RESEARCH ARTICLE

Is groundbreaking research disruptive? The covariation of disruption indices with researchers' self-assessments

Christian Leibel[1,2], Alexander Tekles[3], Dag W. Aksnes[4], and Lutz Bornmann[1]

[1]Science Policy and Strategy Department, Administrative Headquarters of the Max Planck Society, Munich, Germany
[2]Department of Sociology, LMU Munich, Munich, Germany
[3]University of Passau, Passau, Germany
[4]Nordic Institute for Studies in Innovation, Research, and Education (NIFU), Oslo, Norway

an open access journal

Keywords: bibliometrics, CD index, disruption index, novelty, research evaluation

ABSTRACT

Park, Leahey, and Funk (2023) reported a decline in disruptive science since the Second World War, sparking debate about a perceived lack of major scientific achievements. Their study relied on the disruption index, a bibliometric indicator intended to identify disruptive patents and publications. However, this measure may miss substantial efforts contributing to scientific innovation, raising questions about its suitability for identifying major achievements. In this study, we tested whether the two most commonly used variants of the disruption index—the DI_1 and the DI_5—correspond with researchers' self-assessments. Our findings show that researchers view groundbreaking science as combining novelty and importance, but disruption scores do not correspond to authors' self-assessments of their papers' groundbreaking character. This discrepancy may result from measurement weaknesses in the indices or from limitations in using author self-assessment as a benchmark. We conclude that further research is needed on disruption indices before they can be reliably used for research evaluation.

Citation: Leibel, C., Tekles, A., Aksnes, D. W., & Bornmann, L. (2026). Is groundbreaking research disruptive? The covariation of disruption indices with researchers' self-assessments. *Quantitative Science Studies*, *7*, 521–542. https://doi.org/10.1162/QSS.a.462

DOI:
https://doi.org/10.1162/QSS.a.462

Peer Review:
https://www.webofscience.com/api/gateway/wos/peer-review/10.1162/QSS.a.462

Received: 5 February 2025
Accepted: 31 January 2026

Corresponding Author:
Christian Leibel
christian.leibel.extern@gv.mpg.de

Handling Editor:
Vincent Larivière

The MIT Press

1. INTRODUCTION

Theories about scientific progress stress the importance of groundbreaking and transformative research for the evolution of scientific knowledge. According to Thomas S. Kuhn (1962, 1970), the science system repeatedly goes through long periods of incremental progress followed by short periods of scientific revolutions, which disrupt the linear progress of science and completely transform the science system. Similarly, Cole (1983, p. 113) stresses that "knowledge is not a uniform whole," but rather "made up of two principal components," which he refers to as the "core" and the "research frontier." Whereas the core consists of established theories and practice, new knowledge is produced at the research frontier. According to Cole (1983), most new research has little lasting impact, but some select few publications that present new theories, results, or methods manage to become incorporated into the core and thus drive significant scientific progress. For example, Mayor and Queloz (1995) gave rise to a new stream of extrasolar planet research by discovering the first exoplanet around a Sun-like star. Mayor and Queloz (1995) transformed the field of astrophysics by adding a new type of research object to the field's core[1].

[1] The example of Mayor and Queloz (1995) was taken from Bornmann, Wu, and Ettl (2024).

Given the importance of transformative research for the evolution of human knowledge, it is not surprising that a metric by the name of disruption index (here abbreviated: DI_1)[2], which promises to be able to quantify and identify disruptive science, has captured the attention and imagination of researchers. The DI_1 operates on the assumption that transformative research alters citation networks by supplanting prior work (Lin, Li, & Wu, 2026). Leibel and Bornmann (2024), who published an overview of research on the DI_1, mention numerous studies that use the DI_1 to identify transformative papers in their respective fields. Since the introduction of the DI_1 by Funk and Owen-Smith (2017), no fewer than three studies have been published in the influential journal *Nature* that utilize the DI_1 as a measure of transformative research (Lin, Frey, & Wu, 2023; Park et al., 2023; Wu, Wang, & Evans, 2019).

The DI_1 is supposed to distinguish between disruptive and consolidating research. A disruptive paper renders prior knowledge obsolete by creating a new trajectory that diminishes reliance on preceding research. In citation networks, this pattern is reflected when subsequent papers predominantly cite the disruptive paper without referencing its predecessors, signaling a shift away from existing knowledge or standards in research. For example, a disruptive study may introduce a novel method that makes prior techniques less relevant. In contrast, a consolidating paper strengthens and integrates existing knowledge, reinforcing the continuity of prior research. In the citation network, this is reflected by subsequent papers citing both the consolidative paper and its predecessors, indicating complementary integration. Consolidating papers typically refine or synthesize existing ideas, thereby stabilizing the existing knowledge or standards in research. A meta-analysis that synthesizes prior findings and encourages further research within the same framework is a typical example of consolidating research.

The findings of Park et al. (2023) have made waves in and beyond the science community because they suggest that there has been a steady decline of disruptive science and patents in every area of research since the end of the Second World War. Unsurprisingly, this spectacular result triggered a public debate about the apparent stagnation of the global science system. Some media outlets have interpreted the results of Park et al. (2023) as evidence for a lack of "major advances" (Hart, 2023) and "real breakthroughs" (Broad, 2023) in a science system that seems to be running out of steam.

One may question whether the findings of Park et al. (2023) really indicate that the global science system is losing its ability to generate groundbreaking epistemic innovations. A key issue concerns the validity of the indicator—whether it captures the phenomenon it is designed to measure. The DI_1 may not be an adequate measure of epistemic change due to technical or other limitations. As a citation-based measure, the DI_1 suffers from citation inflation when comparing publications from different publication years (Petersen, Arroyave, & Pammolli, 2024, 2025). In addition, the DI_1 is susceptible to data artefacts (Holst, Algaba et al., 2024) and is biased towards publications that cite only few sources of inspiration.

Researchers have raised concerns about the conceptual adequacy of the DI_1 as a measure of groundbreaking research. In a comment on the media response to Park et al. (2023), Kamerlin (2023) points out that the DI_1 seems to ignore the manifoldness of scientific breakthroughs by focusing exclusively on disruptive and consolidating publications. If scientific innovation is to be understood as the process through which (epistemic) problems are

[2] In the literature, the disruption index is also referred to as the "CD index" (Funk & Owen-Smith, 2017) or the "D index" (Ruan, Lyu et al., 2021). Here we use an abbreviation that is consistent with the literature review of Leibel and Bornmann (2024). The DI_1 is calculated with a minimum bibliographic coupling strength of 1. Other variants, like the DI_5, require stronger bibliographic coupling. More details on the definition of the disruption index are presented in Section 3.1.

formulated and subsequently solved, then there may be as many types of scientific breakthroughs as there are types of epistemic problems. The idea that there may be diverse types of groundbreaking research is supported by current research. Based on an in-depth analysis of groundbreaking research articles, Wuestman, Hoekman, and Frenken (2020) developed a sophisticated typology of scientific breakthroughs. Their work shows that there are several distinct types of groundbreaking science. For example, research can be groundbreaking either because it discovers new epistemic problems or because it provides solutions to problems that are well known in the literature. Furthermore, groundbreaking science can either support or question the state-of-the-art literature. In light of the diversity of scientific breakthroughs, one may argue that scientific progress is perhaps not driven exclusively (or even majorly) by disruptive research. Consequently, the DI_1 (and its variants) may have only limited use as an indicator of groundbreaking research.

To properly assess the credibility of empirical studies based on the DI_1 (and its many variants), we need to know—following previous studies on the convergent validity of the DI_1 and its variants—whether the DI_1 is actually able to identify publications that advance (and disrupt) scientific progress. This study aims to answer the question of whether the DI_1 and its variants actually measure what they propose to measure. To this end, we used data from a survey that asked researchers to assess the quality of their own papers. In the first part of our study, we elaborate the current state of research on disruption indices and subsequently contextualize the DI_1 within a typology of scientific breakthroughs. The second part complements the conceptual work by presenting empirical evidence on how the scientific community considers groundbreaking research. The third part addresses the pivotal question of whether groundbreaking science can be characterized by (disruptive) patterns in citation networks.

2. LITERATURE

2.1. Previous Empirical Studies on the Convergent Validity of Disruption Indices

Our study is one of a growing number of papers that use expert and peer judgments to test the convergent validity of the DI_1 and its variants. Convergent validity means that a measure is considered valid if it correlates strongly and positively with expert assessments of the same concept (Kreiman & Maunsell, 2011). So far, papers praised by journal editors (Bittmann, Tekles, & Bornmann, 2022; Bornmann & Tekles, 2021; Deng & Zeng, 2023), highly original papers (Bornmann, Devarakonda et al., 2020; Shibayama & Wang, 2020), scientific breakthroughs announced by the magazine *Science* (Wang, Ma et al., 2023), and Nobel-Prize-winning papers (Wei, Li, & Shi, 2023; Wu et al., 2019) have been used to test the convergent validity of disruption indices. Leibel and Bornmann (2024) summarize the results of the above-mentioned studies in their review of the literature of the DI_1 and its variants. Overall, the previous tests of the convergent validity of disruption indices lead to mixed results. Both the strength and the sign of the correlation between disruption scores and expert judgments vary from study to study. Thus, Leibel and Bornmann (2024, p. 601) conclude that "more research on the validity of disruption scores [...] is needed before the indices can be used in the research evaluation practice."

Our study addresses the need for more research on disruption indices by providing additional empirical results on their convergent validity with researchers' self-assessments. We leverage survey data to examine how the scientific community perceives groundbreaking research and to assess the extent to which these perceptions correspond with disruption scores. Because groundbreaking research shares many characteristics with disruptive research, it provides an informative basis for comparative analysis.

2.2. Is Groundbreaking Research Disruptive?

Previous research on disruption indices found that disruption scores covary with expert ratings (Bornmann et al., 2020) as well as self-assessments (Shibayama & Wang, 2020) of novel publications. Disruption indices also seem to correlate with another bibliometric measure of originality: the *z*-score developed by Uzzi, Mukherjee et al. (2013). Wu et al. (2019) and Park et al. (2023) tested the robustness of their empirical results with several different bibliometric indicators of novel and disruptive science. Both studies found that the DI_1 and the *z*-score produce similar research outcomes. In addition, Wu et al. (2019) report that the DI_1 is able to identify disruptive and developing papers that scholars proposed in a survey. Thus, there is empirical evidence to substantiate the claim that disruption scores function as a valid measure of scientific innovation.

The abovementioned results are contrasted by other findings that lead to less favorable conclusions about the (convergent) validity of disruption indices. In the study of Wang et al. (2023), disruption indices failed to identify scientific breakthroughs. Bornmann and Tekles (2019b) analyzed papers with high disruption scores published in *Scientometrics*. Bornmann and Tekles (2019b) conclude that the DI_1 did not produce satisfactory results because most of the papers with the highest DI_1 values could not be characterized as landmark papers upon manual inspection. Overall, empirical evidence does not conclusively confirm the validity of disruption indices. In addition to the divergent results that disruption indices have produced in empirical studies, there are also conceptual objections to the validity of disruption scores. In our introduction, we pointed out that disruption indices may not capture the diversity of groundbreaking research. To flesh out this argument, we contextualize the concept of disruptive science within the typology of scientific breakthroughs that was developed by Wuestman et al. (2020). In order to construct the typology, Wuestman et al. (2020) analyzed the contents of 335 articles that were included in *Science* magazine's annual announcement of the "Breakthrough of the Year." The title of "Breakthrough of the Year" is given to papers that made original and highly significant contributions to a specific field of research. Wuestman et al. (2020) suggest the following three criteria for the classification of scientific breakthroughs: First, the breakthroughs are either driven by research *questions* or research *objects*. Second, the research questions or research objects are either *known* in the literature or were previously *undocumented*. Third, the research outcomes either *challenge* or *confirm* the findings and hypotheses of previous literature (Table 1).

We consulted the research literature to decide how disruption scores might fit into this typology of scientific breakthroughs. Disruptive research can be either question driven or research object driven, and the questions or objects do not necessarily need to be new ones. According to Lin et al. (2026), disruptive innovation can occur if new literature displaces previous literature by providing alternative (and presumably better) solutions to known epistemic problems. Similarly, Funk and Owen-Smith (2017, p. 793) stress disruptive innovations "challenge the existing order" and overthrow established thinking. In their literature review, Leibel and Bornmann (2024) point out that disruption indices are sometimes linked with Kuhn's theory of paradigm shifts (Kuhn, 1962): Disruptive innovations are those that do not fit within the current paradigm. The common theme among these various interpretations of disruptive

Table 1. Most likely properties of disruptive breakthroughs in the typology of Wuestman et al. (2020)

Question-driven or research-object-driven	New or known question/ research object	Question/research object is against or in line with previous literature
Either	Either	Against literature

research is that disruptive breakthroughs challenge and eventually displace the state-of-the-art literature.

If breakthroughs that depart from or challenge prior literature are the most plausible candidates for disruptive breakthroughs, then the analysis of Wuestman et al. (2020) suggests that such disruptive breakthroughs are exceptionally rare: The vast majority of scientific breakthroughs (77% of 335 papers) are "driven by a known question that is in line with theory" and only 11% of *Science* magazine's "Breakthrough of the Year" articles question the state-of-the-art literature. In light of the evidence presented by Wuestman et al. (2020), it seems likely that disruptive research accounts for only a small portion of scientific breakthroughs. Given that disruption indices only distinguish between disruptive and consolidating research, it is possible that they fail to identify the vast majority of groundbreaking publications.

In summary, the current research literature on scientific novelty and scientific breakthroughs does not provide conclusive evidence regarding the validity of disruption indices. Thus, we pose it as an open question whether disruption scores are convergently valid with researchers' self-assessments of groundbreaking research. It is worth noting that many previous validation studies did not take into account the DI_1's susceptibility to data-induced biases because the full extent of the DI_1's technical limitations was uncovered only recently by Holst et al. (2024) and Petersen et al. (2024). In our study, we mitigate data-induced biases by conducting a robustness check with a selective group of publications for which the quality of the bibliometric data is exceptionally high.

3. METHODS

3.1. Measuring Disruptive Innovation via Discontinuity in Citation Networks

The DI_1 is related to measures of betweenness centrality (Freeman, 1977; Gebhart & Funk, 2023) and its calculation is grounded in bibliographic coupling. Bibliographic coupling links papers that cite the same references. If, for example, publication A and publication B both cite publication C, then there is a bibliographic coupling link between A and B. As Figure 1

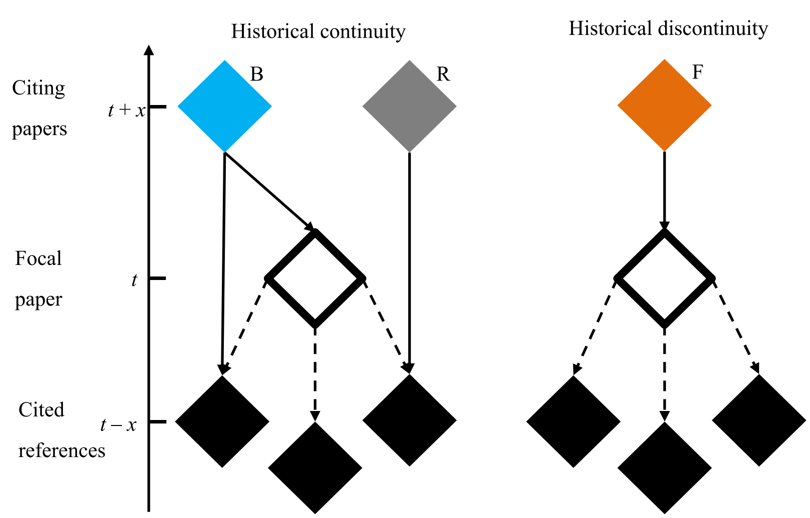

Figure 1. How the DI_1 measures historical continuity and historical discontinuity in citation networks. Citing papers of type B and type R indicate continuity. Citing papers of type F indicate discontinuity. The illustration is based on Funk and Owen-Smith (2017) and shows two idealized citation networks.

illustrates, the DI_1 adds a historical component to the analysis of bibliographic coupling by identifying the links between a focal paper (FP) and subsequent publications. The DI_1 differentiates between three cases regarding bibliographic coupling between the FP and subsequent publications. Papers of type R (standing for "reference") only cite the FP's cited references[3], but not the FP itself. Papers of type B (standing for "both") cite both the FP and at least one of the FP's cited references. Papers of type F (standing for "focal") cite only the FP and none of its cited references.

Leydesdorff and Bornmann (2021) and Leydesdorff, Tekles, and Bornmann (2021) point out that bibliographic coupling reflects historical continuity. Papers of type B and type R represent historical continuity because they connect the more recent research ($t + x$) to research that predates the FP ($t - x$). Bibliographic coupling links between the FP and subsequent publications indicate that the FP did not diminish the importance of previous research, which suggests that the FP is consolidating. By contrast, papers of type F represent historical discontinuity because they contain no bibliographic coupling links with the FP. The research that predates the FP seems to be no longer relevant for future authors, which implies that the FP is disruptive. The DI_1 is calculated by subtracting the amount of citing papers of the FP that indicate historical continuity (N_B) from the number of papers that indicate historical discontinuity (N_F) and dividing by the total number of citing papers in the citation network. The DI_1 is equivalent to the following ratio:

$$DI_1 = \frac{N_F - N_B}{N_F + N_B + N_R}$$

Research on the DI_1 has shown that the disruption score depends on two factors: It obviously depends on the citing behavior of the FP's citing papers, but it also depends substantially on the citing behavior of the FP's authors (Bornmann & Tekles, 2019b). In an illustrative case study, Liu, Zhang, and Li (2023) show that citing just a single highly cited reference may lead to a massive increase in (N_R), which, in turn, may severely diminish the FP's DI_1 score. Thus, the DI_1 rewards papers for not taking much inspiration from well-established streams of research. The dominant role of highly cited references in the calculation of the DI_1 could create an incentive for researchers to purposely avoid citing important sources of inspiration in order to manipulate the disruption scores of their own work (Yang, Hu et al., 2023).

Bornmann et al. (2020) proposed the DI_5 as an alternative to the DI_1. The DI_5 considers only those B-type citing papers that cite at least five of the FP's cited references. As a result, the DI_5 focuses on citing papers that rely more strongly on the research cited by the FP. Papers that only cite a few highly cited references are not considered in the calculation of the DI_5. Compared to DI_1 scores, DI_5 scores should be less dependent on the citing behavior of the FP's authors and instead put more weight on the citing behavior of future research ($t + x$). The DI_5 is equivalent to:

$$DI_5 = \frac{N_F - N_B^5}{N_F + N_B^5 + N_R}$$

In addition to the DI_5, there are numerous other modified variants of the DI_1. Leibel and Bornmann (2024) provide an overview of the growing number of disruption indices variants.

[3] We calculated N_R such that it only includes papers that were published in the year following the FP's publication or later.

In this study, we use the DI_1 and the DI_5. We selected the DI_1 because it is the initially proposed and most commonly used indicator in the literature, and we chose the DI_5 because it showed the best performance out of all variants in previous validation studies.

3.2. Survey

Our study relies on a combination of survey data and bibliometric data. The survey was conducted by Aksnes, Piro, and Fossum (2023) and was sent out in January 2022 to a sample of 1,250 scientists. A total of 592 individuals responded, resulting in a response rate of 47%. The individuals were preselected based on the following criteria:

1. They had published at least one highly cited article (top 10 percentile rank), one article with an intermediate number of citations (within top 10–50 percentile rank), and one less-cited article (bottom 50 percentile rank) between 2015 and 2018 as either the first or the last author.
2. They were affiliated with institutions in Norway at the time when the survey was conducted.

The survey asked researchers to rate a randomly selected paper of theirs from each of the three above-mentioned citation categories (high, intermediate, or low) in regard to four dimensions of research quality: scientific importance, novelty/originality, solidity, and societal impact (Table 2). In addition, the respondents were asked to state whether they would consider the respective article as "groundbreaking research." The survey intentionally did not define "groundbreaking" in order to capture respondents' own interpretations of the term, which we may vary.

The respondents rated the quality of their publications on a scale from 1 (lowest) to 5 (highest). As only a few researchers gave their own work a rating of "1" or "2" on any of the quality

Table 2. Survey questions and answer alternatives

Question	Answer alternatives
How do you regard the ***scientific importance*** of the research reported in these articles (e.g., in terms of new discoveries/findings, theoretical developments, new analytic techniques)?	1 (low)–2–3–4–5 (high)-Cannot say/Not relevant
How do you regard the ***novelty/originality*** of the research reported in these articles (e.g., in terms of topic addressed, research question, methodology, and results)?	
How do you regard the ***solidity*** of the research reported in these articles (i.e., validity and certainty/reliability of the methods and results reported)?	
As far as you know, have the research/results presented in these articles had any ***societal impact*** (i.e., effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia)?	Yes-No-Don't know
Would you consider your article as "groundbreaking research"?	
Please characterize the main contribution(s) from these articles (multiple answers possible)	Theoretical-Empirical-Methodological-Review-Cannot say/Not relevant

dimensions, we binarily recoded the answer so that the ratings "1," "2," and "3" are coded as "low" and the ratings "4" and "5" are coded as "high". We also combined the "No" and "Don't know" answers to the question asking about societal impact because it is not possible to know with certainty that research has not had any societal impact (the researcher may simply not be aware of it). Therefore, it is not so important to distinguish between "No" and "Don't know" in this case (Aksnes et al., 2023). The survey also provides information about whether the publication has a methodological, theoretical, or empirical focus.

3.3. Study Design

In this section, we explicate the assumptions about the causal structure of our bibliometric and survey data that informed our research design. To this end, we used directed acyclic graphs (DAGs) following the recommendations of Elwert (2013) and Rohrer (2018). Figure 2 shows a visual representation of the causal structure of our data.

The DAG on the left side of Figure 2 concerns the relationship between (four dimensions of) research quality and groundbreaking research. We assume that researchers consider the quality of a publication when deciding whether or not the publication can be considered as groundbreaking. Because the assessments depend (at least partially) on subjective judgments, it is likely that the association between research quality and groundbreaking research is confounded by (unobserved) author characteristics. For example, some researchers might have a more favorable opinion of their own work than others. The multi-level structure of our data enables us to get rid of person-specific heterogeneity by performing fixed effects regressions (Allison, 2009; Brüderl & Ludwig, 2014). By using author fixed effects, we avoid potential omitted variables bias. To estimate which dimensions of research quality are predictors of groundbreaking research, we run fixed effects logistic regression.

The DAG on the right side of Figure 2 illustrates how we test the convergent validity of disruption scores with researchers' self-assessments. The test rests on the assumption that researchers possess the expertise required to accurately assess whether or not a publication is groundbreaking. This assumption cannot be tested empirically because the true epistemic features of research publications are unobservable. The empirical test of convergent validity requires the elimination of confounders. Confounders could cause an association between disruption scores and self-assessments independent of a publication's epistemic properties. When selecting the set of control variables, we focused on features of publications that are likely to

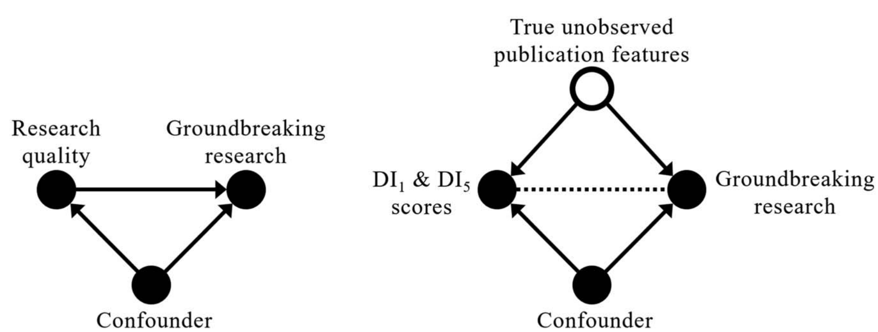

Figure 2. Overview of the assumed causal structure of our data. Arrows symbolize causation, and the dotted line represents the correlation between researcher's self-assessments and disruption scores.

affect citation patterns. Because publications accrue (disruptive) citations over time, we control for publication years. Both the self-assessments of researchers and the disruption scores of publications may depend on discipline-specific evaluation and citation cultures. Therefore, we control for discipline (at the publication level). Additionally, it seems likely that different types of contribution are also cited differently. More specifically, Leahey, Lee, and Funk (2023) show that methodological papers tend to produce more disruptive citation patterns than theoretical or empirical papers. Thus, we control for the type of contribution.

We include the (logarithmic) number of references cited by the FP's authors as a control variable to account for the technical limitations of disruption indices. Because the DI_1 and DI_5 are measures of betweenness centrality, they tend to assign extreme scores to papers with very few cited references (Liu et al., 2023; Ruan et al., 2021). Consequently, it is difficult to compare the disruption scores of papers with different numbers of cited references. Due to the skewed distribution of the number of linked references, the variable enters the regression in logarithmic specifications. Lastly, we eliminate the influence of author characteristics on our estimates via author fixed effects. We assume that our control variables are either confounders (because they are correlated with both the self-assessments and disruption scores) or "neutral controls" (Cinelli, Forney, & Pearl, 2022, p. 1077)—because they only correlate with disruption scores. Neutral controls are control variables that do not affect the coefficient of groundbreaking research on DI_1 and DI_5 scores, but may help to improve the precision of the statistical models.

3.4. Sample Description

The original sample by Aksnes et al. (2023) consisted of 1,776 publications by 592 authors. In our analysis, we had to exclude a number of publications because of the limitations of our bibliometric data and the limitations of the disruption indices. A total of 221 publications could not be included in the analysis because they are not indexed in the Max Planck Society's in-house version of the Web of Science (WoS, Clarivate). In addition, we removed publications that were classified as reviews by their authors in the survey. Disruption scores may not be valid for reviews: In spite of being consolidating by nature, reviews may receive high disruption scores if they are often cited independently of their cited references. Reviews may disrupt citation couplings, but only because citing authors prefer not to cite the source for a piece of research twice: by citing both the primary and the secondary literature.

We removed publications for which no meaningful disruption scores are possible due to a lack of data. Uncited publications were removed because their disruption score is zero by definition. The probability of consolidating citations (measured by N_B and N_R) depends on the number of references cited by an FP. As a result, the DI_1 and DI_5 tend to assign high scores to FPs with very few cited references regardless of whether these publications are actually disruptive. The issue of artificially inflated disruption scores is further compounded by the fact that bibliometric databases provide only limited coverage of scientific publications. An FP may have only a small number of cited references, not because it cites very few documents, but because the cited documents are not indexed in the respective bibliometric database. High disruption scores that are assigned to FPs due to a lack of coverage of cited references are data artefacts with no meaningful interpretation. Holst et al. (2024), Macher, Rutzer, and Weder (2024), and Ruan et al. (2021) demonstrate that artificially inflated disruption scores resulting from data artefacts are common and may severely bias research outcomes.

In the WoS data, we can quantify the number of missing cited references for every FP. Linked references are indexed in WoS as source items, whereas missing references are not indexed as source items and cannot be considered in the calculation of disruption scores.

An FP's total number of cited references is the sum of its linked references and its missing references. In order to mitigate potential coverage-related biases, we restricted our analyses to FPs with at least 10 cited references. Disruption scores may be biased not only by publications with a low *absolute number* of linked references, but also by publications with a low *proportion* of references indexed in the database. If a high proportion of an FP's total number of cited references is missing, this may result in the systematic underestimation of N_B and N_B. To ensure the robustness of our analyses against this potential threat, we performed a robustness check with a highly selective sample of publications: In the robustness check, we only included papers with at least 10 citations, at least 10 linked references and a minimum of 75% references indexed in the database. In both the main analysis and in the robustness check, we only considered publications by respondents who authored at least two publications in the sample. The inclusion of respondents with only one publication in the fixed effects estimation could lead to an overestimation of statistical significance.

The final sample consists of 1,214 publications by 459 authors. Table 3 shows descriptive statistics. Figure 3 shows that most of the papers in our sample receive disruption scores that are very close to zero. The middle 50% of the research articles in our sample receive DI_1 scores between 0 and −0.005. The extreme concentration of disruption scores is a common finding in the literature and has been reported for many other data sets (Bittmann et al., 2022; Bornmann & Tekles, 2021; Leydesdorff et al., 2021). To avoid small decimal values and improve the readability of tables, we multiplied the DI_1 and DI_5 scores by 100 in subsequent tables and figures. Because disruption scores are based on citation data, they may change over time as long as an FP keeps accruing citations. Consequently, disruption scores may not reach a stable value until a few years after publication. Bornmann and Tekles (2019a) recommend a citation window of at least 3 years for the calculation of disruption scores. In our study, disruption scores are calculated with a 5-year citation window. As our bibliometric data are from calendar week 17 of 2024 and the most recently published papers in our sample are from 2018, we cannot use a citation window of more than 5 years for papers from 2018.

Table 3. Descriptive statistics

Variable	Min	Max	*M*	*SD*
Publication year	2015	2018	2,016.585	1.109
Number of linked references	10	157	35.518	18.372
Number of missing references	0	112	12.563	11.944
DI_1	−14.388	16.129	−0.407	1.130
DI_5	−9.506	28.571	0.075	1.232
Groundbreaking research	0	1	0.269	
Scientific importance	0	1	0.687	
Societal impact	0	1	0.261	
Novelty/Originality	0	1	0.726	
Solidity	0	1	0.871	

Note: n = 1,214; M = mean; SD = standard deviation. DI_1 and DI_5 scores were multiplied by 100 to avoid small decimal values.

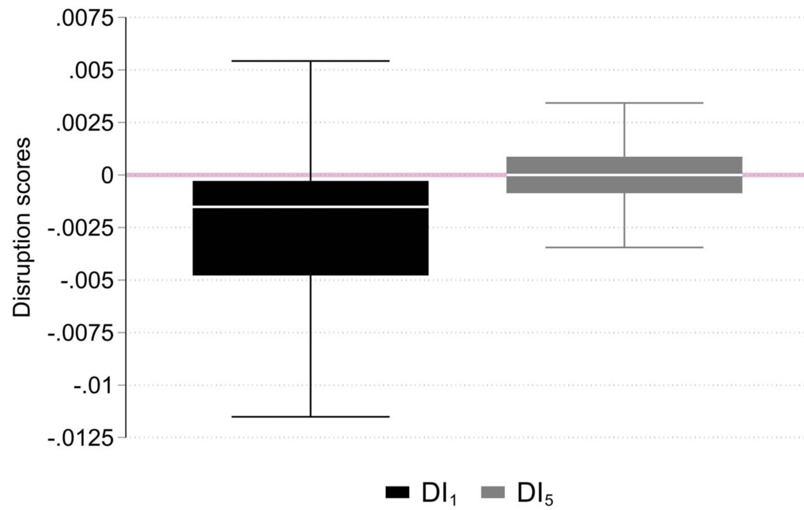

Figure 3. Box plots of DI_1 and DI_5 scores. The boxes show the interquartile range (IQR), which is the middle 50% of the data. The median is displayed by the white line. The whiskers extend to 1.5 × IQR. Values beyond 1.5 × IQR are not displayed in the figure to focus on the most common values.

Table 4 shows the distribution of standard and groundbreaking publications across disciplines and types of works. WoS's coverage of Norwegian scholarly and scientific output is substantially better in the health sciences and natural sciences compared to the social sciences and the humanities. For example, Aksnes and Sivertsen (2019) show that WoS covers 87% of Norwegian scientific output in the health sciences, but only 23% of the output in the humanities. As we only include papers with sufficient coverage of their cited references, the sample is mainly comprised of papers from disciplines with better WoS coverage. As a result, the majority (about 58%) of the papers in our sample belong to medicine or the natural sciences. Papers from the humanities or the social sciences constitute less than 12% of the sample.

Table 5 shows the distribution of (groundbreaking) publications by type of contribution. The type of contribution was assigned to the FPs by their authors in the survey. If authors assigned more than one type of contribution to an FP, its type of contribution is coded as "multiple."

3.5. Regression Models

We calculated two fixed effects regression models in this study (see Eqs. 1 and 2). In both models, α_i and a_i refer to the author-specific error terms. i is an index for the authors (i = 1, …, N_{authors}) and j is an index for their publications (j = 1, …, $N_{\text{publications}}$). Eq. 1 shows the fixed effects logistic regression model that predicts the likelihood that a paper is considered as

Table 4. Distribution of (groundbreaking) publications across disciplines

	Health	Hum.	Med.	Nat. sci.	Soc. sci.	Tech.	Total
Groundbreaking							
No	177	10	263	247	92	98	887
Yes	38	4	96	93	36	60	327
Total	215	14	359	340	128	158	1,214

Note: Hum = Humanities, Med. = Medicine, Nat. sci. = Natural sciences, Soc. sci. = Social sciences, Tech. = Technology.

Table 5. Distribution of (groundbreaking) publications by type of contribution

	Theoretical	Empirical	Methodological	Multiple	Total
Groundbreaking					
No	101	445	84	257	887
Yes	50	117	40	120	327
Total	151	562	124	377	1,214

groundbreaking depending on four dimensions of research quality. *Quality* refers to the dummy variables for high values on scientific importance, novelty/originality, societal impact, and solidity. Eq. 2 explicates the fixed effects regression model that tests whether disruption scores covary with researchers' self-assessments. *C* are the control variables.

$$\text{logit}\big(\text{P}\big(\text{Groundbreaking research}_{ij} = 1\big)\big) = \alpha_i + \beta_1 \text{Scientific importance}_{ij} + \beta_2 \text{Novelty}_{ij} + \beta_3 \text{Solidity}_{ij} + \beta_4 \text{Societal impact}_{ij} + \epsilon_{ij} \quad (1)$$

$$\text{Disruption score}_{ij} = b_0 + a_i + b_1 \text{Groundbreaking research}_{ij} + \beta_C C_{ij} + \epsilon_{ij} \quad (2)$$

4. RESULTS

4.1. Which Aspects of Research Quality Make the Difference Between Standard and Groundbreaking Research?

We performed logistic regressions to analyze how the scientific community considers groundbreaking research in terms of research quality dimensions. Table 6 (in the Appendix) shows the results from a standard logistic regression and Table 7 (in the Appendix) shows the results from the fixed effects logistic regression. The standard logistic regression does not consider author characteristics. The fixed effects logistic regression controls for author characteristics (e.g., gender) via author-specific error terms. In Figure 4, the estimates from both models are converted

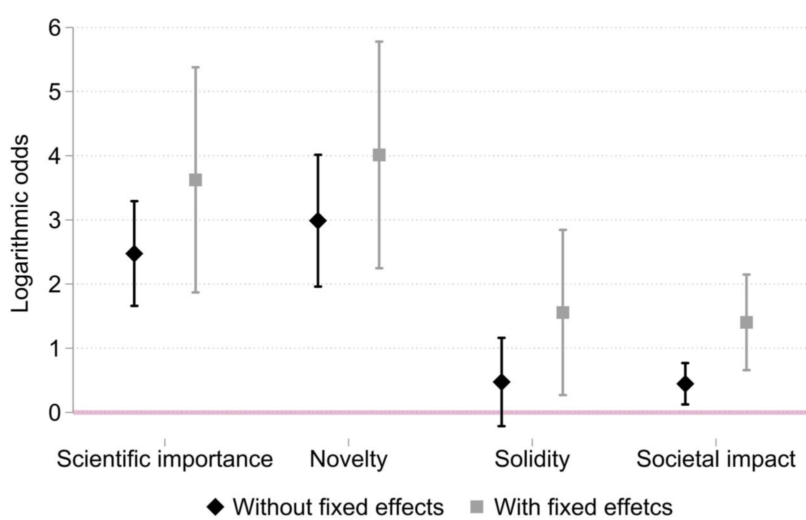

Figure 4. Coefficient plot: logarithmic odds from multivariate logistic regression models with and without fixed effects. Coefficients show the increase in the logarithmic odds that a paper is considered as groundbreaking compared to the reference category.

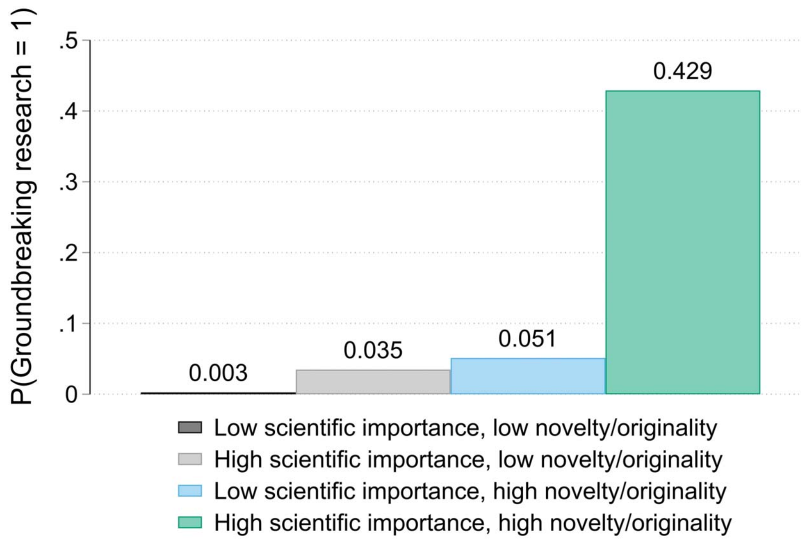

Figure 5. Predicted probabilities from a logistic regression model (without fixed effects). The bars show the probability that a paper is considered as groundbreaking research depending on scientific importance and novelty/originality.

into logarithmic odds and displayed side by side to enable a visual comparison of the results. The estimates from the fixed effects logistic regression are larger than the estimates from the standard logistic regression and have large confidence intervals because this model discards a lot of data: It considers only those authors with at least one standard publication and at least one groundbreaking publication.

In both models, scientific importance and novelty/originality are more strongly associated with groundbreaking research than solidity and societal impact. The differences between the logarithmic odds that a paper is considered as groundbreaking suggest the following: Scientific significance and aspects of novelty, but not societal impact or methodological solidity, make the difference between groundbreaking and standard research. For the purpose of a more tangible interpretation of our results, Figure 5 shows the predicted probability that an FP is considered as groundbreaking depending on its scientific importance and novelty[4]. The probabilities are calculated with the standard logistic regression model (without fixed effects). The probability to be considered as groundbreaking is at least 38% percentage points higher for FPs that are both highly original and highly important compared to papers that received a low rating on at least one of the two quality dimensions. Papers with a low rating on either “novelty/originality” or “scientific importance” have no more than a 5% chance of being considered as groundbreaking research. It seems that novelty and exceptional scientific importance are necessary, but not sufficient, features of groundbreaking research.

4.2. Convergent Validity of DI_1 and DI_5 Scores With Researchers’ Self-Assessments

We tested the convergent validity of the DI_1 and the DI_5 with researchers’ self-assessments by running multivariate fixed effects regressions of the dummy variable “groundbreaking

[4] The figures in Section 4 were created in Stata with the graphic schemes provided by Bischof (2017). In addition, we used *coefplot* and *estout* provided by Jann (2005, 2014) to create coefficient plots and regression tables.

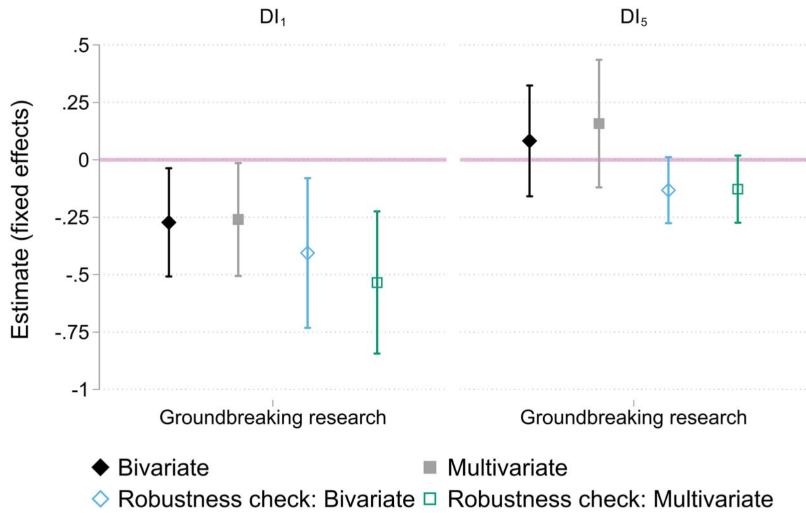

Figure 6. Coefficient plot: coefficients and confidence intervals from eight fixed-effects regressions of "groundbreaking research" on DI_1 and DI_5 scores. The DI_1 and DI_5 scores were multiplied by 100 to avoid small decimal values.

research" on disruption scores. The results of the main analysis and the robustness check are displayed in Tables 8 and 9 (in the Appendix). Figure 6 shows the coefficients and confidence intervals from the bivariate and multivariate models as well as the results from the robustness check. The coefficients indicate whether the average disruption scores of groundbreaking research are higher or lower than the average disruption scores of standard research. The focus is on the sign and the size of the coefficients from the multivariate models.

Across all models, we found that the average DI_1 scores of groundbreaking papers are lower than the average DI_1 scores of standard publications. This negative association is statistically significant. The DI_5 performs only slightly better than the DI_1. Although in the main analysis, the DI_5 displays a positive, but statistically not significant association with groundbreaking research, the coefficients of DI_5 scores on groundbreaking research are consistently small (and possibly negligible) across all models. Both the DI_1 and the DI_5 perform substantially worse in the robustness check in comparison to the main analysis. When the sample is restricted to publications with at least 10 linked references, 10 citations and a minimum 75% proportion of linked cited references, the DI_5 is no longer positively correlated with researchers' self-assessments.

In summary, none of the models we tested predicted a robust and relevant positive association between disruption scores and self-assessments of researchers[5]. Our results suggest that the DI_1 and DI_5 are mostly unable to distinguish with sufficient precision between groundbreaking and standard research. It seems that one cannot expect high disruption scores if research has been characterized as groundbreaking by researchers.

[5] We performed additional analyses with less common variants of the disruption index: The mDI_1 and mDI_5 proposed by Funk and Owen-Smith (2017) as well as the DI_1^{noR} and DI_5^{noR} proposed by Wu and Yan (2019). We applied the same regression model we used to test the convergent validity of the DI_1 and DI_5. The results from the additional indicators, which performed either similar or worse than the DI_1 and DI_5, are displayed in Figures 7 and 8 in the Appendix.

5. DISCUSSION

Funk and Owen-Smith (2017) and Wu et al. (2019) convincingly argued that measures of citation impact miss certain key features of technological and epistemic innovations. The fact that a paper is highly cited does not reveal whether it consolidates or disrupts the status quo. In this study, we found that the converse may also be true: The fact that a paper is (or is not) disruptive (as measured by the DI_1 and the DI_5) provides little information about the importance of its contribution to scientific progress, as assessed by the authors themselves. Our results show that groundbreaking research is characterized by high scientific importance and their originality, but not by elevated disruption scores. In this study, papers perceived as groundbreaking are, on average, no more disruptive than standard publications.

A possible explanation for our findings is that disruption indices capture only a (very) small subset of groundbreaking research papers—namely, those that displace prior literature (Lin et al., 2026). Most scientific breakthroughs extend prior research trajectories rather than challenge them outright and therefore do not produce *disruptive* citation patterns. Future research could thus investigate the link between the epistemic content of publications and the citations patterns they generate.

The properties of our data have implications for the external and internal validity of our empirical results. External validity is limited by the fact that our sample only includes researchers affiliated with Norwegian research institutions. By design, the survey oversampled experienced scientists who have published highly cited papers (Aksnes et al., 2023). The internal validity of our study rests on the assumption that researchers accurately assessed the epistemic properties of their own work in the survey. Self-assessments come with both advantages and disadvantages in comparison to the independent expert judgments used in previous studies. On the one hand, researchers are intimately familiar with their own work, and thus possess the knowledge required to assess the scientific merit of their contributions. On the other hand, self-assessment might be more biased towards positive judgments than evaluations by independent experts. As only 19 publications in our sample were assessed by more than one respondent, we could not calculate a meaningful empirical estimate of interrater reliability.

Our study relies on not only the validity of our survey data but also the accuracy of our bibliometric data. To minimize the risk that our results might be biased by WoS's incomplete coverage of the science literature, we performed a robustness check with a subsample of high-quality data. Both the DI_1 and the DI_5 performed worse in the robustness check than in the main analysis, thus confirming the impression that the validity of disruption indices depends on the reliability and quality of the underlying data.

The development of the DI_1, which combines citing and cited data of FPs to assess the disruptive and consolidating nature of research, represents an influential innovation in the field of scientometrics. Since its introduction, numerous studies have examined the properties and validity of the DI_1 and have proposed a variety of alternative or modified indices. Our study reveals that the DI_1 and DI_5 scores do not correspond to self-assessments of authors about the groundbreaking nature of their papers. This discrepancy may reflect measurement limitations of the indices themselves, or limitations of authors' self-assessments as an evaluative instrument.

Although groundbreaking and disruptive research share many similar characteristics, they are not identical. Disruptiveness highlights the displacement of established knowledge or practices, whereas "groundbreaking" is broader and includes major or foundational contributions without necessarily displacing prior work. As the questionnaire did not provide a specific definition of "groundbreaking," respondents may have applied varying interpretations when answering. Similarly, the interpretation and labeling of bibliometric indices as measures of "disruptiveness" could be debated, and other related concepts may also be relevant. Thus, the issue of comparability and construct validity is complex, with our approach assessing only one dimension of it.

Based on our results and the current state of research on disruption indices, it does not seem advisable to use the DI_1 or the DI_5 as measures of groundbreaking scientific achievements in research evaluation practice. Further research is needed to determine whether—and how—bibliometric metadata can be used to capture epistemic change.

ACKNOWLEDGMENTS

An earlier version of this paper was presented at the 28th International Conference on Science, Technology and Innovation Indicators. We thank the conference participants for their helpful comments and suggestions.

AUTHOR CONTRIBUTIONS

Christian Leibel: Conceptualization, Formal analysis, Methodology, Validation, Visualization, Writing—original draft, Writing—review & editing. Alexander Tekles: Software, Writing—review & editing. Dag W. Aksnes: Investigation, Writing—review & editing. Lutz Bornmann: Conceptualization, Writing—review & editing.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

Part of this research was funded by the Research Council of Norway, grant number 256223 (the R-QUEST Centre).

This study uses data from the IVSDB, which is developed through a collaboration between the Administrative Headquarters of the Max Planck Society and the Information Retrieval Service for the Institutes of the Chemistry, Physics, and Technology (IVS-CPT) Section of the Max Planck Society. The data are drawn from a bibliometric database provided by the German *Kompetenznetzwerk Bibliometrie*, which has been funded since December 2008 by the Federal Ministry of Research, Technology and Space of Germany (BMFTR) and is currently supported under grant numbers 16WIK2101A, 16WIK2101B, and 16WIK2101C (see https://www.bibliometrie.info/).

DATA AVAILABILITY

Data are not available. The participants of this study did not give written consent for their data to be shared publicly. Bibliographic record data cannot be released due to copyright/license restrictions.

REFERENCES

Aksnes, D. W., Piro, F. N., & Fossum, L. W. (2023). Citation metrics covary with researchers' assessments of the quality of their works. *Quantitative Science Studies*, *4*(1), 105–126. https://doi.org/10.1162/qss_a_00241

Aksnes, D. W., & Sivertsen, G. (2019). A criteria-based assessment of the coverage of Scopus and Web of Science. *Journal of Data and Information Science*, *4*(1), 1–21. https://doi.org/10.2478/jdis-2019-0001

Allison, P. D. (2009). *Fixed effects regression models*. Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781412993869

Bischof, D. (2017). New graphic schemes for Stata: Plotplain and plottig. *Stata Journal*, *17*(3), 748–759. https://doi.org/10.1177/1536867X1701700313

Bittmann, F., Tekles, A., & Bornmann, L. (2022). Applied usage and performance of statistical matching in bibliometrics: The

comparison of milestone and regular papers with multiple measurements of disruptiveness as an empirical example. *Quantitative Science Studies*, *2*(4), 1246–1270. https://doi.org/10.1162/qss_a_00158
Bornmann, L., Devarakonda, S., Tekles, A., & Chacko, G. (2020). Are disruption index indicators convergently valid? The comparison of several indicator variants with assessments by peers. *Quantitative Science Studies*, *1*(3), 1242–1259. https://doi.org/10.1162/qss_a_00068
Bornmann, L., & Tekles, A. (2019a). Disruption index depends on length of citation window. *Profesional de la Informacion*, *28*(2), e280207. https://doi.org/10.3145/epi.2019.mar.07
Bornmann, L., & Tekles, A. (2019b). Disruptive papers published in *Scientometrics*. *Scientometrics*, *120*, 331–336. https://doi.org/10.1007/s11192-019-03113-z
Bornmann, L., & Tekles, A. (2021). Convergent validity of several indicators measuring disruptiveness with milestone assignments to physics papers by experts. *Journal of Informetrics*, *15*(3), 101159. https://doi.org/10.1016/j.joi.2021.101159
Bornmann, L., Wu, L., & Ettl, C. (2024). The use of ChatGPT for identifying disruptive papers in science: A first exploration. *Scientometrics*, *129*, 7161–7165. https://doi.org/10.1007/s11192-024-05176-z
Broad, W. J. (2023). What happened to all of science's big breakthroughs? *New York Times*, January 17. https://www.nytimes.com/2023/01/17/science/science-breakthroughs-disruption.html
Brüderl, J., & Ludwig, V. (2014). Fixed-effects panel regression. In H. Best & C. Wolf (Eds.), *The Sage handbook of regression analysis and causal inference* (pp. 327–358). Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781446288146.n15
Cinelli, C., Forney, A., & Pearl, J. (2022). A crash course in good and bad controls. *Sociological Methods & Research*, *53*(3), 1071–1104. https://doi.org/10.1177/00491241221099552
Cole, S. (1983). The hierarchy of the sciences? *American Journal of Sociology*, *89*(1), 111–139. https://doi.org/10.1086/227835
Deng, N., & Zeng, A. (2023). Enhancing the robustness of the disruption metric against noise. *Scientometrics*, *128*, 2419–2428. https://doi.org/10.1007/s11192-023-04644-2
Elwert, F. (2013). Graphical causal models. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 245–273). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-6094-3_13
Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, *40*(1), 35–41. https://doi.org/10.2307/3033543
Funk, R. J., & Owen-Smith, J. (2017). A dynamic network measure of technological change. *Management Science*, *63*(3), 791–817. https://doi.org/10.1287/mnsc.2015.2366
Gebhart, T., & Funk, R. J. (2023). A mathematical framework for citation disruption. *arXiv*. https://doi.org/10.48550/arXiv.2308.16363
Hart, R. (2023). Where are all the scientific breakthroughs? Forget AI, nuclear fusion and mRNA vaccines, advances in science and tech have slowed, major study says. *Forbes*, January 4. https://www.forbes.com/sites/roberthart/2023/01/04/where-are-all-the-scientific-breakthroughs-forget-ai-nuclear-fusion-and-mrna-vaccines-advances-in-science-and-tech-have-slowed-major-study-says/
Holst, V., Algaba, A., Tori, F., Wenmackers, S., & Ginis, V. (2024). Dataset artefacts are the hidden drivers of the declining disruptiveness in science. *arXiv*. https://doi.org/10.48550/arXiv.2402.14583
Jann, B. (2005). Making regression tables from stored estimates. *Stata Journal*, *5*(3), 288–308. https://doi.org/10.1177/1536867X0500500302
Jann, B. (2014). Plotting regression coefficients and other estimates. *Stata Journal*, *14*(4), 708–737. https://doi.org/10.1177/1536867X1401400402
Kamerlin, S. C. L. (2023). The perceived decline of "disruptive" science and technology. *EMBO Reports*, *24*(3), e56876. https://doi.org/10.15252/embr.202356876, PubMed: 36749594
Kreiman, G., & Maunsell, J. H. R. (2011). Nine criteria for a measure of scientific output. *Frontiers in Computational Neuroscience*, *5*, 48. https://doi.org/10.3389/fncom.2011.00048, PubMed: 22102840
Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
Kuhn, T. S. (1970). Reflections on my critics. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965* (Vol. 4, pp. 231–278). Cambridge University Press. https://doi.org/10.1017/CBO9781139171434.011
Leahey, E., Lee, J., & Funk, R. J. (2023). What types of novelty are most disruptive? *American Sociological Review*, *88*(3), 562–597. https://doi.org/10.1177/00031224231168074
Leibel, C., & Bornmann, L. (2024). What do we know about the disruption index in scientometrics? An overview of the literature. *Scientometrics*, *129*, 601–639. https://doi.org/10.1007/s11192-023-04873-5
Leydesdorff, L., & Bornmann, L. (2021). Disruption indices and their calculation using web-of-science data: Indicators of historical developments or evolutionary dynamics? *Journal of Informetrics*, *15*(4), 101219. https://doi.org/10.1016/j.joi.2021.101219
Leydesdorff, L., Tekles, A., & Bornmann, L. (2021). A proposal to revise the disruption index. *Profesional de la Informacion*, *30*(1), e300121. https://doi.org/10.3145/epi.2021.ene.21
Lin, Y., Frey, C. B., & Wu, L. (2023). Remote collaboration fuses fewer breakthrough ideas. *Nature*, *623*(7989), 987–991. https://doi.org/10.1038/s41586-023-06767-1, PubMed: 38030778
Lin, Y., Li, L., & Wu, L. (2026). The disruption index measures displacement between a paper and its most cited reference. *Quantitative Science Studies*, *7*, 229–239. https://doi.org/10.1162/QSS.a.409
Liu, X., Zhang, C., & Li, J. (2023). Conceptual and technical work: Who will disrupt science? *Journal of Informetrics*, *17*(3), 101432. https://doi.org/10.1016/j.joi.2023.101432
Macher, J. T., Rutzer, C., & Weder, R. (2024). Is there a secular decline in disruptive patents? Correcting for measurement bias. *Research Policy*, *53*(5), 104992. https://doi.org/10.1016/j.respol.2024.104992
Mayor, M., & Queloz, D. (1995). A Jupiter-mass companion to a solar-type star. *Nature*, *378*(6555), 355–359. https://doi.org/10.1038/378355a0
Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, *613*(7942), 138–144. https://doi.org/10.1038/s41586-022-05543-x, PubMed: 36600070
Petersen, A. M., Arroyave, F. J., & Pammolli, F. (2024). The disruption index is biased by citation inflation. *Quantitative Science Studies*, *5*(4), 936–953. https://doi.org/10.1162/qss_a_00333
Petersen, A. M., Arroyave, F. J., & Pammolli, F. (2025). The disruption index suffers from citation inflation: Re-analysis of temporal CD trend and relationship with team size reveal discrepancies. *Journal of Informetrics*, *19*(1), 101605. https://doi.org/10.1016/j.joi.2024.101605
Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42. https://doi.org/10.1177/2515245917745629

Ruan, X., Lyu, D., Gong, K., Cheng, Y., & Li, J. (2021). Rethinking the disruption index as a measure of scientific and technological advances. *Technological Forecasting and Social Change*, *172*, 121071. https://doi.org/10.1016/j.techfore.2021.121071
Shibayama, S., & Wang, J. (2020). Measuring originality in science. *Scientometrics*, *122*, 409–427. https://doi.org/10.1007/s11192-019-03263-0
Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468–472. https://doi.org/10.1126/science.1240474, PubMed: 24159044
Wang, S., Ma, Y., Mao, J., Bai, Y., Liang, Z., & Li, G. (2023). Quantifying scientific breakthroughs by a novel disruption indicator based on knowledge entities. *Journal of the Association for Information Science and Technology*, *74*(2), 150–167. https://doi.org/10.1002/asi.24719
Wei, C., Li, J., & Shi, D. (2023). Quantifying revolutionary discoveries: Evidence from Nobel prize-winning papers. *Information Processing & Management*, *60*(3), 103252. https://doi.org/10.1016/j.ipm.2022.103252
Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, *566*(7744), 378–382. https://doi.org/10.1038/s41586-019-0941-9, PubMed: 30760923
Wu, Q., & Yan, Z. (2019). Solo citations, duet citations, and prelude citations: New measures of the disruption of academic papers. *arXiv*. https://doi.org/10.48550/arXiv.1905.03461
Wuestman, M., Hoekman, J., & Frenken, K. (2020). A typology of scientific breakthroughs. *Quantitative Science Studies*, *1*(3), 1203–1222. https://doi.org/10.1162/qss_a_00079
Yang, A. J., Hu, H., Zhao, Y., Wang, H., & Deng, S. (2023). From consolidation to disruption: A novel way to measure the impact of scientists and identify laureates. *Information Processing & Management*, *60*(5), 103420. https://doi.org/10.1016/j.ipm.2023.103420

APPENDIX

Table 6. Odds ratios from a multivariate logistic regression of four dimensions of research quality on "groundbreaking research"

Variable	OR	*SE*	95% CI *LL*	95% CI *UL*	*p*
Scientific importance					
Low	Reference				
High	11.89	4.951	5.26	26.89	<0.001
Novelty/originality					
Low	Reference				
High	19.84	10.409	7.09	55.48	<0.001
Solidity					
Low	Reference				
High	1.61	0.564	0.81	3.20	0.177
Societal impact					
No	Reference				
Yes	1.56	0.258	1.13	2.16	0.007
Pseudo R^2	0.23				
N	1,214				

Note: OR = odds ratio; *SE* = standard error; CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

Table 7. Odds ratios from a multivariate fixed effects logistic regression of four dimensions of research quality on "groundbreaking research"

Variable	OR	*SE*	95% CI *LL*	95% CI *UL*	*p*
Scientific importance					
Low	Reference				
High	37.50	33.55	6.49	216.56	<0.001
Novelty/originality					
Low	Reference				
High	55.30	49.78	9.47	322.82	<0.001
Solidity					
Low	Reference				
High	4.74	3.12	1.31	17.19	0.018
Societal impact					
No	Reference				
Yes	4.07	1.55	1.93	8.58	<0.001
N	583				

Note: OR = odds ratio; *SE* = standard error; CI = confidence interval; *LL* = lower limit; *UL* = upper limit. The sample was restricted to papers from authors with at least one groundbreaking publication and at least one standard publication.

Table 8. Results from two separate fixed effects regressions of "groundbreaking research" on DI_1 and DI_5 scores

	DI_1			DI_5		
Variable	Estimate	*SE*	*p*	Estimate	*SE*	*p*
Groundbreaking research						
No	Reference			Reference		
Yes	−0.26	0.13	0.039	0.16	0.14	0.264
Type of contribution						
Theoretical	Reference			Reference		
Empirical	−0.31	0.22	0.168	−0.30	0.16	0.061
Methodological	0.17	0.32	0.596	0.21	0.38	0.574
Multiple	−0.55	0.34	0.107	−0.61	0.33	0.063
Discipline						
Health	Reference			Reference		
Humanities	0.26	0.32	0.407	0.29	0.21	0.168
Medicine	0.04	0.14	0.768	−0.10	0.19	0.597
Natural sciences	0.13	0.15	0.402	−0.07	0.14	0.645
Social sciences	0.42	0.28	0.144	−0.00	0.14	0.996
Technology	0.43	0.33	0.195	0.64	0.38	0.094
Publication year						
2015	Reference			Reference		
2016	0.17	0.11	0.126	−0.01	0.12	0.905
2017	0.13	0.11	0.228	0.09	0.13	0.516
2018	0.20	0.12	0.093	0.11	0.11	0.285
Logarithmic number of linked references	−0.18	0.12	0.144	−0.60	0.16	<0.001
Constant	0.30	0.48	0.524	2.34	0.57	<0.001
R^2	0.05			0.08		
N	1,214			1,214		

Note: *SE* = standard error. The sample was restricted to papers with at least 10 linked cited references. The DI_1 and DI_5 scores were multiplied by 100 to avoid small decimal values.

Table 9. Robustness check. Results from two separate fixed effects regressions of "groundbreaking research" on DI_1 scores and DI_5 scores

	DI_1			DI_5		
Variable	Estimate	*SE*	*p*	Estimate	*SE*	*p*
Groundbreaking research						
No	Reference			Reference		
Yes	−0.53	0.16	<0.001	−0.13	0.07	0.088
Type of contribution						
Theoretical	Reference			Reference		
Empirical	−0.30	0.39	0.443	−0.27	0.28	0.327
Methodological	−0.19	0.44	0.670	−0.49	0.35	0.164
Multiple	−0.35	0.37	0.339	−0.32	0.25	0.214
Discipline						
Health	Reference			Reference		
Humanities	0.29	0.42	0.485	0.11	0.11	0.332
Medicine	0.52	0.30	0.085	−0.01	0.11	0.922
Natural sciences	0.35	0.39	0.377	−0.04	0.16	0.788
Social sciences	3.97	1.87	0.035	0.39	0.49	0.427
Technology	0.59	0.80	0.457	0.46	0.33	0.165
Publication year						
2015	Reference			Reference		
2016	−0.09	0.15	0.532	−0.10	0.08	0.189
2017	−0.27	0.21	0.211	−0.05	0.09	0.564
2018	−0.02	0.13	0.901	0.09	0.07	0.172
Logarithmic number of linked references	0.42	0.20	0.034	−0.18	0.11	0.096
Constant	−2.00	0.93	0.033	0.95	0.54	0.082
R^2	0.20			0.10		
N	416			416		

Note: *SE* = standard error. The sample was restricted to FPs with at least 10 linked references and at least 10 citations. Additionally, at least 75% of the FP's cited references need to be linked references. The DI_1 and DI_5 scores were multiplied by 100 to avoid small decimal values.

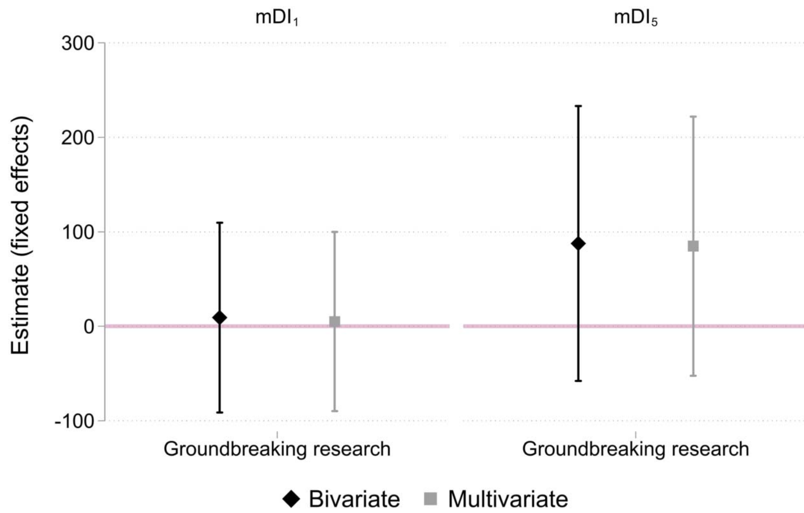

Figure 7. Coefficient plot: coefficients and confidence intervals from four fixed-effects regressions of "groundbreaking research" on mDI_1 and mDI_5 scores. The mDI_1 and mDI_5 scores were calculated by multiplying the DI_1 and DI_5 scores of the FP with its citation count 5 years after publication. For the sake of consistency, the mDI_1 and mDI_5 scores were multiplied by 100. The mDI_1 was proposed by Funk and Owen-Smith (2017) and gives more weight to highly cited publications.

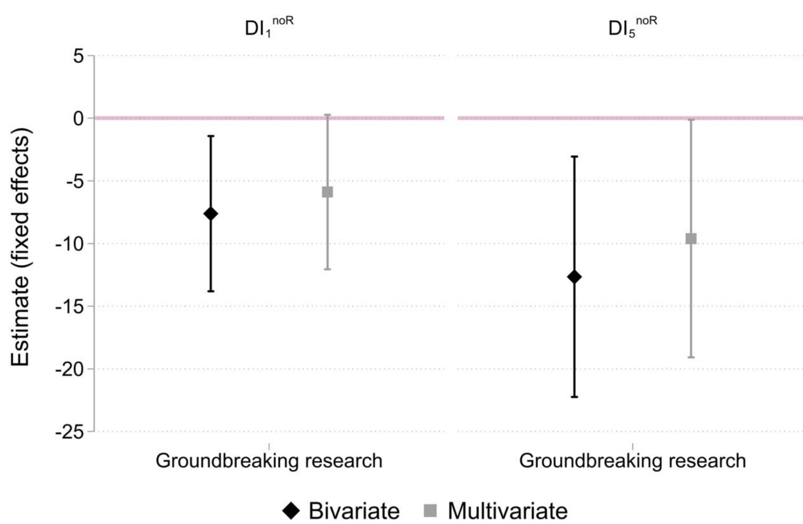

Figure 8. Coefficient plot: coefficients and confidence intervals from four fixed effects regressions of "groundbreaking research" on DI_1^{noR} and DI_5^{noR} scores. The DI_1^{noR} and DI_5^{noR} scores were multiplied by 100 to avoid small decimal values. The DI_1^{noR} was proposed by Wu and Yan (2019). The DI_1^{noR} is identical to the DI_1 but does not contain N_R.