/package=AlgDesign
[17] https://github.com/stlemke/metrics-conjoint/

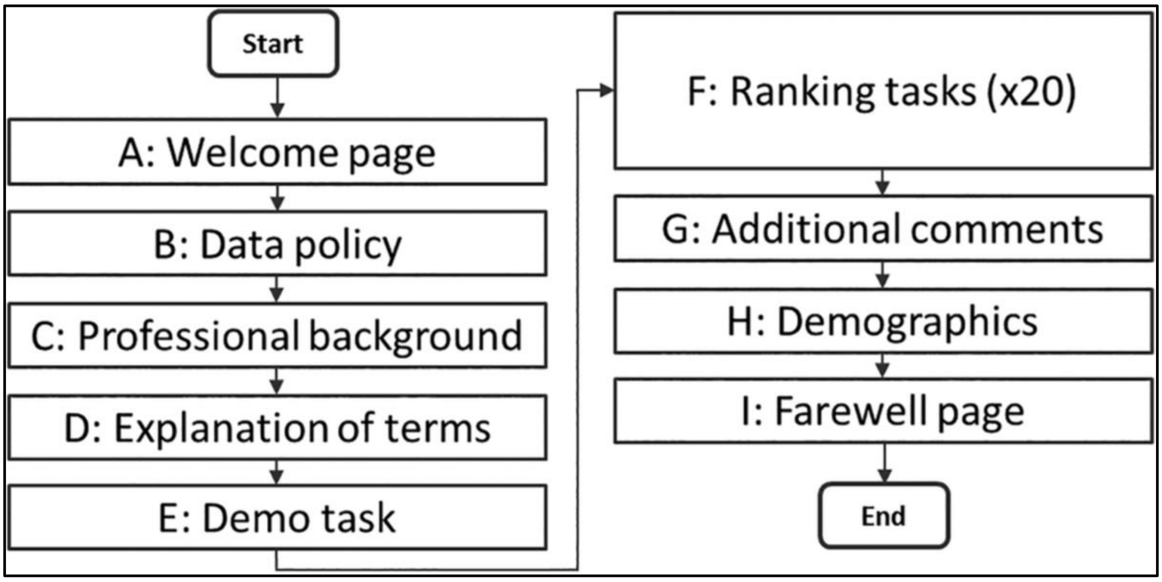


Figure B2: Schematic representation of experiment.

Figure B3 shows an example for a task in our experiment. Each task started with the same explanatory text seen in Figure B3. Below this text, a set of three fictitious publication profiles from our fractional factorial design was presented as boxes. All boxes could be moved and rearranged per drag-and-drop to achieve a desired order. As soon as all three publications had been allocated to the three slots on the right, a participant could move on to the next task by clicking the continue-button. Just like the order in which attributes were listed on publication profiles was randomized between participants, the order of publication profiles in each task was randomized as well to rule out giving an advantage to certain publications by always listing them first.

The software was tested iteratively with multiple rounds of feedback coming from about a dozen of colleagues from both within and without of our research team, to ensure it being as self-explaining and technically sound as needed. To also make sure that our way of processing input data would later allow us to perform all planned steps of analysis (see below) as intended, we simulated later steps once by automatically generating hypothetical input data for 30 fictitious participants. After successful simulation of data analysis, this software-generated input data was discarded.

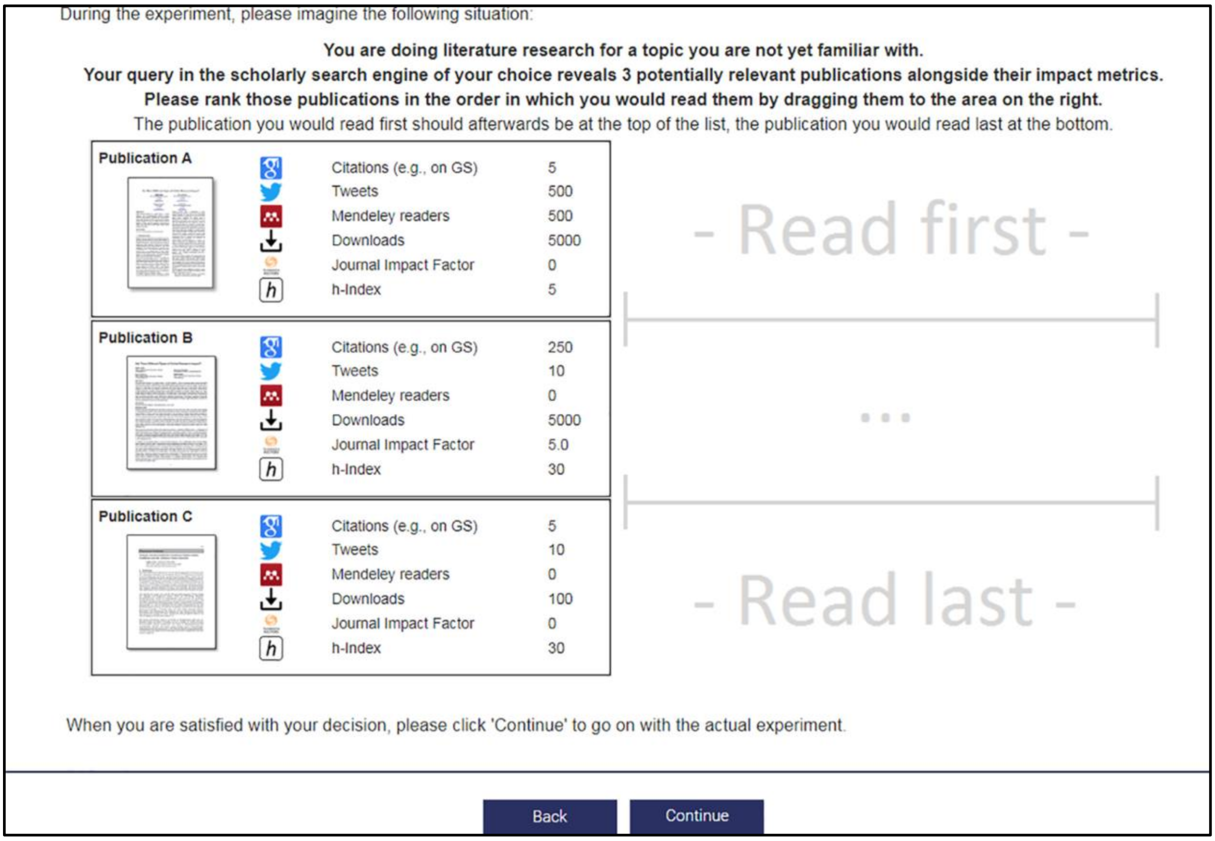


Figure B3: Example ranking task from the experiment.

Dissemination

For the experiment's dissemination, we relied on a subsample of the respondents from a survey among researchers we had conducted in the summer of 2018 (Lemke et al., 2019). Said survey had been sent to authors of social science-related papers found on RePEc and Web of Science, as well as to a mailing list maintained by the *ZBW Leibniz Information Centre for Economics*. The latter consisted of about 12,000 email addresses of researchers, with a strong focus on economists from German-speaking parts of Europe. Of the survey's 2,083 respondents, 938 had agreed to be contacted again for further user studies from our project and therefore received invitations to this experiment half a year later. The dissemination of invitation links, which would allow participants to access the experiment's website, was done via email between November 29 and December 10, 2018. Data collection was carried out till January 15, 2019. As an incentive, participants could optionally partake in a random drawing of fifteen 30€-vouchers for Amazon after completing the experiment.

Data Analysis

Rank orders of publications entered by participants during the experiment were regarded as sequences of choices (Vermeulen et al., 2011). So if in a task participant *P* had ranked the three publications *A*, *B*, and *C* in the order $C > B > A$, for the purposes of our analysis this input would be treated as two parts of information: "between *A*, *B*, and *C*, participant *P* chose *C*" and "between *A* and *B*, participant *P* chose

B". The decision data from all participants (barring each participant's first two tasks, which were regarded as warm-up tasks as described above) was fed into a logit model using R, with the six indicators included in the experiment as independent variables and the binary outcome of a publication being ranked above its competitors as dependent variable.

In a subsequent step, we performed a cluster analysis to identify groups of participants with similar ranking behaviors. To be able to do so, we first for every participant transformed all of their ranking choices to numerical vectors. On these vectors, we then used *k*-means clustering. For every one of the participant groups identified through our cluster analysis we then again computed individual regression models, like we had done before for all participants combined.

Through survey segments right before and after the experiment's ranking tasks, we collected participants' demographics as well as further information on their usual strategies during literature research and on their notions about research indicators. Both from previous literature (Nicholas, Jamali, et al., 2020; Niu & Hemminger, 2012; Tenopir et al., 2011, 2009, 2016) as well as from personal experience we had reason to believe that in real literature research scenarios the respective researchers would often determine their orders of preference based on more complex heuristics, involving more criteria than the six indicators we could ask for in our experiment. As for instance Tenopir et al. (2011) have shown, there are in particular several qualitative aspects researchers look out for when deciding what to read. To not disregard such aspects, particular attention during the analysis went into the responses to two free text questions:

- Right **before** the ranking tasks, we asked every participant: "*When doing literature research, how do you usually determine which search results to read first? Are there publication features you are looking out for?"*
- Right **after** the ranking tasks, we again showed every participant what he or she had responded to the first free text question and asked: *"Now after having finished the experiment, would you like to add anything to your previous answer?"*

The responses to both free text questions were coded manually by one author (S. L.) and grouped by topics they referred to.

2.2.5 Results

In this section, we present our participants' reported demographics, regression models of the choices they made during our experiment, results from the cluster analysis based on said choices, and an analysis of the free text answers our participants gave.

Demographics

In total, 247 of the 938 researchers we had invited participated in our experiment by completing at least its first survey page, meaning a response rate of 26%; 204 participants finished the experiment completely. Figure B4 shows the number of participants that completed a respective page transition.

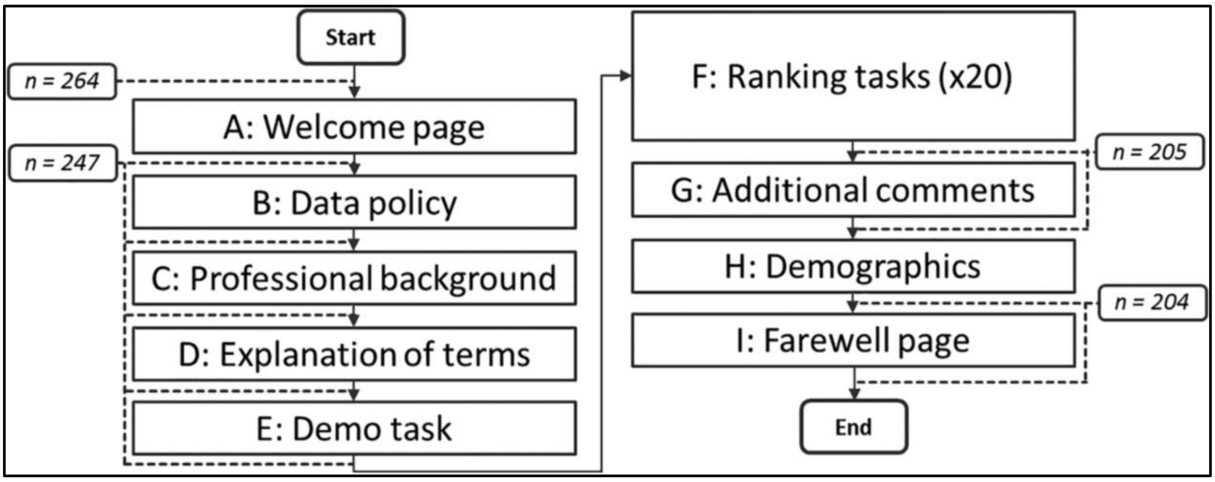


Figure B4: Schematic representation of experiment with numbers of completing participants.

Table B2 shows the participants' stated demographics concerning their discipline, professional role, country of affiliation, and gender.

Table B2: Participants' demographics

Discipline	*Percentage*	*Professional Role*	*Percentage*
Economics	61%	PostDoc/Senior Researcher	22%
Social Sciences	21%	Professor	18%
Other	9%	Assistant Professor	16%
Engineering/Technology	3%	Associate Professor	16%
Life Sciences	3%	PhD Student	9%
Arts/Humanities	2%	Research Assistant + PhD Student	9%
Law	1%	Other	7%
Medicine	<1%	Research Assistant	2%
		Student/Student Assistant	1%

Country of Affiliation	*Percentage*	*Gender*	*Percentage*
Germany	34%	Male	68%
USA	11%	Female	29%
Italy	8%	I prefer not to answer	3%
UK	6%		
Spain	3%		
France	3%		
Poland	3%		
Other (number of other countries = 31)	32%		

The participants' median year of birth is 1980, with a standard deviation of 10.22 years and a range from 1946 to 1993 (a single response of “1900” to the question for year of birth was discarded as implausible). Regarding reported years of academic experience, the median is at 11 years, with a standard deviation of 8.75 years and a range from 2 to 43 years.

First Regression Model

In total, 4,222 comparison tasks were completed by our respondents, up to 20 by each one of them. As we discard the first two tasks solved by every respondent as warm-up tasks, 3,774 evaluable tasks remain. Because every task completion actually consists of two separate choices as described earlier, the amount of evaluable choices to analyze is 7,548.

Table B3 shows the parameters of the logit model created on basis of those 7,548 choices. All estimated coefficients are standardized to respective attributes' standard deviations to facilitate comparisons despite the attributes' differing absolute ranges.

We see that for every one of the six indicators an increase also leads to an increase in the respective article's likelihood of being ranked higher than its competitors. The strongest effect per standard deviation have *citations* - increasing an article's citations by one standard deviation increases its log odds of being preferred to competitors by 0.607. The second-highest effect has the *Journal Impact Factor* (0.468), followed after a considerable gap by *downloads* (0.247). The remaining three indicators all perform similar to each other with coefficients close to 0.160.

Table B3: Coefficients for logit model of all participants' ranking data

Variable	*Estimate*	*Std. Error*	*p*
(Intercept)	-0.419	0.016	<.001
Citations (e.g., on Google Scholar)	0.607	0.017	<.001
Journal Impact Factor	0.468	0.016	<.001
h-Index	0.160	0.017	<.001
Downloads	0.247	0.016	<.001
Tweets	0.159	0.016	<.001
Mendeley readers	0.157	0.017	<.001

Most Helpful Indicator

Right after completing the 20 ranking tasks, participants were asked which one of the six indicators they would find *"most helpful as a tool for deciding which publications to read"*. Figure B5 shows which shares of participants chose which indicator.

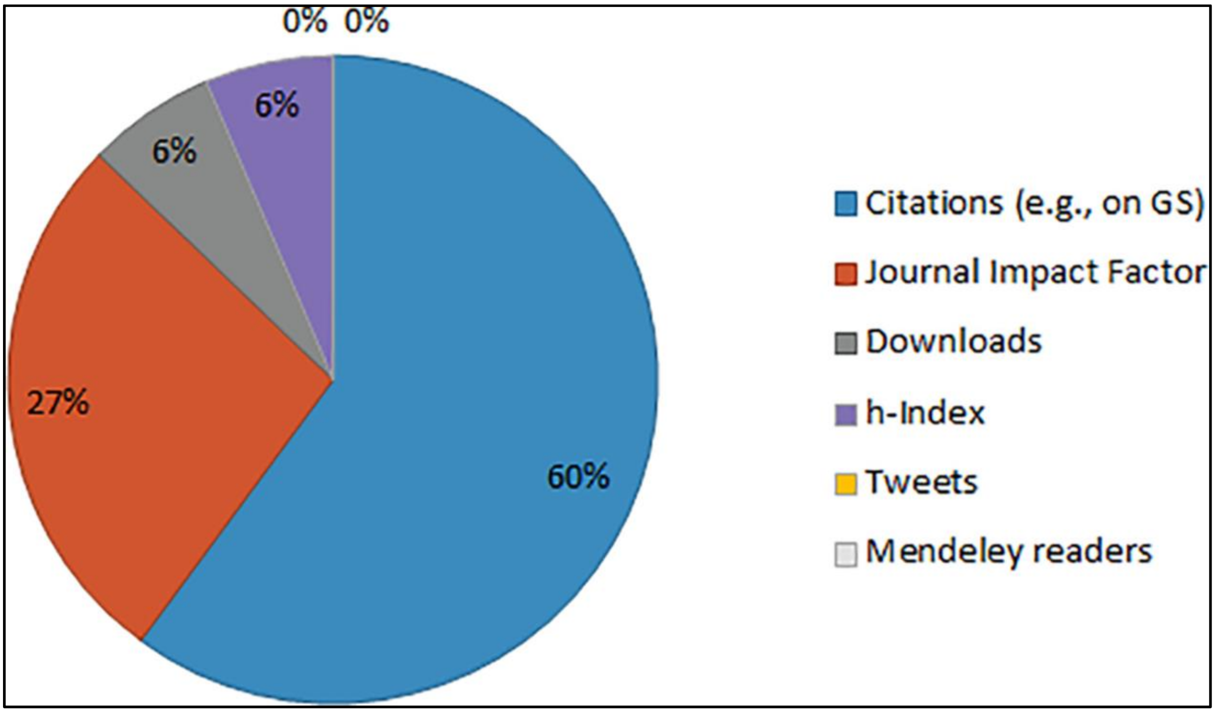


Figure B5: Indicators from the experiment selected as most useful for deciding which publications to read (n = 203).

The rank order of indicators' perceived usefulness achieved this way largely confirms the results from the regression, with most participants assessing citations as being most useful, followed by the Journal Impact factor. Equal numbers of participants chose either downloads or h-index as the most useful indicator. Remarkably, not a single participant chose Mendeley readers or Tweets as preferred indicator, despite our earlier observation that both of these indicators have similarly strong effects on an article's likelihood of being preferred over its competitors as the h-index. A possible explanation for this could be that the two social media-based indicators are only ever used when consulting the preferred indicators does not lead to a clear decision - in these cases the altmetrics could serve as “tiebreakers”.

Clusters of Participants and Their Regression Models

To be able to see whether there are groups of participants that can be distinguished by their ranking strategies, we performed a cluster analysis based on our dataset of ranking choices. Data from participants who had not completed all 20 tasks of the experiment was discarded for the cluster analysis, leaving us with 205 respondents to include. We started our *k*-means clustering with a *k*-value of 2, which we gradually increased until the addition of further clusters would lead to some very small groups including less than 20 respondents per cluster. We this way found $k = 4$ to be a satisfying configuration for getting distinct clusters of significant sizes. The resulting clusters C1 to C4 consisted of 65, 67, 20, and 53 respondents respectively. Detailed tables of each cluster's demographics can be found in this article's supplementary material. The descriptive statistics indicate that the demographic differences between clusters are overall low.

Table B4 summarizes the coefficients of the logit models we estimated for the four individual clusters, just as we had done before for all participants combined.

In the following we briefly characterize the four clusters' ranking behaviors:

Cluster 1: “bibliometrics-believers” - this fairly large group of participants trusts in citations and their derivatives, but seems very skeptical of all offered usage- and altmetrics.

Cluster 2: “Impact Factor-fixated” - this equally large group assigns very high value to the Journal Impact Factor. Apparently, when having to rank based on article- or author-level metrics, no indicator is rejected completely; also noteworthy is this group's comparatively high acceptance of Mendeley counts.

Cluster 3: “usage evidence enthusiasts” - the smallest of all clusters, these participants seem to trust in usage metrics (downloads) and traditional citations as indicators for relevance. Also, they are the group with the lowest interest in non-article-level indicators.

Cluster 4: “open-minded citation users” - like the participants from cluster 3, these cluster's members also have a stronger focus on article-level indicators (and particularly citations) than most other participants, although they do not seem to reject any one indicator completely.

Table B4: Coefficients for cluster-wise logit models

Variable	*Estimates*			
	C1	*C2*	*C3*	*C4*
n	65	67	20	53
Citations	0.853***	0.302***	0.332***	0.935***
Journal Impact Factor	0.486***	0.808***	0.146**	0.248***
h-Index	0.227***	0.167***	0.086	0.171***
Downloads	0.052	0.188***	0.603***	0.455***
Tweets	0.097**	0.171***	0.148**	0.286***
Mendeley readers	0.083**	0.274***	0.104*	0.167***
Cluster description	*"bibliometrics-believers"*	*"IF-fixated"*	*"usage evidence enthusiasts"*	*"open-minded citation users"*

*** $p < .001$; ** $p < .01$; * $p < .05$.

Free Text Responses

A substantial number of participants also provided answers to two free text questions we asked before and after the ranking task-part of the experiment. Not counting non-topical answers (such as "no comment"), 205 participants responded to the question *"When doing literature research, how do you usually determine which results to read first? [...]"*, which participants had been asked right before the start of the ranking tasks. Another 132 topical responses were given to the question *"Now after having finished the experiment, would you like to add anything to your previous answer?"*, which was shown immediately after a participant had finished their 20 ranking tasks, alongside the response that person had given to the first question if applicable. All responses were coded regarding individual selection criteria referred to in them. Table B5 shows these criteria, together with the numbers of responses they have been mentioned in before and after the ranking tasks.

The free text responses confirm that there is a variety of qualitative features researchers tend to look at when deciding about an article's relevance for themselves, for example, its title, abstract, authors, topics, and date of publication. But also quantitative indicators seem to play an important part: citation counts were mentioned as used by many participants both before and after the experiment, download counts and an author's h-index were confirmed to be of interest by several researchers after they had finished the ranking tasks. The most frequently mentioned selection criteria of all was the journal an article is published in, although it is often not possible to tell from the responses whether respective respondents base their judgments about a journal's appeal on quantitative indicators like the Journal Impact Factor or on qualitative criteria.

Table B5: Selection criteria used during literature research as mentioned in free text responses

Criteria	*Mentioned before*	*Mentioned after*
Journal (Prestige/Ranking/Impact Factor)	86	39
Title	50	1
Citation counts	47	61
Abstract	47	1
Authors	42	9
Date of publication/Recency	42	10
Topical proximity	40	1
Keywords	25	-
Other	20	8
Reference-relations (e.g. “cited by”-criteria)	16	3
Publisher/Online source	11	-
Order of appearance in search engine	11	-
Availability/Access	11	-
Content properties (e.g. article length, methodology)	10	1
Publication type	6	-
Downloads	-	32
H-index	-	24
Mendeley readership counts	-	2
Tweets	-	4

Apart from mentioning selection criteria for articles, some respondents used their free text answer after the experiment to express criticism of specific indicators. Table B6 shows the numbers of occurrences of such comments. The comparatively high frequency of arguments against altmetrics is in line with our previous observations of participants' reluctant use of them, although citations are the only indicator that did not provoke any specific criticism at all.

Table B6: Arguments against indicators stated in free text responses after the experiment

Comment	*Mentioned after*
Argument against tweets	11
Argument against downloads	5
Argument against Mendeley readership counts	3
Argument against h-index	2
Argument against Journal Impact Factor	1

As a rough check of whether respondents' choices during ranking tasks had been in accordance with the criteria they had explicitly reported to look out for, we also examined free text responses for each of the four respondent clusters (Table B4) individually. Comparing the percentages of respondents from clusters that explicitly mentioned certain metrics as relevant selection criteria with the respective clusters' regression coefficients as reported in Table B4 revealed an overall high degree of congruence between ranking behavior and free text responses. More detailed results of this consistency analysis can be found in the supplementary material.

2.2.6 Discussion

We conducted an interactive experiment to investigate researchers' usage and preferences regarding quantitative indicators when assessing literature's relevance on a micro-level. The regression models based on participants' ranking choices as well as the survey answers revealed clear preferences for bibliometric indicators, first and foremost citation counts, followed by the Journal Impact Factor and usage metrics in form of download counts. While the author-based h-index and the two altmetrics included in the experiment exhibited similar effect sizes in our main regression analysis, both selection- and free text-based survey responses suggested a particularly widespread wariness toward the use of altmetrics. Our clustering of participants based on their ranking data indicated that several groups of indicators users that follow different strategies regarding their use of indicators for relevance assessment can be distinguished. The analysis of free text responses suggested that in practice researchers inspect both quantitative and qualitative properties of research articles to decide which publications to read, in line with previous studies on researchers' reading decisions (Nicholas, Jamali, et al., 2020; Tenopir et al., 2011).

Comparing our results to the conjoint analysis of article characteristics that researchers value by Tenopir et al. (2011), we can see that every qualitative characteristic they included in their model also in some form came up in the free text responses to our study, confirming these characteristics' general relevance. If we compare the rankings Tenopir et al. (2011) obtained through their conjoint analysis with the frequencies with which individual characteristics came up in our free text responses, we can see differences: while the respondents of Tenopir et al. (2011) put particular emphases on articles' topics and matters of online accessibility, for our respondents an article's publication venue seems to be of importance. An explanation for our respondents' lower focus on aspects of availability could be the high share of economists in our sample, who due to their prevalent reliance on openly accessible working papers might less regularly experience difficulties regarding article availability.

The free text responses we obtained evince that quantitative metrics for research assessment do play a role in researchers' everyday decisions. Moreover, our study shows that many researchers are considerably more open to the use of bibliometrics and in some cases usage metrics as indicators of scientific relevance than to the use of altmetrics. These findings are in line with observations made in previous surveys and interviews that revealed a critical stance many researchers have regarding both altmetrics as relevance indicators and social media platforms as channels for scholarly communication (Aung et al., 2019; Lemke et al., 2019; Nicholas, Herman, et al., 2020). The rank order *bibliometrics* > *usage metrics* > *altmetrics* also coincides with findings by Miles et al. (2018) about academic librarians' familiarity with different types of research impact indicators. It stands to reason that also for researchers their reluctance to use web-based indicators can at least in part be explained by their lesser familiarity with them compared to citation-based metrics (Aung et al., 2019; Bakker et al., 2019; DeSanto & Nichols, 2017; Lemke et al., 2019). Also, citation-based indicators are the type of indicator that (still) counts the most in the academic reward system.

Clearly, a certain amount of caution against basing judgments about individual articles' scientific relevance on any quantitative indicator is advisable - in recent years several high-profile statements have been publicized by experts warning of purely quantitative micro-level assessments as an exemplary form of indicators' misuse (Cagan, 2013; Hicks et al., 2015; Wilsdon et al., 2015). It might therefore be considered unfortunate that the free text responses our participants gave about how they usually determine an article's relevance strongly indicate that their widespread skepticism against altmetrics does not translate to citations and their derivatives. Also, as Ma & Ladisch (2019) have shown, does even a lack of trust in an indicator's objectivity not prevent researchers from using it for certain assessments. This underlines the importance of establishing a level of “metrics literacy” among researchers regardless of discipline that allows them to gauge various impact metrics' respective scopes, strengths, and limitations and avoid misinterpretations (Ma & Ladisch, 2019; see also Rousseau & Rousseau, 2017). For advocates of altmetrics, who aim to broaden the arsenal of tools used for impact measurement in an effort of mitigating the power given to one single indicator, our results show that there is still a lot of work ahead. Additional efforts of informing researchers about and familiarizing them with alternative indicators will be necessary to enable them to make use of the various web-based complements to bibliometrics that exist today.

Limitations of the Study

A limitation of our study lies in its sample, which had a strong focus on social sciences and in particular economics. We assume that many researchers' acceptance of certain metrics as relevance indicators will be affected considerably by their discipline, given the well-documented substantial differences regarding certain metrics' applicability to different fields (Hicks et al., 2015; Thelwall, 2018a). Results from this study can therefore not be generalized to other disciplines. Also, potential self-selection bias and the experimental setting might have led to an overemphasis on comparatively tech-savvy participants as well as on those with high interest in the topics of research assessment and/or impact indicators.

Moreover, although we aimed to base our choices of attributes and levels on established guidelines for conducting conjoint experiments where it seemed feasible, these choices remain arbitrary to a degree. One implication of this is that we might have missed indicators of particular value for our target group. Especially for fields with a high reliance on non-journal article publication formats, as is the case in many social sciences and humanities, the future inspection of perceived values of altmetrics based on formats like gray literature, books, or syllabi might be insightful. Another limitation concerns the choice of levels, as despite our measures taken to confront our participants with a balanced experimental design, we cannot rule out that the attributes' individual level ranges affected our outcomes. For instance, an attribute manifestation perceived as disproportionately valuable could lead participants to ignore other attributes when exposed to it. The perhaps most risky example in our experiment was the highest Journal Impact Factor level of 30 - while most researchers should be able to envision examples for respective articles by thinking about multidisciplinary mega-journals, for mono-disciplinary journals of many fields this value would constitute extreme outliers. Our use of a fractional factorial design should however reduce the overall potential for one outstanding attribute manifestation to distort the results. Another approach would have been to base attribute levels on real articles, although this might come at the cost of making differences between articles harder to distinguish for the participants (see also the advice by Green & Srinivasan [1978] on defining levels in conjoint experiments).

Another aspect that is not accounted for in an experiment like ours is the comparative difficulty with which certain information is obtainable in reality. For instance, on a platform like *Google Scholar* citation counts and even a first author's h-index might be accessible with comparative ease, while the less common altmetrics might be considerably harder to obtain. So while the experiment's assumption, all indicators would be readily available during search, has for instance led us to find that h-index and tweet mentions are overall valued to similar degrees, in reality the h-index might still exert a stronger influence on reading decisions due to its easier availability.

Finally, in our experiment's survey part we did not explicitly state that participants can assume topical relevance for all hypothetical literature finds to be given. While we do not expect this fact to have meaningful implications for the findings of regression models or cluster analysis, it might have influenced the free text responses, as they included several properties that inform about topical relevance (e.g., title, abstract, topical proximity). For future applications of this study's methodology we would recommend to clarify this aspect in the questionnaire.

2.2.7 Conclusions and Future Work

Our study's findings indicate that quantitative indicators are a part of many researchers' practices when initially assessing literature for its relevance, albeit next to a variety of qualitative aspects like for instance topical relevance or accessibility. We have seen distinct groups of users to value different indicators to different degrees, although overall, traditional citation-based indicators overshadowed altmetrics and usage metrics in this regard. Our results inform various stakeholders interested in providing their users with helpful information for deciding on literature, for example, providers of literature search engines, publication databases, or scholarly publishers.

As noted above, additional work should go into the analysis of disciplinary differences regarding acceptance and use of different indicators. Also, a bottleneck in an approach like ours is imposed by the limited number of product characteristics that can be observed at a time. Although we hope to have covered the most relevant six, there obviously are more quantitative indicators that would be interesting to analyze regarding their influence on researchers' consumption behavior. And even for qualitative characteristics, like those collected in the survey part of our study, utilities for potential readers could be estimated, as Tenopir et al. (2011) demonstrated.

Furthermore, the NISO use cases (National Information Standards Organization (NISO), 2016) offer several starting points for expansions to this study. It would for instance be interesting to see whether researchers' acceptance and use of different types of indicators changes when the task at hand is not about evaluation of search results, but about selecting metrics to showcase their own achievements.

Our study introduced new methods to the field of bibliometrics and provides empirical evidence on how indicators guide ranking processes of readers. A conjoint analysis as performed here is an elaborate endeavor that initially takes a lot of preparation. We hope to considerably facilitate various steps of conceptualization and data collection for future studies by making our software, which should be easily adaptable to a multitude of settings and research questions, openly available. In our own continuation of the study presented here, we next will use this approach and software to pursue the question of how certain prevalent methods of visualizing metrics data affect users' perception of research products' relevance.

2.2.8 Author Contributions

SL implemented the software used for the experiments, conducted the online survey process, coded the free text answers, performed the statistical analyses, documented and published the source code and datasets, and prepared this manuscript’s first draft; IP acquired funding for the research project. All authors contributed conception and design of the study, contributed to manuscript revision, and read and approved the submitted version.

2.2.9 Acknowledgments

We thank our colleagues in the *metrics project and at ZBW for their helpful feedback on early versions of the software used in the experiments, as well as the DFG for funding our project (grant number 314727790). Also, we wish to thank all researchers who helped us by participating in our online experiment.

2.3 Interim Conclusions

The results of Study A have given us an overview over the nature and diversity of thoughts and concerns researchers commonly have on the use of impact metrics. Among the most frequently stated difficulties we have seen are a lack of familiarity with the indicators and their methodologies, missing trust regarding their consistent reliability, and doubts regarding what it is that different metrics primarily measure. While these problems can to varying degrees apply to all types of metrics included in Study A, the participants' reservations seemed to weigh particularly heavy in regard to altmetrics. Study B has confirmed several of the observations made in Study A and provided further evidence on the status quo of indicator usage by individual researchers in day-to-day situations.

Together, Studies A and B (along with the literature reviewed) provide us with a picture of where particularly altmetrics stand regarding their usefulness for individual researchers: right now they still constitute a family of indicators that is frequently regarded with skepticism, and its usefulness restricted substantially by many suspected weaknesses. On the one hand, many of the stated concerns represent justified and valid indications of metrics' existing limitations - for instance, metrics' known inherent biases and restricted cross-discipline comparability are aspects that in practical applications should always be considered (aforementioned guidelines like Hicks et al. [2015] or Cagan [2013] provide valuable summaries of several particularly common problems occurring in indicator-based research evaluation, as well as hints on how to circumvent them). A considerable share of the participants' responses in the two studies, however, points towards another deficiency - one that also concerns the using researchers themselves and should be able to be addressed through means of information: frequently, missing knowledge about how different indicators emerge and what actually affects their magnitude impedes their interpretation substantially and lets them appear opaque.

In an endeavor to reduce this opaqueness surrounding metrics and to approach profound answers to the question of what it is that impact metrics actually measure, the following chapter shall therefore shed light on a sparsely studied factor with a substantial potential effect on impact metrics that also came up repeatedly during the user studies presented in Chapter 2, namely how aspects of external science communication shape indicators.

Chapter 3: Research Metrics and External Science Communication

In the scholarly discourse on the relationship between science and journalistic media, these two domains are prevalently considered to be separate systems with distinct objectives - while science strives for the production of true knowledge, journalistic media acts as the transmitter or translator of said scientific knowledge to the general public (Nielsen, 2009). Still, there are vocal arguments implying that the principles that determine what is covered in journalism might substantially affect *how* science aims to fulfill its own objective. The theory of the *medialization of science* (Weingart, 1998) argues that over the past decades an increased coupling between science and media was noticeable, and that this coupling has two facets: first, the representation of science within media changes, as it gets more and more into the focus of mass media; second, the scientific community itself changes, as it more and more tries to adapt modes of selection and presentation that are common in mass media (Schäfer, 2008).

In light of the tendencies proclaimed by the theory of science's medialization, the question arises whether similar feedback effects can be detected in the sphere of scientometrics. Does the media sphere substantially influence the instruments that the scientific sphere uses to evaluate itself? I.e., do the selection processes conducted by outer-academic agents like journalists or PR officers affect metrics used as indicators for the impact of science? If that was the case, this association could be a severe impairment of such indicators' usefulness in evaluation exercises, and it would back up a concern voiced occasionally against many forms of research metrics, namely that they to a significant extent are the result of promotion and dissemination efforts (see also Section 2.1.5). Or, as an author on the blog *sciencemetrics.org* pointedly expressed her suspicions regarding a specific example from altmetrics: "Measuring impact by counting mentions on Twitter is like measuring consumer appetite by counting billboards".[18]

Within this chapter, two further self-contained studies will be presented, which have the primary objective of investigating and quantifying the empirical substance behind the aforementioned assumption of a potential effect between media activities and research metrics. Study C examines the in this regard previously unstudied subject of embargo e-mails. Study D expands on the findings of Study C by transferring similar research questions to a more complex model involving additional factors, a substantially larger data sample, and a more sophisticated methodology in the form of path analysis.

[18] At the time of writing, the respective blog is no longer accessible - which actually serves as a convenient example to illustrate web-based metrics' limitation of temporal instability due to unforeseeable deletions of online content (see also sections 1.3 and 4.1.5).

3.1 Study C - Research Articles Promoted in Embargo E-Mails Receive Higher Citations and Altmetrics

3.1.1 Foreword

Study C presents an examination of the potential association between research articles being promoted in embargo e-mails and the articles' later research metrics. It represents a step towards achieving an understanding of the interplay between selection processes happening in external science communication and article impact, and with embargo e-mails examines a format of science promotion that has barely been studied before. The results should provide an approximation of the nature and magnitude of associations between the two studied spheres, findings which will then be deepened in Study D (Section 3.2).

The manuscript of this study was published at *Scientometrics*[19]. It was co-authored with Max Brede, Sophie Rotgeri, and Isabella Peters and an outcome of the research project *MeWiKo*, funded by the German Federal Ministry of Education and Research (grant number 01PU17018). I wish to thank my co-authors for their support, two anonymous reviewers at *Scientometrics* and Clemens Blümel for their highly productive criticism, as well as the colleagues at *MeWiKo* for their continuous feedback during the study's early stages.

[19] Full reference: Lemke, S., Brede, M., Rotgeri, S., & Peters, I. (2022). Research Articles Promoted in Embargo E-Mails Receive Higher Citations and Altmetrics. *Scientometrics, 127,* 75-97. https://doi.org/10.1007/s11192-021-04217-1

3.1.2 Abstract

In order to be able to provide thorough and timely coverage on the most recent scientific research, science journalists frequently rely on embargoed information sent to them by publishers of scientific journals. In such embargo e-mails, publishers purposefully bring selected upcoming releases to the journalists' attention a few days in advance of their publication. Little is known on how this early highlighting of certain research articles affects their later citations or altmetrics. We present an exploratory case study with the aim of assessing the effects of such promotion activities on scientific articles' bibliometric and altmetric indicators. In a treatment-control design, we analyze citation counts and eight types of altmetrics of 715 articles published between 2016 and 2017 whose DOIs have been mentioned in embargo e-mails and compare these to articles from the same journal issues that have not been highlighted in embargo e-mails. Descriptive statistics and Mann-Whitney-U tests reveal significant advantages for promoted articles across all regarded metrics three to four years after their publication. Particularly large differences can be seen regarding numbers of mentions in mainstream media, in blogs, on Twitter, and on Facebook. Our findings suggest that scholarly publishers exert significant influence over which research articles will receive attention and visibility in various (social) media. Also, regarding utilizations of metrics for evaluative purposes, the observed effects of promotional activities on indicators might constitute a factor of undesirable influence that currently does not receive the amount of consideration in scientometric assessments that it should receive.

3.1.3 Introduction

Staying informed about new developments in their field is an integral part of most researchers' everyday work. For them, the most common media from which to learn about new research findings will usually be academic journals, books, or conferences, depending on the field at hand. However, just like the general public, researchers also consume mass media, in which mediators from outside of the scientific community – typically journalists – communicate recent research findings (Kiernan, 2003a; Phillips et al., 1991).

In addition to universities and research institutions, which, amongst others, distribute new research findings via press releases, scholarly publishers occupy a key role in the dissemination of science via mass media. They regularly provide journalists with prepared summaries of selected new research articles in advance of their publication, so that the journalists have some days to prepare their coverage on said articles (Kiernan, 1997). These in advance-summaries are sent to the journalists *under embargo*, meaning their contents are only allowed to be published further after a date specified by the scholarly publisher has passed. This way, releases of research articles and their journalistic coverage are synchronized. So while journalists benefit from the embargo system by being enabled to publish well-prepared reports on recent science at the earliest possible date, publishers benefit from increased control over the timing of their articles' publicity. Moreover, through the selective provision of embargo information, publishers exercise a strong influence on *which* scientific findings can be covered timely and comprehensively in newspapers, television, radio broadcasts, and other forms of mass media (Kiernan, 1997, 2003a). Kiernan (1997) provides a historical overview over the embargo system's origins and, from reviewing past literature, concludes that it gives editors of scientific journals considerable power over what is regarded as scientific news and when mass media can report on it. Although in recent years, new platforms for the early publication of research (e.g., preprint repositories) might subvert this power by providing journalists with additional ways of obtaining early insights on recent findings, the embargo system is still in place as one of the essential channels for scholarly publishers to distribute new research to journalists (see also Franzen, 2012; Kiernan, 2003b). In a survey sent to the editors of 120 medical journals, Kiernan (2014) found 67% of the respondents to report that their journal would regularly offer journalists some kind of press material on new publications under embargo. Asked for reasons that would justify the embargo of journal articles, surveyed editors most commonly stated that embargos would be helpful to ensure that media coverage coincides with respective articles' publication and that they would help to ensure high-quality press coverage.

Several past studies examined the relationship between external science communication (i.e., the communication of research via channels not primarily aimed at other researchers, e.g., in news media) and how it affects research articles' later scientific impact, i.e. in terms of the number of citations. Phillips et al. (1991) for instance examined whether articles from the *New England Journal of Medicine* that had been featured in the *New York Times* (NYT) have received higher citation numbers than similar articles that had not been featured. They found that the former group did receive significantly higher citations, particularly in the first year after publication. Interestingly, this was not the case for research articles featured in NYT stories published during a period of strike, in which respective NYT issues were not distributed. For Phillips et al. (1991) these results support the 'publicity hypothesis', which assumes that media coverage genuinely increases a scientific article's visibility and thereby likelihood of being cited (as opposed to the 'earmark hypothesis', which hypothesizes that media coverage merely 'earmarks' outstanding articles which would have received many citations anyway). Kiernan (2003a, p. 4) suspected that Phillips et al. might have "ascribed an elite status to *Times* coverage of scholarly research that may not exist", by not addressing that coverage by other media outlets might also influence scientists' reliance on certain research articles. To address this, Kiernan (2003a) did an analysis additionally taking into account the effects of coverage in twenty-four daily newspapers and the evening broadcasts of three major television networks from the United States. The author found NYT coverage to not correlate significantly with citation rates once coverage by television and other newspapers is taken into account, suggesting that the NYT does not have unique influence as a disseminator of news about research to scholars. Even more recently, Fanelli (2013) tried to verify the publicity hypothesis for journal articles featured in British and Italian newspapers and found that the publicity effect is much stronger for English media, while publicity effects from Italian media are primarily local, i.e. mainly affect Italian authors.

Similarly, some studies specifically analyzed press releases as an instrument used by research institutions and scholarly publishers to communicate new research. In a quantitative content analysis of press releases, related journal articles, and news items on biomedical and health sciences, Sumner et al. (2014) found exaggerations in news to be strongly associated with exaggerations in press releases, concluding that improving accuracy of press releases might be a promising approach to reducing misleading news on health sciences. In a later study, Sumner et al. (2016) further investigated the relationship between mentions of caveats and exaggerations in journal articles and press releases and journalistic uptake. The study's findings suggest that press releases are frequently the source of both stated caveats and exaggerations concerning the reported research, but neither of the two seems to significantly affect the likelihood of news coverage. Stryker (2002) coded a sample of 95 journal articles from medical sciences for characteristics related to newsworthiness, examining their relationship to the articles' later amount of newspaper coverage, additionally considering whether respective articles were featured in press releases. The author found both newsworthiness and press release coverage to predict later newspaper coverage. In a similar vein, two previous studies had shown that about 80% and 84% respectively of research articles that get newspaper coverage had appeared in a press release beforehand (de Semir, Ribas, & Revuelta, 1998; Entwistle, 1995). Stryker (2002) however notes that the apparent effect of press releases on later newspaper coverage is reduced substantially when controlling for factors of newsworthiness. Complementing previous studies on the relationship between media coverage and citation rates, Chapman, Nguyen, & White (2007) did an analysis of the association between receiving a press release and the later citations and usage metrics of research articles from the journal *Tobacco Control*, finding press-released articles to receive more web hits, pdf downloads and citations than their counterparts without press releases.

On another note, concerning the point of intersection of science press releases and altmetrics, Bowman & Hassan (2019) examined *EurekAlert!* – an online science news service maintained by the *American Association for the Advancement of Science* – regarding the way scientific research is referenced in its news releases and regarding its presence across social media platforms. They found EurekAlert! to be the second most active source of Altmetric.com data for news releases, while showing only minimal activity on social media platforms. Moreover, Bowman & Hassan (2019) found that only a small share (18%) of EurekAlert! news releases referenced scientific research by DOI.

To briefly summarize, our review of past research on the relationship between external science communication and promoted articles' later metrics has shown that several case studies indicate an association between being selected for a press release or covered by news media and higher citations (Chapman et al., 2007; Fanelli, 2013; Kiernan, 2003a; Phillips et al., 1991) as well as usage metrics (Chapman et al., 2007) for respective research articles. In this exploratory case study, we aim to add to this body of research by (1) describing a previously under-analyzed, usually hard to obtain format of research promotion in the form of scholarly publishers' embargo e-mails to journalists, as well as by (2) also examining the relationship between such promotion and articles' later altmetrics.

Thus, in this exploratory case study we examine whether research articles mentioned in publishers' embargo e-mails sent to journalists differ regarding the bibliometric and altmetric indicators they receive, compared to articles without said mention. To gain insights as to which fields are typically represented in embargo e-mails, we additionally analyze mentioned articles' distribution across journals. We call this study exploratory, as it also represents the first steps to a higher-level goal of our ongoing research project, which is to arrive at a better understanding of the nature and extent of external science communication and its impact on research assessment practices.

Both citation-based indicators and altmetrics are often used as proxies for scientific productivity or relevance (Adie, 2016; Aksnes, Langfeldt, & Wouters, 2019; Waltman, 2016). Oftentimes respective evaluations – at least if they are not conducted in a remarkably knowledgeable and careful way – start on the premise that as long as two articles originate from similar fields of research, are published in outlets of comparable renown and are of similar scientific quality, they will have roughly the same probability of getting cited. Thus, as long as the most decisive factors like field of research, publication type, and publication venue are controlled for, citations would be a useful proxy for scientific quality or relevance. From numerous past studies we do know however that factors affecting numbers of citations are manifold and diverse (see Tahamtan, Afshar, & Ahamdzadeh [2016] for a review). We argue that, under the assumption that the publicity hypothesis as described by Phillips et al. (1991) holds true, research's promotion in external science communication would be such a factor that so far did not receive the scrutiny it would deserve. If articles' individual metrics can be shaped by publishers' promotional operations regardless of their scientific quality or merit, it seems of high importance to be able to inform users of such metrics about the degrees to which this might be the case.

The inclusion of altmetrics in this kind of study is of particular interest for two reasons. First, an argument frequently made in favor of altmetrics is that they might reflect a different form of impact than citations (Bornmann, 2014; Priem, Piwowar, & Hemminger, 2012; Weller, Dröge, & Puschmann, 2011; Wouters & Costas, 2012). Therefore, it is of interest to see whether altmetric indicators behave differently than citations in relation to press promotion, as this knowledge could help to further delineate altmetrics' potential benefit as complements to bibliometric indicators. Second, compared to citations, altmetrics more strongly blur the line between being the result and being part of promotional activities. Altmetrics that are frequently considered as indicators for received attention in many cases also directly include the amount of promotion undertaken to increase exactly this attention. For instance, measuring the tweets that mentioned a certain scientific paper will oftentimes mean to also measure tweets by publishers or involved authors, which were solely sent to advertise said paper (see also Haustein, Peters, Sugimoto, Thelwall, & Larivière, 2014). To be able to achieve a better understanding of this particular limitation of altmetrics, it therefore seems necessary to keep close watch over the relationship between manifestations of different altmetric indicators and promotional activities in external science communication, like for instance the embargo e-mail promotion considered in this study. While this study is not designed to conclusively settle the question regarding the degree to which altmetrics are comprised of advertising, it shall provide first numerical evidence to better describe how the promotion of articles to journalistic channels and the activity surrounding respective articles on platforms used as altmetric sources are connected.

3.1.4 Methods and Data

We compare two groups of articles regarding the metrics they received since their publication in 2016 to 2017 – a treatment and a control group. Both groups coincide in regard to their individual articles' journals and dates of publication, but the treatment group consists of articles that have been mentioned in embargo e-mails, while the articles from the control group have not. We measure attention as citations as measured by Web of Science and in the form of eight types of altmetrics provided by Altmetric.com. Publishers' embargo e-mails are no openly accessible data source – there is no such thing as a public archive gathering them, and building up a dataset of embargo e-mails is not trivial as usually accredited journalists are their only recipients. We rely on data from the *Science Media Center Germany (SMC)* (see also Rödder, 2015) to identify research articles with mentions in embargo e-mails. The SMC is a non-profit and editorially independent institution that supports journalists in reporting on science-related topics. One of its most important services is to send out statements by scientific experts commenting on new scientific findings while they are still under embargo. The SMC started its work in 2016 and since then has accumulated an archive of over 90,000 e-mails containing press materials, 2,638 of whom were identified as embargo e-mails. These embargo e-mails contain information about one or more upcoming articles in either one journal or several journals belonging to the same publisher. The SMC was founded by experienced science journalists, who specifically aim to monitor as many journals as possible that publish articles under embargo. Therefore, we can be confident that the SMC's archive contains a large part of all embargo e-mails sent to journalists. However, the SMC does have a focus on topics deemed "public issues" by its staff, e.g. topics affecting particularly large parts of society like medicine, climate change, or artificial intelligence. Hence, the archive and the algorithms extracting the embargo e-mails might be biased towards these topics. Also, the archive and the software doing the extraction were not constructed with a scientific analysis in mind, but rather as a tool to support the resident journalists. Due to these characteristics of the SMC's archive, we consider our undertaking to be a case study, as we cannot guarantee our sample's representativeness for the entirety of scholarly publishers' embargo e-mails. We hope, however, that the case study will shed light on the general characteristics of embargo e-mails and their relationship to citations and altmetrics.
We identified mentions of research articles in the extracted embargo e-mails by searching the archive for Crossref's recommended regular expression for DOIs.[20] This is a fairly strict criterion for determining whether an article appeared in an embargo e-mail, as in many cases in embargo e-mails articles are not referred to by identifiers. However, a less strict search based on articles' metadata

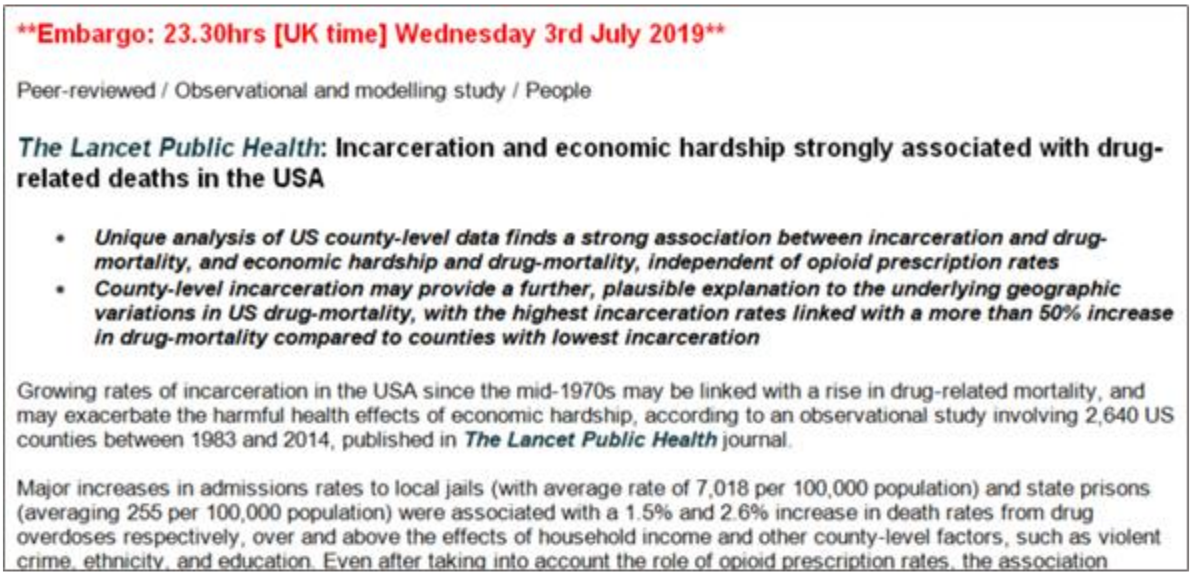
[20] https://www.crossref.org/blog/dois-and-matching-regular-expressions/

seemed hardly to be an option – typically, titles used in embargo e-mails are preliminary or shortened, the amount of detail with which other metadata like author names are included varies substantially between publishers. Furthermore, for our study high precision seemed to be of particular importance – after all, the total population of articles that get promoted in an embargo e-mail will surely be much smaller than the population of articles that do not get this specific kind of promotion. Thus, using less restrictive criteria for determining if an article has been referenced in an embargo e-mail comes with a high risk of adding false-positives to our treatment group. The Figures C1 and C2 show exemplary excerpts of two embargo e-mails sent by different publishers to illustrate the e-mails' heterogeneity regarding structure and level of detail of the information included.

As we wanted to perform citation analysis of articles mentioned in embargo e-mails, to achieve appropriate citation windows we only considered articles that according to Crossref were published either in 2016 or 2017, the first two years since the SMC started archiving embargo e-mails. In addition to this treatment group of articles published in 2016 or 2017 that were mentioned in embargo e-mails from the SMC's archive, we constructed a control group of comparable articles that had not been mentioned in embargo e-mails. For every article in the treatment group, this control group consisted of one randomly chosen article published in the same month, year, and journal and that itself was not already part of neither treatment nor control group. Publication dates, journal affiliations, and DOIs of control group articles were obtained from Crossref's REST API[21]. For further analysis, journals' field categories were retrieved manually via web search, with Web of Science's Master Journal List[22] as a primary source, and then matched to Web of Science research areas[23]. Quantitative data on journal-level was obtained from Incites Journal Citation Reports[24] (JCR).

We obtained citation data for all articles from both groups from the Competence Centre for Bibliometrics[25], which hosts bibliometric databases (which we from here on will call 'CCB databases') built on data from Web of Science. Altmetric data for all articles was obtained from Altmetric.com[26]. All queries were made in July 2021. To maximize recall, DOI-based queries were also performed for all-uppercase and all-lowercase transformations of the DOIs. It should be noted that the CCB databases are updated once per year in summer to Web of Science's state of calendar week 17, which means that the citation data retrieved in this study reflects the state of April 2020. Bibliometric and altmetric data was subsequently analyzed in R (R Core Team, 2020).

[21] https://github.com/CrossRef/rest-api-doc
[22] https://mjl.clarivate.com/
[23] https://images.webofknowledge.com/WOKRS535R76/help/WOS/hp_research_areas_easca.html
[24] https://jcr.clarivate.com/
[25] http://www.bibliometrie.info/
[26] https://www.altmetric.com/



Embargo: 23.30hrs [UK time] Wednesday 3rd July 2019

Peer-reviewed / Observational and modelling study / People

The Lancet Public Health: **Incarceration and economic hardship strongly associated with drug-related deaths in the USA**

- ***Unique analysis of US county-level data finds a strong association between incarceration and drug-mortality, and economic hardship and drug-mortality, independent of opioid prescription rates***
- ***County-level incarceration may provide a further, plausible explanation to the underlying geographic variations in US drug-mortality, with the highest incarceration rates linked with a more than 50% increase in drug-mortality compared to counties with lowest incarceration***

Growing rates of incarceration in the USA since the mid-1970s may be linked with a rise in drug-related mortality, and may exacerbate the harmful health effects of economic hardship, according to an observational study involving 2,640 US counties between 1983 and 2014, published in ***The Lancet Public Health*** journal.

Major increases in admissions rates to local jails (with average rate of 7,018 per 100,000 population) and state prisons (averaging 255 per 100,000 population) were associated with a 1.5% and 2.6% increase in death rates from drug overdoses respectively, over and above the effects of household income and other county-level factors, such as violent crime, ethnicity, and education. Even after taking into account the role of opioid prescription rates, the association

Figure C1: Excerpt of an embargo e-mail by a scholarly publisher - example A.

PLOS ONE

In the **July 17** release:

- Link Between Workplace Sexual Harassment and Women's Negative Self-Views May Be Weakening
- Stone tool changes could reveal how Mesolithic hunter-gatherers responded to changing climate *(image)*
- Endangered Bornean Orangutans Survive in Managed Forests, Decline Near Oil Palm Plantations *(image)*
- Protected area designation effective in reducing, but not completely preventing, land cover changes in Europe

Titles Only:

- People more likely to be supportive of new power lines being put up in their area if lines carry wind or solar-generated electricity, rather than gas or coal-generated electricity
- How switching short car trips in New Zealand to cycling and walking might improve population health and reduce CO2 emissions: A modelling study

EMBARGO: July 17, 2019
11 AM Pacific / 2 PM Eastern Time

Peer-reviewed Observational study People

Figure C2: Excerpt of an embargo e-mail by a scholarly publisher - example B.

The regular expression search for DOIs in the SMC's e-mail archive initially retrieved 1,638 unique DOIs. After both automatic and manual cleansing of the data, 1,456 unique DOIs remained (the removed records for the most part consisted of false-positives returned by the extraction script, i.e. strings that did not include a real DOI known to Crossref). Of these, 715 referred to articles which had been first published in 2016 or 2017, according to Crossref. These DOIs form our treatment group.

3.1.5 Results

To get an impression of the disciplines represented in embargo e-mails of the sample used, we have a look at the journals the 715 articles from our treatment group were published in (for the control group these numbers would be the same, as it contains exactly one counterpart from the same journal for every article from the treatment group). In total the articles were published across 78 different journals, with the numbers of articles per journal ranging from 1 to 77 articles *(PLoS Neglected Tropical Diseases* being the most heavily represented journal). The most prevalent publishers in the sample were *PLoS* (357 articles), *BMJ* (198 articles) and *Springer Science and Business Media LLC* (102 articles). Figure C3 shows how many journals are represented in the sample by how many individual articles each. These first findings indicate a heavily skewed distribution regarding the representation of individual publishers and journals in the embargo e-mails, with few high-profile outlets accounting for a large share of the communication captured by our sample.

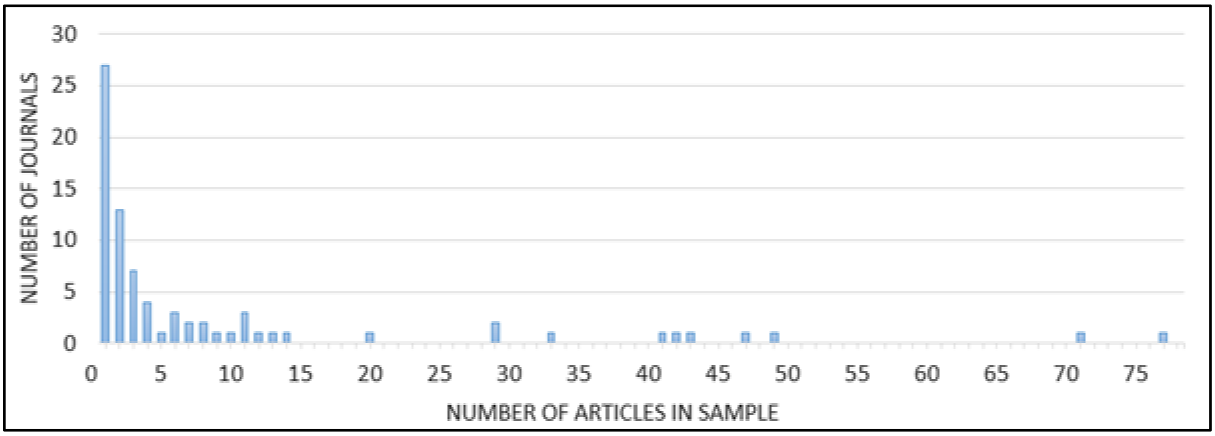


Figure C3: Distribution of treatment group articles across journals.

Examining the 78 journals' field categories as reported on the Web of Science Master Journal List reveals a strong representation of life sciences & biomedicine – 71 of the 78 journals primarily cover topics from this research area. Three of these cases are additionally categorized as multidisciplinary, another three as also explicitly covering fields from social sciences. Six of the remaining 7 journals without explicit affiliation to life sciences & biomedicine were identified as multidisciplinary and most notably refer to certain prominent mega-journals (e.g., *Nature*, *PLoS ONE*, or *Proceedings of the National Academy of Sciences*). The last remaining journal primarily publishes articles from the field of psychology and is therefore categorized as primarily covering social sciences, when applying the Web of Science research areas. Out of the 78 journals in our sample, 9 were not found in the Incites JCR database (*BMC Hematology*, *BMC Psychology*, *BMC Research Notes*, *BMJ Case Reports*, *Drugs and Therapeutics Bulletin*, *Heart Asia*, *Injury Epidemiology*, *Marine Biodiversity Records*, and *Trauma Surgery & Acute Care Open*). Hence, we see that most but not all journals that send embargo e-mails are also indexed by the Web of Science.

The observed dominance of journals from life sciences and biomedical domains is in line with previous studies on the prevalence of certain scientific fields and topics in journalistic media like newspapers (see e.g., Elmer, Badenschier, & Wormer, 2008) or in scientific press releases (Hahn & Lemke, 2020).

Due to the heavy representations of life science- and multidisciplinary journals in our sample, we will regard these two categories in particular detail for the rest of our analysis.

Table C1 shows, for the journals in our sample with JCR records, journal impact factors (JIF), total numbers of cites and numbers of citable items, in 2016, our first year of observation, and in 2020, the most recent year with available JCR data. The table shows large ranges across all parameters for both life science and multidisciplinary journals, indicating that the sample contains journals that implement highly diverse publication patterns. It should be noted that regarding numbers of citable items especially the native open access outlets of prominent publishers (e.g., *Scientific Reports*, *Nature Communications*, *PLoS ONE*) stand out as featuring particularly high numbers, which are often also tied to high numbers of total cites. However, regarding total cites also some high-profile (non-open access) journals like *Nature* or the *New England Journal of Medicine* constitute outliers to the top.

Table C1: Descriptive data on journals that published the articles in our sample of embargo e-mails

Indicator	*Life Sciences & Biomedicine (n=71)*				*Multidisciplinary (n=9)*			
	Min	*Mean*	*Max*	*SD*	*Min*	*Mean*	*Max*	*SD*
JIF ‘16	1.48	9.01	72.41	12.07	2.81	12.80	40.14	11.13
JIF ’20	2.15	11.80	91.25	16.36	2.52	13.38	49.96	13.84
Total cites ‘16	957	26,584	315,143	50,146.39	9,495	260,400	671,254	268,743.23
Total cites ‘20	140	38,260	464,351	73,702.64	1,017	407,735	915,925	368,105.76
Citable it. ‘16	54	293	1,998	332.57	172	6,351	22,077	8,726.17
Citable it. ‘20	15	330	3,266	481.69	121	5,432	21,222	7,405.05

To account for the substantial heterogeneity regarding expected average citations of the journals in our sample, we will also perform selected subsequent steps of analysis separately for articles published in “higher impact journals” (HIJ) and “lower impact journals” (LIJ). To define these two groups we apply the method of characteristic scores and scales as introduced by Glänzel & Schubert (1988) with $k = 2$, based on the journals’ impact factors as reported in 2020. This procedure classifies 17 journals from our sample as HIJ, 51 journals as LIJ, while 10 journals remain unclassified due to them not having a journal impact factor in 2020. On an article-level, our treatment and control group each consist of 115 (16.08%) articles from HIJ, 581 (81.26%) articles from LIJ, and 19 (2.66%) articles from journals without journal impact factor in 2020.

If we measure the distribution of research areas across the 715 individual articles of our treatment group, we find that 594 (83.08%) of the articles were published in journals explicitly associated with life

sciences & biomedicine, while 165 (23.08%) articles were published in multidisciplinary journals. Only 17 (2.38%) articles were published in one of the nine journals not covered by the Incites JCR.
Figure C4 shows the relations between articles from the two most prevalent discipline categories in our sample, life sciences & biomedicine and multidisciplinary, and their respective representation in higher and lower impact journals as a Venn diagram.

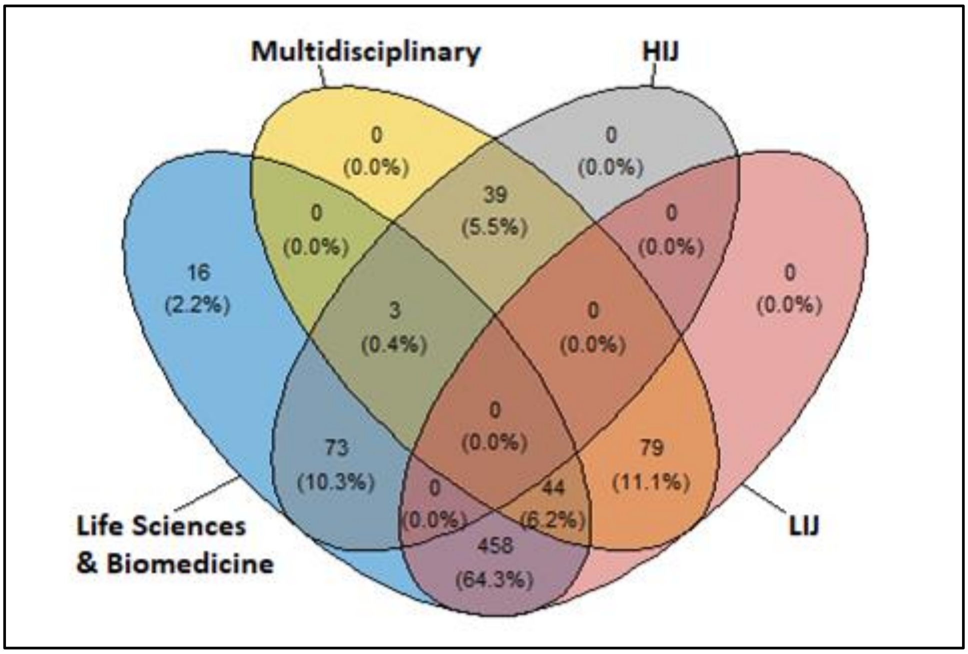


Figure C4: Venn diagram of relations between articles of the treatment group published in life sciences-, multidisciplinary-, lower impact-, and higher impact journals.

For 683 (95.52%) of the 715 articles from the treatment group Web of Science citation counts could be retrieved from the CCB databases, as was the case for 680 (95.10%) control group DOIs (the majority of DOIs for which no citation count could be retrieved belongs to articles from journals that were either not indexed in Web of Science or had only been indexed after the CCB databases' latest update). On Altmetric.com, records were found for 714 (99.86%) treatment group articles, i.e. the article was mentioned at least once on one of the platforms tracked by Altmetric.com, and for 677 (94.69%) of the articles in the control group. In analyses of altmetric counts, articles without records on Altmetric.com were assumed to have values of 0 across all altmetrics, as this should be the only regular circumstance under which an article with a valid DOI does not have a record on Altmetric.com.
It has been shown that on various types of altmetric sources only very few articles receive any mentions at all, which can complicate their usage (Thelwall, 2018a). To get a first rough idea of individual altmetric indicators' suitability for comparisons between the two groups of articles, the shares of articles in both groups that did not receive any mention in respective altmetric sources as provided by Altmetric.com's API are presented in Table C2.

Table C2: Numbers (shares) of articles in both groups that received zero mentions on respective altmetric sources

Altmetric source	*Treatment (n=715)*	*Control (n=715)*
No Twitter mentions	7 (0.98%)	45 (6.29%)
No Facebook mentions	195 (27.27%)	447 (62.52%)
No Wikipedia mentions	615 (86.01%)	683 (95.52%)
No mainstream media mentions	57 (7.97%)	536 (74.97%)
No blog mentions	208 (29.09%)	566 (79.16%)
No Reddit mentions	555 (77.62%)	645 (90.21%)
No Youtube mentions	635 (88.81%)	704 (98.46%)
No Mendeley readers	1 (0.14%)	38 (5.31%)

The shares of articles that received no attention at all are on all platforms higher in the control group. Particularly high differences (percentage-wise) between both groups can be found for mentions in mainstream media and on blogs. On Wikipedia, Reddit, and Youtube, the majority of articles from both groups did not receive any mentions. Group-wise comparisons based on these indicators should therefore be interpreted with caution, as only few articles are responsible for all measured differences.

To arrive at a more precise picture of the data's structure, we next inspect correlations between the metrics in our sample. Figure C5 shows the Spearman rank correlations (pairwise complete observations) for the metrics data of the combined set of 1,430 articles from treatment and control group.

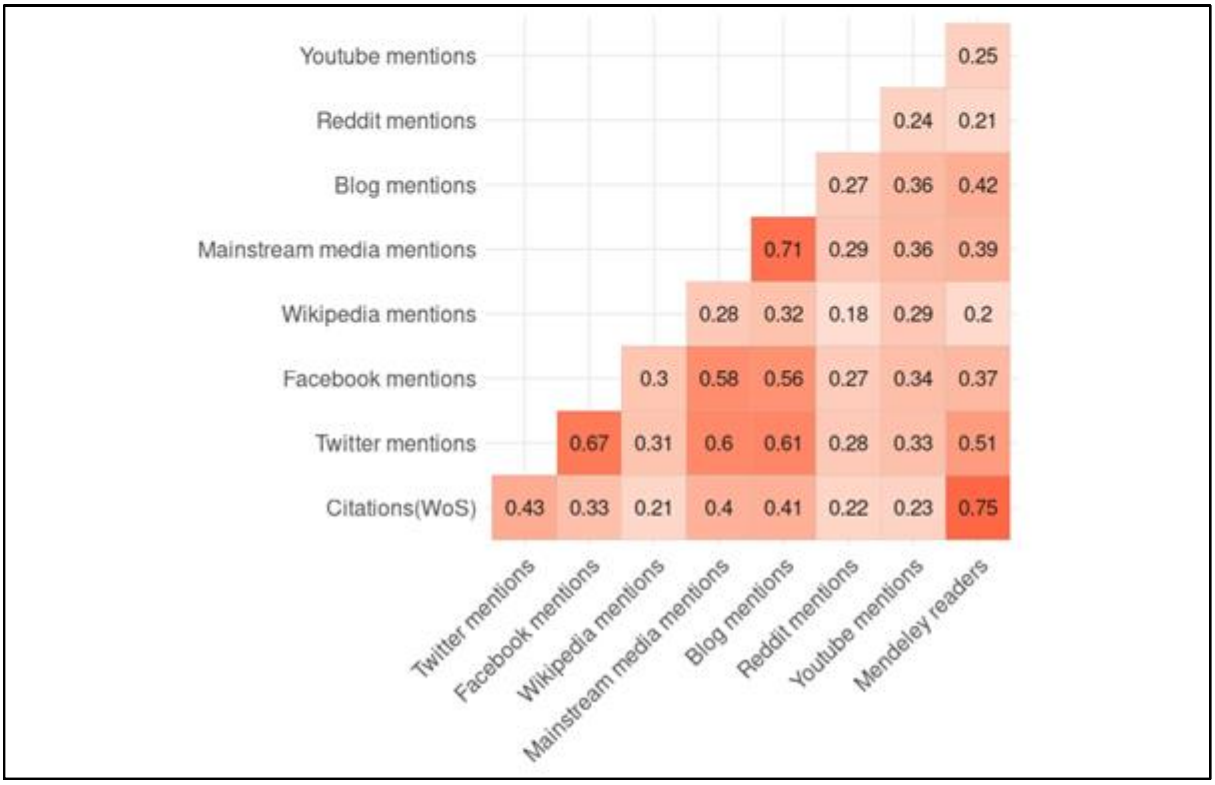


Figure C5: Spearman rank correlations of sample articles' metrics (n=1,430); all correlations are statistically significant with $p < 0.001$.

Among mostly weak to moderate correlations (ranging from 0.18 to 0.43), some strong correlations stand out: first, mentions on Twitter, Facebook, mainstream media, and blogs all correlate strongly with each other; second, Mendeley readers correlate strongly both with Twitter mentions and citations from Web of Science for the articles in our sample.

Table C3 shows the means and standard deviations of citations and altmetric counts that articles from both groups received between their publication in 2016 or 2017 respectively and our data collection. Comparing the means between treatment and control group reveals higher average counts per article from the treatment group in every single metric. Regarding most metrics the counts for treatment group articles are on average between 2 to 6 times higher than for control group articles. The most extreme relative difference is measured for mainstream media mentions, where average values among the treatment group are more than 10 times as high as among the control group.

We cannot infer from Table C3 if both groups are to different extents characterized by extreme outliers, although high standard deviations in comparison to respective means indicate generally wide spreads of individual values. To get a better understanding of the data's distributions, the 5% trimmed mean and median of all indicators per group have been calculated, as shown in Table C4.

Table C3: Means and standard deviations of bibliometric and altmetric indicators across both groups

Indicator	*Mean*		*Standard deviation*	
	Treatment	*Control*	*Treatment*	*Control*
Citations (WoS)	26.37	14.63	56.50	27.80
Twitter mentions	114.82	24.23	393.37	49.59
Facebook mentions	5.78	0.92	19.76	2.68
Wikipedia mentions	0.33	0.06	1.22	0.32
Mainstream media mentions	26.12	2.52	45.62	12.67
Blog mentions	3.03	0.48	6.34	1.87
Reddit mentions	0.35	0.13	1.03	0.58
Youtube mentions	0.25	0.03	1.11	0.26
Mendeley readers	130.44	74.44	263.11	105.62

Table C4: Outlier-robust measures of bibliometric and altmetric indicators across both groups

Indicator	*5% trimmed mean*		*Median*	
	Treatment	*Control*	*Treatment*	*Control*
Citations (WoS)	17.74	10.08	12	8
Twitter mentions	61.65	16.02	33	9
Facebook mentions	3.07	0.55	2	0
Wikipedia mentions	0.13	0.00	0	0
Mainstream media mentions	19.20	0.74	9	0
Blog mentions	2.15	0.22	1	0
Reddit mentions	0.20	0.05	0	0
Youtube mentions	0.07	0.00	0	0
Mendeley readers	92.29	59.13	70	47

Using outlier-robust measures overall does not reduce the previously observed advantages of treatment group articles substantially. On the contrary, for mentions on mainstream media and blogs the relative advantages of the treatment group rise even further. The medians of zero for Wikipedia, Reddit and

Youtube mentions confirm once more what we have already seen in Table C2: for articles from both groups, mentions on these platforms are the exception rather than the rule.
Figure C6 depicts the two groups' distributions explored in Table C3 and Table C4 graphically, as boxplots.

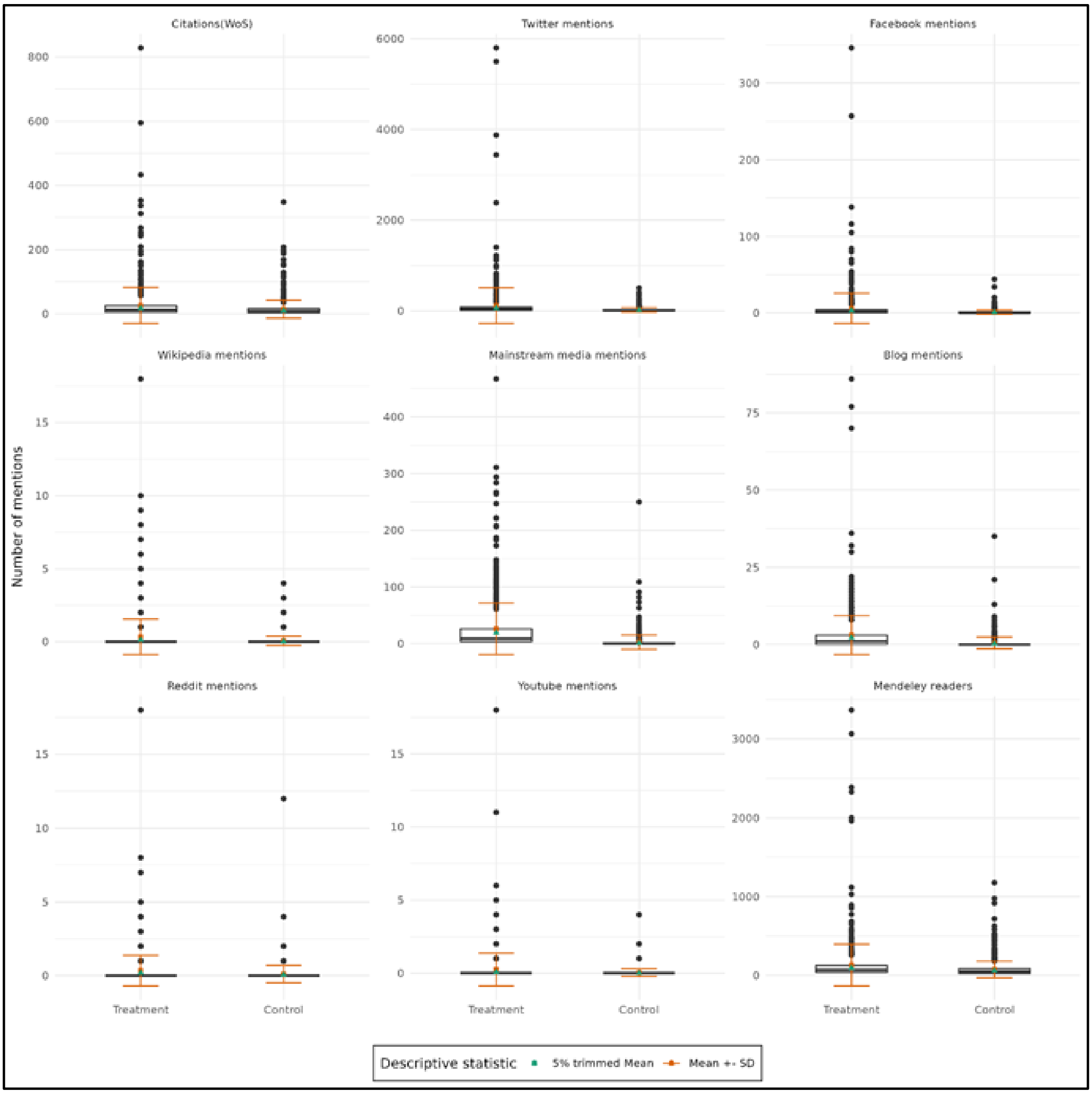


Figure C6: Boxplots of the two groups' metrics.

Next, we calculate means, 5% trimmed means, and standard deviations individually for the two discipline categories prevalent in our sample, 'life sciences & biomedicine' and 'multidisciplinary', to explore whether articles from the discipline-wise diverse journals behave differently from the articles published in more mono-thematic journals. Table C5 shows the three measures for articles published in life sciences- and biomedicine-journals (n=594 per group), Table C6 shows the respective measures for articles published in multidisciplinary journals (n=165 per group).

Table C5: Descriptive statistics for articles from life sciences & biomedicine-journals

Indicator	*Mean*		*5% trimmed mean*		*Standard deviation*	
	Treatment	*Control*	*Treatment*	*Control*	*Treatment*	*Control*
Citations (WoS)	21.36	12.38	16.11	9.21	37.90	23.14
Twitter mentions	104.14	22.87	61.52	15.54	318.82	45.70
Facebook mentions	5.21	0.83	2.85	0.53	19.42	2.31
Wikipedia mentions	0.23	0.05	0.07	0.00	1.03	0.29
Mainstream media mentions	25.98	1.69	19.01	0.47	45.80	7.70
Blog mentions	2.48	0.30	1.86	0.16	4.57	0.78
Reddit mentions	0.28	0.10	0.17	0.04	0.68	0.57
Youtube mentions	0.19	0.02	0.05	0.00	0.97	0.20
Mendeley readers	106.84	67.67	84.47	56.64	157.47	80.68

Table C6: Descriptive statistics for articles from multidisciplinary journals

Indicator	*Mean*		*5% trimmed mean*		*Standard deviation*	
	Treatment	*Control*	*Treatment*	*Control*	*Treatment*	*Control*
Citations (WoS)	43.66	22.20	28.67	15.92	90.09	36.78
Twitter mentions	163.93	37.41	84.91	27.56	557.58	61.57
Facebook mentions	7.26	1.31	3.86	0.75	18.46	3.54
Wikipedia mentions	0.71	0.14	0.41	0.05	1.70	0.52
Mainstream media mentions	25.25	5.36	18.72	2.33	43.23	21.83
Blog mentions	5.24	1.16	3.74	0.58	10.20	3.57
Reddit mentions	0.65	0.20	0.39	0.11	1.75	0.58
Youtube mentions	0.42	0.07	0.17	0.00	1.41	0.40
Mendeley readers	219.86	105.62	139.83	80.53	453.82	159.88

In both treatment and control group, our sample of articles from multidisciplinary journals on average performs better across almost all considered metrics – which, considering the larger relative share of articles from high impact journals among the group of multidisciplinary articles seen in Figure C4, may not come as too much of a surprise. Higher average counts for multidisciplinary articles might to a considerable extent be explained by the fact that our small selection of multidisciplinary journals is mostly comprised of fairly high-profile journals like *Nature* or *PLoS ONE*, while our larger sample of journals covering life sciences & biomedicine represents a more mixed range. An interesting exception are the mainstream media mentions of treatment group articles from life sciences & biomedicine, which are slightly higher than those of the articles with embargo e-mail promotion that were published in multidisciplinary journals. This might indicate that, regarding coverage in mainstream media, articles from life sciences & biomedicine derive an even larger relative benefit from the treatment of being promoted in an embargo e-mail, than the articles from multidisciplinary journals.

In the next step, we once more calculate means, 5% trimmed means and standard deviations, this time separately for the group of articles published in higher impact journals (n=115; Table C7) and articles published in lower impact journals (n=581; Table C8).

Mean values for both treatment and control group articles are across all metrics substantially higher if we only consider articles from higher impact journals, compared to articles from the lower impact journals. Also, just like in the previous comparisons, mean values are across all metrics higher for treatment than for control groups.

Table C7: Descriptive statistics for articles from higher impact journals

Indicator	*Mean*		*5% trimmed mean*		*Standard deviation*	
	Treatment	*Control*	*Treatment*	*Control*	*Treatment*	*Control*
Citations (WoS)	67.01	32,44	50.80	24.22	112.41	58.54
Twitter mentions	299.98	46.73	183.39	32.85	711.22	85.29
Facebook mentions	17.77	2.15	10.81	1.24	43.42	5.58
Wikipedia mentions	0.50	0.17	0.30	0.09	1.19	0.49
Mainstream media mentions	40.94	8.98	32.83	4.26	62.43	28.12
Blog mentions	5.80	1.48	4.27	0.81	10.42	4.16
Reddit mentions	0.60	0.18	0.42	0.08	1.23	0.63
Youtube mentions	0.68	0.12	0.31	0.01	2.19	0.59
Mendeley readers	297.37	118.57	210.77	87.77	511.15	209.80

Table C8: Descriptive statistics for articles from lower impact journals

Indicator	*Mean*		*5% trimmed mean*		*Standard deviation*	
	Treatment	*Control*	*Treatment*	*Control*	*Treatment*	*Control*
Citations (WoS)	18.31	11.24	14.06	9.47	30.75	14.11
Twitter mentions	79.88	19.70	45.25	13.95	286.88	36.25
Facebook mentions	3.47	0.69	2.24	0.45	8.67	1.52
Wikipedia mentions	0.30	0.04	0.10	0.00	1.25	0.28
Mainstream media mentions	22.95	1.31	16.29	0.40	41.20	5.67
Blog mentions	2.52	0.29	1.80	0.15	5.12	0.80
Reddit mentions	0.31	0.12	0.18	0.05	1.00	0.58
Youtube mentions	0.17	0.01	0.05	0.00	0.73	0.12
Mendeley readers	100.12	67.40	79.27	59.17	164.71	67.39

As a last step, Mann-Whitney-U tests are conducted to check for statistical significance of the observed mean differences. Effect sizes are reported in form of *r* values (Fritz, Morris, & Richler, 2012) and calculated following the implementation from R-package rcompanion[27]. We start with calculating effect sizes for the treatment and the control group in their entirety (Table C9) and then repeat the process for article groups distinguished by discipline (Table C10) and impact factor of their journal (Table C11), analogous to our previous steps. Table C9 shows the respective *U* statistics, Bonferroni-adjusted *p*-values and effect sizes *r* for the differences regarding the entire treatment and control group.

The observed differences between treatment and control group are statistically significant regarding all nine examined metrics. According to guidelines reported by Fritz et al. (2012), *r* values of .1 indicate a small effect, values of .3 a medium effect, and values of .5 a large effect. Large effects were measured for mainstream media mentions and blog mentions, a large to medium effect was measured for Twitter mentions and Facebook mentions. Regarding the remaining metrics, measured effect sizes were medium to small.

[27] https://www.rdocumentation.org/packages/rcompanion/versions/2.3.25/topics/wilcoxonR

Table C9: Significances and effect sizes of mean differences between treatment and control group, based on all articles in the sample; all *p*-values are Bonferroni-adjusted

Indicator	*U*	*p*	*r*
Citations (WoS)	286952	<0.001	0.199
Twitter mentions	384232	<0.001	0.436
Facebook mentions	372105	<0.001	0.415
Wikipedia mentions	280352	<0.001	0.167
Mainstream media mentions	453776	<0.001	0.695
Blog mentions	395562	<0.001	0.518
Reddit mentions	288520	<0.001	0.175
Youtube mentions	280270	<0.001	0.197
Mendeley readers	321304	<0.001	0.222

Table C10: Significances and effect sizes of mean differences between treatment and control group, based on publishing journals' disciplines; all *p*-values are Bonferroni-adjusted

Indicator	*Life Sciences & Biomedicine*			*Multidisciplinary*		
	U	*p*	*r*	*U*	*p*	*r*
Citations (WoS)	199262	<0.001	0.212	16262	0.031	0.187
Twitter mentions	263026	<0.001	0.426	19650	<0.001	0.384
Facebook mentions	253555	<0.001	0.400	20132	<0.001	0.432
Wikipedia mentions	189060	<0.001	0.137	16092	<0.001	0.236
Mainstream media ment.	319713	<0.001	0.731	22376	<0.001	0.567
Blog mentions	273427	<0.001	0.528	21066	<0.001	0.496
Reddit mentions	198811	<0.001	0.181	16124	0.005	0.214
Youtube mentions	190684	<0.001	0.185	15588	0.001	0.232
Mendeley readers	221190	<0.001	0.220	17337	<0.001	0.237

Table C11: Significances and effect sizes of mean differences between treatment and control group, based on publishing journals' impact factors; all *p*-values are Bonferroni-adjusted

Indicator	*Higher Impact Journals*			*Lower Impact Journals*		
	U	*p*	*r*	*U*	*p*	*r*
Citations (WoS)	8490.5	<0.001	0.322	193312	<0.001	0.161
Twitter mentions	11304	<0.001	0.613	249472	<0.001	0.414
Facebook mentions	10838	<0.001	0.560	240629	<0.001	0.393
Wikipedia mentions	7456	0.585	0.164	184373	<0.001	0.173
Mainstream media ment.	11044	<0.001	0.586	304403	<0.001	0.728
Blog mentions	10506	<0.001	0.525	260371	<0.001	0.522
Reddit mentions	7907	0.013	0.239	189012	<0.001	0.166
Youtube mentions	7754.5	0.008	0.248	182752	<0.001	0.190
Mendeley readers	9491.5	<0.001	0.377	204550	<0.001	0.183

The discipline-specific effect sizes seen in Table C10 suggest that across most indicators, the association between embargo e-mail promotion and an article's later metrics does not differ substantially between the two discipline categories under observation, with most differences between respective effect sizes only ranging between 0.02 and 0.05. However, one difference stands out particularly in this regard: for articles from life sciences & biomedicine, the positive association with mainstream media coverage seems even higher (r=0.731) than for articles from multidisciplinary journals (r=0.567). This backs up a similar observation we made during the comparison of Tables C5 and C6.

More evident differences between effect sizes can be seen in Table C11 when comparing articles from higher impact journals to those from lower impact journals. Regarding most metrics (citations, Mendeley readers, mentions on Twitter, Facebook, Reddit, and Youtube), for articles from higher impact journals being promoted in an embargo e-mail is associated with stronger positive effects than for articles from lower impact journals. Regarding coverage in mainstream media, however, articles from lower impact journals seem to profit even more from the embargo e-mail promotion.

3.1.6 Discussion

We compared the bibliometric and altmetric counts of articles which had been mentioned in publishers' embargo e-mails to journalists (treatment group) with those of articles from the same journals and publication months without such mentions (control group). We observed statistically significant advantages for treatment group articles across all nine examined metrics. Particularly strong effects were measured regarding mentions in mainstream media and on blogs, followed by Twitter mentions, Facebook mentions, Mendeley reader counts, and citations. Still significant small to medium effects were measured for mentions on Youtube, Wikipedia, and Reddit, although only small shares of both article groups received any mentions on these platforms at all. Our observations when differentiating between articles based on their journals' disciplines and impact factors suggested that the observed effects for the most part do not depend substantially on discipline, while across most metrics, articles from higher impact journals do seem to derive even larger comparative advantages from embargo e-mail promotion than those published in lower impact journals. A remarkable exception to these observations concerns mainstream media mentions, where measured effects are stronger for articles from life sciences & biomedical and lower impact journals than for multidisciplinary and higher impact journals. A possible explanation for this might be the fact that articles from higher impact journals tend to get a fairly high chance of being represented in mainstream media in any case (see also Table C7), regardless of them receiving additional promotion in embargo e-mails. Presumably, many science journalists regularly scan the prominent high impact journals for interesting articles anyway, while articles in lower impact journals might be more reliant on specific promotion to catch the attention of a significant number of journalists.

Multiple possible effects could explain our findings. First, one could assume embargo e-mails to evoke a publicity effect, in line with findings by Phillips et al. (1991) and Fanelli (2013) for newspaper articles. Our results indicate that articles mentioned in embargo e-mails do in fact get significantly more attention on mainstream media and social media platforms. Under the assumption of the publicity hypothesis, these findings would imply that scholarly publishers do indeed exert substantial influence on which articles (and thereby topics) get presence in the mass media, which would be in accordance with suppositions by Kiernan (1997). Moreover, the attention-increasing effect of embargo e-mail promotion apparently goes beyond the media outlets addressed by the e-mails themselves, as is evidenced by promoted articles' higher counts of scientific citations and Mendeley readers. In other words, the attention attracted by embargo e-mails in the public sphere of news media seems to also radiate into the scholarly sphere, which is represented by citations and Mendeley.

The validity of the publicity hypothesis for embargo e-mails would have further implications with regard to impact metrics as a means of research assessment. The use of citations as indicators for academic influence is underpinned by the theory that scholarly publishing is – at least for the most part – its own system, separate for instance from public media. If the publicity hypothesis holds true, it curtails the validity of this assumption. If promotional tools such as embargo e-mails could be used to push the metrics for publications independently from their scientific qualities, this would pose a limitation for the usefulness of such indicators. The promotion of selected articles by scholarly publishers would constitute a substantial interference with the scientific reward system and would add another bias to those already known in the context of citation-based evaluations (Hicks et al., 2015; Tahamtan et al., 2016). And unlike in peer review, the scholarly publishers are not compelled to apply established criteria of good scholarly practice when selecting which articles to promote in embargo e-mails, nor to make their criteria transparent.

However, our approach does not allow us to make definite statements about causalities between events behind different metrics, e.g., whether an article's mainstream media mentions typically motivate a substantial share of its scientific citations, or whether by and large its inherent qualities determine both academic and media uptake. Therefore, assuming embargo e-mails to be the decisive entry point into a chain of different media that an article then passes through with its publicity increasing along the way is not the only possible explanation of our findings. Instead, we could also assume a variation of the earmark hypothesis (Phillips et al., 1991) to be the determining factor behind our observations. In this case it would not be the science journalists, but the scholarly publishers and/or editors before them who on their own identify particularly noteworthy articles as being worthy to be included in embargo e-mails. And because journalists, researchers, and other readers later independently reach similar conclusions about which articles are noteworthy, the same articles that get chosen for embargo e-mails also get referenced more in mainstream media, on social media, in other scientific articles, and so on. Surely, although the degree of this hypothesis' validity may be unknown, it is to be expected that the selection of articles for embargo e-mail promotion does not happen at random but follows a variety of criteria (e.g., newsworthiness, prominence of the authors and institutions involved, scientific quality, etc.), some of which might also explain higher expected citations and/or altmetrics. To which degrees the promoted articles' metric advantages can be explained by which of the two hypotheses discussed in this article – publicity or earmark – remains an open question that will require further research. What this study could show, however, is the existence and magnitude of the advantage in metrics that embargo e-mail promoted articles can be expected to have.

For the scholarly community and its use of citations (and sometimes altmetrics) as indicators of scientific relevance, validity of the earmark hypothesis would surely be less problematic than validity of the publicity hypothesis. In any case, metrics' susceptibility to numerous intrinsic and extrinsic factors (Tahamtan et al., 2016) will continue to be an issue that needs to be communicated with utmost transparence to their users, especially in light of increasing commitments to Open science. Furthermore, the various interests and opportunities to exert influence of the diverse stakeholders involved in science communication should continue to be discussed and made more transparent, as should the relationships between authors, publishers, and scientific reputation system.

As existing research on embargo e-mails' role in the scientific communication system is relatively sparse, many different research streams could proceed from our results. One important aspect that should be investigated are the aforementioned possible causalities between metrics events. For instance, how common is the scenario of researchers first reading about scientific articles in mainstream- or social media, and because of that deciding to read (and later cite) the scientific article itself? Or the other way around: which role do scientific citations play for science journalists when they choose sources to back up their articles?

Furthermore, it should be examined whether mentions in embargo e-mails are merely a specific facet of broader promotional activities by the scholarly publishers. Such promotional activities could also have more direct effects on some of the metrics examined in this study – it should for instance be researched which role publishers' postings play regarding individual articles' mentions on social media platforms, like Twitter or Facebook. Examples for publishers directly engaging in social media postings to promote their own publications are manifold; see Figure C7 for two typical examples of promotional tweets by publishers. The strong correlations we witnessed between mainstream media and various social media mentions (see Figure C5) could also be an indicator of coordinated marketing efforts spanning multiple channels. Partly these correlations can be explained by the permeability and content overlap between respective platforms – many journalistic news articles also get tweeted and/or shared on Facebook the moment they are released – while the existence of a 'publicity snowball effect', where certain creators of social media content just copy stories that already received coverage on other platforms, might also play into these correlations. It remains to be analyzed which share of the total amount of postings mentioning scientific publications publishers account for, whether postings, tweets, and embargo e-mails are all parts of purposefully coordinated promotional activities for selected articles, and which role cross-platform transitions play in the distribution of embargo e-mail promoted research.

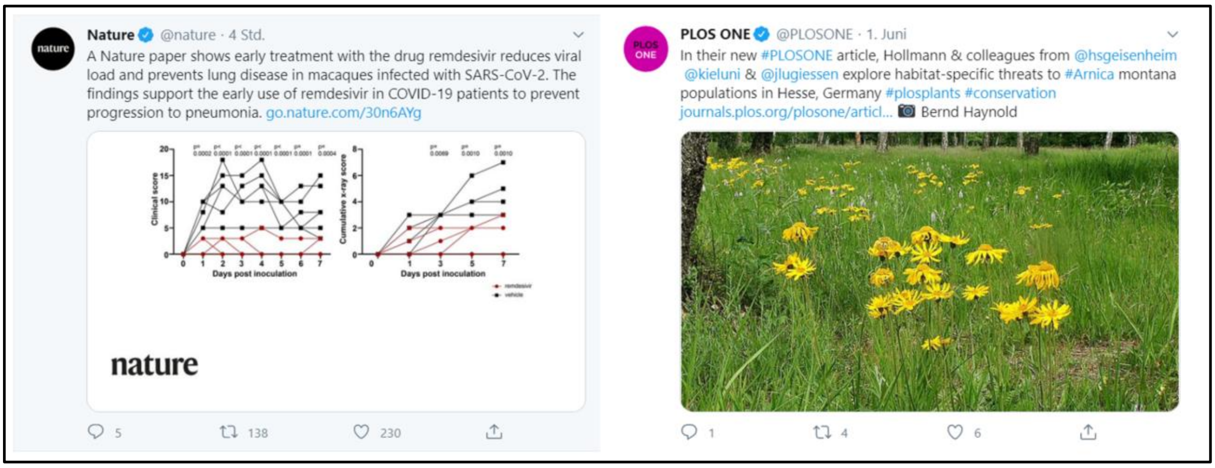


Figure C7: Publishers' tweets promoting research articles

Another worthwhile subject for further studies would be the comparison of the effects of embargo e-mail mentions on metrics to potential similar effects of different forms of external science communication, e.g. in press releases. Helpful insights would be provided by studying whether the promotion of certain articles leads to "spillover effects" of increased attention towards thematically similar articles without explicit promotion. Finally, it would be interesting to analyze the long-term temporal developments of the effects of article promotion examined in this study.

Although only indirectly related to our initial research question (and maybe peculiar to the sample of embargo e-mails used in this study), we have seen that articles from life sciences and biomedicine seem to dominate publishers' communication in embargo e-mails. Moreover, our observations suggest that articles from life sciences-journals might benefit even stronger from being promoted in embargo e-mails than those from multidisciplinary journals regarding visibility in mainstream media. We cannot safely generalize this particular finding to the larger population of scientific articles though, as with regard to the journals represented in them, our samples were uneven between the two subject categories. This observation does however point towards another interesting area for future research.

Our study comes with some limitations. First, as noted before, we cannot guarantee our sample taken from the SMC's e-mail archive to be representative of the entirety of all embargo e-mails sent by scholarly publishers to journalists. Besides, while access to the SMC's e-mail archive allowed us to identify article DOIs that have definitely been mentioned in publishers' embargo e-mails, it is virtually impossible to prove for control group articles that they have not been featured in similar forms of external communication, as not all such communication is recorded somewhere. Also, multidisciplinary journals constituted a considerable share of our sample – due to their multidisciplinary nature, the respective segment of our control group could differ topically from the respective segment of the treatment group. In addition, our choice of databases for citations and altmetrics (CCB and Altmetric.com) does of course introduce its own limitations – for instance, regarding the question what constitutes as a 'mainstream media platform', we rely on Altmetric.com's list of monitored outlets. For a recent analysis of four major altmetric data providers' methodological peculiarities, see Zahedi & Costas (2018). Finally, our restrictive way of determining articles' appearances in embargo e-mails by looking for explicit DOIs probably means that our treatment group is not representative of all embargo e-mail activities performed by scientific publishers in 2016 and 2017 – certain publishers might be more inclined to state DOIs in their e-mails than others (see also the findings Bowman & Hassan [2019] made regarding news releases on *EurekAlert!*). How such potential publisher-specific strategies of promoting their articles affect impact measurements in different ways would be yet another interesting subject for future extensions to this research.

The exploratory case study presented in this article provides first insights on the relationship between publishers' embargo e-mails and the attention that research articles promoted within them receive. Further research is necessary to explain the causes of the observed effects. The difficult accessibility of embargo e-mails will continue to be a major challenge for such endeavors.

3.1.7 Author Contributions

IP, MB, and SL contributed conception and design of the study; SR conducted the acquisition of data on embargo e-mails; MB and SL performed the statistical analyses and provided the figures; IP acquired funding for the research project; SL retrieved data from Altmetric.com, Crossref, the Competence

Centre for Bibliometrics (CCB) database, and Web of Science, performed the coding of journals to research areas and prepared this manuscript's first draft; all authors contributed to manuscript revision, read and approved the submitted version.

3.1.8 Acknowledgments

We thank Altmetric.com and the Competence Centre for Bibliometrics for providing us with access to their APIs and databases, respectively.

3.2 Study D - Path model of the interplay between the promotion and the received attention of research articles

3.2.1 Foreword

Study C provided first insights regarding how external science communication in the form of embargo e-mails might potentially affect research indicators, as well as first quantifications of the magnitude of such effects on various metrics, albeit based on a rather small sample of publications. Study D will expand on the basic research interests laid out in Study C by investigating them within a more sophisticated model, which is based on a substantially larger set of empirical data. Not only shall this model additionally consider another format of external science communication (namely press releases), but also shall it account for the supposedly multi-stage character of the interdependencies between external science communication and attention in different environments as captured by different research metrics.

At the time of writing, the manuscript of Study D is under review for publication in the open access edited volume *The Science-Media-Interface: On the relation between internal and external science communication*, to be published in the book series "Knowledge & Information - Studies in Information Science" by De Gruyter Saur.[28] The study was co-authored with Athanasios Mazarakis and Isabella Peters and another outcome of the research project *MeWiKo*, funded by the German Federal Ministry of Education and Research (grant number 01PU17018). I wish to thank my co-authors as well as my colleagues at *MeWiKo* for their support.

[28]Full reference: Lemke, S., Mazarakis, A., & Peters, I. (under review). Path model of the interplay between the promotion and the received attention of research articles. In Broer, I., Lemke, S., Mazarakis, A., Peters, I., & Zinke-Wehlmann, C. (Ed.), *The Science-Media Interface: On the relation between internal and external science communication*. Berlin: De Gruyter. ISBN 978-3-11-077636-2.

3.2.2 Abstract

Existing systematic analyses of the associations between visibility that research articles receive within different formats of external science communication (e.g., press releases, embargo e-mails, or news stories) and their later impact metrics are mostly restricted to rather specific case studies, despite these studies' recurring findings of substantial potential effects. This study aims to enable a consolidating and more comprehensive perspective on the interplay between research articles' promotion within press releases and embargo e-mails, their publishing journal's prestige, as well as their received attention in mainstream media, on Twitter, and their academic impact as proxied by citations. To achieve this goal, we use the method of path analysis to specifiy models that manifest a range of hypotheses motivated by literature and theory on the relationships that may exist between these variables. We estimate and evaluate our models based on a dataset of 67,581 research articles, which we construct through a combination of empirical data from Web of Science, Altmetric.com, EurekAlert!, and the Science Media Center Germany. The resulting model essentially confirms the conformity of the hypotheses we derived from past literature with the large set of empirical observations within our sample. More specifically, our findings highlight the considerable associations between promotion in press releases and embargo e-mails and the attention research articles can be expected to receive on both mainstream- and social media.

3.2.3 Introduction

Citation analysis – i.e., the analysis of citations to scientific publications as indicators for the latters' academic influence – assumes that the likelihood of an article getting cited correlates with relevant inherent qualities, e.g., its scientific rigor, its novelty, or the significance of its findings. While certain studies argue that citations can indeed be useful predictors for such inherent qualities (see Aksnes et al. [2019] for an in-depth discussion), scientometric research has revealed numerous factors without apparent relation to quality that also affect an article's likelihood of being cited. Being informed of their existence and effects is crucial to assess individual citation analysis applications' validity and potential limitations.

In their review of literature examining factors affecting citations, Tahamtan et al. (2016) divide 28 such factors into the categories of paper related, journal related and author related factors, albeit acknowledging that this selection is not exhaustive. As the three categories identified by Tahamtan et al. (2016) indicate, a large portion of the research focused on how respective research findings are communicated within the academic sphere, e.g., how decisions regarding publication formats or publication venues affect citations. These decisions can, for the largest part, be considered aspects of *internal* science communication or *scholarly* communication, i.e., the communication of research findings primarily targeted at an academic audience. However, another less analyzed set of issues affecting citations results from how research is featured and processed in channels of *external* science communication, i.e., in media targeted at stakeholders outside the scientific community.

Considering the scientific journal article as the exemplary unit of observation, such processing by public media may begin even before said article's publication: many scholarly journals regularly disseminate advance information on upcoming issues to science journalists in an arrangement known as an *embargo* (Kiernan, 1997). Briefly worded, this arrangement provides registered journalists with early access to yet unpublished research findings in exchange for their pledge to not pass on that information before a specified embargo date has passed. The embargo system serves both involved parties well: for the science journalists, embargoed information allows them to timely prepare their coverage on new findings while also providing a certain assurance that other journalistic outlets' respective stories will not leapfrog their own – provided those other outlets do not break the embargo date, of course. For the scholarly publishers, the embargo system provides an opportunity to highlight specific publications and topics to the media as well as a certain control over the respective coverage's timing (Kiernan, 2014). As the embargo information given to journalists usually requires prior registration and thus is not openly accessible, it remains difficult to assess how this specific form of promotion affects the attention individual articles later receive, let alone their probability of getting cited.

Other types of interventions in external science communication that serve the purpose of directing attention to certain articles are less difficult to track. One of the most common tools used by press officers of scholarly publishers, journals or research institutes to promote selected publications is the press release, described by Carver (2014, p. 2) as "essentially a short news article written in a journalistic style that explains a newly published scientific result in a common and not too specialized language". While research on the relationship between press release promotion and articles' later citations is rather scarce, some case studies indicate a positive association between the two (Chapman et al., 2007; Fanelli, 2013; Lemke, 2020), although Fanelli (2013) found this association to become negligible once respective articles' media coverage was controlled for.

Speaking of media coverage, a further body of case studies examined how mentions in newspapers affect scientific publications' later citations (Dumas-Mallet, Garenne, Boraud, & Gonon, 2020; Fanelli, 2013; Kiernan, 2003a; Phillips et al., 1991). Phillips et al. (1991) found articles from the *New England Journal of Medicine (NEJM)* to receive significantly more citations if these had been featured in the *New York Times (NYT)* – although this advantage could not be detected for NEJM articles that had been chosen for coverage in NYT issues that in the end were not disseminated due to the NYT being on strike. This finding backs up what is called the *publicity hypothesis*, which attributes articles' increase in citations associated with press coverage to the concomitant increase in visibility; as opposed to the *earmark hypothesis*, which explains higher citation counts for press-covered research articles with the assumption that journalistic agents merely apply similar criteria in their decisions on which research to cover as researchers do when deciding which research to cite (Phillips et al., 1991). Kiernan (2003a) added to the work by Phillips et al. (1991) by additionally regarding how coverage from twenty-four daily newspapers and several evening broadcasts of major U.S.-television networks affects citation rates. The author found that the NYT's influence on citation rates is not unique, as NYT coverage did not correlate significantly with citation counts once the author controlled for coverage by other newspapers and television. In his study of the association between newspaper coverage and citations, Fanelli (2013) also found regional effects to play a substantial role, as the apparent positive effect of newspaper coverage on citations was stronger for English media than for Italian media, which primarily only affected authors from Italy. More recently, in their analysis of the citation advantages of 162 biomedical association studies reported in newspapers from six specific countries, Dumas-Mallet et al. (2020) found the strength of the observed effects to depend on the influence of the covering newspaper as well as on the number of published press articles. Moreover, they found the positive correlation between newspaper coverage and citation counts to be most significant for research articles published in journals with lower impact factors.

As the examples of embargo e-mails, press releases, newspapers, and television shows illustrate, the landscape of sources of external science communication with potential effects on research impact is vast, heterogeneous, and at times opaque. Several current developments add to this intricacy: ongoing professionalization of institutes' own research communication, as well as increasing commitments to Open Science and the "third mission" of higher education (Berghaeuser & Hoelscher, 2020) eradicate formerly existing boundaries between scientists, journalists, and public audiences and lead to the establishment of new tools and formats of science communication, many of which enable more direct, bidirectional exchanges of and about research content (Liang et al., 2014). In this vein, an additional layer of complexity also results from the increasing digitalization of journalistic media, the advent of social media, and the continually blurring line between these two spheres. In a reciprocal give and take, news stories are posted to and might evoke public discussions on social media platforms like Facebook or Twitter, while journalists also use these platforms to gather news in the first place (Hermida, 2012). Although altmetrics (Priem et al., 2010) provide flexible technical means to track the attention individual research publications attract on various online domains, modeling the relationship between the attention received, for instance, on Twitter, mentions received in the news, and academic citations remains complicated because of such chicken-or-egg dilemmas.

While numerous past studies put spotlights on certain types of research promotion in external science communication and respective research articles' later impact metrics, what is lacking are more comprehensive models that provide entry points to understanding the interdependencies working between the various interventions made to increase research's publicity and the attention observable within different spheres of media and academia as a whole. The present analysis represents a step towards closing that research gap.

3.2.4 Research Aim and Model

This study aims to disentangle the complexity of relationships between various formats of science communication and research impact by formulating and testing path models comprised of several variables that capture particularly relevant manifestations of research articles' impact and uptake in external science communication. We derive hypotheses from literature about the interplay between attention research articles receive within press releases, embargo e-mails, mainstream media, social media, their citations within the academic sphere, and their publication venue's prestige and apply the method of path analysis (Regorz, 2021; Streiner, 2005) to see to which degree these hypotheses can be confirmed through the testing of models based on a large set of empirical data. We choose the method of path analysis because of suspected multi-level interdependencies between the different indicators to be included in our models and rely on its implementation from the R-package *lavaan*[29]. As an extension of the statistical method of multiple regression, path analysis allows us to test more complex models in which certain variables simultaneously affect and are affected by others.

The model we start our analysis with (Figure D1) is motivated by past empirical research and theory and based on six such variables, which we explain in detail after listing them first:

- research articles' numbers of mentions in mainstream news media (*En1*);
- their numbers of mentions on Twitter, as a prototypical example of a social media platform that is broadly used in academic contexts (*En2;* Tahamtan & Bornmann, 2020);
- their numbers of (academic) citations (*En3*);
- the (binary) variable of whether the articles have been featured in a publisher's embargo e-mail (*Ex1*);
- the (binary) variable of whether the articles have been promoted in a press release (*Ex2*);
- the prestige of their publishing journal, proxied by the median number of citations received by articles within said journal during the three years preceding the observed article's publication (*Ex3*). We use median-based impact factors instead of the more commonly seen mean-based impact factors to at least partly account for problems resulting from the latters' skewed distributions; see Kiesslich, Beyreis, Zimmermann, & Traweger (2021) for a detailed discussion of this issue.

Our starting model will be fitted to a large set of empirical data, then evaluated for its fit, and afterwards, if appropriate, respecified and reevaluated till no substantial enhancements appear possible with the data at hand. This process of model estimation serves two primary purposes: first, it serves as a test of whether the range of hypotheses about the interplay between certain events of external science communication and article metrics drawn from past empirical research conforms with a large set of

[29] https://cran.r-project.org/web/packages/lavaan/index.html

empirical observations; and second, it shall provide a comprehensive view on how the hypothesized interactions compare to each other.

Figure D1 shows the model we start this analysis with. We assume mentions in embargo e-mails and press releases to be exogenous variables as in almost all cases these events will happen before or very shortly after the promoted article's publication – making it implausible to assume that respective press officers' or editors' decisions could be affected by any of the endogenous variables in our model, which all accumulate later. As a third exogenous variable, we include the publishing journal's prestige, as a large body of research has found this to be a crucial factor concerning an article's expected citations (see the review by Tahamtan et al. [2016] for an overview over such studies), which therefore cannot be omitted in an endeavor of convincingly modeling the attention articles receive in academic and media spheres. Furthermore, it seems reasonable to assume that also in science journalism there is a common awareness of what the most prestigious journals within a respective covered field of research are – perhaps even of journal impact factors as indicators for such prestige. This makes it likely that journal prestige would be an important variable to explain articles' expected media presence as well (which is also suggested by results from previous case studies; Dumas-Mallet et al., 2020).

To briefly illustrate the effects assumed in our model with a fictitious example: imagine we have an article that was published in 2018 in the *New England Journal of Medicine* (NEJM), an indubitably highly prestigious journal of its field (into our model's estimation, this prestige would in this case enter as the median of citations that NEJM-articles received over the years of 2015 to 2017, as our article was published in 2018). Based on past studies (see below), we believe this prestige to have a positive effect on the likelihoods with which journalists will choose said article for their stories, users will post about it on social media, and researchers will cite it in their works. The article's presence on mainstream and social media will likely be higher, if the article also received promotion within an embargo e-mail or a press release – and the media presence itself will have a positive effect on the likelihood of the article being cited within academic publications as well.

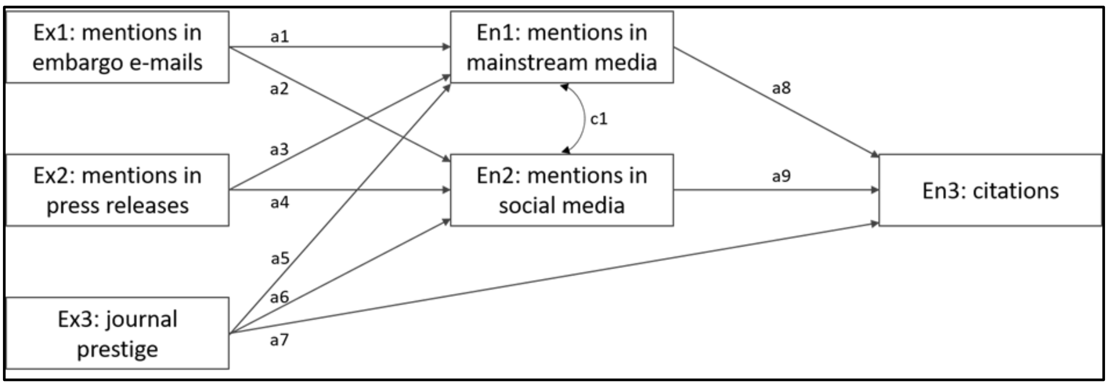


Figure D1: Path model with three exogenous variables (*Ex1-3*), three endogenous variables (*En1-3*), nine assumed effects between these (*a1-9*) and one assumed covariance (*c1*).

The reasoning behind the supposed effects in our model (*a1-9*) is based on positive associations found in past empirical studies:

- Such suggesting that a research article that is featured within an embargo e-mail will be more likely to be mentioned in mainstream and social media (*a1* and *a2*) have been found in a case-control study by Lemke, Brede, Rotgeri, & Peters (2022). Similarly, the case studies by Fanelli (2013), Lemke (2020), and Stryker (2002) all found positive associations between articles being promoted in press releases and their later mainstream media mentions (*a3*). Regarding indicators for online attention on the other hand, the findings by Chapman et al. (2007) and Lemke (2020) suggest that press release promotion is associated with an increase of these as well (*a4*).
- The positive correlation between journal prestige (*a7*), most commonly represented by journal impact factors, and citations has been confirmed by numerous studies (see Tahamtan et al., 2016). Correlations between journal prestige and certain altmetrics (*a5* and *a6)* have, for instance, been found by Li & Thelwall (2012) and Thoma et al. (2015). Studies on the positive correlation between mentions in news media and future citations (*a8*) have already been discussed in detail in this article's introduction (Dumas-Mallet et al., 2020; Fanelli, 2013; Kiernan, 2003a; Phillips et al., 1991). The (varying) potential of social media-based altmetrics to predict later citations (*a9*) is proposed by another rich body of studies (see, for instance, Konkiel et al., 2016; Priem et al., 2012; Thelwall & Nevill, 2018).

While cases are imaginable in which high numbers of citations for an article precede it being mentioned in news media or on social media, we for now consider this a less likely case due to findings that indicate that most attention around research articles in social and news media usually happens soon after their publication (Brainard, 2022; Waltman et al., 2021), while the majority of a scientific article's citations typically occur two to seven years later (Schloegl & Gorraiz, 2010). Similarly, a positive direct effect of having a press release on expected later citations would be thinkable. However, the findings by Fanelli (2013) indicate that this effect likely is already covered to a large extent by the combined effect of positive associations between press release promotion and mainstream media mentions (*a3*) and the positive correlation between mentions in news media and future citations (*a8*). Finally, regarding the presumably complex reciprocal relationship of social media content spawning news content and vice versa – as for example supported by correlations found by Haustein et al. (2015) or Lemke et al. (2022) – we do not assume a unidirectional causal effect, but model the relationship as a covariance (*c1*) instead.
We use maximum likelihood estimation with robust standard errors and a Satorra-Bentler scaled test statistic (Satorra & Bentler, 2010) for model estimation as citation and altmetrics data are usually not normally distributed. For the evaluation of models' global fit, we also consult the (robust) comparative

fit index (CFI), root mean squared error of approximation (RMSEA), and standardized root mean squared residual (SRMR) and apply established cutoff criteria proposed by Hu & Bentler (1999). For evaluations of models' local fit and to identify potential improvements, we calculate modification indices with a cutoff value of 10 (Regorz, 2021). All statistical analyses are performed in R (R Core Team, 2020).

3.2.5 Data Sampling

To analyze the relationships between mentions of research articles in external science communication and their performance regarding citations and altmetrics, we start with data obtained from *EurekAlert!*[30]. EurekAlert! is a platform for the distribution of research-related press and news releases that was set up by the *American Association for the Advancement of Science* in 1996. The platform enables publishers, universities, research institutes, corporations, government agencies, and eligible organizations that engage in scientific research to disseminate press releases to journalists and the public against payment of submission fees. With over 14,000 registered journalists from more than 90 countries, EurekAlert! has become the most prominent multilingual platform of its kind (Vrieze, 2018). Or, as Vrieze (2018) put it: right now, EurekAlert! has become for science press and news releases "what *Google* is for searching and *Amazon* for online shopping".

We focus our analysis on research articles published between 2016 and 2018 to balance the timeliness of the research analyzed in our study with the ability to obtain meaningful citation windows. Thus, the starting point for our dataset is data on 84,194 press releases provided to us by EurekAlert! that were published on the platform between 2016 and 2018. EurekAlert! press releases on new research articles frequently contain a DOI link to the respective article – this enables us to extract 41,937 DOIs. Of these identifiers, 34,055 refer to a valid Web of Science record with publication type journal, document type of either article or review, and a publication year between 2016 and 2018. These 34,055 DOIs form the starting point for our bibliometric analysis.

As a next step, we enrich this data with information about the presence of research publications in publishers' embargo e-mails to journalists. For this, we rely on data from the *Science Media Center Germany (SMC)*. The SMC is an editorially independent non-profit institution with the mission of supporting journalists in reporting on science-related topics. As one service contributing to this mission, the SMC regularly sends out comments by scientific experts on new research findings that are still under embargo. To be able to provide this service, the SMC aims to monitor as many scholarly journals that send out embargo e-mails as possible. Since 2016, the SMC accumulated 2,638 ingoing e-mails identified as embargo e-mails from scholarly publishers. Each of these embargo e-mails contains information about one or more upcoming articles from one or more journals belonging to the same publisher.

[30] https://www.eurekalert.org/

As we rely on DOIs to track articles' citations and altmetrics, we query the SMC's e-mail archive for embargo e-mails containing references to scientific articles via DOI. This led to 953 articles with Web of Science records of document type *article* or *review*, publication type *journal*, and publication year 2016, 2017, or 2018 that also have been promoted in an embargo e-mail between 2016 and 2018 with reference to a DOI. Merging these articles with our dataset of 34,055 articles promoted on EurekAlert! enlarges our sample to 34,413 articles that received promotion in an embargo e-mail and/or a press release on EurekAlert!. Table D1 shows to which extent the two regarded types of promotional activities overlap within our dataset. The two events are not statistically independent from each other (Fisher's exact test, two-sided, $p < .001$).

Table D1: Contingency table of promotion in embargo e-mails and on EurekAlert! for research articles in our sample.

		Featured in embargo e-mail?	
		No	Yes
Featured on Eurekalert!?	No	0	358
	Yes	33,460	595

We also added 'control group articles' to our dataset – articles which, to our knowledge, did not receive any promotion in EurekAlert! press releases or publishers' embargo e-mails, but otherwise should have been published under comparable circumstances as our 'treatment group articles'. To do so, we match every article from the treatment group (i.e., the group of articles that received the 'treatment' of getting promoted within an embargo e-mail, an EurekAlert! press release or both) to one random article from Web of Science that was published in the same publication year and with the same ISSN, but which is not part of the treatment or control group yet. We again restrict our matching to the Web of Science document types *article* and *review*, to avoid matching research articles with, for instance, editorials, letters to the editor, etc..

For articles from multidisciplinary journals (e.g., PLOS ONE, Nature, Science), this procedure might lead to suboptimal matchings of articles from domains with highly heterogeneous citing or publication behaviors – it would, for instance, be questionable to match a biomedical article from PLOS ONE with a sociology-related article from PLOS ONE. For articles published in journals classified as *multidisciplinary* in Web of Science (24.65% of our sample), we therefore apply an additional step, inspired by the matching methodology described in Fraser, Momeni, Mayr, & Peters (2020). To do so, we reclassify these articles and all potential control group articles' disciplines based on the most frequently cited Web of Science subject categories among their references and subsequently use concurrence among these new classifications as an additional matching criterion for articles from multidisciplinary journals.

It should be noted that for a relatively small number of articles from our sample (3.62%), our control group matching procedure does not return a valid match fulfilling the criteria explained above. Thus, a total of 67,581 unique publications serves as our dataset for model estimation.

For bibliometric analyses we use data provided by the *Competence Centre for Bibliometrics* (CCB). The CCB administers databases based on Web of Science, which are updated annually. The bibliographic and citation data used in this study therefore reflects the status of Web of Science from April 2020. The altmetric data used in this study (i.e., articles' numbers of mentions in mainstream media and on Twitter) was obtained via the API of *Altmetric.com* in November 2021.

3.2.6 Results

Before model estimation, we briefly examine some of the articles' metadata to achieve an understanding of our dataset's composition.

In total, 3,419 individual journals are represented within our dataset, the most frequently represented journals being *Nature Communications* (5.38% of all articles), *Scientific Reports* (4.26%), *PNAS* (3.96%), *PLOS ONE* (2.62%) and *Nature* (2.57%). The frequency of the remaining journals follows in a long tail distribution, with most journals (1,450) only being represented by a single article each within our treatment and control group. Applying traditional Web of Science subject categories, a total of 241 different categories is represented among our data. Table D2 shows the ten most strongly represented journals and Web of Science subject categories from our sample and their respective shares. The outstandingly high share of the category Multidisciplinary Sciences (15.32%) appears to back up what the examination of most commonly represented journals had also shown – namely, that prominent high-impact journals like Nature, PNAS, or PLOS ONE account for a large share of the research that gets featured in press releases and/or embargo e-mails. Also, the look at the most heavily represented categories suggests a particularly substantial representation of research dealing with biomedical subjects.

Four of the six variables in our model are metric: articles' citation counts, numbers of mentions in tweets, numbers of mentions in mainstream media (MSM mentions), and the median of citations articles within the respective journal received during the past three years (abbreviated as JCM or journal citation median). Table D3 shows descriptive statistics of our sample regarding these variables.

Table D2: Most frequent journals and Web of Science subject categories among our sample.

Journal	*Freq.*	*Subject Category*	*Freq.*
Nature Communications	5.38%	Multidisciplinary Sciences	15.32%
Scientific Reports	4.26%	Cell Biology	4.95%
PNAS	3.96%	Biochemistry & Molecular Biology	3.97%
PLOS ONE	2.62%	Neurosciences	2.89%
Nature	2.57%	Materials Science, Multidisciplinary	2.54%
Science	2.24%	Chemistry, Multidisciplinary	2.41%
Science Advances	1.33%	Biology	2.35%
eLife	1.14%	Ecology	2.33%
Cell Reports	1.11%	Medicine, General & Internal	2.19%
Physical Review Letters	1.05%	Environmental Sciences	2.03%

Table D3: Descriptive statistics of metric variables within the model.

	Citations	*Tweet mentions*	*MSM mentions*	*JCM*
Minimum	0	0	0	0
Mean	26.92	39.25	8.70	12.37
Median	12	7	2	8
Maximum	13,715	9,951	533	876
Standard deviation	106.98	150.13	20.51	14.10

According to our research aim and method we estimate a path model based on our publication data and the specification outlined in Figure D1. To account for the considerable differences in their variances, the four metric variables included in our model (citations, tweet mentions, mainstream media mentions, and journal citation median) were all standardized via z-transformation (subtraction of mean and division by standard deviation for each observation) before model estimation.

We evaluate our model's global fit before we consult the estimated model’s coefficients. The χ^2 test for exact model fit is significant (χ^2 = 11.503, *df* = 2, *p* = .003). Comparative fit index (CFI), root mean squared error of approximation (RMSEA), and standardized root mean squared residual (SRMR) for our model are shown in Table D4.

Table D4: Global fit indices for our first model.

	CFI	*RMSEA*	*SRMR*
Model 1	1.000	0.010	0.003

Applying cutoff criteria recommended by Hu & Bentler (1999), the CFI > .95, RMSEA < .06, and SRMR < .08 all indicate an already relatively good fit between our hypothesized model and observed data. Significance of the χ^2 test is undesirable, however, so we are interested in whether further substantial improvements to the model are possible. Using the lavaan package's function *modindices*, we calculate modification indices to see if and how the change of existing or addition of further effects to our model could further increase its fit with the data. The suggested model change associated with the highest expected model improvement is the replacement of the unidirectional effect *a8* by a covariance – so the abandonment of the assumption that mainstream media mentions have a mostly unidirectional effect on citations in favor of a model in which no clear causal direction between citations and media mentions is assumed. As such a non-unidirectional relation might make sense from a theoretical standpoint as well, we respecify the model accordingly and again assess its global fit through a χ^2 test (χ^2 = 2.161, df = 2, p = .339) and the calculation of fit indices shown in Table D5.

Table D5: global fit indices for our second model.

	CFI	*RMSEA*	*SRMR*
Model 2	1.000	0.001	0.001

Both χ^2 test and fit indices indicate an improvement in global model fit compared to the first model. As another iteration of modification index calculation does not reveal any further model changes that would be linked to substantial expected improvements (applying our cutoff value of 10 for modification indices), we continue with the interpretation of the model's coefficients. Figure D2 shows the resulting model along with its path coefficients and covariances, and Table D6 provides additional statistics. Table D7 shows R^2-values for the three endogenous variables within the model.

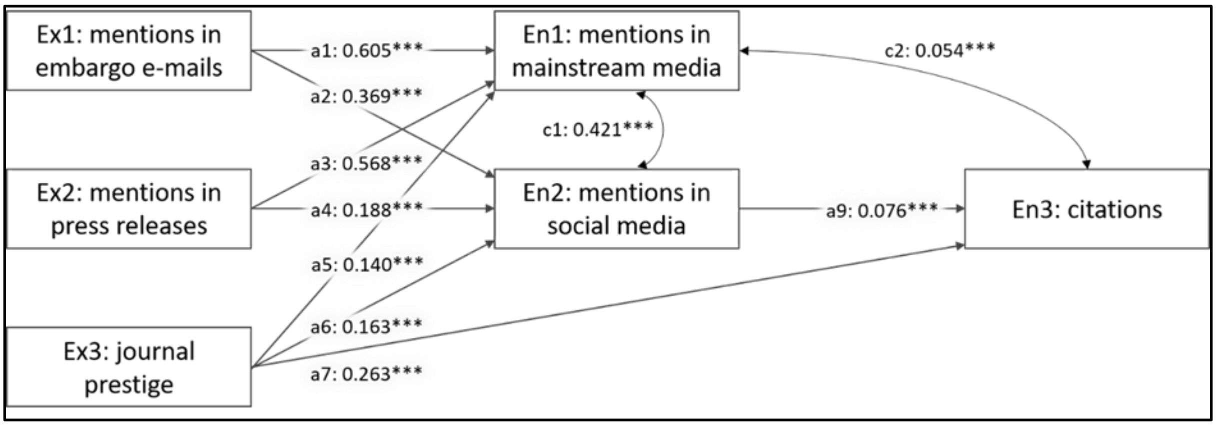


Figure D2: Final path model with path coefficients; *** indicate $p < .001$.

Table D6: Final path model's parameter estimates.

	Estimate	*Standard error*	*z-value*	*p*
a1	0.605	0.062	9.773	<.001
a2	0.369	0.061	6.075	<.001
a3	0.568	0.007	77.265	<.001
a4	0.188	0.008	24.859	<.001
a5	0.140	0.010	14.608	<.001
a6	0.163	0.011	15.248	<.001
a7	0.263	0.012	22.336	<.001
a9	0.076	0.014	5.385	<.001
b1	0.421	0.036	11.699	<.001
b2	0.054	0.015	3.661	<.001

Table D7: Endogenous model variables' R^2-values.

	Citations	*Tweet mentions*	*MSM mentions*
R^2	0.081	0.037	0.104

3.2.7 Discussion

The obtained fit indices and significances of effects suggest a good fit between the model we hypothesized based on findings from past case studies (Chapman et al., 2007; Dumas-Mallet et al., 2020; Fanelli, 2013; Kiernan, 2003; Konkiel et al., 2016; Lemke, 2020; Lemke et al., 2022; Li & Thelwall, 2012; Phillips et al., 1991; Priem et al., 2012; Stryker, 2002; Tahamtan et al., 2016; Thelwall & Nevill, 2018; Thoma et al., 2015) and the actual publication data at hand. Thus, the first stated purpose of our analysis, which was to test whether our hypotheses about the interactions between events of external science communication and article metrics drawn from past studies conform with a large set of empirical observations, can be considered as fulfilled.

Regarding the effects of exogenous variables on various media mentions, our final model suggests that promotion within embargo e-mails seems to affect articles' expected mainstream media presence to a slightly larger (0.605) but similarly high degree as promotion within a press release (0.568). Furthermore, considering social media presence, the comparatively higher strength of embargo e-mail promotion as a predictor for later mentions becomes even more apparent (0.369, opposed to 0.188 from press releases). As it might seem counterintuitive that embargo e-mails would to a greater extent contribute directly to an article's visibility on various media than the more openly accessible format of a press release, these findings might also suggest that embargo e-mail promotion (as tracked by the Science Media Center Germany) is – compared to press release promotion – reserved for even more elite research publications, which due to innate qualities not represented in our model (e.g., particular originality, societal value, or some form of provocativeness, to just name a few possibilities) will likely attract more media attention on their own. If the selection of articles for embargo e-mail promotion indeed typically follows more rigorous criteria than the selection for press releases, this finding might also be considered a hint towards the validity of the earmark hypothesis as suggested by Phillips et al. (1991). However, further research on publishers', journals', and institutions' criteria behind the selection of research articles for both these forms of promotion would be necessary to solidify this hypothesis.

Our third exogenous variable, journal prestige, proves to be a significant predictor for citations (0.263), mainstream media mentions (0.140), and Twitter mentions alike (0.163) – the direct effect on citations, however, is most substantial. What might be considered a surprising finding are the fairly weak relationships between both mainstream- and social media mentions and citations (0.054 and 0.076 respectively). Possibly, controlling for journal prestige (which, as we have seen, is also a strong predictor for mentions in both forms of media) already accounts for most of the citation advantage expected from increased media presence; this interpretation is in contrast to findings by Dumas-Mallet et al. (2020) though, who for their biomedical sample found the expected citations of articles from lower impact journals to benefit particularly strongly from media mentions. Another interesting and perhaps surprising insight concerns the observation that led to our model respecification, namely that a model without assumed directional effect of mainstream media mentions on citations fits the empirical data better than our initial model, where this effect was present. This finding might hint at a more bidirectional relationship between academic impact and media attention than past case studies, which in line with Phillips et al.'s (1991) publicity hypothesis often focused on how media exposure might increase citations without much considering the opposite phenomenon of outstanding research evoking media attention, suggested.

Also, although not at the center of our inquiry, our look at the discipline- and journal-wise composition of our sample of articles promoted in either press releases or embargo e-mails indicates that large shares of them were published in multidisciplinary high-impact journals and cover subjects from life sciences, which would confirm observations of media coverage of science made in previous studies (e.g., Elmer et al., 2008; Lemke, Sakmann, Brede, & Peters, 2021).

Overall, our findings support the existence of statistically significant associations between the promotion of research in science PR (i.e., embargo e-mails and press releases) and impact metrics that past case studies had found for individual parts of our model's components and smaller, more restricted samples of scientific articles (Dumas-Mallet et al., 2020; Fanelli, 2013; Kiernan, 2003; Lemke, 2020; Lemke et al., 2022; Phillips et al., 1991). It thus underlines the importance of further, more in-depth research on the selection criteria with which PR officers and science journalists decide on which research to cover (see for instance Badenschier & Wormer, 2012; Broer, 2020), as these criteria might diverge substantially from the characteristics that metrics-based indicators are supposed to reflect when used for evaluative purposes.

3.2.8 Conclusions

We specified a path model of the interplay between two prevalent measures of external science communication, journal prestige, presence in mainstream and on social media, as well as citations and tested the model against a large set of empirical data from Web of Science, Altmetric.com, EurekAlert! and the SMC Germany. The empirical results confirmed the significance of the effects assumed in the model and signaled substantial associations between the three exogenous variables and articles' expected later impact in both media and academia. In particular, the results highlighted the considerable potential effects of embargo e-mails and press releases on (social) media attention and of journal prestige on citations.

Our findings add to existing case studies on associations between media coverage of research and said research's impact by taking a more comprehensive perspective than past studies, which mostly focused on fewer variables, and by analyzing a large sample of articles from a wide variety of journals. Also, to the best of our knowledge, there has neither been a comparative analysis of the effects of embargo e-mails and press releases before, nor an application of path analysis in a large-scale bibliometric analysis like ours. For the interpretation of the results, readers should however keep in mind that our sampling approach started with articles featured on either EurekAlert! or within an embargo e-mail tracked by the SMC Germany and that our study therefore remains a case study whose generalizability might be limited by its sample. Moreover, with its limited number of variables considered, our study can only serve as a starting point for disentangling the complex relationships and effects between the systems of science communication and academic reputation.

Our study comes with some further limitations. First, it is virtually impossible to prove for 'control group articles' that these did not ever receive any kind of external promotion similar to a press release or an embargo e-mail that was not tracked by our data sources. However, as both observed kinds of promotion will most likely be the exception rather than the rule among randomly sampled articles as done in this study, we consider it unlikely that this limitation would have impaired our results significantly.

A second limitation results from our reliance on DOIs to identify references on EurekAlert! and in the SMC's database. While the high amount of DOIs found within the two data sources suggests that they are a common way of referencing articles within them, our DOI-based approach means that our data likely underestimates the total shares of articles receiving promotion within press releases or embargo e-mails.

Third, as is the case for many bibliometric and altmetric analyses, results should be interpreted with the data sources used to obtain metrics in this study in mind, as with our reliance on these sources, we also inherit some of their limitations; e.g., the limitation of Web of Science to publications indexed by it, or the limited transparency of what Altmetric.com tracks as mainstream media mentions and what it does not (for a recent assessment of Altmetric.com's news mention data see also Fleerackers et al., 2022).

Future research should work on providing a more detailed picture of the criteria that affect a research article's likelihood of receiving certain forms of promotion in external science communication, as well as investigate which article properties (for instance, considering topics, authors, or publication venues) influence respective promotional activities' effectiveness with regard to impact generation. Furthermore, as citation norms and behaviors as well as the ways media discusses research findings can vary considerably between fields, a worthwhile continuation of this analysis could lie in the specification, estimation, and comparison of discipline-specific models of the interdependencies examined in this study.

Finally, it should be noted that we focused our analysis on six variables that we deemed particularly relevant and that were generated via literature research to explain the phenomena we aimed to explore. However, there certainly are more factors that might affect the likelihoods of research articles getting promoted in external science communication, featured on various media platforms, or cited within other publications, that could in principle be included in an analysis as ours. In fact, we would argue that within a context as intricate as the system of science communication, no matter which number of factors a model considers, there will probably always be more one could add. Nevertheless, the incorporation of further variables would be another promising path for future research that aims to build upon this work to take – for instance, including a parametrization of authors' prestige might lead to new valuable insights on the interplay between promotion and received attention of research articles, to name just one example.

3.2.9 Author Contributions

SL performed the statistical analysis, provided the figures used in the manuscript and wrote its first draft; IP acquired funding for the research project. All authors contributed conception and design of the study, contributed to manuscript revision, and read and approved the submitted version.

3.2.10 Acknowledgments

We thank Sophie Rotgeri for her crucial support regarding the provision of data on embargo e-mails by the *Science Media Center Germany* and on press releases from *EurekAlert!*, as well as Nicholas Fraser for his input regarding the control group matching algorithm for articles from multidisciplinary journals. Moreover, we thank the SMC, EurekAlert! and the *Competence Centre for Bibliometrics* for providing us with access to their databases. This study is part of the research project *MeWiKo* funded by the German Federal Ministry of Education and Research (grant number 01PU17018).

Chapter 4: Discussion

This thesis set out to make two major contributions to the ongoing discussion of different research metrics' potential as impact indicators. In Chapter 2, two user studies were presented with the aim of painting a detailed picture of the status quo of the perception and use of various metrics. This approach should help to assess the maturity of different types of research metrics from the viewpoint of researchers, while the investigation of scholars' concerns about indicators should aid in identifying the needs that better accepted next-generation metrics would need to fulfill. In Section 4.1 within this chapter, the key observations made in Chapter 2 will therefore be condensed, while tangible methods and actions will be proposed on how to disperse the problems users commonly see with current research metrics.

The thesis' second major goal was to contribute to diminishing the frequently described unclarity of what it is that research metrics measure, by targeting a specific but wide research gap within that context: the question of the relationship between research metrics and promotional activities in external science communication, e.g., mentions of research publications within press releases or embargo e-mails. Section 4.2 will consolidate the central observations and arguments made on this issue in Chapter 3, expand on them and derive resulting implications for future use and research of metrics.

Finally, this chapter will close with conclusions that briefly summarize the central practical implications of the thesis, as well as give an overview over the most pressing challenges future research building upon this work should address.

4.1 Perception and Use of Research Metrics

Studies A and B provided detailed information on researchers' thoughts and concerns about the use of various metrics for the assessment of research - knowledge that should inform the development of useful and accepted next-generation metrics, which adequately address existing points of criticism against current indicators.

Both presented user studies have shown that a certain degree of awareness and use of various impact indicators is common for many of the participating researchers. In particular the interviews presented in Study A indicate that such use is usual even for researchers in early career stages. This finding is also in line with a more recent study by Nicholas, Herman, et al. (2020), who found 60.9% of surveyed early career researchers to use citation-based metrics and 41.0% to use altmetrics for at least one given purpose. Such purposes reported in Study A include the utilization of journal impact factors or citation counts as filters during literature research, to identify 'mandatory' references concerning a respective topic, to find particularly progressive research within a field of interest, or to decide on venues in which to publish own works. Researchers' use of metrics in information-seeking activities has since then been confirmed by further case studies (Bakker et al., 2020; Ma & Ladisch, 2019; Nuredini, 2021), while other recent surveys have shown researchers to also use metrics for purposes of grant applications, promotion applications, or to demonstrate achievements in CVs, annual performance reviews, and reports to researchers' institutions (Haddow & Hammarfelt, 2019; Hammarfelt & Haddow, 2018).

The survey presented in Study B confirmed the significant role metrics appear to play in literature selection processes for many researchers - although it should be noted that quantitative metrics are merely one piece of information, next to several qualitative criteria, that researchers commonly consult when deciding on which literature to read (see also Tenopir et al., 2011). Moreover, the cluster analysis performed in Study B highlights that there are various different kinds of metrics users among researchers, which accept and make use of different types of metrics to different degrees. However, regarding the differentiation between traditional citation-based indicators (e.g., citation counts, journal impact factors, or the h-index), web-based usage metrics, and altmetrics, overall Studies A and B (as well as other recent surveys; Aung et al., 2019; Nicholas, Herman, et al., 2020) repeatedly arrived at the conclusion that altmetrics are still far behind bibliometric indicators concerning their renown, acceptance, and use among researchers, with usage metrics (i.e. download and view counts) standing somewhere between the other two families of indicators in these regards (Lemke, Mazarakis, & Peters, 2021; Lemke et al., 2019; Miles et al., 2018).

Many specific concerns were expressed by the studies' participants regarding the use of metrics for research assessments. Some of these concerns reflect and some of them extend the open challenges for scientometric research that have been identified in the past (see also Section 1.3 regarding challenges specifically applying to altmetrics), and together lead to requirements that next-generation metrics would need to fulfill to have the potential to be widely accepted and considered helpful by researchers.

In the following, the concerns identified in user studies and meta-publications (such as for instance the *Leiden Manifesto*) described in Chapter 2 will be condensed and linked to concrete requirements for useful research metrics that they imply. Along with these requirements, hints or recommendations for how to approach them or indications of where need for further research persists will be presented.

4.1.1 Openness and Transparency of Metrics - Provider-Related Aspects

Among the most frequently stated concerns regarding various metrics - ranging from citation-based indicators like the journal impact factor to altmetrics - was the concern that many metrics appear non-transparent to researchers and their meanings therefore opaque. While this problem is multidimensional and also has user- and application-related aspects to it (which will be discussed in further detail in the next subsection), substantial parts of the challenge primarily concern the providers of metric data and - in principle - would not be that complicated to resolve. ‘Providers’ in this context refers to any parties that build, maintain, and provide databases or software tools that offer their users metrics data, e.g., bibliographic databases like Web of Science or Dimensions, altmetric aggregators like Altmetric.com or Plum Analytics, or more specialized discovery tools operating on citation graphs like for instance *Google Scholar* or *scite_*[31].

The necessity for metrics used for evaluation purposes to be transparent has also been a prominent subject of discussion in the high-profile publications of *The Metric Tide Report* (Wilsdon et al., 2015) and the *Leiden Manifesto* (Hicks et al., 2015). Transparency of indicators concerns various steps of their construction: first, the databases used for the derivation of a metric should be openly accessible. Ideally this relates to both the processes said database’s provider uses for its construction (e.g., the algorithms used for the indexation of publication records and the retrieval of cites/mentions of those records), as well as to the resulting metric data that ultimately serves as the corpus that indicators are derived from. Such openness should empower stakeholders - especially those under evaluation - to comprehend (and, if necessary, debate) the exact procedure behind a respective act of evaluation, as well as allow them to reproduce and verify analyses and thus ensure data quality.

Obviously, economic interests often oppose the opening up of databases, as neither Clarivate nor Elsevier have an interest in giving away their products - in this case the contents of the predominant bibliographic databases Web of Science and Scopus - for free. However, in recent years the rise of movements devoted to Open Science also brought with it a growing sensibility for the worth of open bibliographic data, leading to promising initiatives like *I4OC* or *OpenAlex*, who aim to provide aforementioned openness and are continuously improved to become suitable alternatives to the proprietary databases that are still used for the majority of bibliometric analyses (Priem, Piwowar, & Orr, 2022). For providers of altmetric data like Altmetric.com, secrecy of the algorithms used to create and enrich their data products is just as much part of their business models as fee-based access is for

[31] https://scite.ai/

providers of bibliometric data. Nevertheless, there are examples for open approaches towards the collection of altmetrics as well, e.g., *Crossref Event Data* or *Lagotto* (Lin & Fenner, 2013). Right now, the open alternatives to proprietary data sources might not be as mature as the products offered by for-profit providers that have evolved over longer time spans. However, their improvement is a shared endeavor that in principle the whole scholarly community can contribute to, be it by usage, by research, by spreading the word, or through the development of software tools that enhance and facilitate the use of such open databases.

Lastly, especially for providers of altmetric data, another step towards increasing metrics' transparency in the long term lies within promoting and complying to shared codified standards for the collection and provision of metrics data.[32] While the scarcity of such standards within the field of altmetrics is still frequently criticized, the results of the *NISO Alternative Assessment Project* represent a promising start in this regard (National Information Standards Organization (NISO), 2016). The data provider Altmetric.com for instance already declared its compliance to NISO's guidelines, a detailed report listing the concrete implications of this commitment has been published on Altmetric.com's website.[33]

4.1.2 Openness and Transparency of Metrics - User-Related Aspects

Apart from the aforementioned provider-related measures that should be taken to address concerns regarding metrics transparency, there are several aspects closer related to the actual application of metrics, which may also contribute to arriving at overall more transparent processes of metrics use. These aspects involve several stakeholder groups.

The first group with respective obligations are evaluators that develop or calculate compound indicators to be used in research assessments. There are several reasons for which more sophisticated indicators involving additional steps of processing (e.g., field- or time-based normalization, see also *Susceptibility to Inherent Biases* below) are usually preferable to raw count data in research evaluation scenarios. But just like the technical processes and sources behind research indicators should be open and transparent, so should be the indicators and rules behind their generation themselves. In line with Hicks et al. (2015), to maintain comprehensibility, evaluators therefore have to balance processing meant to increase the robustness of their indicators with simplicity. An exemplary, specific recommendation in this regard comes from the final report of the **metrics* project, which recommends to generally avoid the use of indicators that aggregate count data from heterogeneous sources (such as the *Altmetric Attention Score*), as such aggregations obscure the semantic differences between their individual components (Lemke, Zagovora, et al., 2020). The calculation and use of compound indicators should follow clear protocols, which are documented in peer-reviewed literature, and which where necessary get updated and

[32] See for instance https://www.projectcounter.org/.

[33] https://staticaltmetric.s3.amazonaws.com/uploads/2016/06/Altmetric-NISO-Data-compliance-report.pdf

improved based on new insights made in scientometric research (Hicks et al., 2015; Wilsdon et al., 2015).

Another group that can help to dissolve common issues of metrics' non-transparency are institutions of higher education - by offering at least basic training regarding the interpretation and use of impact metrics for prospective researchers, regardless of their discipline. Multiple case studies - including Study A from Chapter 2 - have revealed many researchers to experience a kind of dilemma regarding the use of impact metrics (Hammarfelt & Haddow, 2018; Lemke et al., 2019). On the one hand, they in several situations feel pressured to rely on or report impact metrics, on the other hand researchers are often not well-equipped enough with knowledge on those same metrics to fully comprehend their methodologies, biases, and weaknesses. Calls for more widespread dissemination of 'metrics literacy' or 'metrics-wiseness', especially targeting doctoral students, have been expressed by several scientometric experts (Haustein, 2020; Rousseau & Rousseau, 2017). Confronting students early in their scientific careers with the basic properties and most problematic caveats of commonly used research indicators would reduce insecurities regarding their adequate interpretation and sensitize the scientific community's next generation to detect and prevent cases of indicator misuse. In addition to institutions of higher education, information infrastructures like libraries, scholarly publishers, or scholarly search engines should play a role in disseminating basic competencies on valid uses of research indicators, e.g., by making sure that metric data incorporated into their products is enriched with or links to context information that helps users understand the metrics' limitations.

Openly accessible resources that provide such context information or basic competencies regarding research metrics (and that teachers of metrics literacy could draw from) are plentiful. Easily digestible pieces are for example offered by the *Parthenos project*[34] or the *Metrics Toolkit*[35], while the principles stated in the *Leiden Manifesto*, the *San Francisco Declaration on Research Assessment*, or *The Metric Tide Report* could serve as helpful guidelines for the design of more in-depth courses on the subject (Cagan, 2013; Hicks et al., 2015; Wilsdon et al., 2015).

Finally, a particular obligation to increase impact metrics' transparency concerns the scientometric research community, of course. The better description of what both bibliometric and altmetric indicators actually measure and which outer influences shape them is a wide and multifaceted area of ongoing research within science of science (see also Section 1.3 for examples of research on this matter; Tahamtan & Bornmann, 2020). With its second part, this thesis itself addresses a particularly large research gap within that area - the unclear relationship between research indicators and external science communication - the core results of which will be summarized in Section 4.2.

[34] https://training.parthenos-project.eu/sample-page/intro-to-ri/research-impact/

[35] https://www.metrics toolkit.org

4.1.3 Susceptibility to Manipulations or Gaming

A concern voiced particularly frequently in conjunction with altmetrics is the suspicion that metrics could easily be manipulated via automated bots that are used to artificially increase scores. Case studies indicate that automated postings or bots could indeed have a serious impact on altmetric counts (Haustein, Bowman, Holmberg, et al., 2016; Robinson-Garcia et al., 2017), although, as discussed in Section 1.3, this influence will vary greatly between different types of metrics. Depending on the source platform at hand, some precautions against such distortions can be implemented by providers of altmetric data. For instance, software tools exist to identify automated accounts on social media like Twitter (see for example Chavoshi, Hamooni, & Mueen, 2016) with reasonable precision, which could be incorporated into data collection pipelines to increase the resulting data quality.

Also, it should be noted that the *altmetrics manifesto* explicitly mentions the relative ease with which journal impact factors can be gamed as one reason for the need for new indicators - altmetrics on the other hand should impede deliberate manipulations, as they help to spread measurements over a larger, more diverse set of indicators (Priem et al., 2010). This ties in with principle 8 from the *Leiden Manifesto*, which recommends to base evaluations on multiple indicators to provide more pluralistic and robust pictures of research impact (Hicks et al., 2015). Following this principle whenever feasible is a measure that users of metrics can take to reduce the risk of significant distortions caused by deliberate manipulations of indicators - and also, doing so directly addresses another comment that came up frequently in Study A's survey: namely, that metrics can appear little trustworthy on their own, while the use of a combination of different metrics can help to increase this trustworthiness.

4.1.4 Susceptibility to Inherent Biases

Another group of concerns encountered in Chapter 2's user studies relates to several biases that quantitative metrics tend to be subject to. A frequently stated example are the variances between expected citation- or mention counts for articles from different scientific disciplines, which have been documented extensively by past studies both for citations as well as for several altmetrics (see for example Gorraiz et al., 2014; Haustein, Costas, et al., 2015; Tahamtan et al., 2016). Within bibliometrics, certain best practices have emerged because of these variances - most importantly, that article metrics should best be normalized by subject field and publication age. Another method to enable more robust comparisons between articles are weightings based on percentiles - i.e., instead of comparing absolute counts, articles can be compared on the basis of the percentile to which they belong in the citation distribution of their field (e.g., the top 1%, 5%, or 10%; Hicks et al., 2015). The same principles can and should also be applied to altmetrics and usage metrics.

Other biases, for instance related to authors' genders (Larivière, Ni, Gingras, Cronin, & Sugimoto, 2013) or superficial article properties like their titles (Zagovora et al., 2018), might be more difficult to account for in metric analyses. Here a major responsibility again lies with the scientometric research community to detect, quantify, and communicate such biases to raise a solid degree of sensitivity for them among metrics users. The catalog of questions to ask before utilizing altmetrics in comparative scenarios presented by Zahedi (2018, p. 168) represents an example for how scientometric research can equip metrics users with easily accessible guidelines to support them in avoiding unintentional misuse of metric data.

4.1.5 Further Challenges for Metrics

The previous subsections summarized the most severe concerns that came to light in the user studies and publications discussed in Chapter 2, alongside recommendations of how to resolve respective difficulties. While the survey-based approach taken to identify these concerns helped to detect what the most pressing issues regarding metrics usage are from a user perspective, further challenges for metrics research have been defined in the literature. Most of these are of more technical nature and concern phenomena that might not be directly visible for occasional metrics users, which might explain their absence from the responses in Studies A & B. Nevertheless, what follows is a brief listing of these remaining challenges, along with hints about how to mitigate them.

One unsolved issue that usually affects citation-based metrics to a limited but web-based metrics to a much more substantial degree are the inconsistencies between data providers regarding metrics for the same entities (Jobmann et al., 2014), due to differences in indexing and retrieval strategies. While in some situations having access to similar metrics constructed with different retrieval strategies can be beneficial, the inconsistencies between providers' data can impede study replications (Zahedi, 2018). A definite solution to this difficulty is not in sight - what can be done to minimize replication problems, however, is for providers of data collection software to disclose their retrieval methods as fully as possible, and for users of metrics data to always clearly document the utilized data's provenience.

Another technical issue is metrics' discrimination against certain types of scholarly objects. This problem affects citation-based metrics and altmetrics in very different ways: for citation-based metrics, the limitations are usually caused by the restrictive indexing of bibliographic databases like Web of Science, which only considers a fixed range of publication venues (mostly consisting of academic journals, which puts disciplines in which journal articles are no common form of publication at a disadvantage). Altmetrics on the other hand have from the start been promoted as being more easily applicable to diverse formats of scholarly outputs. Here however limitations are caused by altmetric data collectors' reliance on persistent identifiers. Altmetric.com for instance only tracks mentions of scholarly objects that are referable by DOI, which could lead to biases against fields where these are less common, like the social sciences or humanities (Haustein, 2016). Such discrimination is another problem that most likely will not soon be overcome - still, technical enhancements of the data collection software utilized by providers as well as overall increasing use of DOIs (Wilsdon et al., 2015) might help to slowly reduce this problem over time.

A further problem much more relevant for altmetrics than for citation-based indicators is temporal instability. While the disappearance of a citation (for instance in the case of retractions) should be a relatively rare phenomenon, the disappearance of a publication's online mentions (e.g., due to social media postings being deleted) may be much more common. This further complicates the reproducibility of altmetric analyses. A mitigation of this problem might lie in providing timestamps for longitudinal statistics (Haustein, 2016).

As this section has shown, there are many open opportunities to improve the landscape of research metrics and enable indicators that address pressing concerns users have with the existing supply. These opportunities involve stakeholders from various spheres of the scholarly domain, e.g., the commercial sphere (i.e. retailers of databases and software tools that provide metrics data), the administrative sphere (i.e. research administrations that conduct evaluation exercises utilizing metrics), or the policy sphere (i.e. committees and bodies that compile the guidelines that lead institutional usage of and education about research metrics). The empirical foundation that all of these spheres should draw from is provided by the scientometric research sphere. A main challenge for this stakeholder group is to broaden and strengthen the conceptual frame in which applications of research metrics take place. Central is the question of what it is that different metrics allow to measure. While significant amounts of research efforts went into providing insights in this regard (Bornmann & Daniel, 2008; Tahamtan et al., 2016; Tahamtan & Bornmann, 2020), several research gaps remain to be filled (see also Section 1.3). One particularly understudied research gap concerns the question of the degree to which research metrics are a product of promotional efforts taking place in external science communication, which therefore was examined as this thesis' second main contribution (Chapter 3). In the upcoming Section 4.2, the pivotal observations and arguments made in relation to this contribution will be consolidated and reflected upon regarding their practical implications.

4.2 Associations between Research Metrics and External Science Communication

The two studies presented in Chapter 3, Study C and Study D, set out to examine and quantify the associations between mentions of research in two common formats of external science communication - press releases and embargo e-mails - and promoted articles' later impact metrics. These analyses should broaden our understanding of the modes of operation of these two promotional formats, as well as provide insights into whether associations between external science communication and research metrics can potentially be an issue - and if so, deliver first quantitative estimations of its magnitude. The following section will discuss and expand on the findings regarding science's representation in press releases and embargo e-mails, summarize the gathered insights on correlations between such promotional formats and research metrics, and look at the implications that these findings have for the research and use of metrics.

4.2.1 Representation of Disciplines in External Science Communication

One of the first insights from both studies concerned the kind of research that gets representation within embargo e-mails and/or press releases, i.e., the research publications that publishers' or institutions' press officers select to receive these types of promotion. It became apparent that prominent multidisciplinary high-impact journals (e.g., Science, Nature, PNAS, PLOS ONE) and journals primarily covering biomedical and life sciences (e.g., Cell Reports, New England Journal of Medicine, The Lancet) account for an overwhelming share of the publications that receive visibility in these channels. The predominance of medicine/health and biology is in line with the topics that tend to dominate science coverage in media worldwide, followed by science news related to environment, technology or behavioral sciences (Badenschier & Wormer, 2012; Bauer, 2000; Bucchi & Mazzolini, 2003; van Rooyen, 2002). To facilitate the derivation of such biases' implications, it is worthwhile to at least briefly think about the circumstances and mechanisms that might cause them.

The presence of individual journals on press release platforms like EurekAlert! or IDW-Online will to some degree be determined by the journals' individual extents of PR efforts - it can for instance be assumed that aforementioned prominent high-impact journals will usually have more resources available to promote their publications than comparatively little-known journals with much smaller readerships. One could therefore argue that discipline-related biases among science press releases might to an extent just be a consequence of which disciplines the journals with the largest PR budgets and best professional networks happen to cover, without any deliberate decisions being involved that lead to certain disciplines receiving more press release promotion. However, preliminary case studies suggest that the majority of press releases on EurekAlert! and IDW-Online do not come from scholarly publishers or journals, but directly from universities and other research institutions (Hahn & Lemke, 2020; see also Bowman & Hassan, 2019). As these institutions in most cases produce research from a range of fields, it seems likely that the discipline-wise biases among science press releases, for instance towards biomedical subjects, are not just the result of economical preconditions among journals, but also of PR departments' conscious choices.

Furthermore, the strong topical foci within the supply of science press releases and embargo e-mails will likely influence considerably which science topics have a chance to appear in journalistic media. After all, science journalists may in many cases have little incentive (or time) to look beyond PR officers' offerings to grant topics from without of those foci coverage in their journalistic output (Vrieze, 2018). Adopting this hypothesis, the primary role of the science journalist would be to extract the most relevant information from the incoming stream of science news releases and embargo e-mails, in accordance with the journalistic theory of gatekeeping (or *gatewatching*; Bruns, 2005). What enters the stream, however, would mostly be determined by the creators of press releases and embargo e-mails, who would therefore wield considerable power over how and which scientific results are portrayed in journalism.

However, not only the *supply* of press releases, but also media consumer's *demand* for certain topics - taken up by journalists, who pass it on to science PR - might be a decisive factor regarding which research gets visibility in press releases and embargo e-mails. The fact that certain (for instance health-related) topics tend to "sell well" in mainstream media presumably also influences which kinds of publications publishers and institutions highlight with particularly high frequency in their press releases and embargo e-mails - after all, it is in their own interest to provide journalists with findings that the latter will likely pick up in their stories.

Of course, the supply-based and the demand-based approach to an explanation as to whose influence determines which disciplines receive representation in external science communication do not contradict each other and the respective mechanisms likely are at work simultaneously. And there is at least one further factor with potential influence on a discipline's presence in external science communication: that discipline's size, i.e., the overall amount of publications it produces. To arrive at a clearer picture of the degrees to which science PR's supply, media consumers' demand, or such further external factors shape the representation of science in the media, future research should examine in greater detail how research institutions and publishers select articles for their promotional activities (see also Kiernan, 2014). A better knowledge of the central factors that these selection processes are typically based upon would help to identify more precisely which specific domains and kinds of research may be most heavily affected by external promotion's influence on metrics, i.e., reveal where the critical reassessment of metrics use in light of their susceptibility to such distortions is most crucial.

4.2.2 Correlations of Metrics and External Promotion

The findings from both studies presented in Chapter 3 support substantial associations between the presence of research within press releases or embargo e-mails and their impact metrics. Most evident became the association between the two observed types of promotion and the articles' later mentions within mainstream media. But also metrics with a less obvious connection to the journalistic domain, which is targeted directly by press releases and embargo e-mails, exhibited considerable correlations with article mentions in those formats. In Study C, blog mentions stood out as another altmetric heavily associated with embargo e-mail mentions, followed by mentions on Twitter and Facebook, readerships on Mendeley, and Web of Science citations. Study D confirmed the effects of press releases and embargo e-mails on mainstream media mentions, Twitter mentions and citations, while also the significant role of journal prestige for the expected later mentions and citations of an article became apparent.

Two essential hypotheses to explain the correlation between a research publication's promotion in external science communication and its later metrics exist:

- the publicity hypothesis (i.e., the increased visibility directly or indirectly caused by the appearance of a scientific publication in press releases, embargo e-mails, and journalistic media increases its likelihood of being cited or interacted with online) and
- the earmark hypothesis (i.e., inherent characteristics of certain research publications simultaneously increase their likelihood of being chosen for coverage in external science communication and for being cited or interacted with online).

The combined findings of Studies C and D as well as of the related literature reviewed within them suggest that both hypothesized effects might apply concurrently. Particularly Phillips et al.'s (1991) case study on articles selected for coverage within the New York Times, and their later citation advantage depending on whether respective New York Times issues were actually published, provides convincing evidence for the significance of the publicity hypothesis with regard to newspaper mentions and citations. Another argument for the publicity hypothesis in this regard is provided by Fanelli (2013), who found features of research publications in British newspapers to be associated with substantially larger citation advantages than features in Italian newspapers - Fanelli (2013) argues that because the potential scientific audience for English news reporting will be much larger than the one for Italian news, these results indicate the greater importance of publicity effects compared to earmark effects. Finally, in their own study on newspaper coverage's positive association with mentioned research publications' later citations, Dumas-Mallet et al. (2020) additionally found a positive correlation between the number of different news articles reporting on a single publication and the magnitude of that publication's citation advantage, providing a further argument for the significance of publicity effects.

Given the findings of case studies that indicate a considerable overlap between the agents that are responsible for publications' academic citations (i.e., researchers) and those that are responsible for a considerable share of their mentions in online fora (also researchers; Birkholz et al., 2015; Sugimoto et al., 2017; Tsou et al., 2015), it seems reasonable to assume that the publicity effect that Phillips et al. (1991), Fanelli (2013), and Dumas-Mallet et al. (2020) revealed for academic citations also applies to many altmetrics in a similar way. After all, if researchers become inclined to cite certain publications due to being exposed to them within mainstream media, it seems natural that those researchers would also become more likely to mention the respective publications online, download them, bookmark them, and so on.

But there are strong arguments to be made for the significance of the earmark hypothesis as well. From the perspective of researchers, the plausibility of the earmark hypothesis can be explained with Weingart's (1998) theory of the medialization of science. One facet of said medialization concerns the observation that researchers more and more adopt modes of presentation and topic selection from mass media - following this hypothesis, it stands to reason that over time the kind of research that is preferably conducted by and receives attention within the scholarly community becomes more and more congruent with the kind of research journalists preferably report about (Dumas-Mallet et al., 2020). Another basis to argue for the plausibility of the earmark hypothesis is provided by the overlap between known news factors for science journalism (Badenschier & Wormer, 2012) and factors known to influence citations (Tahamtan et al., 2016). To give just two examples for such overlap, Badenschier & Wormer (2012) mention *[authors'] influence/scientific power* and *reputation of a scientific journal* as relevant news factors in science journalism, while the effects of author reputation and journal reputation have also been shown to be factors with significant effect on articles' expected citations (Jiang, He, & Ni, 2013). Such evidence suggests that journalistic selection processes as well as researchers' selection processes indeed follow shared criteria without necessarily influencing each other directly, just as the earmark hypothesis proposes.

Following this reasoning it seems likely that both publicity and earmark hypothesis apply, the degrees of which may vary depending on different scientific disciplines, different channels of journalistic coverage, different types of external promotion, etc.. To more precisely determine the degrees to which both hypotheses apply, as well as the effects such contextual factors have in this, thorough qualitative studies on the processes and criteria that cause researchers to cite, PR officers to promote, and journalists to cover certain publications before others could lead to valuable new insights (see also Badenschier & Wormer, 2012; Bartlett, Sterne, & Egger, 2002; Broer, 2020; Stryker, 2002; Sumner et al., 2016, 2014).

4.2.3 Implications for Applications of Research Metrics

As we have seen, the system of science communication's complexity does make it extraordinarily difficult to establish credible claims about concrete causalities behind the associations between external science communication and research metrics that were analyzed in Chapter 3. Following the argumentation in subsections 4.2.1 and 4.2.2, however, it stands to reason that the actions of science PR exert at least some influence over impact metrics' expected developments. Moreover, what could be shown in Chapter 3 is the substantial magnitude of positive correlations between research articles' mentions in press material and their later metrics, virtually across the whole range of examined indicators. These findings reveal that the *potential* extent of external science communication's interference regarding manifestations of impact metrics is large and does not only apply to citations, but also to altmetrics - and there in many cases to an even larger degree. To resume just some exemplary findings from Study C's results regarding the association between embargo e-mail promotion and metrics, we had for instance found articles with respective promotion to on average receive about twice as many citations, about five times as many Twitter mentions, and almost ten times as many mentions in mainstream media compared to articles without known promotion in embargo e-mails. Results as these underline (1) the importance of further, close investigations of these relationships between impact metrics and external science communication, as well as (2) the necessity for the scholarly community to critically reflect the implications of such potential influence of marketing efforts for the use of indicators in research assessment.

Altmetrics, as has been expounded in Chapter 1 of this thesis, are still lacking a robust fundament of theory and the issue of what individual altmetrics measure is still subject of ongoing, vibrant discussions. At the core of these discussions have often been questions as to which target groups' actions are reflected within altmetrics, e.g., whether altmetrics are suitable to reflect impact among practitioners from outside of academia (Mohammadi, Thelwall, Haustein, & Larivière, 2015) or among the wider society in general (Tahamtan & Bornmann, 2020; see also Section 1.3). This thesis' findings highlight that also endeavors of quantifying the degree to which individual altmetrics are shaped by various kinds of external promotion are a necessary requirement to arrive at a solid understanding of altmetrics' meanings, and the results of studies C and D provide first and much needed empirical input on this issue.

For citation-based indicators, which have been widely used as indicators for academic impact for decades now, the implications of considerable associations between external promotion and articles' citation counts might be more severe. The idea of citations as indicators for a scholarly work's academic impact is fundamentally based on Robert Merton's normative-meritocratic theory of citing, which regards citations as "atoms of peer recognition" (Small, 2004, p. 72). This theory postulates that the act of citing resembles the payment of an intellectual debt, i.e., the citation reflects that the cited work was a relevant influence during the creation of the citing work. The validity of the normative-meritocratic theory of citing has been questioned by many scholars for several different reasons before (see MacRoberts & MacRoberts [2018] for a recent overview); citations to a considerable extent being the consequence of research's promotion to and within science journalism would constitute another such reason that did not receive much attention from the citations-using scholarly community so far. The finding that external science communication can cause a substantial amount of citations, more or less regardless of cited publications' actual worth for the citing work, would diminish their usefulness as indicators for relevance significantly. After all, PR departments and science journalism have no obligation to be transparent or apply other criteria of good scholarly practice in their decisions about which publications to feature in press releases, embargo e-mails, or journalistic works. Furthermore, it will often be the researchers themselves who exert considerable influence over whether (and which of) their publications get the chance to be featured in press releases, either as editors of the press release-issuing scholarly journals, or by making respective suggestions to their institutions' PR departments. The fact that researchers might have such considerable influence over which publications get press-released makes a following publicity effect on the metrics that are used to evaluate the same researchers' productivity even more problematic.

It should be noted once more that the severity of this problem depends on the degree to which the publicity hypothesis holds true, precise quantifications of which have yet to be made (Phillips et al., 1991). If on the other hand the citation advantages of press released articles could entirely be attributed to earmark effects, the basic assumptions of evaluative citation analysis would not be infringed. Considering the aforementioned findings by Phillips et al. (1991), Fanelli (2013), and Dumas-Mallet et al. (2020), however, a rejection of the publicity hypothesis does not seem justified - thus, increased caution regarding potential publicity effects on research metrics seems advisable for researchers and other users of the latter.

To establish preconditions that allow users to express such caution practically in their evaluative uses of research metrics will be a complex task, but it also constitutes a highly promising avenue for future research at the intersection of scientometrics and applied computer science. The best practices for citation use stated in prominent guidelines such as the Leiden Manifesto (Hicks et al., 2015) regularly highlight the importance of normalizing citation-based indicators by scientific field and time of publication, to diminish the advantages certain works might have over others due to differing publication- and citation practices between fields, or due to longer time windows in which citations could accumulate. Developing a complementary normalization for research publications' received external promotion, as a tool to estimate how individual publications would have performed if their received promotion would have been comparable, might be valuable to enable new, more nuanced metric-based comparisons. Altmetric data aggregators like Altmetric.com, which already also trace mentions of research publications on press release platforms like EurekAlert!, can be seen as a first approach towards a collection of the data needed for such a normalization. Still, many further conceptual and technological questions remain to be solved by the scientometric research community to possibly arrive at a robust methodology for a publicity-based normalization of impact indicators - e.g., regarding which sources (such as EurekAlert!) to track as relevant platforms of external promotion, how to account for the substantially varying reach of different sources, how to model aspects like supposed temporal declines of individual promotional activities' effectiveness, and many more.

Also, in the interest of openness and transparency of research evaluation it would be beneficial to increase the efforts going into the development of non-proprietary databases of research mentions in external science communication, analogous to how initiatives like I4OC or OpenAlex strive for better open infrastructures of citation data. The most promising starting point in this regard might be Crossref Event Data[36], although its range of tracked data sources for now appears far more limited than that tracked for instance by Altmetric.com. It thus remains an ongoing challenge for the scholarly community to further develop such openly accessible databases of research mentions in external science communication.

[36] https://www.crossref.org/services/event-data/

4.3 Conclusions

The elaborations within this thesis have cast a light upon several challenges that metric research in general and altmetrics research in particular are confronted with. The use of metrics as indicators for scientific relevance is (and has always been) subject of vibrant controversies (MacRoberts & MacRoberts, 2018), some of which were reflected within the user studies presented in Chapter 2. Chapter 3 on the other hand provided a look into some of the intricate interdependencies within the multi-faceted communication and impact system of science, which might let the task of acquiring robust concepts of what individual impact metrics measure appear daunting. However, developing a common theoretical foundation for the interpretation of metrics is a prerequisite to decrease their ambiguities, turn vague assumptions about their capabilities and limitations into informed assessments, and ultimately make indicators as transparent and informative as possible. The studies presented within this thesis contribute a range of empirical insights to inform the advancement of metrics theory. Moreover, over the past chapters several opportunities to mitigate some of the problems researchers commonly associate with the use of metrics for research assessment have been revealed.

The perhaps most overt takeaway from this thesis' user studies is that a significant amount of researchers' concerns frequently associated with impact metrics could be resolved by providing more far-reaching offerings of education about and information on research metrics, especially for early-career researchers. Institutions of higher education, libraries, and providers of metrics data can all play a part in this, either by making research metrics a subject of existing PhD programs, or by offering additional specific seminars on the topic. A broader knowledge about metrics among the next generation of scholars should reduce irresponsible indicator use in multiple relevant areas, as in many cases these scholars will also outside of their roles as researchers fill positions which vitally shape what metrics' future usage and significance for both individual and institutional use cases will look like - e.g., in roles as supervisors, journal editors, lecturers, administrators, policy advisors, or in appointment committees. Furthermore, a departure from metrics-based methods in assessments of researchers' scientific productivity is not to be expected any time soon. Therefore, equipping researchers with basic competencies on the indicators that will later be used for their very own evaluation, to enable them to scrutinize and if necessary dispute those methods, is also a matter of fairness (see also Hicks et al., 2015; Rousseau & Rousseau, 2017).

Another recurring takeaway from the studies presented in this thesis concerns the importance of open data and infrastructures in contexts of research metrics. Such openness refers to the services and methods that are used to obtain metrics, as well as to the resulting metric data itself. Not only assists this openness in achieving higher levels of transparency and reproducibility of research indicators, it also facilitates the development of new services and tools on top of citation or altmetric data, and opens up new opportunities for scientometric research.[37] More particular arguments for the value of open *altmetric* data were encountered in the discussion within subsection 4.2.3, as achieving greater transparency about and traceability of activities undertaken to promote research (which themselves can be considered a specific kind of altmetric events) would also improve the possibilities to more precisely describe the relationships between such promotion and research metrics in general (e.g., the significance of publicity effects regarding metrics' manifestations). Promising starting points for establishing open infrastructures of citation and altmetric data exist (e.g., I4OC, OpenAlex, Crossref Event Data, Lagotto), although in several aspects these alternatives are not as mature and established as their proprietary counterparts are yet. However, the change towards open infrastructures and databases being the default rather than the exception in research metrics is an endeavor for large parts of the scientific community to contribute to, e.g., by actively using and promoting open metric data. The findings can also be understood as a call to action to the software developing part of the scholarly community, which should increase efforts regarding the development of services and tools that facilitate the utilization of open metric data.

To conclude, the main profiteers of this thesis' results can be summarized as follows:

(1) developers of software for the gathering, processing, and provision of metrics, as well as those of systems harnessing metrics for secondary purposes, like for instance recommender systems or search engines;
(2) suppliers of research works such as libraries, scholarly publishers, or research repositories, who aim to provide their users with informative indicators;
(3) research administrations, which rely on metrics in evaluative contexts;
(4) scientometric researchers, who seek a more substantiated understanding of research metrics and aim to progress the development of their theoretical foundation.

All of these groups are addressed by the findings on needs with regard to improved research metrics and the respective actions to be taken that were consolidated in Section 4.1. This thesis' results and discussions on the relationship between research metrics and external science communication on the other hand (condensed in Section 4.2) should especially inform the scientometric research community in its endeavors of further comprehending and specifying the semantic richness of different indicators.

With regard to future research, the complexity witnessed during the analyses of the interplay between certain formats of external science communication and various research metrics demonstrates the limitations of purely quantitative approaches towards understanding the communication and impact system of science. While the mostly quantitative approach followed by the studies presented in Chapter 3 provided detailed empirical data on directions and magnitudes of potential effects between different kinds of science promotion and research metrics, more qualitative examinations of involved agents'

[37] See also https://i4oc.org/#faqs.

(i.e., science journalists, PR officers, citing researchers) interests and opportunities to exert influence within the system of science communication should in future be undertaken to complement this thesis' findings. Especially the discussions concerning possible causalities behind observed associations (subsection 4.2.2) have underlined that purely quantitative views will not be sufficient to describe the intricate relationships between (external) science communication and impact metrics precisely.
Altmetrics harbor considerable potential to make pathways of the attention, usage, and impact a scientific work receives or causes observable, und thus to provide new perspectives on the interactions within the systems of science communication. In this regard, altmetrics traverse traditional distinctions between external communication of science and scientific impact measurement. Descriptive examples for this are mentions in social or mainstream media, which since the early days of altmetrics as a concept have widely been assessed as new, complementing means for impact measurement, while they at the same time could also be considered indicators of a work's received promotion. To achieve more profound clarity to which degree which types of altmetrics can validly indicate which of these properties, further future research should in particular investigate (1) to what extent the same or different agents typically are responsible for the mentions of a work on different platforms (i.e. metrics sources), and (2) which pathways the attention individual works receive typically takes across platforms, i.e., do occurrences of one type of metric event typically spark occurrences of specific other types of metric events?

The extent and density of unsolved issues and questions related to the validity of metrics as research indicators encountered in this thesis may feel sobering and call their usefulness for research evaluation purposes into question (Zahedi, 2018). And indeed does also the in this thesis particularly closely examined potential sensitivity of metrics regarding marketing efforts highlight the persistent necessity for the scientometric community to continue to deepen the understanding of metrics to achieve more solid prerequisites for their application in evaluative scenarios.
Still, it would be shortsighted to conclude from the available quantitative research metrics' flaws that the scholarly community would be better off by abandoning them completely. On the contrary - never before have such versatile means been available to draw diverse and differentiated pictures of where and how research findings have been used and possibly made an impact. The broad range of bibliometrics, usage metrics, and altmetrics can potentially enable astoundingly detailed insights on researches' utilization, and not further developing the tools to leverage the diversity of existing metrics would equal a tremendous omission of opportunities. What will certainly remain imperative in the use of metrics, however, is care in choosing the appropriate set of indicators depending on the research question or evaluation context at hand, prudence and a certain humbleness in their interpretation, as well as a consciousness of the fundamental limitations of quantitative measurements of research impact.

Bibliography

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

Adie, E. (2016). The rise of altmetrics. In A. Tattersall (Ed.), *Altmetrics: A Practical Guide for Librarians, Researchers and Academics* (pp. 67–82). Facet. https://doi.org/10.29085/9781783301515.005

Aizaki, H., & Nishimura, K. (2008). Design and Analysis of Choice Experiments Using R: A Brief Introduction. *Agricultural Information Research*, *17*(2), 86–94. https://doi.org/10.3173/air.17.86

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *SAGE Open*, *9*(1), 2158244019829575. https://doi.org/10.1177/2158244019829575

Aksnes, D. W., & Rip, A. (2009). Researchers' perceptions of citations. *Research Policy*, *38*(6), 895–905. https://doi.org/10.1016/j.respol.2009.02.001

Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the world wide web: Methodological approaches to 'webometrics.' *Journal of Documentation*, *53*(4), 404–426. https://doi.org/10.1108/EUM0000000007205

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

Aung, H. H., Erdt, M., & Theng, Y.-L. (2017). Awareness and usage of altmetrics: A user survey. *Proceedings of the Association for Information Science and Technology*, *54*(1), 18–26. https://doi.org/10.1002/pra2.2017.14505401003

Aung, H. H., Zheng, H., Erdt, M., Aw, A. S., Sin, S.-C. J., & Theng, Y.-L. (2019). Investigating familiarity and usage of traditional metrics and altmetrics. *Journal of the Association for Information Science and Technology*, *70*(8), 872–887. https://doi.org/10.1002/asi.24162

Badenschier, F., & Wormer, H. (2012). Issue Selection in Science Journalism: Towards a Special

Theory of News Values for Science News? In Simone Rödder, M. Franzen, & P. Weingart (Eds.), *The Sciences' Media Connection –Public Communication and its Repercussions* (pp. 59–85). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-2085-5_4

Bakker, C., Bull, J., Courtney, N., DeSanto, D., Langham-Putrow, A., McBurney, J., & Nichols, A. (2019). How Faculty Demonstrate Impact: A Multi-Institutional Study of Faculty Understandings, Perceptions, and Strategies Regarding Impact Metrics. *Association of College and Research Libraries (ACRL) Conference*. Retrieved from https://scholar.valpo.edu/ccls_fac_presentations/20

Bakker, C., Cooper, K., Langham-Putrow, A., & McBurney, J. (2020). Qualitative Analysis of Faculty Opinions on and Perceptions of Research Impact Metrics. *College & Research Libraries*, *81*(6), 896. https://doi.org/10.5860/crl.81.6.896

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond citations: Scholars' visibility on the social Web. *ArXiv:1205.5611 [Physics]*. Retrieved from http://arxiv.org/abs/1205.5611

Bartlett, C., Sterne, J., & Egger, M. (2002). What is newsworthy? Longitudinal study of the reporting of medical research in two British newspapers. *BMJ*, *325*(7355), 81–84. https://doi.org/10.1136/bmj.325.7355.81

Bauer, M. (2000). "Science in the Media" as cultural indicator: contextualizing surveys with media analysis. In *Between Understanding and Trust* (1st ed.). Routledge.

Bellis, N. D., Wouters, P., Day, R. E., Furner, J., Gingras, Y., McCain, K. W., … Rosen, R. (2014). *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (1st edition; B. Cronin & C. R. Sugimoto, Eds.). Cambridge, Massachusetts: The MIT Press.

Berghaeuser, H., & Hoelscher, M. (2020). Reinventing the third mission of higher education in Germany: Political frameworks and universities' reactions. *Tertiary Education and Management*, *26*(1), 57–76. https://doi.org/10.1007/s11233-019-09030-3

Birkholz, J. M., Seeber, M., & Holmberg, K. (2015). Drivers of Higher Education Institutions' Visibility: A Study of UK HEIs Social Media Use vs. Organizational Characteristics. *Proceedings of the 2015 International Society for Scientometrics and Informetrics*, 502–513.

Istanbul, Turkey. Retrieved from https://www.issi-society.org/proceedings/issi_2015/0502.pdf

Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, *36*(1), 2–72. https://doi.org/10.1002/aris.1440360102

Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics*, *8*(4), 895–903. https://doi.org/10.1016/j.joi.2014.09.005

Bornmann, L. (2016). Scientific Revolution in Scientometrics: The Broadening of Impact from Citation to Societal. In *Theories of Informetrics and Scholarly Communication* (pp. 347–359). De Gruyter Saur. Retrieved from https://www.degruyter.com/document/doi/10.1515/9783110308464-020/html

Bornmann, L., & Daniel, H.-D. (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., & Haunschild, R. (2016a). How to normalize Twitter counts? A first attempt based on journals in the Twitter Index. *Scientometrics*, *107*(3), 1405–1422. https://doi.org/10.1007/s11192-016-1893-6

Bornmann, L., & Haunschild, R. (2016b). To what extent does the Leiden manifesto also apply to altmetrics? A discussion of the manifesto against the background of research into altmetrics. *Online Information Review*, *40*(4), 529–543. https://doi.org/10.1108/OIR-09-2015-0314

Bornmann, L., & Haunschild, R. (2017). Measuring field-normalized impact of papers on specific societal groups: An altmetrics study based on Mendeley Data. *Research Evaluation*, *26*(3), 230–241. https://doi.org/10.1093/reseval/rvx005

Bornmann, L., & Haunschild, R. (2018). Normalization of zero-inflated data: An empirical analysis of a new indicator family and its use with altmetrics data. *Journal of Informetrics*, *12*(3), 998–1011. https://doi.org/10.1016/j.joi.2018.01.010

Bornmann, L., Haunschild, R., & Adams, J. (2019). Do altmetrics assess societal impact in a comparable way to case studies? An empirical test of the convergent validity of altmetrics based on data from the UK research excellence framework (REF). *Journal of Informetrics*, *13*(1), 325–340. https://doi.org/10.1016/j.joi.2019.01.008

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*(11), 2215–2222. https://doi.org/10.1002/asi.23329

Bowman, T. D., & Hassan, S.-U. (2019). *Science News and Altmetrics: Looking at EurekAlert!*. Presented at the altmetrics19 Workshop. Stirling, Scotland. Retrieved from http://altmetrics.org/wp-content/uploads/2019/10/Bowman_altmetrics19_paper_6.pdf

Brainard, J. (2022). Riding the twitter wave. *Science*, *375*(6587), 1344–1347. https://doi.org/10.1126/science.abq1541

Broer, I. (2020). Rapid reaction: Ethnographic insights into the Science Media Center and its response to the COVID-19 outbreak. *Journal of Science Communication*, *19*(5), A08. https://doi.org/10.22323/2.19050208

Bruns, A. (2005). *Gatewatching: Collaborative Online News Production*. Peter Lang.

Bucchi, M., & Mazzolini, R. G. (2003). Big science, little news: Science coverage in the Italian daily press, 1946-1997. *Public Understanding of Science*, *12*(1), 7–24. https://doi.org/10.1177/0963662503012001413

Burnard, P. (1991). A method of analysing interview transcripts in qualitative research. *Nurse Education Today*, *11*(6), 461–466. https://doi.org/10.1016/0260-6917(91)90009-Y

Butler, L. (2005). What Happens When Funding is Linked to Publication Counts? In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems* (pp. 389–405). Dordrecht: Springer Netherlands. https://doi.org/10.1007/1-4020-2755-9_18

Butler, L. (2007). Assessing university research: A plea for a balanced approach. *Science and Public Policy*, *34*(8), 565–574. https://doi.org/10.3152/030234207X254404

Cagan, R. (2013). The San Francisco Declaration on Research Assessment. *Disease Models & Mechanisms*, *6*(4), 869–870. https://doi.org/10.1242/dmm.012955

Carver, R. B. (2014). Public communication from research institutes: Is it science communication or public relations? *Journal of Science Communication*, *13*(3), C01. https://doi.org/10.22323/2.13030301

Chapman, R. G., & Staelin, R. (1982). Exploiting Rank Ordered Choice Set Data within the Stochastic Utility Model. *Journal of Marketing Research*, *19*(3), 288–301. https://doi.org/10.2307/3151563

Chapman, S., Nguyen, T. N., & White, C. (2007). Press-released papers are more downloaded and cited. *Tobacco Control*, *16*(1), 71–71. https://doi.org/10.1136/tc.2006.019034

Chavoshi, N., Hamooni, H., & Mueen, A. (2016). DeBot: Twitter Bot Detection via Warped Correlation. *2016 IEEE 16th International Conference on Data Mining (ICDM)*. https://doi.org/10.1109/ICDM.2016.0096

Collins, K., Shiffman, D., & Rock, J. (2016). How Are Scientists Using Social Media in the Workplace? *PLOS ONE*, *11*(10), e0162680. https://doi.org/10.1371/journal.pone.0162680

Cress, P. E. (2014). Using Altmetrics and Social Media to Supplement Impact Factor: Maximizing Your Article's Academic and Societal Impact. *Aesthetic Surgery Journal*, *34*(7), 1123–1126. https://doi.org/10.1177/1090820X14542973

Cronin, B., Snyder, H. W., Rosenbaum, H., Martinson, A., & Callahan, E. (1998). Invoked on the Web. *Journal of the American Society for Information Science*, *49*(14), 1319–1328. https://doi.org/10.1002/(SICI)1097-4571(1998)49:14<1319::AID-ASI9>3.0.CO;2-W

Cronin, B., & Sugimoto, C. R. (2015). *Scholarly metrics under the microscope: From citation analysis to academic auditing*. Medford, NJ: Information Today. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.23676

de Semir, V., Ribas, C., & Revuelta, G. (1998). Press Releases of Science Journal Articles and Subsequent Newspaper Stories on the Same Topic. *JAMA*, *280*(3), 294. https://doi.org/10.1001/jama.280.3.294

DeSanto, D., & Nichols, A. (2017). Scholarly Metrics Baseline: A Survey of Faculty Knowledge, Use, and Opinion about Scholarly Metrics. *College & Research Libraries*, *78*(2), 150. https://doi.org/10.5860/crl.78.2.150

Desrochers, N., Paul-Hus, A., Haustein, S., Costas, R., Mongeon, P., Quan-Haase, A., Bowman, T. D., Pecoskie, J., Tsou, A., & Larivière, V. (2018). Authorship, citations, acknowledgments and visibility in social media: Symbolic capital in the multifaceted reward system of science. *Social Science Information*, *57*(2), 223–248. https://doi.org/10.1177/0539018417752089

Dumas-Mallet, E., Garenne, A., Boraud, T., & Gonon, F. (2020). Does newspapers coverage influence the citations count of scientific publications? An analysis of biomedical studies. *Scientometrics*, *123*(1), 413–427. https://doi.org/10.1007/s11192-020-03380-1

Elmer, C., Badenschier, F., & Wormer, H. (2008). Science for Everybody? How the Coverage of Research Issues in German Newspapers Has Increased Dramatically. *Journalism & Mass Communication Quarterly*, *85*(4), 878–893. https://doi.org/10.1177/107769900808500410

Entwistle, V. (1995). Reporting research in medical journals and newspapers. *BMJ : British Medical Journal*, *310*(6984), 920. https://doi.org/10.1136/bmj.310.6984.920

Fanelli, D. (2013). Any publicity is better than none: Newspaper coverage increases citations, in the UK more than in Italy. *Scientometrics*, *95*(3), 1167–1177. https://doi.org/10.1007/s11192-012-0925-0

Fleerackers, A., Nehring, L., Maggio, L. A., Enkhbayar, A., Moorhead, L., & Alperin, J. P. (2022). *Identifying science in the news: An assessment of the precision and recall of Altmetric.com news mention data*. Zenodo. https://doi.org/10.5281/zenodo.6366635

Franzen, M. (2012). Making Science News: The Press Relations of Scientific Journals and Implications for Scholarly Communication. In Simone Rödder, M. Franzen, & P. Weingart (Eds.), *The Sciences' Media Connection –Public Communication and its Repercussions* (pp. 333–352). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-2085-5_17

Fraser, N., Momeni, F., Mayr, P., & Peters, I. (2020). The relationship between bioRxiv preprints, citations and altmetrics. *Quantitative Science Studies*, *1*(2), 618–638. https://doi.org/10.1162/qss_a_00043

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology. General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338

Garfield, E., & Sher, I. H. (1963). New factors in the evaluation of scientific literature through citation indexing. *American Documentation*, *14*(3), 195–201. https://doi.org/10.1002/asi.5090140304

Garfield, E. (1972). Citation Analysis as a Tool in Journal Evaluation. *Science*, *178*(4060), 471–479. https://doi.org/10.1126/science.178.4060.471

Glänzel, W. (1996). A bibliometric approach to social sciences. National research performances in 6 selected social science areas, 1990-1992. *Scientometrics*, *35*(3), 291–307.

Glänzel, W., & Gorraiz, J. (2015). Usage metrics versus altmetrics: Confusing terminology? *Scientometrics*, *102*(3), 2161–2164. https://doi.org/10.1007/s11192-014-1472-7

Glänzel, W., & Schubert, A. (1988). Characteristic scores and scales in assessing citation impact. *Journal of Information Science*, *14*(2), 123–127. https://doi.org/10.1177/016555158801400208

Gorraiz, J., Gumpenberger, C., & Schlögl, C. (2014). Usage versus citation behaviours in four subject areas. *Scientometrics*, *101*(2), 1077–1095. https://doi.org/10.1007/s11192-014-1271-1

Green, P. E., & Srinivasan, V. (1978). Conjoint Analysis in Consumer Research: Issues and Outlook. *Journal of Consumer Research*, *5*(2), 103. https://doi.org/10.1086/208721

Green, P. E., & Wind, Y. (1975). New Way to Measure Consumers' Judgments. *Harvard Business Review*. Retrieved from https://hbr.org/product/new-way-to-measure-consumers-judgments/75404-PDF-ENG

Gross, P. L. K., & Gross, E. M. (1927). College Libraries and Chemical Education. *Science*, *66*(1713), 385–389. https://doi.org/10.1126/science.66.1713.385

Haddow, G., & Hammarfelt, B. (2019). Quality, impact, and quantification: Indicators and metrics use by social scientists. *Journal of the Association for Information Science and Technology*, *70*(1), 16–26. https://doi.org/10.1002/asi.24097

Hahn, O., & Lemke, S. (2020). *An Exploration of Scientific Press Releases in the Context of Altmetrics*. Presented at the altmetrics20 Workshop. https://doi.org/10.5281/zenodo.4446908

Hammarfelt, B., & Haddow, G. (2018). Conflicting measures and values: How humanities scholars in Australia and Sweden use and react to bibliometric indicators. *Journal of the Association for Information Science and Technology*, *69*(7), 924–935. https://doi.org/10.1002/asi.24043

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

Haustein, S. (2020). *Metrics Literacies research project mapped to Knowledge two Action (K2A) framework*. Zenodo. https://doi.org/10.5281/zenodo.4029695

Haustein, S., Bowman, T. D., & Costas, R. (2016). Interpreting 'Altmetrics': Viewing Acts on Social Media through the Lens of Citation and Social Theories. In Cassidy R. Sugimoto (Ed.), *Theories of Informetrics and Scholarly Communication*. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110308464-022

Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2016). Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter. *Journal of the Association for Information Science and Technology*, *67*(1), 232–238. https://doi.org/10.1002/asi.23456

Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing Social Media Metrics of Scholarly Papers: The Effect of Document Properties and Collaboration Patterns. *PLOS ONE*, *10*(3), e0120495. https://doi.org/10.1371/journal.pone.0120495

Haustein, S., Larivière, V., Thelwall, M., Amyot, D., & Peters, I. (2014). Tweets vs. Mendeley readers: How do these two social media metrics differ? *It - Information Technology*, *56*(5). https://doi.org/10.1515/itit-2014-1048

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

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

Haustein, S., Sugimoto, C., & Larivière, V. (2015). Guest editorial: Social media in scholarly communication. *Aslib Journal of Information Management*, *67*(3). https://doi.org/10.1108/AJIM-03-2015-0047

Hermida, A. (2012). Social Journalism: Exploring how Social Media is Shaping Journalism. In *The*

Handbook of Global Online Journalism (pp. 309–328). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118313978.ch17

Hicks, D. (2005). The Four Literatures of Social Science. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (pp. 473–496). Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-2755-9_22

Hicks, D., Stahmer, C., & Smith, M. (2018). Impacting Capabilities: A Conceptual Framework for the Social Value of Research. *Frontiers in Research Metrics and Analytics*, *3*. Retrieved from https://www.frontiersin.org/article/10.3389/frma.2018.00024

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature News*, *520*(7548), 429. 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 of the United States of America*, *102*(46), 16569–16572. https://doi.org/10.1073/pnas.0507655102

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Illingworth, S., & Allen, G. (2020). *Effective Science Communication (Second Edition): A practical guide to surviving as a scientist*. IOP Publishing. https://doi.org/10.1088/978-0-7503-2520-2

Jiang, J., He, D., & Ni, C. (2013). The correlations between article citation and references' impact measures: What can we learn? *Proceedings of the American Society for Information Science and Technology*, *50*(1), 1–4. https://doi.org/10.1002/meet.14505001162

Jin, T., Duan, H., Lu, X., Ni, J., & Guo, K. (2021). Do research articles with more readable abstracts receive higher online attention? Evidence from Science. *Scientometrics*, *126*(10), 8471–8490. https://doi.org/10.1007/s11192-021-04112-9

Jobmann, A., Hoffmann, C. P., Künne, S., Peters, I., Schmitz, J., & Wollnik-Korn, G. (2014). Altmetrics for large, multidisciplinary research groups: Comparison of current tools. *Bibliometrie - Praxis Und Forschung*, *3*(1), 1–19.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Khazragui, H., & Hudson, J. (2015). Measuring the benefits of university research: Impact and the REF in the UK. *Research Evaluation*, *24*(1), 51–62. https://doi.org/10.1093/reseval/rvu028

Kiernan, V. (1997). Ingelfinger, Embargoes, and Other Controls on the Dissemination of Science News. *Science Communication*, *18*(4). (world). https://doi.org/10.1177/1075547097018004002

Kiernan, V. (2003a). Diffusion of News about Research. *Science Communication*, *25*(1), 3–13. https://doi.org/10.1177/1075547003255297

Kiernan, V. (2003b). Embargoes and Science News. *Journalism & Mass Communication Quarterly*, *80*(4), 903–920. https://doi.org/10.1177/107769900308000410

Kiernan, V. (2014). Public relations practices at medical journals. *Learned Publishing*, *27*(1), 5–13. https://doi.org/10.1087/20140102

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

Kirchhoff, K., Capurro, D., & Turner, A. M. (2014). A Conjoint Analysis Framework for Evaluating User Preferences in Machine Translation. *Machine Translation : MT*, *28*(1), 1–17. https://doi.org/10.1007/s10590-013-9140-x

Konkiel, S. (2013). Tracking citations and altmetrics for research data: Challenges and opportunities. *Bulletin of the American Society for Information Science and Technology*, *39*(6), 27–32. https://doi.org/10.1002/bult.2013.1720390610

Konkiel, S., Madjarevic, N., & Rees, A. (2016). *Altmetrics for Librarians: 100+ tips, tricks, and examples*. Altmetric. Retrieved from https://www.altmetric.com/libraries-ebook/ http://dx.doi.org/10.6084/m9.figshare.3749838

Könneker, C. (2017). Wissenschaftskommunikation in vernetzten Öffentlichkeiten. In H. Bonfadelli, B. Fähnrich, C. Lüthje, J. Milde, M. Rhomberg, & M. S. Schäfer (Eds.), *Forschungsfeld Wissenschaftskommunikation* (pp. 453–476). Wiesbaden: Springer Fachmedien.

https://doi.org/10.1007/978-3-658-12898-2_24

Kousha, K., & Thelwall, M. (2016). Can Amazon.com reviews help to assess the wider impacts of books? *Journal of the Association for Information Science and Technology*, *67*(3), 566–581. https://doi.org/10.1002/asi.23404

Kramer, B., & Bosman, J. (2015). 400+ Tools and innovations in scholarly communication. Retrieved February 16, 2017, from Google Docs website: https://docs.google.com/spreadsheets/d/1KUMSeq_Pzp4KveZ7pb5rddcssk1XBTiLHniD0d3nDqo/edit?usp=embed_facebook

Kramer, B., & Bosman, J. (2016). Innovations in scholarly communication—Global survey on research tool usage. *F1000Research*, *5*, 692. https://doi.org/10.12688/f1000research.8414.1

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

Lemke, S. (2020). *The Effect of Press Releases on Promoted Articles' Citations and Altmetrics*. Presented at the Metrics 2020: ASIS&T Virtual Workshop on Informetrics and Scientometrics Research. https://doi.org/10.5281/zenodo.4351360

Lemke, S. (2022). *The role of impact metrics in researchers' literature selection processes*. Invited talk presented at the Wiley Research APAC Webinars. Retrieved from https://live-web-webinars-wileyresearch.pantheonsite.io/webinar/role-impact-metrics-researchers-literature-selection-processes/

Lemke, S., Bräuer, P., & Peters, I. (2021). *Does the General Public Share Research on Twitter? A Case Study on the Online Conversation about the Search for a Nuclear Repository in Germany*. Glückstadt: Werner Hülsbusch. Retrieved from https://epub.uni-regensburg.de/44940/

Lemke, S., Brede, M., Rotgeri, S., & Peters, I. (2022). Research Articles Promoted in Embargo E-Mails Receive Higher Citations and Altmetrics. *Scientometrics*, (127), 75–97. https://doi.org/10.1007/s11192-021-04217-1

Lemke, S., Mazarakis, A., & Peters, I. (2021). Conjoint analysis of researchers' hidden preferences for bibliometrics, altmetrics, and usage metrics. *Journal of the Association for Information*

Science and Technology, *72*(6), 777–792. https://doi.org/10.1002/asi.24445

Lemke, S., Mehrazar, M., Mazarakis, A., & Peters, I. (2018). Are There Different Types of Online Research Impact? *Proceedings of the 81st ASIS&T Annual Meeting: Building an Ethical & Sustainable Information Future with Emerging Technology*, 282–289. Silver Springs, MD, USA: American Society for Information Science. Retrieved from https://www.asist.org/wp-content/uploads/2018/12/Final-81st-Annual-Meeting-Proceedings.pdf

Lemke, S., Mehrazar, M., Mazarakis, A., & Peters, I. (2019). "When You Use Social Media You Are Not Working": Barriers for the Use of Metrics in Social Sciences. *Frontiers in Research Metrics and Analytics*, *3*, 39. https://doi.org/10.3389/frma.2018.00039

Lemke, S., Nuredini, K., & Peters, I. (2020). PLOS Article Level Metrics. In R. Ball (Ed.), *Handbook Bibliometrics* (pp. 235–244). Munich: De Gruyter Saur.

Lemke, S., Sakmann, J., Brede, M., & Peters, I. (2021). Exploring the Relationship between Qualities of Press Releases to Research Articles and the Articles' Impact. *Proceedings of the 2021 International Conference on Scientometrics and Informetrics*, 639–644. Leuven, Belgium.

Lemke, S., Zagovora, O., Weller, K., Orth, A., Beucke, D., Stropel, J., & Peters, I. (2020). **metrics—Recommendations from the DFG *metrics project for "Measuring the Reliability and Perceptions of Indicators for Interactions with Scientific Products."* https://doi.org/10.18452/22242.2

Li, X., & Thelwall, M. (2012). F1000, Mendeley and traditional bibliometric indicators. *In Proceedings of the 17th International Conference on Science and Technology Indicators*, 451–551.

Liang, X., Su, L. Y.-F., Yeo, S. K., Scheufele, D. A., Brossard, D., Xenos, M., … Corley, E. A. (2014). Building Buzz: (Scientists) Communicating Science in New Media Environments. *Journalism & Mass Communication Quarterly*, *91*(4), 772–791. https://doi.org/10.1177/1077699014550092

Lin, J., & Fenner, M. (2013). Altmetrics in Evolution: Defining and Redefining the Ontology of Article-Level Metrics. *Information Standards Quarterly*, *25*(2), 20. https://doi.org/10.3789/isqv25no2.2013.04

Louviere, J. J., Flynn, T. N., & Carson, R. T. (2010). Discrete Choice Experiments Are Not Conjoint Analysis. *Journal of Choice Modelling*, *3*(3), 57–72. https://doi.org/10.1016/S1755-5345(13)70014-9

Ma, L., & Ladisch, M. (2016). Scholarly Communication and Practices in the World of Metrics: An Exploratory Study. *Proceedings of the 79th ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives Through Information & Technology*, 132:1-132:4. Silver Springs, MD, USA: American Society for Information Science. Retrieved from http://dl.acm.org/citation.cfm?id=3017447.3017579

Ma, L., & Ladisch, M. (2019). Evaluation complacency or evaluation inertia? A study of evaluative metrics and research practices in Irish universities. *Research Evaluation*, *28*(3), 209–217. https://doi.org/10.1093/reseval/rvz008

MacRoberts, M. H., & MacRoberts, B. R. (2018). The mismeasure of science: Citation analysis. *Journal of the Association for Information Science and Technology*, *69*(3), 474–482. https://doi.org/10.1002/asi.23970

Mandavilli, A. (2011). Peer review: Trial by Twitter. *Nature*, *469*(7330), 286–287. https://doi.org/10.1038/469286a

McCullough, D. (2002). A user's guide to conjoint analysis. *Marketing Research*, *14*(2), 18–23.

McKiernan, E. C., Schimanski, L. A., Nieves, C. M., Matthias, L., Niles, M. T., & Alperin, J. P. (2019). *Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations*. PeerJ Inc. https://doi.org/10.7287/peerj.preprints.27638v2

Mehrazar, M., Kling, C. C., Lemke, S., Mazarakis, A., & Peters, I. (2018). Can We Count on Social Media Metrics?: First Insights into the Active Scholarly Use of Social Media. *Proceedings of the 10th ACM Conference on Web Science*, 215–219. New York, NY, USA: ACM. https://doi.org/10.1145/3201064.3201101

Merton, R. K. (1948). The Self-Fulfilling Prophecy. *The Antioch Review*, *8*(2), 193–210. https://doi.org/10.2307/4609267

Miles, R., Konkiel, S., & Sutton, S. (2018). Scholarly Communication Librarians' Relationship with Research Impact Indicators: An Analysis of a National Survey of Academic Librarians in the

United States. *Journal of Librarianship and Scholarly Communication*, *6*(1). https://doi.org/10.7710/2162-3309.2212

Moed, H. F. (2016). Altmetrics as traces of the computerization of the research process. In C. R. Sugimoto (Ed.), *Theories of informetrics and scholarly communication. A Festschrift in Honor of Blaise Cronin* (pp. 360–371). Berlin: De Gruyter.

Moed, H. F. (2017). *Applied Evaluative Informetrics* (H. F. Moed, Ed.). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-60522-7_1

Moed, H. F. (2018). Assessment and support of emerging research groups. *FEMS Microbiology Letters*, *365*(17). https://doi.org/10.1093/femsle/fny189

Mohammadi, E., Thelwall, M., Haustein, S., & Larivière, V. (2015). Who reads research articles? An altmetrics analysis of Mendeley user categories. *Journal of the Association for Information Science and Technology*, *66*(9), 1832–1846. https://doi.org/10.1002/asi.23286

Mohammadi, E., Thelwall, M., Kwasny, M., & Holmes, K. L. (2018). Academic information on Twitter: A user survey. *PLOS ONE*, *13*(5). https://doi.org/10.1371/journal.pone.0197265

Nacke, O. (1979). Informetrie: Ein neuer Name für eine neue Disziplin. Begriffsbestimmung, Wissensstand und Entwicklungsprinzipien. *Nachrichten Für Dokumentation*, *30*(6), 219–226.

National Information Standards Organization (NISO). (2016). *NISO RP-25-2016, Outputs of the NISO Alternative Assessment Project* (p. 86) [Output Report]. Baltimore. Retrieved from NISO website: http://www.niso.org/apps/group_public/document.php?document_id=17091&wg_abbrev=altmetrics

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

Nicholas, D., Herman, E., Jamali, H. R., Abrizah, A., Boukacem-Zeghmouri, C., Xu, J., … Świgon, M. (2020). Millennial researchers in a metric-driven scholarly world: An international study. *Research Evaluation*, *29*(3), 263–274. https://doi.org/10.1093/reseval/rvaa004

Nicholas, D., Herman, E., Jamali, H. R., Osimo, D., Pujol, L., & Porcu, F. (2015). *Analysis of* *emerging reputation and funding mechanisms in the context of open science 2.0*. Luxembourg: Publications Office of the European Union. Retrieved from https://publications.jrc.ec.europa.eu/repository/handle/JRC94952

Nicholas, D., Jamali, H. R., Herman, E., Watkinson, A., Abrizah, A., Rodríguez-Bravo, B., … Polezhaeva, T. (2020). A global questionnaire survey of the scholarly communication attitudes and behaviours of early career researchers. *Learned Publishing*, *33*(3), 198–211. https://doi.org/10.1002/leap.1286

Nicholas, D., & Rowlands, I. (2011). Social media use in the research workflow. *Information Services & Use*, *31*(1–2), 61–83. https://doi.org/10.3233/ISU-2011-0623

Nielsen, K. H. (2009). In quest of publicity: The science—media partnership of the Galathea Deep Sea Expedition from 1950 to 1952. *Public Understanding of Science*, *18*(4), 464–480. https://doi.org/10.1177/0963662507083529

Niu, X., & Hemminger, B. M. (2012). A study of factors that affect the information-seeking behavior of academic scientists. *Journal of the American Society for Information Science and Technology*, *63*(2), 336–353. https://doi.org/10.1002/asi.21669

Nuredini, K. (2021). *Altmetrics for Digital Libraries. Concepts, Applications, Evaluation and Recommendations*. Berlin: Logos Verlag. https://doi.org/10.30819/5309

Pautasso, M. (2012). Publication Growth in Biological Sub-Fields: Patterns, Predictability and Sustainability. *Sustainability*, *4*(12), 3234–3247. https://doi.org/10.3390/su4123234

Peters, I., Jobmann, A., Eppelin, A., Hoffmann, C. P., Künne, S., & Wollnik-Korn, G. (2014). Altmetrics for large, multidisciplinary research groups: A case study of the Leibniz Association. *Assessing Libraries and Library Users and Use. Part II: Altmetrics - New Methods in Assessing Scholarly Communication and Libraries: Issues Applications, Results*. Zadar, Croatia. Retrieved from http://www.leibnizopen.de/suche/handle/document/115088

Phillips, D. P., Kanter, E. J., Bednarczyk, B., & Tastad, P. L. (1991). Importance of the Lay Press in the Transmission of Medical Knowledge to the Scientific Community. *New England Journal of Medicine*, *325*(16), 1180–1183. https://doi.org/10.1056/NEJM199110173251620

Piwowar, H., & Priem, J. (2012). A new framework for altmetrics. Retrieved July 4, 2019, from

Impactstory blog website: http://blog.impactstory.org/31524247207/

Priem, J. (2014). Altmetrics. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (1st edition, pp. 263–288). Cambridge, Massachusetts: The MIT Press.

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. *ArXiv:1203.4745 [Cs]*. Retrieved from http://arxiv.org/abs/1203.4745

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *Proceedings of the 26th International Conference on Science, Technology and Innovation Indicators*. Presented at the 26th International Conference on Science, Technology and Innovation Indicators (STI 2022). https://doi.org/10.48550/arXiv.2205.01833

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto. Retrieved January 23, 2017, from http://altmetrics.org/manifesto/

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Regorz, A. (2021). *Einführung in die Pfadanalyse mit R/lavaan*. Bochum.

Rijcke, S. de, Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use—A literature review. *Research Evaluation*, *25*(2), 161–169. https://doi.org/10.1093/reseval/rvv038

Robinson-Garcia, N., Costas, R., Isett, K., Melkers, J., & Hicks, D. (2017). The unbearable emptiness of tweeting—About journal articles. *PLOS ONE*, *12*(8). https://doi.org/10.1371/journal.pone.0183551

Rödder, S. (2015). Science Media Centres and public policy. *Science and Public Policy*, *42*(3), 387–400. https://doi.org/10.1093/scipol/scu057

Rousseau, R., & Ye, F. Y. (2013). A multi-metric approach for research evaluation. *Chinese Science Bulletin*, *58*(26), 3288–3290. https://doi.org/10.1007/s11434-013-5939-3

Rousseau, S., & Rousseau, R. (2017). Being metric-wise: Heterogeneity in bibliometric knowledge.

El Profesional de La Información, *26*(3), 480–487.

Rowlands, I. (2018). What are we measuring? Refocusing on some fundamentals in the age of desktop bibliometrics. *FEMS Microbiology Letters*, *365*(8). https://doi.org/10.1093/femsle/fny059

Satorra, A., & Bentler, P. M. (2010). Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic. *Psychometrika*, *75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Schäfer, M. S. (2008). Medialisierung der Wissenschaft? Empirische Untersuchung eines wissenschaftssoziologischen Konzepts / “Medialization” of Science? Empirical Assessment of a Sociological Concept. *Zeitschrift für Soziologie*, *37*(3), 206–225. https://doi.org/10.1515/zfsoz-2008-0302

Schloegl, C., & Gorraiz, J. (2010). Comparison of citation and usage indicators: The case of oncology journals. *Scientometrics*, *82*(3), 567–580. https://doi.org/10.1007/s11192-010-0172-1

Sivertsen, G. (2016). Publication-Based Funding: The Norwegian Model. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research Assessment in the Humanities: Towards Criteria and Procedures* (pp. 79–90). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-29016-4_7

Sivertsen, G., & Larsen, B. (2012). Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: An empirical analysis of the potential. *Scientometrics*, *91*(2), 567–575. https://doi.org/10.1007/s11192-011-0615-3

Small, H. (2004). On the shoulders of Robert Merton: Towards a normative theory of citation. *Scientometrics*, *60*(1), 71–79. https://doi.org/10.1023/B:SCIE.0000027310.68393.bc

Sosteric, M. (1999). Endowing Mediocrity: Neoliberalism, Information Technology, and the Decline of Radical Pedagogy. *Radical Pedagogy*, *1*(1), 1–41.

Streiner, D. L. (2005). Finding Our Way: An Introduction to Path Analysis. *The Canadian Journal of Psychiatry*, *50*(2), 115–122. https://doi.org/10.1177/070674370505000207

Stryker, J. E. (2002). Reporting Medical Information: Effects of Press Releases and Newsworthiness on Medical Journal Articles’ Visibility in the News Media. *Preventive Medicine*, *35*(5), 519–530. https://doi.org/10.1006/pmed.2002.1102

Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology*, *68*(9), 2037–2062. https://doi.org/10.1002/asi.23833

Sumner, P., Vivian-Griffiths, S., Boivin, J., Williams, A., Bott, L., Adams, R., … Chambers, C. D. (2016). Exaggerations and Caveats in Press Releases and Health-Related Science News. *PLOS ONE*, *11*(12), e0168217. https://doi.org/10.1371/journal.pone.0168217

Sumner, P., Vivian-Griffiths, S., Boivin, J., Williams, A., Venetis, C. A., Davies, A., … Chambers, C. D. (2014). The association between exaggeration in health related science news and academic press releases: Retrospective observational study. *BMJ*, *349*. https://doi.org/10.1136/bmj.g7015

Tahamtan, I., Afshar, A. S., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics*, *107*(3), 1195–1225. https://doi.org/10.1007/s11192-016-1889-2

Tahamtan, I., & Bornmann, L. (2020). Altmetrics and societal impact measurements: Match or mismatch? A literature review. *El Profesional de La Información*, *29*(1). https://doi.org/10.3145/epi.2020.ene.02

Tenopir, C., Allard, S., Bates, B. J., Levine, K. J., King, D. W., Birch, B., … Caldwell, C. (2011). Perceived value of scholarly articles. *Learned Publishing*, *24*(2), 123–132. https://doi.org/10.1087/20110207

Tenopir, C., King, D. W., Spencer, J., & Wu, L. (2009). Variations in article seeking and reading patterns of academics: What makes a difference? *Library & Information Science Research*, *31*(3), 139–148. https://doi.org/10.1016/j.lisr.2009.02.002

Tenopir, C., Levine, K., Allard, S., Christian, L., Volentine, R., Boehm, R., … Watkinson, A. (2016). Trustworthiness and authority of scholarly information in a digital age: Results of an international questionnaire. *Journal of the Association for Information Science and Technology*, *67*(10), 2344–2361. https://doi.org/10.1002/asi.23598

Thelwall, M. (2017a). Are Mendeley reader counts useful impact indicators in all fields? *Scientometrics*, *113*(3), 1721–1731. https://doi.org/10.1007/s11192-017-2557-x

Thelwall, M. (2017b). Do Mendeley reader counts indicate the value of arts and humanities research? *Journal of Librarianship and Information Science*, 0961000617732381. https://doi.org/10.1177/0961000617732381

Thelwall, M. (2018a). Altmetric Prevalence in the Social Sciences, Arts and Humanities: Where are the Online Discussions? *Journal of Altmetrics*, *1*(1), 4. https://doi.org/10.29024/joa.6

Thelwall, M. (2018b). Early Mendeley readers correlate with later citation counts. *Scientometrics*, *115*(3), 1231–1240. https://doi.org/10.1007/s11192-018-2715-9

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. *PLOS ONE*, *8*(5), e64841. https://doi.org/10.1371/journal.pone.0064841

Thelwall, M., & Nevill, T. (2018). Could scientists use Altmetric.com scores to predict longer term citation counts? *Journal of Informetrics*, *12*(1), 237–248. https://doi.org/10.1016/j.joi.2018.01.008

Thoma, B., Sanders, J. L., Lin, M., Paterson, Q. S., Steeg, J., & Chan, T. M. (2015). The Social Media Index: Measuring the Impact of Emergency Medicine and Critical Care Websites. *Western Journal of Emergency Medicine*, *16*(2), 242–249. https://doi.org/10.5811/westjem.2015.1.24860

Tian, Y., Wen, C., & Hong, S. (2008). Global scientific production on GIS research by bibliometric analysis from 1997 to 2006. *Journal of Informetrics*, *2*(1), 65–74. https://doi.org/10.1016/j.joi.2007.10.001

Tsou, A., Bowman, T. D., Ghazinejad, A., & Sugimoto, C. R. (2015). Who tweets about science? *Proceedings of the 2015 International Conference on Scientometrics and Informetrics*. Presented at the ISSI2015, Istanbul, Turkey.

Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature*, *512*(7513), 126–129. https://doi.org/10.1038/512126a

van Rooyen, C. (2002). *A Report on Science and Technology Coverage in the SA Print Media* (p. 22). South Africa: Department of Journalism, University of Stellenbosch. Retrieved from Department of Journalism, University of Stellenbosch website:

https://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=16CEFB03D33737257219DF5B29E41ACD?doi=10.1.1.180.5586&rep=rep1&type=pdf

Vermeulen, B., Goos, P., & Vandebroek, M. (2011). Rank-order choice-based conjoint experiments: Efficiency and design. *Journal of Statistical Planning and Inference*, *141*(8), 2519–2531. https://doi.org/10.1016/j.jspi.2011.01.019

Vrieze, J. de. (2018). EurekAlert! Has spoiled science news. Here's how we can fix it. Retrieved August 14, 2020, from VWN - Vereniging voor Wetenschapsjournalistiek en -communicatie Nederland website: https://www.vwn.nu/?p=2225/

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 365–391. https://doi.org/10.1016/j.joi.2016.02.007

Waltman, L., Pinfield, S., Rzayeva, N., Oliveira Henriques, S., Fang, Z., Brumberg, J., … Swaminathan, S. (2021). *Scholarly communication in times of crisis: The response of the scholarly communication system to the COVID-19 pandemic* [Report]. Research on Research Institute. https://doi.org/10.6084/m9.figshare.17125394.v1

Weingart, P. (1998). Science and the media. *Research Policy*, *27*(8), 869–879. https://doi.org/10.1016/S0048-7333(98)00096-1

Weller, K., Dröge, E., & Puschmann, C. (2011). Citation analysis in Twitter. In: Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences. *CEUR-WS.Org, Tilburg University*.

White, H. D., Boell, S. K., Yu, H., Davis, M., Wilson, C. S., & Cole, F. T. H. (2009). Libcitations: A measure for comparative assessment of book publications in the humanities and social sciences. *Journal of the American Society for Information Science and Technology*, *60*(6), 1083–1096. https://doi.org/10.1002/asi.21045

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Willmott, H. C. (2011). *Journal List Fetishism and the Perversion of Scholarship: Reactivity and the ABS List* (SSRN Scholarly Paper No. ID 1753627). Rochester, NY: Social Science Research

Network. Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=1753627

Wilsdon, J., Bar-Ilan, J., Frodeman, R., Lex, E., Peters, I., & Wouters, P. (2017). *Next-generation metrics: Responsible metrics and evaluation for open science*. Luxembourg: Publications Office.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., … Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. doi: 10.13140/RG.2.1.4929.1363. Retrieved from http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/

Wouters, P., & Costas, R. (2012). *Users, Narcissism and the Control: Tracking the Impact of Scholarly Publications Inthe 21st Century*. Utrecht: Stichting Surf. Retrieved from http://research-acumen.eu/wp-content/uploads/Users-narcissism-and-control.pdf

Xia, F., Su, X., Wang, W., Zhang, C., Ning, Z., & Lee, I. (2016). Bibliographic Analysis of Nature Based on Twitter and Facebook Altmetrics Data. *PLOS ONE*, *11*(12). https://doi.org/10.1371/journal.pone.0165997

Xia, F., Wang, W., Bekele, T. M., & Liu, H. (2017). Big Scholarly Data: A Survey. *IEEE Transactions on Big Data*, *3*(1), 18–35. https://doi.org/10.1109/TBDATA.2016.2641460

Yu, H. (2017). Context of altmetrics data matters: An investigation of count type and user category. *Scientometrics*, *111*(1), 267–283. https://doi.org/10.1007/s11192-017-2251-z

Zagovora, O., Weller, K., Janosov, M., Wagner, C., & Peters, I. (2018). What increases (social) media attention: Research impact, author prominence or title attractiveness? *STI 2018 Conference Proceedings*. https://doi.org/10.31235/osf.io/mwxye

Zahedi, Z. (2018). *Understanding the value of social media metrics for research evaluation*. Leiden University. Retrieved from https://hdl.handle.net/1887/67131

Zahedi, Z., & Costas, R. (2018). General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. *PLOS ONE*, *13*(5). https://doi.org/10.1371/journal.pone.0197326

Zahedi, Z., Fenner, M., & Costas, R. (2014). How consistent are altmetrics providers? Study of 1000

PLOS ONE publications using the PLOS ALM, Mendeley and Altmetric. Com APIs. *Altmetrics 14. Workshop at the Web Science Conference Bloomington, USA*. Bloomington. https://doi.org/10.6084/m9.figshare.1041821

Zhang, L., & Wang, J. (2021). What affects publications' popularity on Twitter? *Scientometrics*, *126*(11), 9185–9198. https://doi.org/10.1007/s11192-021-04152-1

Zuccala, A. A., Verleysen, F. T., Cornacchia, R., & Engels, T. C. E. (2015). Altmetrics for the humanities: Comparing Goodreads reader ratings with citations to history books. *Aslib Journal of Information Management*, *67*(3), 320–336. https://doi.org/10.1108/AJIM-11-2014-0152

Workshare of Co-authored Publications

Details on the doctoral candidate’s workshare regarding co-authored manuscripts included in this dissertation.

Study A

Authors: Steffen Lemke, Maryam Mehrazar, Athanasios Mazarakis, & Isabella Peters

Candidate’s estimated workshare:

Conceptualization	Planning	Execution	Manuscript preparation
30%	50%	60%	80%

Details: SL performed the statistical analysis, coded the survey free text answers according to their themes and wrote the first draft of the manuscript; MM and SL conducted the interviews, implemented and supervised the online survey and provided the figures used in the manuscript; IP acquired funding for the research project. All authors contributed conception and design of the study, contributed to manuscript revision, and read and approved the submitted version.

Study B

Authors: Steffen Lemke, Athanasios Mazarakis, & Isabella Peters

Candidate’s estimated workshare:

Conceptualization	Planning	Execution	Manuscript preparation
50%	70%	90%	80%

Details: SL implemented the software used for the experiments, conducted the online survey process, coded the free text answers, performed the statistical analyses, documented and published the source code and datasets, and prepared this manuscript’s first draft; IP acquired funding for the research project. All authors contributed conception and design of the study, contributed to manuscript revision, and read and approved the submitted version.

Study C

Authors: Steffen Lemke, Max Brede, Sophie Rotgeri, & Isabella Peters

Candidate’s estimated workshare:

Conceptualization	Planning	Execution	Manuscript preparation
50%	80%	80%	80%

Details: IP, MB, and SL contributed conception and design of the study; SR conducted the acquisition of data on embargo e-mails; MB and SL performed the statistical analyses and provided the figures; IP acquired funding for the research project; SL retrieved data from Altmetric.com, Crossref, the Competence Centre for Bibliometrics (CCB) database, and Web of Science, performed the coding of journals to research areas and prepared this manuscript's first draft; all authors contributed to manuscript revision, read and approved the submitted version.

Study D

Authors: Steffen Lemke, Athanasios Mazarakis, & Isabella Peters

Candidate's estimated workshare:

Conceptualization	Planning	Execution	Manuscript preparation
80%	90%	90%	80%

Details: SL performed the statistical analysis, provided the figures used in the manuscript and wrote its first draft; IP acquired funding for the research project. All authors contributed conception and design of the study, contributed to manuscript revision, and read and approved the submitted version.

Steffen Lemke - Resumé

E-mail: s.lemke@zbw.eu; steffen.t.lemke@gmail.com
URL: http://blog.stlemke.de/
ORCID iD: https://orcid.org/0000-0002-3506-7083
GitHub: https://github.com/stlemke

Professional Experience

Since 2022	Research assistant/PhD candidate at *Kiel University*
2017 - 2022	Research assistant/PhD candidate at *ZBW Leibniz Information Centre for Economics*
2014 - 2016	Student assistant at *ZBW Leibniz Information Centre for Economics*

Higher Education

2013 - 2016	Master of Science in Information Systems at Kiel University
2010 - 2013	Bachelor of Science in Information Systems at Kiel University

Academic Service & Community Engagement

- member of the organizing committee of the *Altmetrics Workshop Series*
- special session co-organizer and chair at the *STI2021 Conference*
- reviewer for several pertinent journals and conferences, e.g., *Scientometrics*, *JASIST*, *PLoS ONE*, *Profesional de la Información*, the *ISSI Conference*, the *STI Conference*, the *ASIST Annual Meeting*, the *ACM Web Science Conference*, etc.

Teaching

SS 2021	Project Management tutorials
SS 2020	Information Systems - Fundamental Competencies tutorials
SS 2019	Project Management tutorials
WS 18/19	Information and Knowledge Management tutorials
SS 2018	Information Systems - Fundamental Competencies tutorials
WS 17/18	Project Management tutorials
SS 2017	Information and Knowledge Management tutorials

Further Training

06/2021	Science communication training with *sciencehoch3*
01/2020	Time- and self-management seminar with *Jochen Schlicht Leadership Development*
07/2019	European Summer School for Scientometrics at *Katholieke Universiteit Leuven*
04/2019	Fair Data & Software workshop at *TIB Leibniz Information Centre for Science and Technology*
09/2018	Open Citations workshop at *Università di Bologna*

03/2018 Basic program ”Didactics in Higher Education“ at *Kiel University*

02/2018 Writing Academic Texts in English seminar with *Crow Consulting*

09/2017 Alternative Indicators Summer School at *University of Wolverhampton*

07/2017 GESIS Methods Seminar Statistics at *GESIS Leibniz Institute for the Social Sciences*

03/2015 2nd Barcamp Science 2.0 at *ZBW Leibniz Information Centre for Economics*

11/2014 5th Student Workshop for Research in the Information Sciences at *Humboldt University Berlin*

Publications

Journal Articles

Lemke, S., Brede, M., Rotgeri, S., & Peters, I. (2022). Research Articles Promoted in Embargo E-Mails Receive Higher Citations and Altmetrics. *Scientometrics, 127,* 75-97. https://doi.org/10.1007/s11192-021-04217-1

Lemke, S., Mazarakis, A., & Peters, I. (2021). Conjoint Analysis of Researchers‘ Hidden Preferences for Bibliometrics, Altmetrics and Usage Metrics. *Journal of the Association for Information Science and Technology (JASIS&T), 72*, 777-792. https://doi.org/10.1002/asi.24445

Lemke, S., Mehrazar, M., Mazarakis, A., & Peters, I. (2019). “When you use social media, you are not working“: Barriers for the Use of Metrics in Social Sciences. *Frontiers in Research Metrics and Analytics, 3*(39), 1-18. https://doi.org/10.3389/frma.2018.00039

Monographs & Book Chapters

Lemke, S., Mazarakis, A., & Peters, I. (under review). Path model of the interplay between the promotion and the received attention of research articles. In Broer, I., Lemke, S., Mazarakis, A., Peters, I., & Zinke-Wehlmann, C. (Ed.), *The Science-Media Interface: On the relation between internal and external science communication*. Berlin: De Gruyter. ISBN 978-3-11-077636-2.

Lemke, S., Zagovora, O., Weller, K., Orth, A., Beucke, D., Stropel, J., & Peters, I. (2020). **metrics – Recommendations from the DFG *metrics Project for „Measuring the Reliability and Perceptions of Indicators for Interactions with Scientific Products“*. Deutsche Initiative für Netzwerkinformation e. V.. https://doi.org/10.18452/22242

Lemke, S., Nuredini, K., & Peters, I. (2020). PLOS Article Level Metrics. In Ball, R. (Ed.), *Handbook Bibliometrics*. Munich: De Gruyter Saur. ISBN 978-3-11-064661-0. https://doi.org/10.1515/9783110646610

Nuredini, K., Lemke, S., & Peters, I. (2020). Social Media and Altmetrics. In Ball, R. (Ed.), *Handbook Bibliometrics*. Munich: De Gruyter Saur. ISBN 978-3-11-064661-0. https://doi.org/10.1515/9783110646610

Papers in Conference Proceedings (peer reviewed)

Lemke, S., Witthake, A., & Peters, I. (2022). Altmetrics for German medical research: what leads to research articles achieving policy impact?. *Proceedings of the 26th International Conference on Science, Technology and Innovation Indicators (STI 2022)*, 7-9 September 2022, Granada, Spain. https://doi.org/10.5281/zenodo.6645109.

Lemke, S., Sakmann, J., Brede, M., & Peters, I. (2021). Exploring the Relationship between Qualities of Press Releases to Research Articles and the Articles' Impact. *Proceedings of the 18th International Conference on Scientometrics & Informetrics (ISSI 2021)* (p. 639–644). Leuven, Belgium. https://kuleuven.app.box.com/s/kdhn54ndlmwtil3s4aaxmotl9fv9s329

Lemke, S., Bräuer, P., & Peters, I. (2021). Does the General Public Share Research on Twitter? A Case Study on the Online Conversation about the Search for a Nuclear Repository in Germany. *Proceedings of the 16th International Symposium for Information Science (ISI 2021)* (p. 94-114), 8-10 March 2021, Regensburg, Germany. Glückstadt, Germany: vwh. https://doi.org/10.5283/epub.44940

Hahn, O., Lemke, S., Mazarakis, A., & Peters, I. (2020). Which Visual Elements Make Texts Appear Scientific? An Empirical Analysis: Welche Layout-Elemente lassen Texte wissenschaftlich erscheinen? Eine empirische Untersuchung. In *Mensch und Computer 2020 (MuC'20)*, 6-9 September 2020, Magdeburg, Germany. New York, NY, USA: ACM. https://doi.org/10.1145/3404983.3410014.

Lemke, S., & Peters, I. (2019). Coping with Altmetrics' Heterogeneity – A Survey on Social Media Platforms' Usage Purposes and Target Groups for Researchers. *Proceedings of the 17th International Conference on Scientometrics & Informetrics* (p. 2320-2325). Rome, Italy.

Lemke, S., Mehrazar, M., Mazarakis, A., & Peters, I. (2018). Are There Different Types of Online Research Impact? *Proceedings of the 81st ASIS&T Annual Meeting of the Association for Information Science & Technology* (p. 282-289). Vancouver, Canada.

Mehrazar, M., Kling, C. C., Lemke, S., Mazarakis, A., & Peters, I. (2018). Can We Count on Social Media Metrics? First Insights into the Active Scholarly Use of Social Media. *Proceedings of the 10th ACM Conference on Web Science – WebSci18* (p. 215-219). New York, NY, USA: ACM.

Lemke, S., & Mazarakis, A. (2017). Analyse wissenschaftlicher Konferenz-Tweets mittels Codebook und der Software Tweet Classifier. In T. Köhler, E. Schoop, & N. Kahnwald (Ed.), *GeNeMe17 – Gemeinschaften in Neuen Medien* (p. 95–104). Dresden, Germany: TUDpress.

Mazarakis, A., Lemke, S., & Peters, I. (2016). Tweets and Scientific Conferences: The Use Case of the Science 2.0 Conference – Revisited. In Bernadas, C., & Minchella, D. (Ed.), *Proceedings of the 3rd European Conference on Social Media (ECSM 2016)* (p. 214–222). Reading, UK: Academic Conferences and Publishing International Limited.

Lemke, S., Mazarakis, A., & Peters, I. (2016). Characteristics of Twitter Usage at Scientific Conferences. *Proceedings of the 18th General Online Research Conference (GOR 2016)* (p. 34–35). Dresden, Germany: Deutsche Gesellschaft für Online-Forschung (DGOF).

Posters (peer reviewed)

Hahn, O., & Lemke, S. (2020). An Exploration of Scientific Press Releases in the Context of Altmetrics. Poster and extended abstract available at http://altmetrics.org/altmetrics20/. Presented at *altmetrics20 workshop*, online, 6 November 2020.

Mehrazar, M., Shema, H., Lemke, S., & Peters, I. (2019). Why do researchers from Economics and Social Sciences cite online? Insights from an exploratory survey. Presented at *17th International Conference on Scientometrics & Informetrics*, Rome, Italy, 2-5 September 2019. https://doi.org/10.5281/zenodo.3475019.

Zagovora, O., Mehrazar, M., Lemke, S., Morovatdar, T., Peters, I., & Weller, K. (2018). Which social media interactions indicate positive opinions about cited publications? A comparison of user survey and sentiment analysis. Zenodo. Presented at *Altmetrics Conference (5:AM)*, London, England, 26-27 September 2018. http://doi.org/10.5281/zenodo.2386713.

Lemke, S., Mehrazar, M., Peters, I., Beucke, D., Gottschling, M., Krausz, A., Kusche, M., Lindner, D., Mazarakis, A., Orth, A., Weller, K., & Zagovora, O. (2017). Exploring the Meaning and Perception of Altmetrics. Zenodo. Presented at *Altmetrics Conference (4:AM)*, Toronto, Canada, 26-29 September 2017. http://doi.org/10.5281/zenodo.1037146.

Lemke, S., Mazarakis, A., & Peters, I. (2015). Understanding Scientific Conference Tweets. *Proceedings of the 17th General Online Research Conference (GOR 2015)* (p. 52–53). Cologne, Germany: Deutsche Gesellschaft für Online-Forschung (DGOF).

Workshop Presentations

Lemke, S. (2020). The Effect of Press Releases on Promoted Articles‘ Citations and Altmetrics. Slides available at: https://doi.org/10.5281/zenodo.4118512 Presented at *Metrics 2020: ASIS&T Virtual Workshop on Informetrics and Scientometrics Research*, Pittsburgh, Pennsylvania, USA, 22-23 October 2020.

Lemke, S. (2020). Welche visuellen Elemente lassen Texte wissenschaftlich erscheinen? Presented at *JF:TEC-Jahrestreffen 2020*, Kassel, Germany, 09 September 2020.

Lemke, S. (2019). The *metrics-project’s user studies: how researchers perceive and use metrics. Slides available at: https://doi.org/10.5281/zenodo.2661038. Presented at **Metrics in Transition Workshop*, Goettingen, Germany, 27-28 March 2019.

Lemke, S. (2018). User- and Usage Studies in the *metrics Project. Slides available at: https://doi.org/10.5281/zenodo.1283183. Presented at *COAR *metrics Repository Workshop*, Hamburg, Germany, 14 May 2018.

Lemke, S., Mehrazar, M., Peters, I., & Mazarakis, A. (2017). Evaluating altmetrics acts through their creators – how to advance?. Extended abstract available at http://altmetrics.org/altmetrics17/. Slides available at: http://doi.org/10.5281/zenodo.1037276. Presented at *altmetrics17 workshop*, Toronto, Canada, 26 September 2017.

Lemke, S. (2016). EconStor und das *metrics-Projekt. Presented at *ZB MED Vernetzungsworkshop*, Cologne, Germany, 30 November 2016.

Lemke, S. (2016). Predictive Analytics für die Ökonomie. Presented at *7. Studierenden-Workshop für informationswissenschaftliche Forschung*, Kiel, Germany, 18-19 November 2016.

Lemke, S. (2014). Tweets und Wissenschaftliche Konferenzen. Presented at *5. Studierenden-Workshop für informationswissenschaftliche Forschung*, Berlin, Germany, 14-15 November 2014.

Invited Talks & Panel Discussions

Lemke, S. (2022). The role of impact metrics in researchers' literature selection processes. Invited talk at *Wiley Research APAC Webinars*: https://live-web-webinars-wileyresearch.pantheonsite.io/webinar/role-impact-metrics-researchers-literature-selection-processes/. Associated presentation slides available at https://doi.org/10.5281/zenodo.5873423

Lemke, S. (2020). Attitudes towards altmetrics from German researchers: insights from the *metrics-project. Invited talk at *Altmetric.com Webinar on „Open Access and Altmetrics in Germany"*. London, England, 04 February 2020. Recording available on figshare: https://figshare.com/articles/Open_Access_and_Altmetrics_in_Germany/11836956. Associated presentation slides available at https://doi.org/10.5281/zenodo.3635714

Lemke, S. (2018). Open Metrics: Originators and their Perceptions. Zenodo. Invited talk at *Workshop on Open Metrics*. Uppsala, Sweden, 23 May 2018. Associated presentation slides available at http://doi.org/10.5281/zenodo.1254924

Appearances in Science Blogs or Podcasts

Lemke, S. (2020). Altmetrics: How researchers assess the significance for scholarly impact. Guest post for the *ZBW Mediatalk Blog*: https://www.zbw-mediatalk.eu/2020/02/altmetrics-how-researchers-assess-the-significance-for-scholarly-impact/ Original post in German: https://www.zbw-mediatalk.eu/de/2020/02/altmetrics-so-bewerten-forschende-die-aussagekraft-fuer-den-wissenschaftlichen-einfluss/

Interview about the *metrics-project's user studies for the exhibition Open UP!: https://100jahre.zbw.eu/wp-content/uploads/2019/06/Steffen-Lemke_Einfluss_der_Wissenschaft.mp3

Lemke, S., Peters, I., & Mazarakis, A. (2019). "If you use social media then you are not working" – How do social scientists perceive altmetrics and online forms of scholarly communication?. Guest post for the *LSE Impact Blog*: https://blogs.lse.ac.uk/impactofsocialsciences/2019/03/20/if-you-use-social-media-then-you-are-not-working-how-do-social-scientists-perceive-altmetrics-and-online-forms-of-scholarly-communication/

Interview about Altmetrics on the *Science for Progress-Podcast*: https://www.scienceforprogress.eu/21-altmetrics-a-better-way-to-evaluate-researchers-with-steffen-lemke

Honors & Awards

Best Paper Award at the ASIS&T SIG-MET Workshop 2020 for *The Effect of Press Releases on Promoted Articles‘ Citations and Altmetrics*

Best Poster Award (2nd place) at the General Online Research Conference 2015 for *Understanding Scientific Conference Tweets*

Statutory Declaration

I hereby confirm that - apart from my supervisors' guidance - the content and design of this thesis is my own work and all used sources have been listed and duly acknowledged in the text. The thesis has not been submitted for any other degree or professional qualification. Parts based on jointly-authored publications have been clearly denoted as such and my own respective workshare documented. The thesis has been prepared subject to the Rules of Good Scientific Practice of the German Research Foundation. I confirm that no academic degree has ever been withdrawn from me.

Place, Date Signature of the doctoral candidate