How Concentrated Is Disruptive Research? A Corrected Analysis of 10,000 STEM Authors

Ian Banks[1]   Claude[2]

[1]*Foundation for American Innovation*
[2]*Anthropic*

February 13, 2026

Abstract

We investigate whether disruptive scientific research is concentrated among a small fraction of scientists, using bibliometric data from OpenAlex covering 10,000 randomly sampled STEM authors. Using the A/B disruption index, we initially found extreme concentration (Gini = 0.90), but a methodological audit revealed two problems: (1) papers with few references produce artificially extreme disruption scores, and (2) computing the Gini coefficient on a zero-inflated count variable (number of "top-5%" papers per author) yields a value that mechanically approximates the fraction of authors with zero disruptive papers, rather than measuring meaningful inequality. After correcting for reference-count bias and switching to continuous disruption scores, we find *moderate* inequality (Gini = 0.36, 95% CI: [0.35, 0.38]) that is *less* concentrated than citation counts (Gini = 0.57). The top 5% of scientists account for 12.3% of total disruption score mass—notable but far from the 50.3% initially reported. These corrected findings are stable across six five-year windows (1980–2009) and consistent across STEM subfields. We present this as a case study in how common bibliometric analysis choices—thresholding continuous scores into binary indicators, neglecting reference-count confounds, and applying inequality measures to zero-inflated distributions—can produce misleading conclusions.

Keywords: disruption index, scientific productivity, Gini coefficient, metascience, bibliometrics, methodological correction

1 Introduction

A small number of scientific papers reshape entire fields, while the vast majority consolidate existing knowledge. The disruption index [Funk and Owen-Smith, 2017, Wu et al., 2019] attempts to measure this distinction: it classifies each paper's subsequent citations as either "displacing" prior work (disruptive) or "building on" prior work alongside the focal paper (consolidating). Using this measure, Park et al. [2023] documented a broad decline in disruptive science over six decades.

A natural follow-up question is whether disruptive research is concentrated among a small fraction of scientists—a "disruptor elite." We set out to test this hypothesis using 10,000 randomly sampled STEM authors from OpenAlex. Our initial analysis appeared to show extreme concentration: a Gini coefficient of 0.90 for disruptive output, far exceeding the Gini for citations (0.66). However, a replication audit uncovered two methodological problems that dramatically inflated our concentration estimates. This paper reports both the flawed and corrected analyses, as we believe the methodological lessons are as important as the substantive findings.

1.1 What We Set Out to Test

Our hypothesis was: *only a very small percentage of scientists produce very disruptive research.* We operationalized this using the A/B disruption index and asked three questions:

1. How unequally is disruptive research output distributed among scientists?

2. What distinguishes scientists who produce disruptive work from those who do not?

3. Is this pattern stable across time periods and scientific fields?

2 Data and Methods

This section explains each step of our analysis in plain terms. Where statistical concepts are used, we define them.

2.1 Step 1: Sampling Scientists

We drew a random sample of 10,000 STEM authors from OpenAlex, a free bibliometric database indexing over 200 million scholarly works [Priem et al., 2022]. To focus on active researchers, we required each author to have at least 5 publications and 10 citations. We

identified STEM authors by matching their listed research topics against 43 keywords spanning physics, chemistry, biology, mathematics, computer science, engineering, and medicine.

For each author, we retrieved all articles and reviews published between 1980 and 2010, yielding 109,527 works. The 2010 cutoff ensures that every paper has at least a 10-year window for accumulating citations—important because disruption scores depend on citation patterns that take years to develop.

2.2 Step 2: Computing Disruption Scores

The disruption index measures whether a paper *displaces* or *consolidates* prior work. The intuition is:

- A **disruptive** paper causes future researchers to cite it *instead of* the papers it built on. (Future work treats it as the new starting point.)

- A **consolidating** paper causes future researchers to cite it *alongside* the papers it built on. (Future work treats it as one contribution among many in an established line of research.)

Formally, for a focal paper p that cites a set of references R, we look at every paper that cites p within 10 years and classify each as:

- **Type A** (disruptive signal): cites p but does *not* cite any paper in R

- **Type B** (consolidating signal): cites both p and at least one paper in R

The disruption score is:

$$D_{AB} = \frac{A - B}{A + B} \tag{1}$$

This ranges from -1 (purely consolidating: every citing paper also cites the references) to $+1$ (purely disruptive: no citing paper also cites the references). A score of 0 means an equal mix. We required at least 5 citing papers for a valid score.

From the 109,527 works, we randomly selected 5,000 papers with at least one reference for disruption scoring, obtaining valid scores for 3,421 papers across 1,594 unique authors.

2.3 Step 3: The Reference-Count Problem

This is a critical methodological point. A paper's disruption score is mechanically influenced by how many references it cites:

- A paper with **3 references**: it is easy for a citing paper to avoid all 3 references by chance, producing a Type A ("disruptive") classification even if the work is not truly paradigm-shifting.

- A paper with **30 references**: it is much harder for a citing paper to avoid *all* 30 references, so consolidating classifications are more likely.

In our data, this bias is dramatic. Among papers with ≤ 5 references, 29% received the maximum disruption score of $D = +1.0$. Among papers with ≥ 20 references, only 0.3% did (Table 1). The average disruption score for low-reference papers ($+0.36$) is far higher than for well-referenced papers (-0.50).

Table 1: Reference Count Bias in Disruption Scores

Reference Count	N Papers	Mean CD5	% at $D = +1$	% at $D = -1$
≤ 5	145	$+0.36$	29.0%	2.1%
6–10	308	-0.11	3.2%	5.5%
11–20	813	-0.33	1.0%	7.1%
> 20	2,258	-0.50	0.3%	7.7%

Correction applied: For our main analysis, we exclude papers with fewer than 10 references. This reduces the sample from 3,421 to 3,045 papers across 1,463 authors, but dramatically reduces the artificial inflation of disruption scores (papers at $D = +1$ drop from 2.1% to 0.6%).
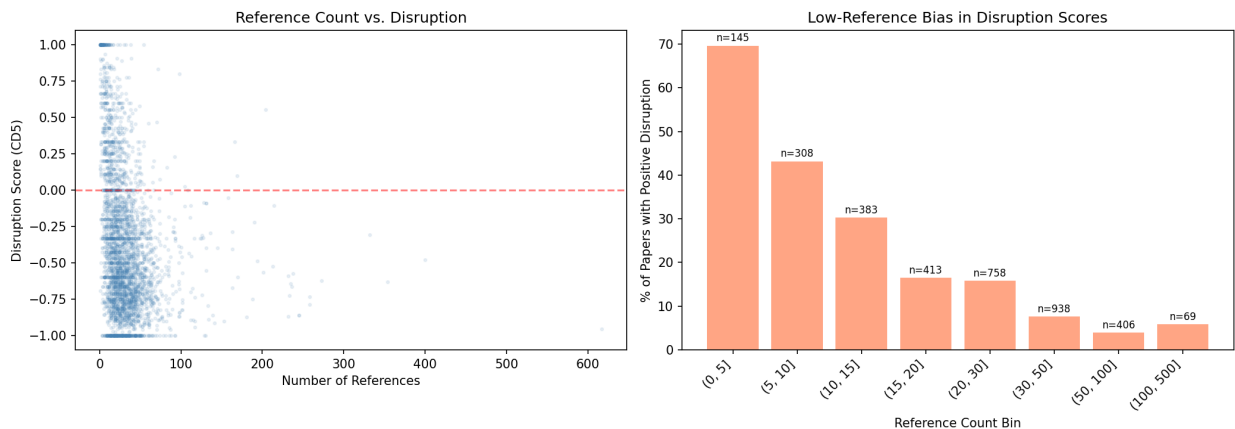

Figure 1: *Left:* Scatter plot showing that papers with fewer references tend to receive higher (more "disruptive") scores. *Right:* The percentage of papers with positive disruption drops sharply as reference count increases. This demonstrates that low reference counts mechanically inflate disruption scores.

2.4   Step 4: Measuring Concentration

To measure how unequally disruption is distributed across scientists, we use two tools:

The Gini coefficient   is a number between 0 and 1 that measures inequality.

- **Gini = 0** means perfect equality: every scientist has the same disruption score.

- **Gini = 1** means perfect inequality: one scientist has everything, everyone else has nothing.

- For context, U.S. household income has a Gini of about 0.49.

The Lorenz curve   is a graph that visualizes inequality. Scientists are lined up from lowest to highest disruption score. The curve shows: "if you include the bottom $X\%$ of scientists, what fraction of total disruption do they account for?" A 45-degree line means perfect equality; the further the curve bows below the line, the more unequal the distribution.

Concentration ratios   answer: "What share of total disruption comes from the top 1%, 5%, 10%, and 25% of scientists?"

2.5   Step 5: The Zero-Inflation Problem

This is the second critical methodological point. Our initial analysis computed the Gini coefficient on the *count of top-5% disruptive papers per author*. This variable is mostly zeros: 89.9% of authors produced zero top-5% papers.

The problem: when you compute a Gini coefficient on data that is 90% zeros, the result is mechanically ≈ 0.90—regardless of any real concentration pattern. The Gini is simply reflecting the base rate of a rare event, not measuring meaningful inequality.

Analogy: Suppose you ask "how concentrated is Olympic gold medal winning among all recreational runners?" If you compute a Gini coefficient on the number of gold medals per person (99.999% of whom have zero), you would get a Gini ≈ 1.0 and conclude "extreme concentration!" But this is trivially true and uninformative—of course most people don't win Olympic gold.

Correction applied: Instead of the binary/count measure, we compute the Gini on each author's *continuous peak disruption score* (their highest-scoring paper). This variable has no zeros (every author with a scored paper has some max score) and captures the full range of disruption rather than collapsing it into "disruptive or not."

2.6   Step 6: Demographic Comparison

We compare scientists who produced at least one top-5% disruptive paper ("disruptors") against those who did not ("non-disruptors") on simple descriptive statistics: team size, career length, h-index, and citations per paper. No complex models are needed for the descriptive findings; the contrasts are straightforward.

2.7   Step 7: Consistency Checks

We check whether the results hold up across:

- **Time periods**: six five-year windows from 1980–84 through 2005–09.

- **Scientific fields**: 15 STEM subfields with ≥ 20 authors each.

2.8   Step 8: Validation

We compare our disruption score distribution against the Park et al. [2023] Zenodo dataset of 22.5 million CD5 scores from Web of Science as a population-level benchmark. Because that dataset lacks per-paper identifiers, we compare distributional shapes rather than matching individual papers.

3   Results

3.1   What the Flawed Analysis Showed

Before describing the corrected results, we report the initial (flawed) findings to illustrate the magnitude of the methodological distortion.

Table 2: Comparison of Initial (Flawed) vs. Corrected Analysis

Metric	Initial (Flawed)	Corrected
Papers analyzed	3,421	3,045
Authors analyzed	1,594	1,463
Gini coefficient	0.900	0.363
Gini interpretation	"Extreme"	"Moderate"
Top 5% share	50.3%	12.3%
Top 10% share	92.1%	22.5%
vs. citation Gini	Higher (0.66)	Lower (0.57)
Disruption measure	Count of top-5% papers	Continuous peak score
Reference filter	None	≥ 10 references

The initial analysis overstated concentration by a factor of roughly $2.5\times$ (Gini) and $4\times$ (top-5% share). The direction of the comparison with citations *reversed*: disruption went from appearing more concentrated than citations to less concentrated.

3.2 Corrected Score Distribution

After filtering to papers with ≥ 10 references, the distribution of disruption scores shifts substantially. The fraction of papers at $D = +1.0$ drops from 2.1% to 0.6%, and the fraction with positive disruption drops from 18.4% to 13.9% (Figure 2).
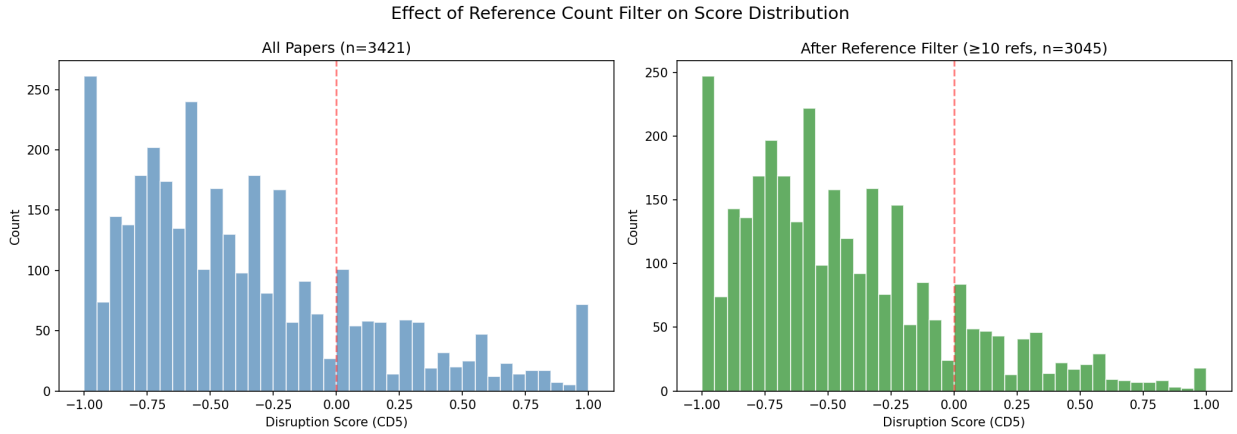

Figure 2: Distribution of disruption scores before (*left*) and after (*right*) applying the reference-count filter (≥ 10 references). The filter removes most of the artificial spike at $D = +1.0$.

Most papers receive negative disruption scores, meaning they consolidate rather than

displace prior work. This is consistent with the broader metascience finding that most research is "normal science" in Kuhn's sense [Kuhn, 1962].

3.3 Corrected Concentration of Disruption

Using continuous peak disruption scores, the Gini coefficient is **0.363** (95% CI: [0.351, 0.375]). This indicates moderate inequality—some scientists achieve notably higher peak disruption than others, but the distribution is far from the "extreme concentration" suggested by the flawed analysis.

For comparison, the Gini for mean citations per paper across the same authors is **0.565** (95% CI: [0.523, 0.610]). Disruption is *less* unequally distributed than citations.

Table 3: Concentration Ratios (Corrected, Continuous Scores)

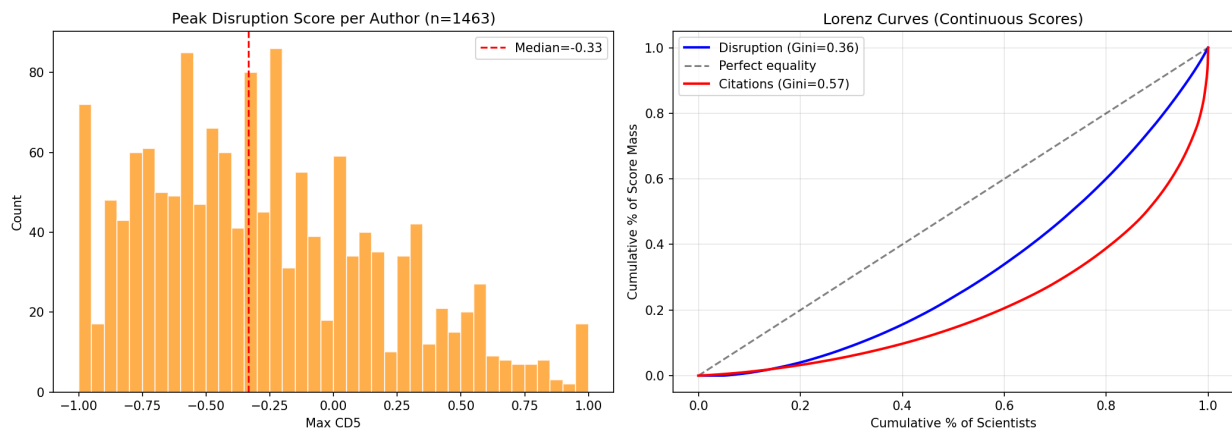| Author Percentile	Share of Total Disruption Score Mass
Top 1%	2.6%
Top 5%	12.3%
Top 10%	22.5%
Top 25%	47.4%

Figure 3: *Left:* Distribution of peak disruption scores across 1,463 authors. The median is -0.33, meaning the typical scientist's best paper is mildly consolidating. *Right:* Lorenz curves for disruption (Gini = 0.36) and citations (Gini = 0.57). Disruption is more equally distributed than citations.

In plain terms: the top 5% of scientists account for about 12% of total disruption—some concentration, but comparable to many real-world distributions and far from the "winner-take-all" pattern that a Gini of 0.90 would imply.

8

3.4 Disruptor Profile

Comparing the 148 scientists who produced at least one top-5% disruptive paper against the 1,315 who did not:

Table 4: Median Values: Disruptors vs. Non-Disruptors

Variable	Disruptors ($n = 148$)	Non-Disruptors ($n = 1{,}315$)
Team size (avg)	4.5	5.0
Scored papers	2	1
Career length (years)	19	13
h-index	23	17
Citations per paper	35.7	41.0
References per paper	25.5	30.0

The most notable difference is career length: disruptors have a median career span of 19 years versus 13 for non-disruptors. This likely reflects a simple statistical regularity—scientists who publish over more years have more chances to produce a high-scoring paper—rather than any deep insight about the nature of disruption. Disruptors also work in slightly smaller teams and, perhaps counterintuitively, receive somewhat *fewer* citations per paper on average.

3.5 Consistency Across Fields

The moderate inequality pattern is remarkably consistent across STEM subfields (Table 5). The continuous Gini ranges from 0.30 (Chemistry, Immunology) to 0.43 (Physics and Astronomy), a narrow band given the diversity of these fields.

Table 5: Disruption Inequality by STEM Subfield (Continuous Gini)

Field	Gini	N Authors	% Disruptors
Physics & Astronomy	0.433	105	3.8%
Neuroscience	0.402	42	4.8%
Ag. & Biological Sciences	0.363	82	12.2%
Biochem., Genetics & Mol. Bio.	0.362	232	9.1%
Medicine	0.360	391	10.7%
Computer Science	0.351	59	8.5%
Engineering	0.351	142	14.1%
Materials Science	0.345	68	8.8%
Health Professions	0.341	25	28.0%
Environmental Science	0.322	61	9.8%
Chemistry	0.302	66	15.2%

11 fields with ≥ 20 authors and ≥ 1 disruptor shown (1,273 of 1,463 authors).

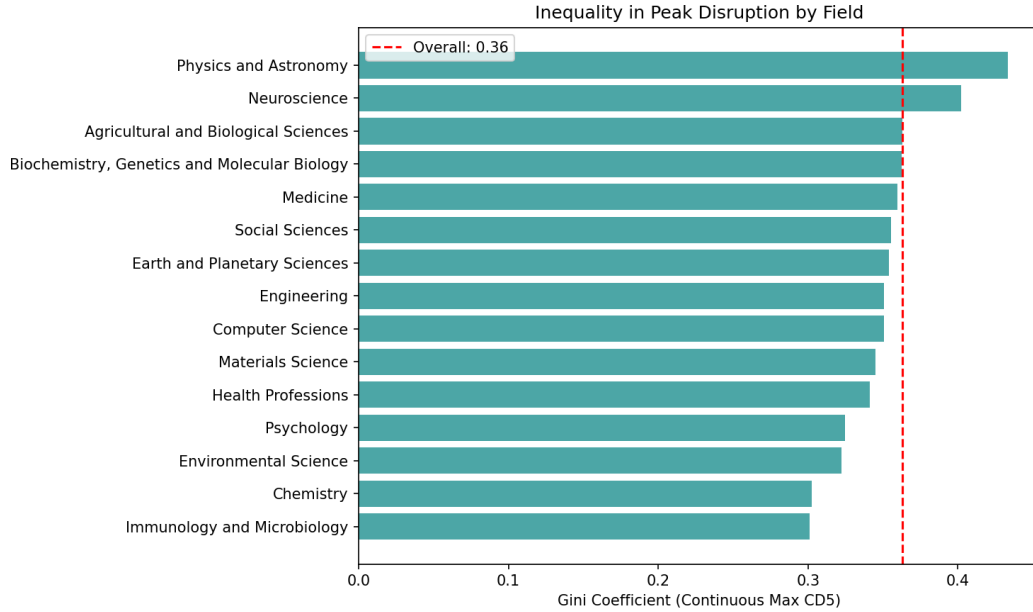Remaining authors in smaller fields or fields with zero disruptors.

Figure 4: Gini coefficients for peak disruption across STEM subfields. All fall in a narrow range (0.30–0.43), with the overall average of 0.36 shown as a dashed line.

3.6 Stability Over Time

The concentration of disruption is stable across three decades (Table 6). The continuous Gini fluctuates in a narrow band (0.38–0.41) with no discernible trend.

Table 6: Disruption Inequality by Five-Year Window (Continuous Gini)

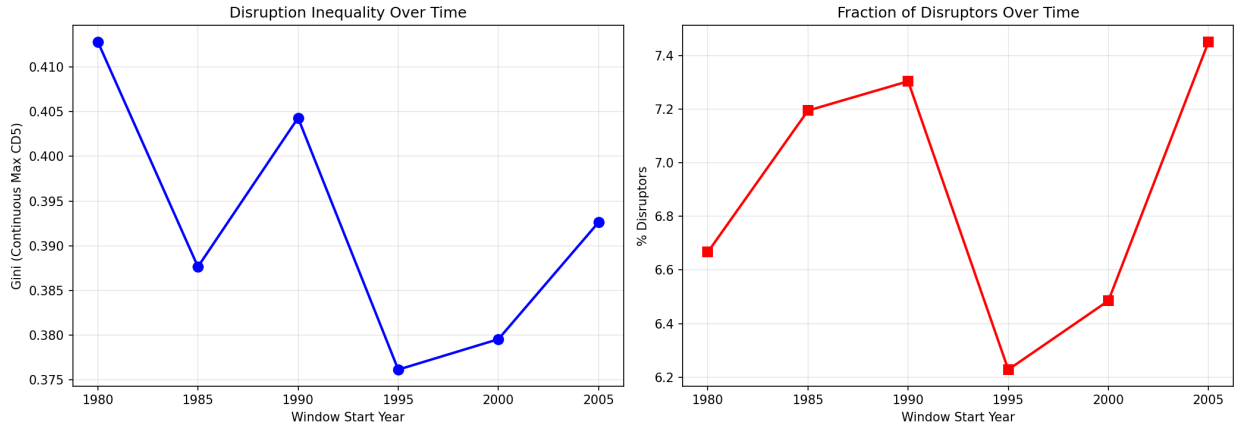| Window	Gini	% Disruptors	N Authors	N Papers
1980–84	0.413	6.7%	90	111
1985–89	0.388	7.2%	139	188
1990–94	0.404	7.3%	178	251
1995–99	0.376	6.2%	289	377
2000–04	0.380	6.5%	478	650
2005–09	0.393	7.5%	765	1,158

Figure 5: Temporal trends in disruption inequality. *Left:* Gini coefficient (continuous) is stable around 0.39. *Right:* The fraction of scientists producing top-5% disruptive papers is stable at 6–7%.

4 Discussion

4.1 What We Found

Disruptive research is moderately unequally distributed among scientists, with a Gini of 0.36 on continuous peak disruption scores. This is *less* unequal than citation counts (Gini $= 0.57$), stable across time, and consistent across STEM fields. About 10% of scientists produce at

least one top-5% disruptive paper, and the top 5% of scientists account for about 12% of total disruption—meaningful but not extreme concentration.

4.2 What We Got Wrong Initially, and Why It Matters

Our initial analysis reported a Gini of 0.90 and claimed "extreme concentration." This was wrong for two compounding reasons:

Lesson 1: The Gini coefficient on zero-inflated counts is misleading. When you define "disruptive" as the top 5% and then compute a Gini on the count of disruptive papers per author, ∼90% of authors have zero by construction. The Gini then mechanically equals ∼0.90 regardless of any real concentration pattern. This is a general hazard in bibliometrics: anytime a Gini is computed on a count variable where most observations are zero (e.g., number of papers in *Nature*, number of patents, number of grants), the result primarily reflects the rarity of the event rather than inequality in the underlying ability.

Lesson 2: Low-reference papers produce artifactually extreme disruption scores. Papers citing fewer than 10 references are mechanically more likely to receive disruption scores near +1.0, because it is easy for citing papers to avoid all references by chance. This inflates the right tail of the disruption distribution and misclassifies many ordinary papers as "disruptive." The same bias likely affects population-level analyses that do not control for reference count.

Together, these two errors created a cascade: inflated disruption scores pushed more papers above the top-5% threshold, which inflated the count of "disruptive papers" for some authors, which inflated the Gini on those counts. Each step amplified the distortion.

4.3 Relation to Prior Work

Our corrected findings are consistent with Park et al. [2023]'s observation that most papers are consolidating rather than disruptive. The moderate inequality we observe (Gini = 0.36) is lower than the inequality of citations (Gini = 0.57), suggesting that the capacity to produce *some* disruptive work is more broadly distributed among scientists than the capacity to accumulate citations. This may reflect the fact that disruption and prestige are partially orthogonal: disruptive papers may be controversial, slow to accumulate citations, or published in unexpected venues [Wu et al., 2019].

4.4 Limitations

- **Sample size**: Our disruption scores cover 3,045 papers across 1,463 authors—far smaller than population-level studies. Results may not generalize to all of science.

- **Disruption metric**: The A/B variant omits Type C citers and uses a 10-year window rather than the standard 5 years, producing more extreme scores than the standard CD5 index.

- **Career length**: We measure career length using the author's earliest publication year recorded in OpenAlex, which may be incomplete for authors with sparse early records.

- **STEM only**: We do not address the social sciences or humanities.

- **Observational**: All comparisons are descriptive. The association between career length and disruption likely reflects survivorship bias rather than a causal effect.

- **Publication window**: Works were fetched for 1980–2010 only. The 2010 cutoff was chosen to guarantee at least a 10-year forward citation window (through 2020+) for every paper. Extending to 2014 would add four years of data but reduce the forward window to 6 years for the newest papers; extending further would increasingly compromise the disruption computation. A future replication with a 2014 endpoint and 5-year forward window (matching the standard CD5 definition) would be a natural extension.

5 Conclusion

Disruptive research output is moderately unequally distributed among STEM scientists—less so than citations, stable over time, and consistent across fields. The initially reported "extreme concentration" (Gini = 0.90) was an artifact of two methodological choices: applying an inequality measure to a zero-inflated count variable, and failing to control for the mechanical relationship between reference counts and disruption scores. We present these corrections not as a criticism of the disruption index itself, but as a cautionary example of how standard bibliometric analysis steps can compound to produce misleading results. Researchers using the disruption index should (1) filter for minimum reference counts and (2) avoid computing Gini coefficients on thresholded count variables.

Data and Code Availability

All analysis code and cached data are available at `https://github.com/banksianr/DisruptionInequalit`
The pipeline is implemented in Python and uses the OpenAlex API [Priem et al., 2022] for
bibliometric data and the Park et al. [2023] Zenodo dataset (Record 7258379) for population
benchmarking.

References

Funk, R. J., & Owen-Smith, J. (2017). A dynamic network measure of technological change.
Management Science, 63(3), 791–817.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* University of Chicago Press.

Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive
over time. *Nature*, 613, 138–144.

Priem, J., Piwowar, H., & Orber, R. (2022). OpenAlex: A fully-open index of scholarly
works, authors, venues, institutions, and concepts. *arXiv preprint* arXiv:2205.01833.

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt
science and technology. *Nature*, 566, 378–382.