



A large-scale validation of the relationship between cross-disciplinary research and its uptake in policy-related documents, using the novel Overton altmetrics database

Henrique Pinheiro , Etienne Vignola-Gagné , and David Campbell 

Science-Metrix, an Elsevier company, Montreal, Quebec, Canada

an open access  journal



Citation: Pinheiro, H., Vignola-Gagné, E., & Campbell, D. (2021). A large-scale validation of the relationship between cross-disciplinary research and its uptake in policy-related documents, using the novel Overton altmetrics database. *Quantitative Science Studies*, 2(2), 616–642. https://doi.org/10.1162/qss_a_00137

DOI:
https://doi.org/10.1162/qss_a_00137

Peer Review:
https://publons.com/publon/10.1162/qss_a_00137

Supporting Information:
https://doi.org/10.1162/qss_a_00137

Received: 2 December 2020
Accepted: 20 April 2021

Corresponding Author:
David Campbell
d.campbell@elsevier.co

Handling Editor:
Ludo Waltman

Copyright: © 2021 Henrique Pinheiro, Etienne Vignola-Gagné, and David Campbell. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



Keywords: altmetrics, cross-disciplinary research, interdisciplinary research, policy outcomes, program evaluation, societal outcomes of research

ABSTRACT

Cross-disciplinary research (multi-/interdisciplinarity) is incentivized by funding agencies to foster research outcomes addressing complex societal challenges. This study focuses on the link between cross-disciplinary research and its uptake in a broad set of policy-related documents. Using the new policy-oriented database Overton, matched to Scopus, logistic regression was used in assessing this relationship in publications from FP7- and H2020-supported projects. Cross-disciplinary research was captured through two lenses at the paper level, namely from the disciplinary diversity of contributing authors (DDA) and of cited references (DDR). DDA increased the likelihood that publications were cited in policy documents, with DDR possibly making a contribution, but only when publications result from the work of few authors. Citations to publications captured by Overton were found to originate in scientific advice documents, rather than in legislative or executive records. Our approach enables testing in a general way the assumption underlying many funding programs, namely that cross-disciplinary research will increase the policy relevance of research outcomes. Findings suggest that research assessments could benefit from measuring uptake in policy-related literature, following additional characterization of the Overton database; of the science-policy interactions it captures; and of the contribution of these interactions within the larger policymaking process.

1. INTRODUCTION

With the increasing emphasis that funding organizations place on the longer-term socioeconomic impacts from research, an increasing number of funding programs promote cross-disciplinary research (XDR), assuming these scientific practices are more likely to fuel such policy-related returns (Gleed & Marchant, 2016; Rylance, 2015). This argument linking cross-disciplinary research and societal outcomes is supported by limited evidence concerning the ability of the first to bring about the second (Chavarro, Tang, & Rafols, 2014) or whether policymakers generally succeed in fostering the first (Garner, Porter et al., 2012). In addition to these issues, existing quantitative measurement strategies for societal outcomes of research tend to be restricted in scope (e.g., patent-based metrics).

Altmetrics methods, in particular, citations to the peer-reviewed literature in policy documents, offers potential to broaden the spectrum of societal outcomes that can be assessed in a robust quantitative manner. Altmetrics capture instances of uptake or mentions towards peer-reviewed publications in a range of potential knowledge transfer contexts, including in journalistic news outlets, Facebook posts, Wikipedia entries, and the like. Recently, path-breaking studies have focused specifically on policy citations towards peer-reviewed publications, specifically those recorded in the *Altmetric.com* database. These studies reported that only low numbers of papers get cited in policy documents. They also reported a high skewness in citation distribution; a concentration of citedness in applied life sciences and social science fields; and technical issues with the *Altmetric.com* database (Bornmann, Haunschild, & Marx, 2016; Haunschild & Bornmann, 2017; Tattersall & Carroll, 2018). Based on detailed content analysis of citing policy documents, Newson and colleagues found multiple instances of research mentions that were not made as formal citations, concluding that “[c]itation rates are likely to provide an underestimation of research use by policy agencies” (Newson, Rychetnik et al., 2018, p. 10).

Despite the potential of altmetric methods, definitive attribution of research outcomes to specific funding programs (such as the instruments promoting cross-disciplinary approaches that interest us) is challenging for several reasons:

1. A lack of suitable data sets to confidently discard confounding factors (such as local and global trends in research systems; or the combination of impacts that comes with combining multiple streams of funding in research) in testing the effects of specific programs using quantitative approaches such as econometric modeling and difference-in-differences (Buenstorf & Koenig, 2020; Hird & Pfothenauer, 2017); for example, data sets related to funding programs specifically building on XDR are often too small to offer adequate statistical power in the complex model specifications required to resolve attribution.
2. A lack of unbiased and well-characterized (altmetrics) data sources, making it difficult to explore/rank the longer-term impacts of research in a quantitative manner.
3. Uncertainty regarding the phenomena captured by altmetrics and their exact association with the goal of increasing societal outcomes from research (Haustein, 2016).

To circumvent these constraints, we devise a methodological framework to assess, in a generic way, the likely effects of research and innovation policy interventions tapping on a specific mechanism (e.g., cross-disciplinary research), rather than aiming to directly assess the effect of a specific intervention. In this way, the proposed approach enables testing the underlying assumption of a specific mechanism rather than a specific intervention, thereby partially circumventing the first limitation mentioned above. To address the other two limitations, we make use of the new Overton database, an altmetrics source dedicated to documenting some of the policy outcomes produced by research.

Our three starting research questions were as follow:

- A. Can altmetrics offer a practical, reliable, generalizable method for capturing societal outcomes of research in the form of informing policymaking?
- B. Do higher degrees of XDR found in peer-reviewed publications increase the odds of these articles being cited in policy documents?
- C. Can policy interventions (mainly funding) promoting XDR increase the odds of resulting findings supporting evidence-based policymaking?

To address Question A, the citation context of a random sample of 50 publications cited by documents indexed in Overton was assessed qualitatively to determine the extent to which

such citations reflect research input into decision making (see Section 3.2 for the corresponding methods). This process was also used in validating the accuracy of Overton's linkages between policy documents and the cited literature.

To address Question B, statistical modeling was performed on a sample of papers funded through the Framework Programmes (FPs) for Research and Technological Development (i.e., FP7 or H2020), and the extent of cross-disciplinarity was captured at the paper level through two lenses: disciplinary diversity of contributing authors (DDA) and disciplinary diversity of cited references (DDR), which tracks diversity of integrated knowledge (see Section 3.3). The DDA is often a target of policy intervention when funders require certain collaborative formats to teams applying to grants targeted for XDR research. It is expected that this diversity will bring to projects the variety of tools required for real-world problem solving, including to tackle policy-relevant problems (Belcher & Hughes, 2020; Rylance, 2015; Schneider, Buser et al., 2019). Disciplinary diversity in references would indicate that this diversity in tools and approaches has been integrated within the intellectual fabric of the publication's text itself. This result, however, could be independent of team composition and may even be achieved by individual researchers that achieve "individual interdisciplinarity" (Calvert, 2010). If DDA and DDR are properly captured, one would thus expect a positive, but not necessarily strong, correlation between them. By selecting papers from projects supported by FP7 or H2020, it was also possible to control for some characteristics of these projects in assessing the association between cross-disciplinary research and subsequent policy uptake. The addition of numerous controls to the model specification enabled moving away from a simple analysis of correlation to the assessment of a causal link.

The model was tested first on a subset of UK papers and then expanded to the broader set of EU papers. By first restricting the analyses to a sample of UK papers, we limited the possible effects of coverage biases in the Overton database, which was used to track citations in policy documents (it is produced by a UK-based company with a better coverage of policy documents from the Anglo-Saxon world, in particular from the United Kingdom).

The analyses were subsequently expanded to a larger pool of about 126,000 FP7/2020-funded papers from all European countries (see Section 4.2.2; see Section 3.3 for the corresponding methods). By testing the replicability of the study findings for the UK sample on this larger EU data set, we intended to assess if Overton's coverage bias may be large enough to distort the conclusions of similar studies covering non-Anglo-Saxon countries, as well as to make this study's conclusions more generally applicable to the broader European context. In doing so, we acknowledge that not all European countries are equally represented in this broader data set. Indeed, the larger players (e.g., France and Germany) weigh heavily here and are still better covered in Overton than the smaller Eastern European countries.

Finally, Question C was answered by combining the findings for Questions A and B. For instance, if a positive answer is obtained for both A and B, it would be fair to argue that funding programs promoting XDR increase the odds of the resulting findings supporting evidence-based policymaking.

Going forward, the term UPRL ("uptake in policy-relevant literature") will be preferred when discussing references from policy-relevant documents towards peer-reviewed publications. This abbreviation takes stock of our findings for Question A (Section 4.1). In brief, the "policy documents" whose citations are captured in Overton were not so much markers of legislative change as markers of scientific advice and evidence synthesis activities targeted towards policymakers. Before moving to a more detailed description of methods (Section 3) and results (Section 4) from the empirical investigation, the literature review is presented in Section 2 to provide an overview

of the evidence currently available on the following chain of assumptions that underpin our research questions:

- science, technology, and innovation (STI) policy increasingly supports cross-disciplinary research (Section 2.1)
- cross-disciplinary research practices can be fostered through funding instruments and other policy interventions (Section 2.2)
- bibliometric indicators offer promise to robustly measure intensity in the deployment of XDR research practices, at scale (Section 2.3)
- cross-disciplinary research leads to improved societal outcomes (Section 2.4)
- altmetrics could offer a robust quantitative strategy to capture societal outcomes from research (Section 2.5)
- altmetrics relying on policy documents citing the scientific literature could offer a robust quantitative strategy to capture societal outcomes from research specifically on the topic of (governmental, NGO, or think tank) decision-making (Section 2.6)

2. LITERATURE REVIEW

This article uses the term *cross-disciplinary research* (XDR) to collectively refer to several research practices and organization modalities that are often referred to as interdisciplinarity research (IDR) in the literature, but also *multidisciplinarity* and even *transdisciplinarity* (Chavarro et al., 2014; van der Hel, 2016).

2.1. Policy Interest in Cross-Disciplinarity

Policy interventions often aim to address complex challenges (e.g., the UN Sustainable Development Goals) requiring input from a broad range of stakeholders. It is commonly assumed that the diversity of stakeholders needed to inform such interventions can span multiple dimensions, such as their activity sector, geographic location, and disciplinary background (Rylance, 2015). With the increasing emphasis that funding organizations place on the longer-term socio-economic impacts from research, an increasing number of funding programs promote scientific collaboration across these dimensions, assuming that it will fuel such returns. The following is a small sample of pre-eminent policies and interventions targeting interdisciplinarity and boundary-spanning collaboration as policy goals:

- the EC FP's Responsible Research & Innovation agenda
- the EC FP's European Research Council (ERC) funding mechanism
- H2020 and especially its Future and Emerging Technologies and Societal Challenges pillars (LERU, 2016)
- EC COST actions, including thematic subfunding programs such as BiodivERsA, which emphasize interdisciplinarity
- multiple US National Science Foundation (NSF) initiatives, including Convergence awards, Research Coordination Networks, Science and Technology Centers, and Synthesis Centers (Hackett, Leahey et al., 2021)
- large areas of research at the US National Institutes of Health (NIH) that fall under the concept of the "translational sciences"
- a small group of multilateral or intergovernmental funders including, notably, the Human Frontier Science Program (HFSP) (Science-Metrix, 2018), the Belmont Forum (Technopolis Group & Science-Metrix, 2020) and Future Earth

- a set of interventions that fall within the category of “excellence” competitive national funding to universities, including the Canada First Research Excellence Fund (CFREF), China’s Double First Class University Plan and its predecessors, Germany’s Excellence Initiative and Excellence Strategy, and Japan’s World Premier International Research Center Initiative

An international survey of national research funding agencies sponsored by the Global Research Council—a multilateral knowledge exchange mechanism for more than 20 national funding councils—found that, although “[m]ost of the funding agencies we interviewed were open in stating that they do not have formal policies relating to interdisciplinarity, [they] do have practices to encourage and support it” (Gleed & Marchant, 2016, p. 8). Support for interdisciplinary research (writ large) can therefore be safely considered a ubiquitous feature of STI policy in 2020.

But what justifies this flurry of policy interventions? Interdisciplinarity is advocated as the preferred tool to realize a number of central policy objectives for governments and societies. As a member of an EC expert committee on Research, Innovation, and Science Policy put it, fostering interdisciplinary research could result in

crossing departmental boundaries and inter-disciplinarity to generate new knowledge of transformative power ... exploit[ing] new types of problem-driven and user-oriented R&D research programmes that go way beyond well-established modes of targeted, incentivised R&D top-down ... Stimulat[ing] disruptive innovations to accelerate value creation across different industries and branches of knowledge through intellectual fusion, combinations and interfaces (Allmendinger, 2015, p. 4).

While the citation above may potentially capture an excessively optimistic view of the outcomes of XDR practices, it is nevertheless indicative of the very high stakes ascribed to these practices in policymaking for science and innovation.

2.2. Fostering of Interdisciplinary Research Practices Through Funding Instruments and Other Policy Interventions

Of the assumptions that underpin the research presented here, perhaps the most fragile is the one that policy interventions can foster increased interdisciplinarity in the research groups they target. For instance, many studies have documented trends towards increased interdisciplinarity in research, but without specifically linking this shift to policy interventions (Dworkin, Shinohara, & Bassett, 2019; Okamura, 2019; Porter & Rafols, 2009) or, as we just did above, they note the multiplication of interdisciplinarity initiatives and narratives originating from policymakers. If proved effective, these initiatives could be very important in fostering cross-disciplinary research, as there is quantitative evidence demonstrating that traditional grant mechanisms tend to be conservative and to shy away from cross-disciplinarity. Bromham and colleagues, examining the interdisciplinarity and multidisciplinary (defined here in the same terms as we did above) intensity of proposals to the Australia Discovery grants, found that interdisciplinarity in proposals was “consistently negatively correlated with funding success.” Multidisciplinary was positively correlated with peer-review scores but at a very small magnitude (Bromham, Dinnage, & Hua, 2016).

Most of the restricted body of work on the policy mechanisms through which funding interventions foster interdisciplinary research is qualitative and based on case studies, often resulting in recommendations for the management of these programs (Lyll, Bruce et al.,

2013; Molas-Gallart, Rafols, & Tang, 2014). Elsewhere, program evaluations have used peer-review panels to assign scores to projects or initiatives and therefore measure achievements in interdisciplinarity in a semiquantitative manner (European Research Council, 2018).

Lyall and colleagues find that despite high-profile initiatives and policy exhortations to engage in interdisciplinarity, transdisciplinarity, and/or knowledge transfer, still only a modest volume of STI policy practices in the United Kingdom meaningfully engage with these approaches in practice (Lyall, Meagher, & Bruce, 2015).

Porter, Garner, and Crowl (2012) have provided one of a few specific evaluations of a policy instrument's effect on levels of interdisciplinary integration within supported scientific projects. The authors characterized the set of publications originating from the US NSF Research Coordination Network. This program aimed to foster novel research networks around interdisciplinary intellectual projects. They found this program to have succeeded in achieving high networking and interdisciplinarity metrics in related papers, although the authors noted that successful applicants to the program already displayed higher scores on these dimensions prior to the support period in comparison to unsuccessful applicants. Similarly, Science-Metrix, using a difference-in-differences approach (the control group was selected using a regression discontinuity design), quantitatively demonstrated a positive association of one of the HFSP's funding mechanisms (i.e., cross-disciplinary fellowships) on the level of interdisciplinarity achieved by its awardees (Science-Metrix, 2018). While both the awardees and control group scored highly prior to funding, HFSP funding appeared to have enabled the former group to maintain its level of interdisciplinarity during funding, whereas this was not the case for the latter. A sustainable and positive effect was also perceptible after funding for awardees who did increase their score by a greater margin than the control group by that time. However, the authors noted the lower reliability of the findings for this group given its size, and most of the other HFSP funding mechanisms did not appear to further increase the level of interdisciplinarity of the awardees. Still, HFSP stood out well relative to other funders for the overall interdisciplinary level of its supported papers.

2.3. Bibliometric Characterization of XDR Research Practices

While Porter, Garner, and Crowl (2012) provide one of the few available examples linking quantitative measures of XDR to funding program interventions, the field of bibliometrics has produced multiple measurement strategies and indicators with an aim to quantitatively assess the degree of XDR achievements in research publications. One important stream of bibliometric studies on XDR practices has emerged in the last 15 years (Stirling, 2007). Initial studies in this stream shared a few core methodological parameters:

- the integration of three core dimensions (variety, balance and disparity) of XDR through the Rao-Stirling formula
- characterizing the disciplinary location of cited references in the publication set of interest as the main bibliometric phenomenon of interest (and used a proxy for XDR knowledge integration)
- common use of Web of Science (WoS), All Science Journal Classification (ASJC), or national evaluation categories as the reference classifications against which to map the diversity of disciplines encountered (but see Rafols and Meyer (2010) for an early bottom-up, citation-network driven approach)

Excellent overviews and summaries of this research stream already have already been produced, notably in Abramo, D'Angelo, and Zhang (2018); we will not recapitulate them here.

Keeping foundational insights from the “Rao-Stirling stream” but also integrating outcomes from important recent studies that have sometimes operated outside this stream yields, in our view, four main strategies to move forward:

- exploring bibliometric phenomena to be used as proxies for attributing disciplinary locations to publications beyond cited references, including:
 - o disciplinary diversity in author affiliations (Abramo et al., 2018; Zhang, Sun et al., 2018; Zuo & Zhao, 2018)—although the roots of this approach predate Rao-Stirling-centered methods)
 - o disciplinary diversity in author prior publications (Moschini, Fenialdi et al., 2020; Zuo & Zhao, 2018)
 - o disciplinary diversity in author prior publications’ cited references (Moschini et al., 2020)
 - o disciplinary diversity in citations received by the publication set of interest (Moschini et al., 2020)
 - o disciplinary diversity in topical clusters of concepts retrieved in publications’ texts (Hackett et al., 2021; Zuo & Zhao, 2018)
- using undirected clustering strategies to create emergent classifications of publications, notably when measuring XDR intensities through citation links or textual clusters (Hackett et al., 2021; Zhang et al., 2018; Zuo & Zhao, 2018); but also Rafols and Meyer (2010)
- assigning publications or researchers to vectors of relative disciplinary engagement (cutting across all categories in a classification) rather than to single disciplines or categories (Adams, Loach, & Szomszor, 2016; Zuo & Zhao, 2018)
- considering of mathematical alternatives to Rao-Stirling to calculate a composite indicator of intensity in XDR practice, including “div,” or the interpretation and analysis of the constituent dimensions of the Rao-Stirling indicator individually (Hackett et al., 2021; Wang, Thijs, & Glänzel, 2015)

2.4. Prior Evidence on Improved Societal Impacts for Cross-Disciplinary Research

By the early 2010s and onwards, there appeared to be “a consensus in the literature that socially relevant research is most often interdisciplinary” (Chavarro et al., 2014). Despite this consensus, broad scope quantitative evidence on the capacity of cross-disciplinary research to produce improved societal outcomes was still sparse (Rylance, 2015). Interdisciplinarity may have become somewhat conflated with the notion of intersectoral collaboration or engagement, which is a precondition of knowledge and technology transfer, the latter themselves being clear instances of societal impact. The vast literature on technology transfer, academic entrepreneurship and “mode 2” research may have underpinned the emergence of a consensus on the societal relevance of interdisciplinarity. Still, there is surprisingly little in the way of overt, generalizable evidence to support this collective assumption, especially in a way that applies to multiple pathways and modalities of interdisciplinary practice.

To review the literature on the contributions of academic entrepreneurship and mode 2 research practices to societal outcomes would be out of scope in the current brief. The research that has focused on a stricter definition of interdisciplinarity has, for its part, mostly relied on case studies. Disciplinary diversity in researchers’ background has been found to be associated with increased chances to engage in entrepreneurship and technology transfer (Deste, Mahdi et al.,

2012). Qualitative research has also shown that stakeholders and users in interdisciplinary projects share the perception that the approach is conducive to generating useful outcomes for these stakeholders' problems, although the extent to which these perceptions were realized was highly dependent on the type of strategies used (Molas-Gallart et al., 2014).

On the quantitative side, Chavarro and colleagues found that, in a set of WoS publication records with at least one coauthor from Colombia, papers with higher scores on certain (but not all) dimensions of interdisciplinarity that they measured were indeed associated with a greater orientation towards local issues (Chavarro et al., 2014). Campbell and colleagues also found that the odds of research uptake in the patent literature was positively and significantly related to the multidisciplinary of research teams on scientific papers, accounting for field of research and a number of additional variables (Campbell, Struck et al., 2017). Wang and Li (2018) found similar results looking at the effect of the scope of integrated knowledge on uptake in patents.

2.5. Altmetrics to Measure Societal Outcomes of Research

In the decade spanning 2010–2020, a novel quantitative research evaluation tool emerged with the launch of databases recording the uptake of journal-based (or proceedings-based) scientific outputs in social media, blogs, news, and educational resources, among other sources. These data, because they are hoped to track usage beyond academic circles as traditionally captured in bibliometric indicators, are often referred to as *alternative data* (or *altmetrics*). Included in the databases' coverage are platforms such Facebook and Twitter, a selection of blogging platforms, journalistic and news websites, Wikipedia, Reddit, Stack Exchange, and library holding databases. These mentions are usually tracked through document identifiers such as DOI, PMID, and the URL of the article.

New altmetrics approaches continue to emerge, as in the case of the Overton policy database that will be deployed in the empirical component of this study. Arguably, other analytical strategies, such as examining citations to scientific publications from patent or clinical guideline records can also be included within the broader definition of altmetrics, especially as the field relates to broad societal outcomes of research (Tahamtan & Bornmann, 2020).

The value of altmetrics mentions to journal articles is that they may capture degrees of readership, uptake, and engagement in an audience that is theoretically not restricted to peers. Such findings could in principle be obtained at scale for a fraction of the levels of effort typically required by qualitative approaches. Expectations for the contributions of altmetrics to decision-making and evaluation have been high, as illustrated by the contentions of an expert group on altmetrics recently convened by the European Commission:

Altmetrics also have potential in the assessment of interdisciplinary research and the impact of scientific results on the society as a whole, as they include the views of all stakeholders and not only other scholars (as with citations). Hence, altmetrics can do a better job at acknowledging diversity (of research products, reflections of impact etc.), providing a holistic view of users as well as providers of scientific products, and enhancing exploration of research results (European Commission Expert Group on Altmetrics, 2017, p. 11).

Further, the same group summarizes the potential advantages of altmetrics as broadness (inclusion of multiple stakeholder types), diversity (type of outputs measured), multifaceted (different signals for a given output), and speed (readership of an article typically taking place faster than the uptake of its findings in ulterior research).

Like citation counts, observations on altmetric mentions or interactions can be processed in multiple ways to compute different indicators. Again, as in computations with citation data, altmetric observations have been shown to be shaped by varying disciplinary features, temporal trends, and database coverage biases, meaning that raw volume counts are almost never useful (Thelwall, 2016). Basic normalization procedures used for citation indicators can also be applied for altmetric indicators (normalization by subfield and by year). Thoughtful interpretation of altmetrics findings should also consider several limitations identified in the specialized literature.

Finally, we note that most altmetrics-based strategies are geared towards the capture of broader societal attention towards scientific publications issued in journals or conference proceedings. Yet it could be argued that nonscientific or hybrid outcomes are increasingly becoming the focus of transdisciplinary, coproductive, or locally oriented research projects (Koier & Horlings, 2015). Altmetrics approaches have still to be convincingly deployed for these kinds of outcomes, and the collective amount of explorative effort conducted to try and do so has been low. Assessment of these kinds of outcomes must—until large-scale efforts for their indexation materialize—make use of qualitative, expert review, or survey methods.

2.6. Measuring Citations from Policy Documents to Measure the Outcomes of Research on Decision-Making

The capacity of a given altmetrics research strategy to effectively capture societal outcomes of research is closely associated with the basic features of the phenomenon recorded through its main data source (Haustein, 2016; Tahamtan & Bornmann, 2020). Altmetric data on Twitter has been shown to be of restricted relevance for capturing deep knowledge transfer or public engagement processes, and instead to capture online buzz around publications. Pulido and colleagues conducted in-depth examinations of the content of Twitter and Facebook posts on scientific articles, with an aim to determine whether these posts provided evidence of societal change achieved through the research (rather than online discussion and interaction strictly) (Pulido, Redondo-Sama et al., 2018). They found that this was only the case in 0.5% of social media mentions to more than 5,000 journal articles from EU-funded projects. Data sets on clinical guideline and patent citations towards publications, on the other hand, are considered to have good precision in measuring activities that are components of important knowledge transfer processes (Thelwall & Kousha, 2015).

There are reasons to argue that measurements of citations in policy documents towards scientific publications would, in principle at least, also feed into a precise indicator of societal outcomes of research. While policy citations, just like scientific citations, are likely to be practiced for a variety of reasons, governmental or quasigovernmental use of scientific results in the formulation and implementation of public policies is widely regarded as a desirable research outcome. This mechanism of knowledge transfer has also been observed and studied before (Bornmann et al., 2016; Tahamtan & Bornmann, 2020).

Preliminary work from a handful of studies that have used the *Altmetric.com* database's policy citation records makes it possible to infer some of the basic features of policy documents as a source of altmetric information. To our knowledge, no work has been produced yet on other altmetric databases covering policy documents, such as Overton.

Haunschuld and Bornmann (2017) have examined policy document citations from the *Altmetric.com* database to a publication set consisting of more than 11.25 million WoS indexed articles issued between 2000 and 2014. They found 0.32% to have at least one policy citation. The set of papers for the year 2005 displayed the highest share of policy citation

(almost 0.5%), indicating potentially much longer lags from publication to citation peak year in comparison to citations from other journal articles. Publication sets in the fields of Agricultural Economics and Policy (2.97%), Tropical Medicine (2.64%) and Economics (2.18%) had the highest chances of receiving policy citations.

Bornmann et al. (2016) examined policy citations from the *Altmetric.com* database towards records in their custom-built set of more than 190,000 papers on climate change. A share of 1.2% of these papers had at least one policy citation in *Altmetric.com*. Of these papers, 78.7% received only one policy citation. The authors found citation peaks to occur between two to 4 years after publication, but those documents with the highest levels of policy citations had citation peaks occurring later than the overall figure.

Tattersall and Carroll (2018) considered policy citations to journal articles published by the University of Sheffield. They report a share of 1.41% of the overall Sheffield publication set to have been cited by at least one policy document. The disciplinary distribution of these citations very much followed what has been reported above for other policy citation studies and studies that capture other altmetric dimensions. Much like in bibliometrics generally, citation distributions were also highly skewed, with only a few articles achieving citation counts above 1. One finding from this team is worrisome: Manually validating 21 policy citations to University of Sheffield articles, they found seven for which attribution to the University of Sheffield or to the Sheffield article was problematic. Another finding from this study that acts as a call for caution is that there were a number of duplicate policy citations in the Sheffield set, sometimes because individual chapters of a full policy report are published separately. Additionally, some of the policy citations were found to originate in journal articles rather than actual government reports.

Newson and colleagues have used a “backward tracing” approach to understanding policy citations, starting from policy documents and trying to characterize how they use citations. They selected a number of Australian policy documents relating to the topic of childhood obesity. These 86 childhood obesity policy documents made 526 unique references to topically relevant research content, of which half were peer-reviewed publications and a fifth were nonpeer-reviewed research publications. They concluded that in many cases (they did not compute a share of the overall citation data set), textual context for the citations does not make it possible to unambiguously attribute impact on the policy process for the research findings cited. As in citations within the scientific community, the purposes and intentions for making a citation appeared diverse. The authors also found multiple instances of mentions to research that were not accompanied by an attendant formal citation, concluding that “[c]itation rates are likely to provide an underestimation of research use by policy agencies and the method has the potential to miss research that was in fact impactful, and place undue importance on cited research” (Newson et al., 2018, p. 10).

3. METHODS

3.1. Data Sources

To perform this study, lists of publications produced through FP7- and H2020-supported projects were obtained from OpenAIRE (<https://www.openaire.eu/> (FP7)) and CORDIS (<https://cordis.europa.eu/projects> (H2020)) and matched to Scopus and Overton. Choosing papers that could be matched to specific grants made it possible to control for some funding characteristics in modeling the relationship between XDR and UPRL (see Section 3.3).

Scopus is a global repository tracking publication of peer-reviewed articles and other scientific communications. The match of FP7- and H2020-supported papers to Scopus was based

on the digital object identifiers that were available in the lists of publications and complemented by a fuzzy matching algorithm building on information such as the author names, publication year, and title.

Overton—a novel database established with an explicit goal to increase the coverage and comprehensiveness of policy-focused altmetrics—was used to track the uptake of publications in a range of policy documents. Overton records are built from combining a broad panel of government sources with web crawling. The base list of governmental sources in Overton includes a long tail of repositories with just a few documents each. The database indexes more than two million policy documents produced by national governmental entities, international governmental organizations and think tanks. While the UPRL dimension presented here was measured mostly from citations towards peer-reviewed publications originating in evidence syntheses, scientific advice reports, and some forms of grey literature documents, Overton does index legislative and executive documents, including governmental white papers or transcripts of parliamentary sessions. While close to 75% of these records are provided by US, UK, and intergovernmental sources, the database also contains more than 100,000 entries from Japan and 70,000 from Germany, to take just some examples. Overton coverage extends to the year 2020. FP7- and H2020-supported papers in Scopus were subsequently matched to Overton using their DOIs.

3.2. Assessing Whether Citations to the Scientific Literature in Overton Reflect Research Input into Policymaking (Question A)

Despite the shortcomings generally identified for altmetrics approaches, uptake in the policy-relevant literature (UPRL) stand as arguably the next best candidate to sit alongside patent citations and clinical guideline citations within the upper tier of comparatively reliable quantitative indicators of societal outcomes (Wilsdon, Allen et al., 2015). Policy mentions stand a high chance of capturing a well-defined societal impact in the form of a scientific contribution to evidence-based policymaking (Bornmann et al., 2016). Citation in policy documents was one of eight indicators rated as highly important for the evaluation of societal outcomes by the stakeholders consulted by Willis, Riley et al. (2017).

Prior studies on policy citations used the *Altmetric.com* database and uncovered a number of issues (see Section 2.5) worth assessing in the context of the new Overton database, which was used for this study. Using a random sample of 50 FP-supported publications cited by documents indexed in Overton, we qualitatively assessed the extent to which UPRL reflect research input into decision making to address Question A. When some of these publications registered more than one citation from a policy document in Overton, only the first citation was assessed. The original, citing, document was retrieved and reviewed to validate the existence and locate the citation to the peer-reviewed publication of interest; assess the overall character and content of the citing document (executive or legislative document; grey literature; affiliations of its authors; publishing organization, etc); and assess referencing practices (format and presentation of citations made, as well as apparent motivations for making a citation) in the documents of interest. Finally, the lag to UPRL and the share of publications with at least one such citation were also computed.

By way of additional validation and characterization of the Overton database, and in collaboration with the Overton team, additional measurements of UPRL, beyond FP7- and H2020-supported papers, were taken across the 174 subfields contained in the Science-Metrix classification (Archambault, Beauchesne, & Caruso, 2011; Rivest, Vignola-Gagné, & Archambault, 2021) in Scopus. This was achieved by taking a random sample of 1,000 peer-reviewed publications

(mostly articles, reviews, and conference papers) in each of those subfields. The samples were restricted to papers published between 2008 and 2016 for comparability with our core data set of FP7- and H2020-supported papers (see Section 3.3). Post 2016, the citation window would also be quite short (see Section 4.1). These publications' DOIs were then queried against the Overton database to retrieve information on their UPRL. The Overton team returned, for each paper in the provided samples, a link to the Overton records that cited it. This data set enabled estimating the degree of UPRL, of publications in the full Scopus database as well as by scientific subfield. Finally, additional qualitative validations of the citation contexts recorded in Overton were also conducted in two random samples within this ancillary experiment: in 50 randomly chosen citation records irrelevant of subfield; and in 50 randomly chosen records for the specific subfield of Development Studies, given the high uptake measured in this specific case.

3.3. Modeling the Relationship Between XDR and UPRL (Question B)

The outcome variable was coded as 1 for papers cited at least once in Overton records and 0 otherwise. The data set was restricted to publications produced until 2016 (inclusively) allowing, at a minimum, for a 4-year policy citation window (publication year plus three). This choice balanced the need to maximize the number of observations with information on the lag from publication to eventual UPRL (see Figure 1 in Section 4.1). Statistical models were tested with and without the publication year as a control, knowing that older papers have a higher chance of having been cited in a policy document. The final data set contained 126,441 papers published between 2008 and 2016. These papers were paired to FP7/H2020 project funding, sometimes more than one, resulting in ~137,000 observations.

Cross-disciplinarity at the paper level was captured through two lenses: DDR (tracks diversity of integrated knowledge) and DDA. The former is equivalent to the integration metrics of Porter and Rafols (2009) relying on Science-Metrix's classification of science (Archambault et al., 2011) to classify a paper's cited references by subfield. While the original version of the classification used journal-level categorization, an updated, hybrid (journal- and article-level) version of the classification has been used here (Rivest et al., 2021). This updated version notably individually redistributes articles in generalist journals into the classification categories with support from deep learning approaches.

The DDA indicator measured diversity as reflected in the prior disciplinary background of a paper's coauthors (team multidisciplinary). Authors were disambiguated using Scopus author IDs, which produce reliable results at scale (Campbell & Struck, 2019). Science-Metrix subfields were assigned to authors based on their prior publications. DDA was designed to increase for teams involving authors from different subfields, particularly where these subfields are not frequently connected in Scopus. This was achieved by adapting the metrics by Porter & Rafols to the disciplinary profile of coauthors (see the supplementary materials for more details on the computation of DDA and DDR). DDA and DDR were normalized by subfield to avoid coverage biases (Campbell, Deschamps et al., 2015). Other diversity indicators were computed, including the share of women authors, number of authors, and number of countries.

The model's specification accounted for additional characteristics of a paper: subfield/year-normalized (intraresearch) citation counts and CiteScore, document type, and average number of prior papers per author. It also accounted for specific characteristics of the papers' research projects that are absorbed by the fixed effects (e.g., research teams' proximity to policymaking, amount funded, main topic of interest).

UPRL being measured as a binary variable, this justified the choice of logistic regression to test the associations between bibliometric variables and UPRL. The estimated coefficients linked the explanatory variables to the odds that a scientific publication impacts policy. The logit function is expressed as

$$\log\left(\frac{p}{1-p}\right) = \alpha_k + \sum_1^q \beta_j x_j + u$$

Conditional logistic regression (Agresti, 2012) was used because papers from the same project may be more similar. Roughly, this specification allowed for each research project to have a different baseline impact on policy. In the equation above, this is represented by the subscript k in the α intercept, which was allowed to be different for each research project.

This model is more suited to accounting for the unobserved characteristics of different projects that could affect estimates in “usual” logistic regressions (if we were able to measure them). For example, the chances of being cited in written policy may be higher if the lead of a research team actively collaborates with policymakers.

Different transformations of the same indicators were tested across different models, to test the robustness of the signal and significance of the estimates and to allow for a better assessment of the effect sizes of each variable. First, the logarithmic forms of highly skewed explanatory variables were used to reduce the effect of outliers (normalized citation counts, normalized CiteScore, DDA, DDR, number of authors/countries, and average number of papers per author). The remaining less skewed variables were not transformed. The second model was based on the original form of all variables. In the third specification the explanatory variables were divided by their standard deviations (except for the variable document type). Therefore, the odds ratios from this specification refer to the changes in the odds of UPRL associated with one standard deviation in the explanatory variable. This allowed for comparisons across variables in Model 3, although caution is advised, as changes of one standard deviation in highly skewed variables are usually less likely. The standardized model (i.e., third specification) was the base of the main observations regarding effect sizes in this study. Further alternative specifications were used to test the robustness of the study’s findings (see Section 4.2.2).

4. RESULTS

4.1. General Statistics on the Level of, and Lag to, UPRL, as well as an Assessment of the Citation Context (Question A)

Prior work reported between 0.32% and 1.41% of publication sets being cited by at least one policy document (Bornmann et al., 2016; Haunschild & Bornmann, 2017; Tattersall & Carroll, 2018). Our results strongly contrast this work, with a figure of 6.0% for the entire data set of FP-funded publications, 8.6% for the subset limited to UK publications, and 5.1% for the subset of nonUK publications. The difference between the three groups could reflect the coverage bias of Overton in favor of the United Kingdom.

While other studies could not provide robust data on intervals to policy citation peak (in terms of share cited), results from our global data set of FP7- and H2020-supported papers shows that this peak may take place around the third year after publication year for 2008–2011 papers (dotted line, Figure 1); for 2012–2015 papers, the peak was between the second and third years after publication year (data not shown). About 50% of the papers cited in policy had received their first citation 3 years after publication year (solid line, Figure 1). These findings also hold true for the subset of papers from the United Kingdom. This suggests that

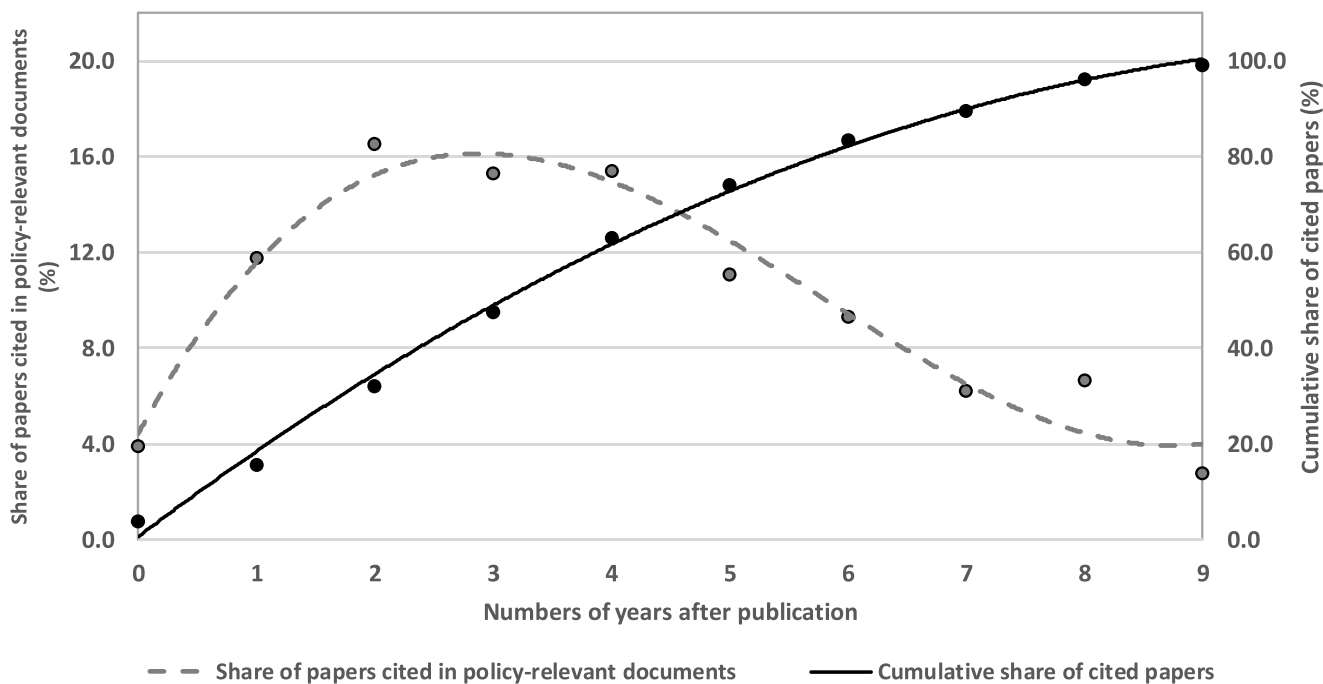


Figure 1. Uptake in policy-relevant documents of FP7- and H2020-funded papers by number of year(s) after publication, 2008–2011 (source: Scopus and Overton).

policy altmetrics could find practical use in midterm or ex post (near the end as opposed to well after) program evaluations, to a greater extent than previously envisaged.

The qualitative assessment of a subset of UPRL found that none of 50 UPRL citations were a “technical false positive,” that is, a mention to the cited publication that could not be retrieved in the original, citing, policy-relevant document. A group of six Overton records of UPRL could be considered “conceptual false positives,” in that the original citing documents were found to really be scientific publications rather than grey literature reports or governmental white papers. These were copies of articles made available online on institutional websites rather than on journal websites. An additional four citations were made from “hybrid” documents authored by academic authors but published by governmental or think tank organizations and whose content was judged to be very similar to that of formal peer-reviewed research output. The remaining 40 citations originated from policy documents that must be considered part of the “regulatory science” or science advisory branches of governance systems. Most of these documents appeared to have been authored by government scientists (sometimes in collaboration with academic scientists or scholars). Many consisted of syntheses and reviews of research findings without explicit conclusions for policy formulation or implementation. So, while these citations should not be interpreted as indicative of advanced policy outcomes of research directly reaching the legislative or executive processes, they can be seen as achievements in contributing to the first stages of these processes, at the intersection between governance and academia. Within the 44 citations that can be considered valid, only four did not make a clear reference to the cited publication in the body of the document’s text. These made use of the reference to provide prior findings, or support theory- or method-building. Finally, only in six cases was the reference made as part of a grouped citation containing multiple

references (four or more citations). Examples of policy-relevant documents found in the sample for the qualitative analysis included:

1. multiple IPCC reports, including an IPCC expert testimony before the U.S. House of Representatives Select Committee on the Climate Crisis
2. a “literature review and horizon scanning” from the U.K. Animal and Plant Health Agency
3. a EC Directorate-General for Employment, Social Affairs and Inclusion research brief conducted by a researcher at the London School of Economics and Political Science
4. a position report from the Ghanaian think tank IMANI Center for Policy and Education
5. an OECD country-level analysis for Sweden on immigration and diversity issues
6. an evidence synthesis written by an international team of researchers for the Arctic Council.
7. an evidence synthesis written for Interpol forensic science managers by a mixture of university researchers and forensics professionals
8. a World Health Organization white paper titled *The global plan to stop TB 2011–2015: transforming the fight towards elimination of tuberculosis*
9. the World Meteorological Organization report *Seamless prediction of the Earth system: from minutes to months*
10. a workshop report on legal, ethical and societal issues related to *Human enhancement and the future of work*, jointly organized by the Royal Society, the Academy of Medical Sciences, and the British Academy. Rather than presenting the presentations made by researchers at the workshop themselves, the committee members (both academics and nonacademics) synthesize observations and arguments across around themes of their own invention

Results on the share of publications with at least one UPRL citation within the random samples of the Scopus database show that the results described above for FP7- and H2020-funded papers likely hold in a broader context. By aggregating (using a weighted average) the results obtained with the random samples taken across all subfields, it was estimated that 5.8% (stability intervals of 5.7%–5.9%) of Scopus records between 2008 and 2016 received at least one UPRL citation in Overton. However, results varied greatly by year as well as by Science-Matrix subfield or domain. Figure 2 illustrates this observation for the main domains in the Science-Matrix classification. Table S5 in the supplementary materials shows that the share of publications with at least one UPRL citation range from 47.0% for Development Studies to 0.1% for Mathematical Physics and Drama & Theater.

Here again, qualitative validation of Overton citations indicates that these measurements somewhat overestimate the proportions of peer-reviewed publications to achieve UPRL. In the sample of 50 random publications taken from all subfields, there was one technical false positive (the cited peer-reviewed publication could not be located in the policy-relevant document), and 12 conceptual false positives. These conceptual false positives included many borderline cases that are unlikely to be correctly classified by automated means alone, including

- three instances of PhD dissertations conducted seemingly in collaboration with governmental agencies, posted on these agencies’ websites and categorized as policy documents by Overton
- two UK NICE medical guidelines where the citation was traced back to the “excluded studies” section of the bibliography
- cases of governmental institutes listing their own publications within their annual reports

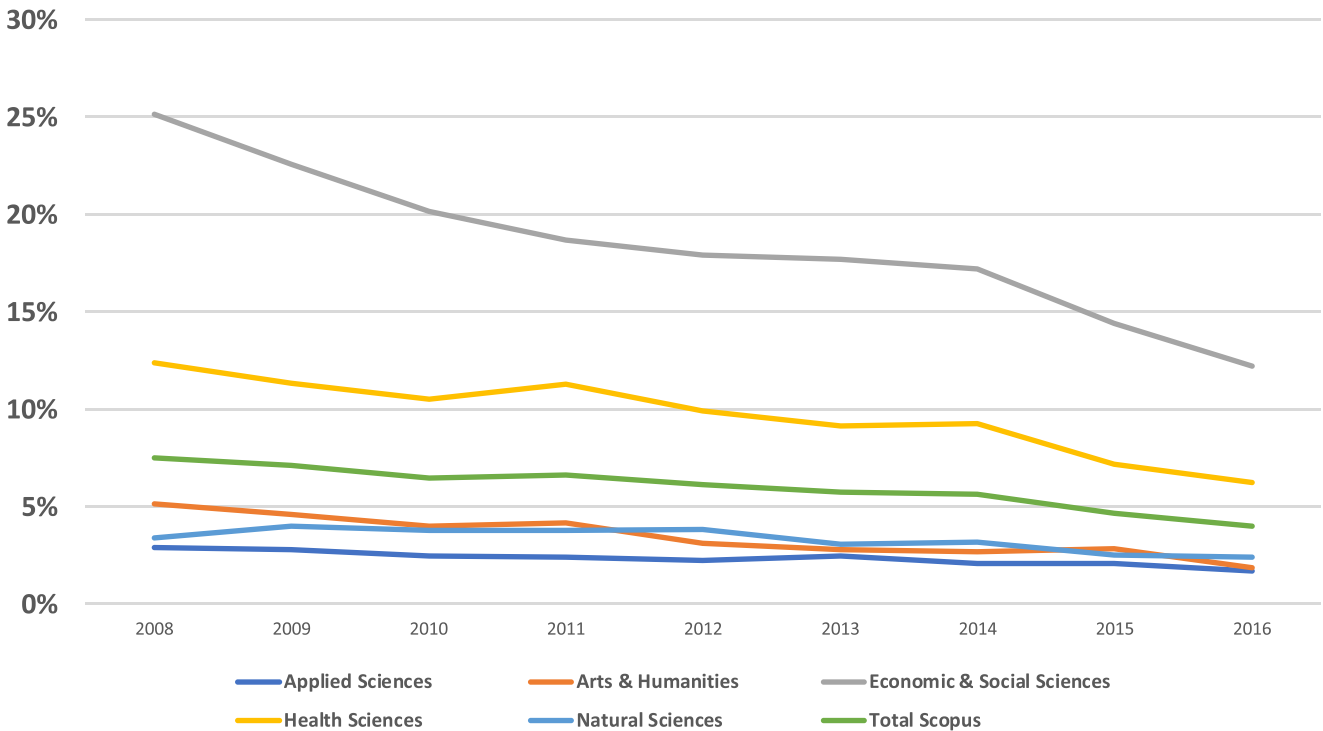


Figure 2. Estimated share of publications with at least one UPRL citation in Overton for the whole of Scopus as well as research domain, 2008–2016 (source: Scopus and Overton).

We would suggest that in some research or evaluation designs, these instances (most notably cases of PhD dissertations that could be confirmed to have been expressly conducted with a clear aim for application in policymaking contexts) may even be considered as true positives.

In the subset of UPRL citations to Development Studies publications, we have 41 validated entries. The other records included instance of a policy-relevant document no longer being available online and disallowing the validation attempt; one technical false positive; and seven conceptual false positives.

4.2. Modeling the Link Between Cross-Disciplinary Research and Uptake in Policy-Relevant Documents (Question B)

4.2.1. FP-supported papers with at least one UK-based author: Main specifications

Table 1 summarizes the main results from statistical modeling for the sample of FP7- and H2020-funded papers with at least one UK-based author. Coefficients were mostly significant at an alpha of 0.01, while some were only so at 0.05. Assessment of effect sizes in terms of probability is challenging, even when using the standardized odds ratios (Model 3). The effect on the probability of being cited depends not only on the change in the explanatory variable, but also on the initial value of the probability (i.e., the relationship between the odds ratio and variation in probability is not linear). Here, we focused on the effect sizes of the standardized coefficients (Model 3) for initial values roughly corresponding to our sample’s share of papers cited in policy documents (i.e., a baseline probability of 10%).

Using this standard, this paper mostly focuses on DDA and DDR coefficients, as they are the ones directly related to our core research question, the others being included as controls.

Table 1. Results of logistic regression for UPRL, of FP-funded publications with at least one UK-based author, 2008–2016 (source: Scopus and Overton)

Variable	Logistic regression for variables in logarithmic form and in level						
	Log-transformed (Model 1)		Level (Model 2)		Level – for one standard dev. (Model 3)		
	Coefficients	Odds ratio	Coefficients	Odds ratio	Coefficients	Odds ratio	Δp (baseline = 10%)
Normalized citation score	0.960*** (0.035)	2.613	0.086*** (0.006)	1.09	0.613*** (0.043)	1.846	7.0%
Normalized CiteScore	0.024 (0.063)	1.024	0.153*** (0.021)	1.166	0.207*** (0.029)	1.23	2.0%
Document type (Review = 1)	0.727*** (0.101)	2.07	0.507*** (0.096)	1.661	0.507*** (0.096)	1.661	5.6%
Multidisciplinary (DDA)	0.234*** (0.058)	1.264	0.109*** (0.026)	1.115	0.129*** (0.031)	1.138	1.2%
Interdisciplinarity (DDR)	0.017 (0.125)	1.017	0.07 (0.089)	1.073	0.032 (0.041)	1.033	0.3%
Number of authors	−0.111** (0.048)	0.895	−0.006*** (0.001)	0.994	−0.780*** (0.176)	0.458	−5.2%
Number of countries	0.163*** (0.063)	1.178	0.074*** (0.011)	1.077	0.279*** (0.040)	1.322	2.8%
Proportion of female authors	0.300** (0.152)	1.35	0.152 (0.143)	1.164	0.033 (0.031)	1.033	0.3%
Avg. number of papers per author	0.107*** (0.040)	1.113	0.003*** (0.001)	1.003	0.115*** (0.039)	1.122	1.1%
Observations	37,897		37,897		37,897		
Log Likelihood	−3,446.58		−3,754.04		−3,754.04		
Wald Test (df = 9)	1,046.660***		528.910***		528.910***		
LR Test (df = 9)	1,345.617***		730.702***		730.702***		
Score (Logrank) Test (df = 9)	1,316.022***		766.147***		766.147***		

Note: Binary logarithms of normalized citation counts, normalized CiteScore, DDA, DDR, number of authors/countries, and average number of papers per author were used in Model 1. Therefore, the odds ratio of these coefficients refers to the variation in odds associated with a twofold change in the explanatory variables. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Results for DDA suggested that bringing together authors from different subfields of science is positively associated with UPRL, with an associated increase in the probability of 1.2 percentage points for each additional standard deviation in this variable (Model 3, Table 1). DDR was not statistically associated with UPRL after controlling for DDA, but it was without controlling for DDA (Table 2, Model 7) (see further discussion of this result in Section 4.2.2). The models, with or without publication year as a control, provided similar results (Supplementary material Table S4 Model 15). Controlling for authors' seniority, prior publication visibility (through the CiteScore indicator), and prior publication citation impact likewise did not substantially alter the previous results (Supplementary material Table S4 Model 16).

4.2.2. Alternative model specifications

Table 2 expands the previous analysis, reporting on the findings from statistical models similar to Model 3 from Table 1. Models 4 and 5 reproduced Model 3, but for different sets of papers (see Table S3 in the supplementary material for odds ratios):

- Europe (Model 4), corresponding to all FP-supported papers, regardless of the authors' affiliation countries
- Europe nonUK (Model 5), corresponding to all papers in the data set except those having at least one UK-based author (i.e., those in the data set for Model 3)

The coefficients for DDA and DDR showed the same signs and statistical significance in the three models. However, the lower coefficient observed for DDA in Model 5 (non-U.K. authors) suggests that multidisciplinary collaboration had less importance in driving UPRL for nonUK publication output, or that the coefficient in this model was affected by the lower coverage of Overton outside the United Kingdom. This latter hypothesis may very well be at play considering the lower share of papers cited in Overton for non-U.K. papers (5.1%) compared to U.K. papers (8.6%).

The coefficients of most of the remaining variables were also comparable across the three data sets used in Models 4 and 5 (Table 2). In two cases (number of authors and average number of papers per author), the coefficients in the non-U.K. data set were no longer statistically different from zero. These two indicators were included as control variables with no prior expectations regarding their signs. As with DDA, these differences may have reflected cross-countries differences in the coefficients or may have resulted from differences in coverage among different countries. The fact that none of the coefficients presented different signs in these different data sets pointed to some degree of robustness in this indicator as one way to capture UPRL.

Models 6, 7, and 8 (Table 2) are based on the same set of U.K. authors from Table 1. However, they were set up to provide additional insights into the association between each variable of XDR (DDA or DDR) and UPRL. The first point to be highlighted is that the association between DDA and UPRL is less dependent on model choice. This association was positive and significant whether DDR was controlled for or not, and in other models provided in the supplementary material (Table S4). The exception was for the model that included an interaction term for DDA and DDR (Table S4 Model 9). This interaction variable, however, introduced excessive collinearity in the model, resulting in estimates that were not statistically significant for DDA, DDR, or DDA*DDR¹.

¹ The variance inflation factors (VIF) in the model with the interaction term were: 6.4 for DDA*DDR, 5.2 for DDA, and 1.68 for DDR. The high VIFs for DDA and DDA*DDR explained the lack of statistical significance for these terms in this model.

Table 2. Results of logistic regression for UPRL of FP-funded publications, 2008–2016: alternative specifications (source: Scopus and Overton)

Variable	Logistic regression for variables in level Models for one standard deviation					
	UK base Model 3	EU Model 4	EU nonUK Model 5	UK_a Model 6	UK_b Model 7	UK_c Model 8
Normalized citation score	0.613*** (0.043)	0.604*** (0.022)	0.655*** (0.028)	0.615*** (0.043)	0.618*** (0.043)	0.612*** (0.043)
Normalized CiteScore	0.207*** (0.029)	0.146*** (0.015)	0.101*** (0.021)	0.208*** (0.029)	0.209*** (0.029)	0.208*** (0.029)
Document type (Review = 1)	0.507*** (0.096)	0.722*** (0.054)	0.813*** (0.071)	0.489*** (0.095)	0.497*** (0.095)	0.510*** (0.096)
Multidisciplinary (DDA)	0.129*** (0.031)	0.101*** (0.016)	0.065*** (0.021)	0.142*** (0.029)		0.132*** (0.032)
Interdisciplinarity (DDR)	0.032 (0.041)	0.021 (0.021)	0.015 (0.025)		0.088** (0.038)	0.014 (0.042)
Number of authors	−0.780*** (0.176)	−0.635*** (0.076)	−0.008 (0.018)	−0.768*** (0.174)	−0.776*** (0.175)	−0.775*** (0.176)
Number of countries	0.279*** (0.040)	0.251*** (0.018)	0.106*** (0.017)	0.277*** (0.040)	0.295*** (0.040)	0.276*** (0.040)
Proportion of female authors	0.033 (0.031)	0.021 (0.016)	0.022 (0.020)	0.03 (0.031)	0.031 (0.031)	0.032 (0.031)
Avg. number of papers per author	0.115*** (0.039)	0.052*** (0.020)	0.035 (0.025)	0.119*** (0.039)	0.101*** (0.039)	0.118*** (0.039)
Fewer than 2 authors (Fewer than 2 authors) * DDR						−0.592* (0.305) 0.230* (0.120)
Observations	37,897	137,419	99,522	38,426	37,897	37,897
Log Likelihood	−3,754.04	−12,927.03	−7,756.73	−3,793.02	−3,762.98	−3,752.12
Wald Test	528.910*** (df = 9)	1,836.020*** (df = 9)	1,004.670*** (df = 9)	534.920*** (df = 8)	510.950*** (df = 8)	532.400*** (df = 11)
LR Test	730.702*** (df = 9)	2,473.643*** (df = 9)	1,413.255*** (df = 9)	738.604*** (df = 8)	712.811*** (df = 8)	734.541*** (df = 11)
Score (Logrank) Test	766.147*** (df = 9)	2,714.351*** (df = 9)	1,410.890*** (df = 9)	777.312*** (df = 8)	748.805*** (df = 8)	769.708*** (df = 11)

Model 7 revealed a positive association between DDR and UPRL when DDA is excluded from Model 3 (Table 2). Model 8 further investigates the association between DDR and UPRL, while accounting for DDA, to reveal conditions under which both XDR variables might be associated with UPRL. Briefly, a variable indicating papers written by one or two authors was interacted with DDR to estimate the correlation between DDR and UPRL within this group of papers, where DDA is not expected to be of major importance². The results showed that, at least for this group of papers (with one or two authors), the model was able to capture a positive association between DDR and UPRL³. It is possible that the existing correlation between DDA and DDR for papers involving a higher number of authors may be affecting the estimated effect of DDR on UPRL in Model 3. Within this hypothesis, the association between DDR and UPRL could not be measured for papers involving more authors due to collinearity between DDA and DDR.

Other specifications were tested to assess the robustness of the estimates for the relationship between XDR and UPRL. Except for the case described in the previous paragraph, the signal and coefficient of DDA remained stable in models including quadratic terms⁴ for DDA, DDR, number of authors, and countries. The results were also stable for models including year fixed effects (in the form of dummies for year of publication of the papers).

5. DISCUSSION AND CONCLUSIONS

5.1. Discussion of Main Findings

While prior studies using quantitative analysis of (what they called) policy citations towards research publications saluted the comparatively sound conceptual basis of the use of citations originating in policy documents and made towards peer-reviewed publications to monitor some of the societal outcomes of research, they found cause for caution in the current infrastructure available for its implementation. Here, the feasibility of this altmetrics approach to the quantitative measurement of societal outcomes of research was assessed in a large data set of publications (~117,000) resulting from FP7 and H2020 projects using Overton—a novel database focused on policy documents and on policy-related outcomes of research.

5.2. Question A

Following a preliminary quantitative and qualitative assessment of Overton data in this context (Question A), the new database appears as an important addition to the quantitative toolbox of altmetrics and other instruments for tracking such societal research outcomes. Indeed, at least one UPRL could be found for as many as 6% of publications from these EU-funded research projects using Overton, a much higher figure than reported in any of the previous studies using alternative data sources, although the publication sets used were admittedly quite differently constructed in each study (Bornmann et al., 2016; Haunschild & Bornmann, 2017; Tattersall & Carroll, 2018). In fact, this figure was as high as 42% and 72% in some subfields (Political Science & Public Administration; and Economics). The value of Overton in the altmetrics toolbox is also supported by the observation of numerous citation peaks between years 2 and 3 after publication, which means this data source shows potential for informing

² By definition, DDA is 0 for single-authored papers. The DDA of papers with two authors may be heavily impacted by copublications between professors and PhD students (which is prone to have low DDA).

³ With the effect estimated by adding the coefficient of DDR (0.014) and (one or two authors) × DDR (0.230) = 0.244; $s_e = 0.1168$, $p = 0.037$.

⁴ To test the possibility of nonlinear associations.

decision-making in a timely manner. Additionally, in an ancillary qualitative evaluation of the reliability of Overton data, it was found that the number of false positives in that database should be low, and that the motivations behind citation acts were generally clear and convincing.

This research has identified evidence synthesis reports, scientific advice reports, and other grey literature as the main type of citing documents retrieved in the Overton database. We certainly cannot argue that these documents capture in-depth societal or policy change that might have been set in motion by the findings contained in the peer-reviewed publications under study. Nevertheless, they likely indicate an incremental gain in the likelihood of policy impact above the cited peer-reviewed articles, an interpretation supported by studies that highlight policymakers' continued appreciation for knowledge syntheses and grey literature reports (Lawrence, 2018; Lawrence, Houghton et al., 2014). To capture definitive evidence of policy impact from research, qualitative or expert assessment methods remain necessary in the future, although it must be noted that even these exercises face their own set of difficulties and are by no means straightforward to conduct and interpret, nor inexpensive (Fowle, Wells et al., 2020).

We fully recognize the complex, nonlinear features of interaction between “evidence-making” and policymaking practices (Brownson, Chriqui, & Stamatakis, 2009). To us, this general argument motivates the use of quantitative proxy indicators that capture increased probabilities for policy–evidence interactions—complemented by careful interpretation—as a useful albeit imperfect addition to the toolbox of societal outcomes measurement. This is especially so given the fact that most executive and legislative documents should not be expected to make clear references, in the style of scientific citations, to specific studies. There are two reasons for this. First, policy impact is expected to derive from a collectively produced body of research and evidence than from any single project taken individually (Weiss, 1979). Secondly, even in those rare cases where legislative or executive change can be pinpointed to specific studies (Warner & Tam, 2012), the association is not often made in the relevant texts in a format that can be automatically retrieved through text mining or other techniques. Typically, mentions of academic research findings in a parliamentary document may take the form of an expert testimony by a participant in a relevant research project, but with no clear mentions toward discrete research articles.

We would argue that given the inherent difficulty of tracking science policy interactions, and given the shortcomings of prior strategies, the addition of even an imperfect quantitative strategy offers a net gain for the toolbox on measuring the societal outcomes of research. Levels of “science advice citations” can be triangulated with findings from qualitative methods, or can be used in mixed-methods strategies to focus more demanding investigations towards a subset of promising developments. On their own, UPRL measurements are likely to indicate groups of publications with a higher pull for policymakers. We contend that they are more likely to contribute to the aggregate body of knowledge that policymakers draw from during complex policy formulation and implementation processes.

5.3. Questions B and C

The assumption that XDR is more likely to foster broad societal outcomes than disciplinary research has been highly prevalent in current policymaking and research, but there is surprisingly little work that directly tests this relationship. In their basic formulation (Model 3), our regression models using the above sample of FP7- and H2020-publications showed that higher UPRL is associated with collaborative XDR (DDA) but not intellectual integration of

disciplines, narrowly defined (DDR). However, complementary analyses using alternative specifications showed that DDR has a positive link to UPRL when DDA is not accounted for, or when both DDA and DDR are included, adding an interaction term between DDR and a binary variable to identify papers with one or two authors.

As previously explained, high DDR can be achieved by single authors or small teams even though the likelihood of high DDR increases with DDA, which itself partly correlates with team size. The correlation between DDR and DDA would provide a potential explanation to the pattern observed in Model 3 (Table 1) where the effect of DDA could be encapsulating those of DDR, meaning that DDR could still have an effect in larger teams. Under this hypothesis, Model 8 (Table 2) would be able to show a relationship between DDR and UPRL for papers with one or two authors as, for this group, lower levels of correlation between DDR and DDA are observed (Pearson coefficient = 0.22 vs. 0.38 for all papers). An alternative explanation for the results from Model 8 would be that the relationship between DDR and UPRL is only valid for papers with one or two authors, with no effect of DDR in larger teams.

Given all the findings (from multiple models) presented here, we see no evidence to reject a potential effect of DDR on UPRL, even in larger research teams, and this relationship should be considered in future research. Overall, the models presented in this paper (and the supplementary material) offer more evidence of an effect of DDA on UPRL and suggest a plausible relationship between DDR and UPRL that may not have been fully uncovered in our models due to the existence of collinearity between the two variables used to measure XDR.

We contend that our positive findings for Question A and Question B provide a basis for selecting instruments that implement DDA when designing research funding programs with an explicit goal to increase societal outcomes, in the form of increased knowledge transfer towards policymakers (Question C). This conclusion holds implications for the conduct and evaluation of policies and programs aiming to support research with an orientation towards societal impact, and especially, towards impact on research-informed policymaking. Based on our findings, if those programs' publications in the aggregate do display higher DDA than prior publications by the same team, then these papers should tendentially exhibit greater UPRL. Of course, not all programs promoting XDR, and even projects within a program, will be equally successful at driving XDR and in generating subsequent policy uptake; XDR is at best expected to be one out of many contributing factors influencing the complex phenomenon driving UPRL, possibly none with a large size effect. Accordingly, even if the answer to Question C is that programs promoting XDR increase the odds of their research outputs being taken up in policy, it should not be used as a justification to bypass actual measurements of a program's achievement in ex-post research evaluations.

We also found that scientific citations correlate quite strongly with UPRL citations, indicating that policymakers may rely on traditional markers of excellence when seeking out scientific evidence to support their activities; or that work with a higher relevance for policy may also tend to be more highly cited. Reverse causality is a relevant hypothesis for the correlation observed between scientific and policy UPRL citations. In many cases, UPRL could even have preceded scientific citations, included in the models mainly to control for a possible indirect link between cross-disciplinarity and UPRL through a higher impact within the scientific community.

5.4. Limitations

Going forward, it will be possible to deploy the Overton database as well as the research strategies implemented here in two core contexts: in quantitative research on XDR, societal

outcomes of research, and/or the development of altmetrics, and also in the applied context of (research and innovation) program evaluation. In the later context, we advocate for UPRL analysis to be used in an investigative manner, to answer well-defined research questions that elucidate major mechanisms of action for the program under review. This use of the metric in benchmarking exercises should provisionally be paired with very cautious interpretation. For instance, our results are subject to a number of limitations:

- UPRL captures only a subset of science policy interactions, with prior reports showing that most knowledge transfer takes place through tacit and local engagement rather than formal channels; nevertheless, the high shares of policy citedness among EU-funded papers reported here may indicate that the importance of formal channels for knowledge transfer towards policy has been underestimated.
- The Overton database has yet to be the subject of sustained investigation in the altmetrics and bibliometrics communities; further work is necessary to better understand the limitations of this data set. Particularly, the citations from policy-relevant documents retrieved for the set of publications examined originated to a large extent from regulatory science or scientific advisory documents rather than executive or legislative documents. This observation indicated that although some societal impact had been achieved by the peer-reviewed publications examined, this impact was located very much in the first steps of the evidence-based policymaking process, rather than in the deeper stages of integration. Note that Overton does contain records on executive and legislative documents, and that our finding may be a function of the specific publication set used here.
- The binary indicator used to represent UPRL does not capture the differences in the number of UPRL citations received by scientific papers. Papers are treated similarly whether they have been cited only once or many times in policy documents. This option was shaped by the low proportion of papers being cited in policy and by the observation that the number of citations received may be a less precise indicator compared to the binary variable chosen, especially in a database that has not been frequently used before in such types of work.
- As described in Section 4.1, the qualitative assessment of Overton showed that some of its citation records did not fall under the concept of UPRL employed in this paper. These false positives and the fact that Overton does not cover all policy-related documents reveal that even the binary indicator of UPRL may carry errors that would introduce noise in the models reported. The most likely consequence of this situation is a reduction in our ability to detect links between explanatory variables and UPRL. Future improvements in Overton or future research should allow for more precise estimates of the relationships reported here.
- Our results indicate that caution should be exerted if using Overton-based metrics in a program evaluation context. First, the results reported are based on a sample of publications from the FPs for Research and Technological Development (i.e., FP7 or H2020). Whether these results are valid for other types of funding initiatives remains open for further research. However, the results should be useful to inform funding bodies and programs with similar characteristics and goals.
- A second limitation of using Overton-based metrics in research evaluations is coverage bias, highlighted by the lower share of papers cited in Overton for non-U.K. papers (5.1%) compared to U.K. papers (8.6%). Similarly, research does not appear to be equally relevant to policy across disciplines, highlighting the importance of accounting for such differences in comparative studies.

- Causal claims about regression coefficients based on observational data sets are usually unlikely. Nevertheless, the models reported here accounted for a considerable selection of confounders and for fixed effects for research projects. It should help to approximate these coefficients to the actual causal relationship, compared to simple measures of correlations that do not account for the effect of confounders. However, the closeness of the reported coefficients and the true causal relationships is hard to be assessed. Triangulation with future work, quantitative or qualitative, should help to validate the findings reported in this paper.

Future research work could test for potential effects, on the XDR-UPRL relationship, of a number of additional factors: shared authorship between scientific publication and policy-relevant documents (i.e., including self-citation); or the use of disciplinary diversity indicators beyond DDA and DDR (Hackett et al., 2021; Leydesdorff, Wagner, & Bornmann, 2019).

ACKNOWLEDGMENTS

The authors wish to thank Euan Adie of Overton for offering generous access to the database and his contributions to its characterization. The study team also acknowledges key contributions from Maxime Rivest in the implementation of the DDA indicator in the Science-Metrix implementation of Scopus. Colleagues and participants in an internal Elsevier Research Analytics & Data Services, as well as Athina Karvounaraki and Tiago Pereira from DG Research & Innovation, provided useful comments on preliminary findings related to this publication. We thank Beverley Mitchell for support in copyediting.

AUTHOR CONTRIBUTIONS

Henrique Pinheiro: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—original draft, Writing—review & editing. Etienne Vignola-Gagné: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing—original draft, Writing—review & editing. David Campbell: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing—original draft, Writing—review & editing.

The three authors consider that their differentiated contributions to this article's research and writing strictly balance out in value and intensity.

COMPETING INTERESTS

The authors are employees of Science-Metrix, an Elsevier company.

FUNDING INFORMATION

This research was funded through mandated work conducted for the European Commission Directorate General for Research and Innovation.

DATA AVAILABILITY

We have added extensive supplementary materials to the article, which should help clarify many of the methodological choices, and provide high-level descriptions of the data sets, used by the research team. Overton is a commercial database for which access was partially secured on a fee-for-service basis by Science-Metrix, as part of its own service offer to the

European Commission DG Research and Innovation. Overton provided additional complementary access for the purposes of this publication. The authors are not able to provide unaggregated Overton data under their license agreement. Article-level normalized findings for interdisciplinarity and multidisciplinarity are proprietary data of Elsevier.

REFERENCES

- Abramo, G., D'Angelo, C. A., & Zhang, L. (2018). A comparison of two approaches for measuring interdisciplinary research output: The disciplinary diversity of authors vs the disciplinary diversity of the reference list. *Journal of Informetrics*, 12(4), 1182–1193. <https://doi.org/10.1016/j.joi.2018.09.001>
- Adams, J., Loach, T., & Szomszor, M. (2016). *Interdisciplinary research: Methodologies for identification and assessment*. Retrieved from Digital Science website: https://figshare.com/articles/Digital_Research_Report_Interdisciplinary_Research_-_Methodologies_for_Identification_and_Assessment/4270289
- Agresti, A. (2012). 16.1 Delta method. In *Categorical Data Analysis* (3rd ed.). New York: John Wiley & Sons.
- Allmendinger, J. (2015). *Quests for interdisciplinarity: A challenge for the ERA and HORIZON 2020. Policy Brief by the Research, Innovation, and Science Policy Experts (RISE)*. Retrieved from <https://ec.europa.eu/research/openvision/pdf/rise/allmendinger-interdisciplinarity.pdf>
- Archambault, É., Beauchesne, O. H., & Caruso, J. (2011). Towards a multilingual, comprehensive and open scientific journal ontology. In B. Noyons, P. Ngulube, & J. Leta (Eds.), *Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics* (pp. 66–77). Retrieved from <https://science-matrix.com/?q=en/publications/conference-presentations/towards-a-multilingual-comprehensive-and-open-scientific>
- Belcher, B. M., & Hughes, K. (2020). Understanding and evaluating the impact of integrated problem-oriented research programmes: Concepts and considerations. *Research Evaluation*, rvaa024. <https://doi.org/10.1093/reseval/rvaa024>
- Bornmann, L., Haunschild, R., & Marx, W. (2016). Policy documents as sources for measuring societal impact: How often is climate change research mentioned in policy-related documents? *Scientometrics*, 109, 1477–1495. <https://doi.org/10.1007/s11192-016-2115-y>, PubMed: 27942080
- Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature*, 534(7609), 684–687. <https://doi.org/10.1038/nature18315>, PubMed: 27357795
- Brownson, R. C., Chriqui, J. F., & Stamatakis, K. A. (2009). Understanding evidence-based public health policy. *American Journal of Public Health*, 99, 1576–1583. <https://doi.org/10.2105/AJPH.2008.156224>, PubMed: 19608941
- Buenstorf, G., & Koenig, J. (2020). Interrelated funding streams in a multi-funder university system: Evidence from the German Exzellenzinitiative. *Research Policy*, 49(3), 103924. <https://doi.org/10.1016/j.respol.2020.103924>
- Calvert, J. (2010). Systems biology, interdisciplinarity and disciplinary identity. In J. N. Parker, N. Vermeulen, & B. Penders (Eds.), *Collaboration in the new life sciences*. Aldershot: Ashgate.
- Campbell, D., Deschamps, P., Côté, G., Roberge, G., Lefebvre, C., & Archambault, É. (2015). Application of an “interdisciplinarity” metric at the paper level and its use in a comparative analysis of the most publishing ERA and non-ERA universities. *20th International Conference on Science and Technology Indicators*. Retrieved from https://science-matrix.com/sites/default/files/science-matrix/publications/campbell_et_al_sti2015_short_paper_final_web.pdf
- Campbell, D., & Struck, B. (2019). Reliability of Scopus author identifiers (AUIDs) for research evaluation purposes at different scales. *Proceedings of the 17th International Conference of the International Society for Scientometrics and Informetrics (ISSI), 2–5 September 2019* (Vol. II, pp. 1276–1287). Retrieved from <https://issi-society.org/publications/issi-conference-proceedings/proceedings-of-issi-2019/>
- Campbell, D., Struck, B., Tippett, C., & Roberge, G. (2017). Impact of multidisciplinary research on innovation. *16th International Conference of the International Society for Scientometrics and Informetrics (ISSI), 16–20 October 2017*. Retrieved from https://www.science-matrix.com/sites/default/files/science-matrix/publications/issi2017_paper_153_d_campbell_impact_multidisciplinarity.pdf
- Chavarro, D., Tang, P., & Rafols, I. (2014). Interdisciplinarity and research on local issues: evidence from a developing country. *Research Evaluation*, 23(3), 195–209. <https://doi.org/10.1093/reseval/rvu012>
- Deste, P., Mahdi, S., Neely, A., & Rentocchini, F. (2012). Inventors and entrepreneurs in academia: What types of skills and experience matter? *Technovation*, 32(5), 293–303. <https://doi.org/10.1016/j.technovation.2011.12.005>
- Dworkin, J. D., Shinohara, R. T., & Bassett, D. S. (2019). The emergent integrated network structure of scientific research. *PLOS ONE*, 14(4). <https://doi.org/10.1371/journal.pone.0216146>, PubMed: 31039179
- European Commission Expert Group on Altmetrics. (2017). *Next-generation metrics: Responsible metrics and evaluation for open science*. Brussels.
- European Research Council. (2018). *Qualitative evaluation of completed projects funded by the European Research Council*. Retrieved from https://erc.europa.eu/sites/default/files/publication/files/ERC_Science_behind_the_projects_FP7-2007-2013.pdf
- Fowle, K., Wells, B., Day, M., Kumar, A., Bess, C., Bingham, B., & Wayman, A. (2020). The program and policy change framework: A new tool to measure research use in low- and middle-income countries. *Research Evaluation*, rvaa017. <https://doi.org/10.1093/reseval/rvaa017>
- Garner, J. G., Porter, A. L., Newman, N. C., & Crawl, T. A. (2012). Assessing research network and disciplinary engagement changes induced by an NSF program. *Research Evaluation*, 21(2), 89–104. <https://doi.org/10.1093/reseval/rvs004>
- Gleed, A., & Marchant, D. (2016). *Interdisciplinarity survey report for the Global Research Council 2016 annual meeting*. Stockport.
- Hackett, E. J., Leahy, E., Parker, J. N., Rafols, I., Hampton, S. E., ... Vision, T. J. (2021). Do synthesis centers synthesize? A semantic analysis of topical diversity in research. *Research Policy*, 50(1). <https://doi.org/10.1016/j.respol.2020.104069>, PubMed: 33390628
- Haunschild, R., & Bornmann, L. (2017). How many scientific papers are mentioned in policy-related documents? An empirical

- investigation using Web of Science and Altmetric data. *Scientometrics*, 110(3), 1209–1216. <https://doi.org/10.1007/s11192-016-2237-2>, PubMed: 28255186
- 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>
- Hird, M. D., & Pfotenhauer, S. M. (2017). How complex international partnerships shape domestic research clusters: Difference-in-difference network formation and research re-orientation in the MIT Portugal Program. *Research Policy*, 46(3), 557–572. <https://doi.org/10.1016/j.respol.2016.10.008>
- Koier, E., & Horlings, E. (2015). How accurately does output reflect the nature and design of transdisciplinary research programmes? *Research Evaluation*, 24(1), 37–50. <https://doi.org/10.1093/reseval/rvu027>
- Lawrence, A. (2018). Influence seekers: The production of grey literature for policy and practice. *Information Services and Use*, 37(4), 389–403. <https://doi.org/10.3233/ISU-170857>
- Lawrence, A., Houghton, J., Thomas, J., & Weldon, P. (2014). *Where is the evidence? Realising the value of grey literature for public policy & practice*. Retrieved from <https://apo.org.au/research/where-evidence-realising-value-grey-literature-public-policy-and-practice>
- LERU. (2016). *Interdisciplinarity and the 21st century research-intensive university*. Retrieved from <https://www.leru.org/files/Interdisciplinarity-and-the-21st-Century-Research-Intensive-University-Full-paper.pdf>
- Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2019). Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. *Journal of Informetrics*, 13(1), 255–269. <https://doi.org/10.1016/j.joi.2018.12.006>
- Lyll, C., Bruce, A., Marsden, W., & Meagher, L. (2013). The role of funding agencies in creating interdisciplinary knowledge. *Science and Public Policy*, 40(1), 62–71. <https://doi.org/10.1093/scipol/scs121>
- Lyll, C., Meagher, L., & Bruce, A. (2015). A rose by any other name? Transdisciplinarity in the context of UK research policy. *Futures*, 65, 150–162. <https://doi.org/10.1016/j.futures.2014.08.009>
- Molas-Gallart, J., Rafols, I., & Tang, P. (2014). On the relationship between interdisciplinarity and impact: Different modalities of interdisciplinarity lead to different types of impact. *Journal of Science Policy and Research Management*, 29(September), 69–89.
- Moschini, U., Fenaldi, E., Daraio, C., Ruocco, G., & Molinari, E. (2020). A comparison of three multidisciplinary indices based on the diversity of Scopus subject areas of authors' documents, their bibliography and their citing papers. *Scientometrics*, 125(2), 1145–1158. <https://doi.org/10.1007/s11192-020-03481-x>
- Newson, R., Rychetnik, L., King, L., Milat, A., & Bauman, A. (2018). Does citation matter? Research citation in policy documents as an indicator of research impact—An Australian obesity policy case-study. *Health Research Policy and Systems*, 16(1), 55. <https://doi.org/10.1186/s12961-018-0326-9>, PubMed: 29950167
- Okamura, K. (2019). Interdisciplinarity revisited: evidence for research impact and dynamism. *Palgrave Communications*, 5(1), 1–9. <https://doi.org/10.1057/s41599-019-0352-4>
- Porter, A. L., Garner, J. G., & Cowl, T. (2012). Research coordination networks: Evidence of the relationship between funded interdisciplinary networking and scholarly impact. *Bioscience*, 62(3), 282–288. Retrieved from <https://www.jstor.org/stable/10.1525/bio.2012.62.3.9>. <https://doi.org/10.1525/bio.2012.62.3.9>
- Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, 81(3), 719–745. <https://doi.org/10.1007/s11192-008-2197-2>
- Pulido, C. M., Redondo-Sama, G., Sordé-Martí, T., & Flecha, R. (2018). Social impact in social media: A new method to evaluate the social impact of research. *PLOS ONE*, 13(8), e0203117. <https://doi.org/10.1371/journal.pone.0203117>, PubMed: 30157262
- Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics*, 82(2), 263–287. <https://doi.org/10.1007/s11192-009-0041-y>
- Rivest, M., Vignola-Gagné, E., & Archambault, E. (2021). Article-level classification of scientific publications: a comparison of deep learning, direct citation and bibliographic coupling. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0251493>, PubMed: 33974653
- Rylance, R. (2015). Grant giving: Global funders to focus on interdisciplinarity. *Nature*, 525, 313–315. <https://doi.org/10.1038/525313a>, PubMed: 26381969
- Schneider, F., Buser, T., Keller, R., Tribaldos, T., & Rist, S. (2019). Research funding programmes aiming for societal transformations: Ten key stages. *Science and Public Policy*, 46(3), 463–478. <https://doi.org/10.1093/scipol/scy074>
- Science-Metrix. (2018). *Review of the Human Frontier Science Program 2018*. Retrieved from <https://www.hfsp.org/node/12547#book/>
- Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15), 707–719. <https://doi.org/10.1098/rsif.2007.0213>, PubMed: 17327202
- Tahamtan, I., & Bornmann, L. (2020). Altmetrics and societal impact measurements: Match or mismatch? A literature review. *El Profesional de La Información*, 29(1), e290102. <https://doi.org/10.3145/epi.2020.ene.02>
- Tattersall, A., & Carroll, C. (2018). What can Altmetric.com tell us about policy citations of research? An analysis of Altmetric.com data for research articles from the University of Sheffield. *Frontiers in Research Metrics and Analytics*, 2(9). <https://doi.org/10.3389/frma.2017.00009>
- Technopolis Group & Science-Metrix. (2020). *Evaluation of the Belmont Forum: Final report*. Retrieved from <https://www.belmontforum.org/wp-content/uploads/2021/03/Belmont-Forum-Evaluation-Report.pdf>
- Thelwall, M. (2016). Web indicators for research evaluation: A practical guide. In *Synthesis Lectures on Information Concepts, Retrieval, and Services* (Vol. 8). <https://doi.org/10.2200/s00733ed1v01y201609icr052>
- Thelwall, M., & Kousha, K. (2015). Web indicators for research evaluation. Part 1: Citations and links to academic articles from the Web. *El Profesional de La Información*, 24(5), 587. <https://doi.org/10.3145/epi.2015.sep.08>
- van der Hel, S. (2016). New science for global sustainability? The institutionalisation of knowledge co-production in Future Earth. *Environmental Science and Policy*, 61, 165–175. <https://doi.org/10.1016/j.envsci.2016.03.012>
- Wang, J., Thijs, B., & Glänzel, W. (2015). Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity. *PLOS ONE*, 10, e0127298. <https://doi.org/10.1371/journal.pone.0127298>, PubMed: 26001108
- Wang, L., & Li, Z. (2018). Effects of knowledge flows from public science to industrial technologies. *Proceedings of the 2018 STI Conference* (pp. 576–588). Retrieved from <https://openaccess>

- .leidenuniv.nl/bitstream/handle/1887/65342/STI2018_paper_109.pdf?sequence=1
- Warner, K. E., & Tam, J. (2012). The impact of tobacco control research on policy: 20 years of progress. *Tobacco Control, 21*(2), 103–109. <https://doi.org/10.1136/tobaccocontrol-2011-050396>, PubMed: 22345232
- Weiss, C. H. (1979). The many meanings of research utilization. *Public Administration Review, 39*(5), 426. <https://doi.org/10.2307/3109916>
- Willis, C. D., Riley, B., Stockton, L., Viehbeck, S., Wutzke, S., & Frank, J. (2017). Evaluating the impact of applied prevention research centres: results from a modified Delphi approach. *Research Evaluation, 26*(2), 78–90. <https://doi.org/10.1093/reseval/rvx010>
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., ... Johnson, B. (2015). *The metric tide: Report of the independent review of the role of metrics in research assessment and management*. <https://doi.org/10.13140/RG.2.1.4929.1363>
- Zhang, L., Sun, B., Chinchilla-Rodríguez, Z., Chen, L., & Huang, Y. (2018). Interdisciplinarity and collaboration: on the relationship between disciplinary diversity in departmental affiliations and reference lists. *Scientometrics, 117*(1), 271–291. <https://doi.org/10.1007/s11192-018-2853-0>
- Zuo, Z., & Zhao, K. (2018). The more multidisciplinary the better? The prevalence and interdisciplinarity of research collaborations in multidisciplinary institutions. *Journal of Informetrics, 12*(3), 736–756. <https://doi.org/10.1016/j.joi.2018.06.006>