International Journal of Computational Research and Development (IJCRD)
International Peer Reviewed - Refereed Research Journal, Website: www.dvpublication.com
Impact Factor: 5.015, ISSN (Online): 2456 - 3137, Volume 9, Issue 2, July - December, 2024

ASSESSING ASSOCIATIONS BETWEEN BIBLIOGRAPHIC CHARACTERISTICS AND SUBSEQUENT CITATION IMPACT ACROSS SCIENTIFIC BENCHMARKING PUBLICATIONS USING OPENALEX

Divya Vishwanath* & Paul Johnson Asamoah**
* Janardhan Rai Nagar Rajasthan Vidyapeeth, Udaipur, Rajasthan, India
** Power Life University College, Accra, Ghana, West Africa

Cite This Article: Divya Vishwanath & Paul Johnson Asamoah, "Assessing Associations Between Bibliographic Characteristics and Subsequent Citation Impact Across Scientific Benchmarking Publications Using OpenAlex", International Journal of Computational Research and Development, Volume 9, Issue 2, July - December, Page Number 111-118, 2024.

Copy Right: © DV Publication, 2024 (All Rights Reserved). This is an Open Access Article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium provided the original work is properly cited.

Abstract:

This analysis examines whether citation averages conceal differences in the distribution of scholarly attention, addressing international debates on research evaluation. A descriptive cross-sectional analysis uses 199 OpenAlex publication records issued between 2010 and 2022, with citations measured during 2023. Distributional summaries, exploratory bibliographic group comparisons, and concentration shares distinguish average attention from typical outcomes. Publications received 822 indexed citations, with the ten most cited accounting for 51.82 percent, while 40.70 percent received none. Across increasing reference count and author count bands, the proportion with zero indexed citations declined consistently, although mean citations did not increase consistently across intermediate groups. Older publication bands contained progressively larger proportions with zero indexed citations, yet the oldest band exhibited a rebound in mean citations alongside a zero median. These contrasts reveal how concentrated recognition can coexist with limited attention across a publication group. The contribution is a distributional qualification of average based interpretations of scientific impact. Research evaluation should combine means, medians, indexed zero proportions, and concentration measures before comparing publication groups or allocating institutional support.

Key Words: Bibliometrics; Citation Concentration; Descriptive Analysis; Research Evaluation; Scholarly Attention

1. Introduction:

Citation averages can conceal a basic distinction in scientific recognition: whether attention reaches many publications or concentrates among a few. This distinction matters when bibliographic indicators inform research assessment, publication comparisons, and judgments about scholarly influence. Two groups can have similar average citation counts while differing substantially in the proportion of publications receiving no citations. Interpreting citation attention therefore requires examining its distribution alongside its average volume. Scientific publications enter an expanding system of knowledge production and referencing. Bornmann and Mutz, 2015, document growth in publications and cited references, establishing the broader context in which scholarly contributions compete for attention. Within this system, citation counts reflect recorded connections between publications, but their interpretation depends on bibliographic characteristics and observation periods. Tahamtan et al., 2016, identify multiple factors associated with citation counts, including characteristics of publications, authors, and publication venues. Their review supports treating citations as a partial indicator of scholarly attention rather than a direct measure of research quality. Reference count provides one observable connection between a publication and the existing literature. More references may indicate broader engagement with prior knowledge, but quantity does not reveal how those sources are combined or used. Uzzi et al., 2013, show that scientific impact relates to the combination of conventional knowledge foundations and unusual knowledge combinations. Their findings distinguish intellectual composition from bibliographic volume. This distinction raises a narrower empirical question: whether publications with more indexed references exhibit broader participation in citation attention, higher average citation counts, or both. Examining these outcomes separately avoids assuming that an extensive reference list necessarily represents a more influential contribution.

Author count captures another dimension of publication production. Larger teams may combine expertise and professional connections, yet participation scale also corresponds to differences in the nature of scientific contributions. Wu et al., 2019, distinguish the developmental contributions associated with larger teams from the disruptive contributions associated with smaller teams. Their evidence challenges a uniform interpretation of team size as a measure of scientific advantage. For citation analysis, the relevant issue is therefore whether attention varies across observed team sizes and whether averages adequately represent that variation. Similar mean citation counts may coexist with different medians and different proportions of publications receiving no indexed citations. Publication age introduces a temporal dimension that further complicates comparison. Earlier publication allows more time for discovery, while changes in research priorities can redirect attention toward newer contributions. Wang et al., 2013, explain citation trajectories through interacting processes involving preferential attachment, aging, and publication fitness. These processes imply that age cannot be interpreted solely as accumulated opportunity. The choice of citation window is consequential: lifetime totals combine attention over different durations, whereas a common annual window records attention during the same calendar year. Annual counts permit comparison of contemporaneous recognition, although publications remain at different stages of their citation trajectories.

The reviewed evidence establishes the relevance of knowledge connections, team structure, and temporal dynamics to scientific impact. It does not establish how these characteristics jointly appear in the descriptive distribution of annual citation attention among the benchmarking publications examined here. The specific issue concerns the relationship between average citation volume and the frequency of receiving no indexed citations across bibliographic groups. The contribution is therefore a bounded measurement comparison rather than the identification of a new causal mechanism. Reference count, author count, and publication age are treated as separate observed characteristics, allowing their distributional patterns to remain visible. This study aims to describe how citation attention during 2023 varies across reference counts, author counts, and publication ages among

sampled scientific benchmarking publications. The central research question asks how these characteristics correspond to differences in average citations, median citations, and the proportion of publications with zero indexed citations. The analysis examines each characteristic separately and assesses whether these summaries produce consistent or contrasting interpretations of scholarly attention.

The empirical setting consists of 199 OpenAlex publication records issued between 2010 and 2022. The accessible frame includes records containing benchmark in the title and at least one indexed reference. A random extraction selected 200 records, from which one record without indexed authorship was excluded. Each retained publication contributes one observation, with citation attention measured during the common calendar year 2023. The analysis uses descriptive statistics, citation concentration measures, and exploratory comparisons across reference, author, and age bands. This design supports transparent examination of observed distributions without attributing differences to independent causal effects. The findings show why this distinction matters. Across increasing reference and author bands, the proportion of publications with zero indexed citations declines consistently, while mean citations do not increase consistently across intermediate groups. Across publication age bands, the proportion with zero indexed citations rises, yet the oldest group exhibits a rebound in average citation attention. These patterns demonstrate that the frequency of receiving attention and its average volume can convey different information within the same sample. The empirical contribution is to make that divergence explicit through a common observation window and complementary distributional summaries.

The practical relevance concerns how research evaluators interpret bibliographic comparisons. Means alone can obscure whether a group contains broadly distributed recognition or a smaller set of publications attracting substantial attention. Reporting medians, indexed zeros, and concentration alongside averages provides a more transparent account of the observed evidence. Although the sample has no geographical restriction, its lexical selection and database coverage limit population generalization. The transferable insight concerns measurement: comparisons of scholarly attention benefit from distinguishing its distribution from its volume before drawing conclusions about reference practices, collaboration, or publication age.

2. Theory, Conceptual Framework and Research Questions Development:

Scientific citation emerges from an interdependent system in which researchers draw on existing knowledge, produce new contributions, and allocate attention through references. Publications compete for recognition within this expanding knowledge system. Bornmann and Mutz, 2015, document growth in scientific publications and cited references, providing a structural basis for understanding competition for scholarly attention. Within this setting, bibliographic characteristics describe how a publication connects to prior research, how many authors contribute to its production, and when it enters the scientific record. This analysis examines 199 publications issued between 2010 and 2022 and measures subsequent citation impact through citations received in 2023. Reference count, author count, and publication age constitute separate observed characteristics rather than indicators of an underlying latent construct. Citation impact represents recorded scholarly attention rather than a direct measure of research quality. Tahamtan et al., 2016, identify multiple influences on citation counts, supporting an interpretation in which bibliographic characteristics provide partial explanations of citation variation. The framework therefore develops research questions about observable associations without assuming that these relationships establish causal effects.

Reference count measures the number of indexed publications cited by a focal publication. Conceptually, references connect a contribution to existing knowledge and position its claims within ongoing scientific discussions. A larger reference count may indicate broader engagement with prior work, potentially increasing the number of research communities that find the contribution relevant. However, the number of references does not reveal how effectively a publication combines or applies existing knowledge. Extensive referencing can accompany either an integrative contribution or a narrowly focused review of established material. The relationship between reference count and subsequent citation impact consequently depends on characteristics that reference quantity alone cannot identify. Uzzi et al., 2013, show that scientific impact relates to the combination of conventional knowledge foundations and unusual knowledge combinations. Their findings support the importance of intellectual connections, but they do not establish that adding references increases citations. This analysis therefore treats indexed reference count as a measurable bibliographic characteristic and leaves its association with subsequent attention open to empirical assessment.

RQ1: How does the number of citations received in 2023 vary with indexed reference count among the sampled publications?

Author count measures the number of indexed authors associated with each publication. Whereas reference count describes connections to existing literature, author count describes the scale of participation in producing a contribution. More authors may provide complementary expertise, distribute research tasks, and connect the publication to broader professional networks. These mechanisms could increase its visibility and relevance. Larger teams also face coordination demands and may pursue research agendas that differ from those of smaller teams. Team size can therefore reflect differences in the nature of scientific work rather than a uniform increase in its subsequent recognition. Additional authors do not necessarily contribute equivalent expertise, effort, or dissemination opportunities. Wu et al., 2019, demonstrate that larger teams tend to develop established scientific trajectories, while smaller teams more often disrupt them. This distinction provides a boundary condition for interpreting author count: participation scale may correspond to different forms of contribution rather than a consistent citation advantage. The available measure captures indexed team size, making the relationship with subsequent citation attention an empirical question.

RQ2: How does the number of citations received in 2023 vary with indexed author count among the sampled publications?

Publication age measures elapsed calendar years between publication and the citation observation year. We calculate this characteristic as 2023 minus the publication year. Age locates each publication within its citation trajectory. Earlier publication provides more time for discovery and incorporation into subsequent research, while more recent publication may align more closely with active research priorities. These mechanisms imply competing expectations for annual citation counts. Recognition can develop gradually, but attention may decline as research priorities change or newer contributions replace earlier work. Because the outcome records citations during one calendar year, greater publication age does not mechanically produce a larger outcome through a longer accumulation period. Wang et al., 2013, model citation dynamics through interacting processes

involving preferential attachment, aging, and publication fitness. Their framework explains why elapsed time can correspond to different citation trajectories. This analysis uses publication age to examine temporal variation without imposing a uniformly positive or negative relationship.

RQ3: How does the number of citations received in 2023 vary across publication ages among the sampled publications?

The conceptual framework connects three observed bibliographic characteristics to one observed citation outcome. Reference count represents indexed connections to prior literature, author count represents indexed participation in research production, and publication age represents temporal position. Each connection denotes an association to be examined rather than a demonstrated causal pathway. The three research questions retain distinct interpretations and allow positive, negative, negligible, or inconclusive associations.

3. Data:

This section documents the source records, sample construction, cleaning rules, and measurement of bibliographic characteristics and citation impact. It establishes a descriptive framework for examining 199 publications using a common citation observation year.

3.1 Data Sources, Population, and Sample Construction:

The analysis uses publication metadata from OpenAlex, accessed through its public works application programming interface. The extraction includes work identifiers, titles, publication dates, publication years, referenced works, authorships, digital object identifiers, and annual citation counts. Persistent work identifiers provide the connection between each analytical observation and its source record. The workbook preserves the extraction query and access date. The records represent a retrieved database snapshot rather than information archived at each publication date. The target population comprises scientific benchmarking publications issued between 2010 and 2022, with citation attention observed in 2023. The accessible sampling frame is narrower: OpenAlex records containing benchmark in the title and at least one indexed reference. This lexical criterion captures publications across disciplines but does not establish substantive membership in a single research field. No geographical restriction was applied, and country or sector classifications were not constructed. Growth in scientific publications and cited references provides context for interpreting citation attention within an expanding literature, as documented by Bornmann and Mutz, 2015.

The extraction requested a random sample of 200 records using seed 42. One record lacked indexed authorship and was excluded, leaving 199 publications. The complete eligible population was not enumerated. The value of 200 in the extraction metadata represents the requested sample and is therefore not reported as the population count. The workbook documents the eligibility and exclusion rules but does not establish their timing relative to inspection of the records. The unit of analysis is an individual OpenAlex publication record. The dataset forms a cross section, with publication cohorts spanning 2010 to 2022 and a common annual citation outcome in 2023. Each publication contributes one observation. Citations received in 2023 constitute an annual flow rather than cumulative citations since publication. The sample contains publications from all 13 eligible publication years, with cohort sizes ranging from 7 to 33 records.

Table 1: Sample construction

Stage	Publications	Observations	Reason or Rule
Accessible eligible frame	Not enumerated	Not enumerated	Title contains benchmark, publication during 2010 to 2022, and at least one indexed reference
Selected records	200	200	Random sample requested using seed 42
Excluded records	1	1	Indexed authorship unavailable
Retained descriptive sample	199	199	Complete values for the three bibliographic measures and the constructed citation outcome

3.2 Data Integration, Cleaning, and Quality Assurance:

All analytical fields originate from the same OpenAlex work records. Construction therefore required no external dataset merger, identifier crosswalk, or frequency conversion. The OpenAlex work identifier serves as the record key. Verification identified 199 distinct identifiers among the 199 retained observations, with no repeated identifier. This check establishes record uniqueness within the extraction, although separate records can represent different versions of related scholarly contributions. Reference count was extracted as the number of identifiers in the referenced works list. Author count was extracted as the number of returned authorship entries. Publication year and the annual citation series were taken from the same record, ensuring temporal alignment at the publication level. No currency conversion, inflation adjustment, or aggregation across reporting frequencies was required. The retained sample contains no missing values for publication year, reference count, author count, publication age, or the constructed citation outcome. Annual citation entries explicitly identify 2023 for 118 publications. For 81 publications, the returned annual series omits 2023. These omissions were coded as zero indexed citations for that year, following the source convention that annual entries with zero citations are omitted. The workbook retains a separate flag distinguishing explicit entries from these indexed zeros.

The one record without authorship was removed rather than assigned an artificial author count. Digital object identifiers are available for 188 retained publications; the remaining 11 retain persistent OpenAlex links. Missing digital object identifiers did not trigger exclusion. Observed values satisfy the implemented measurement boundaries: reference counts and author counts are positive integers, publication years fall within the eligibility window, and citation counts are nonnegative integers. All publication ages reconcile with the difference between 2023 and the recorded publication year. No trimming, winsorization, or outlier deletion was applied. The largest reference, author, and citation counts remain in the dataset because magnitude alone does not establish a data error. The extraction preserves current database descriptions of historical publications and citation years. Bibliographic links can therefore reflect subsequent database corrections rather than the information recorded when a publication first appeared. Interpretation remains sensitive to indexing completeness and disciplinary differences. Tahamtan et al., 2016,

identify multiple determinants of citation counts, reinforcing the need to distinguish observed bibliographic variation from a complete explanation of scientific attention.

3.3 Variable Measurement, Construction, and Validation:

Bibliographic characteristics are represented by three separate observed measures: reference count, author count, and publication age. They are not combined into a composite index. Subsequent citation impact is measured as the number of indexed citations received during 2023. Higher values indicate greater recorded citation attention during that year. Reference count describes the extent of indexed connections to earlier literature. It does not measure the novelty, diversity, or quality of those connections. Uzzi et al., 2013, distinguish conventional knowledge foundations from unusual knowledge combinations, demonstrating why reference quantity and intellectual composition require separate interpretations. Author count describes indexed participation in producing a publication. It does not establish individual contributions or team organization. Wu et al., 2019, associate team size with differences in scientific development and disruption, supporting this distinction between participation scale and contribution type. Publication age was calculated using Equation 1.

$$A_i = 2023 - P_i$$

Here, A_i denotes publication age in calendar years for publication i, and $P_i$denotes its recorded publication year. The measure ranges from 1 to 13 years. Publication year is retained for provenance and cohort description rather than treated as a separate dimension alongside its directly calculated age.

Table 2: Variable dictionary

Variable or Indicator	Analytical Role	Definition and Construction	Source and Identifier	Unit and Frequency	Transformation and Direction	Missing Value Rule	Coverage and Limitation
Citations received in 2023	Outcome	Citation count attached to the 2023 entry in the annual citation series	OpenAlex, counts by year and cited by count	Citations per publication during one calendar year	Untransformed count; higher values indicate more indexed citation attention	Omitted 2023 entry coded as indexed zero and separately flagged	199 publications; reflects source coverage rather than all possible citations
Reference count	Bibliographic dimension	Number of identifiers in the referenced works list	OpenAlex, referenced works	Indexed references per publication record	Untransformed count; higher values indicate more indexed outgoing references	Publications without indexed references were outside the sampling frame	199 publications; does not measure reference quality or intellectual diversity
Author count	Bibliographic dimension	Number of returned authorship entries	OpenAlex, authorships	Indexed authors per publication record	Untransformed count; higher values indicate larger indexed teams	One record without indexed authorship excluded	199 publications; depends on authorship indexing completeness
Publication age	Bibliographic dimension	Difference between 2023 and publication year	OpenAlex, publication year; Equation 1	Calendar years per publication	Higher values indicate older publications	Calculated for every retained record	199 publications; measures age at annual rather than daily resolution

Validation focused on identifier uniqueness, complete analytical fields, admissible numerical values, and reconciliation of calculated age with publication year. No latent construct was estimated. Consequently, internal consistency coefficients and factor analysis were not applied to these directly observed counts.

Table 3: Descriptive statistics

Variable	Valid Observations	Mean	Standard Deviation	Median	Minimum	Maximum
Citations received in 2023	199	4.13	12.48	1	0	147
Reference count	199	33.00	29.97	26	1	280
Author count	199	4.67	5.33	3	1	55
Publication age	199	5.53	3.74	5	1	13

Citation attention is unevenly distributed. The mean exceeds the median, and the standard deviation exceeds the mean. The 81 publications coded with zero indexed citations represent 40.70 percent of the retained sample. Reference count and author count also exhibit differences between their means and medians, indicating that larger values influence the averages. Reporting medians and ranges alongside means therefore preserves information that a single summary statistic would conceal.

3.4 Analytical Framework, Identification, and Reproducibility:

The implemented analysis is descriptive. It establishes the distributions of bibliographic characteristics and annual citation attention without estimating causal effects or asserting statistically significant relationships. All 199 retained publications contribute to each descriptive statistic, so comparisons across variables use a consistent sample. The common citation year separates annual attention from the mechanical accumulation inherent in lifetime citation totals. However, publications differ in age and can occupy different stages of their citation trajectories. Wang et al., 2013, demonstrate the importance of aging and publication specific dynamics in explaining citation histories. Accordingly, differences in citations received during 2023 cannot be interpreted as consequences of publication age alone.

The available measures do not isolate disciplinary composition, publication venue, document type, or prior recognition. Random selection from the accessible frame does not remove these sources of heterogeneity. The title search and indexed reference requirement also restrict the population to which the descriptive findings apply. The reported statistics characterize the retained OpenAlex sample rather than every scientific benchmarking publication. Reproducibility rests on the preserved extraction query, sample seed, persistent work identifiers, raw reference lists, authorship identifiers, annual citation entries, and variable definitions. The workbook separates raw extraction from the analytical dataset and documents the exclusion and zero coding rules. These materials permit reconstruction of the reported measures from the saved records. Because the source database can revise records, repeating the query can produce different membership or values. Reproduction of the reported statistics therefore requires the preserved extraction.

4. Methods:

4.1 Research Design and Analytical Scope:

The analysis uses a descriptive cross-sectional design to examine variation in annual citation attention across bibliographic characteristics. The unit of analysis is an individual publication record. Each of the 199 retained publications contributes one observation, with publication dates spanning 2010 to 2022 and citation attention measured during 2023. The research questions concern differences in citation distributions across indexed reference counts, indexed author counts, and publication ages. A common annual observation window permits comparisons without combining citations accumulated over unequal publication lifetimes. Publication age nevertheless distinguishes records occupying different stages of their citation trajectories. Wang et al., 2013, demonstrate the relevance of aging to citation dynamics, supporting the distinction between contemporaneous attention and cumulative recognition. Publication age follows Equation 1 and remains an observed temporal characteristic rather than an identified causal exposure. The analytical targets are sample means, sample standard deviations, medians, ranges, proportions with zero indexed citations, and citation concentration shares. Comparisons describe variation within the retained sample. They do not estimate independent effects because disciplinary composition, publication venues, document types, and other potential explanations are not held constant. The design therefore answers the research questions through observed distributions rather than causal identification.

4.2 Sample and Measurement Alignment:

Table 1 documents the selection of 200 records and the exclusion of one record without indexed authorship. Table 2 provides the definitions and construction rules for the analytical measures. All 199 retained records contribute to every overall summary and grouping analysis. No additional observations are removed during calculation. The accessible frame comprises OpenAlex records containing benchmark in the title, published during the eligible period, and containing at least one indexed reference. The extraction uses a random sample request with seed 42. Random selection applies within this accessible frame; it does not establish representativeness of every scientific benchmarking publication.

Reference count and author count are extracted as the lengths of the returned referenced works and authorship lists. Citation attention is the indexed count assigned to 2023 in the annual citation series. An omitted entry for that year is coded as zero indexed citations, with a separate flag preserving the distinction between omitted and explicit entries. Publication age is calculated from the recorded publication year. The three bibliographic characteristics remain separate observed measures. No normalization, weighting, or aggregation into a composite score is applied. Their interpretation follows the measurement boundaries of indexed records. Reference count does not establish intellectual diversity, author count does not establish individual contributions, and citation count does not directly establish research quality. These distinctions accord with the multiple influences on citation counts reviewed by Tahamtan et al., 2016.

4.3 Descriptive Calculations and Exploratory Comparisons:

Calculations use Python 3.12.14 and its standard library. The zipfile and xml.etree. Element Tree modules read the saved workbook, while statistics and collections support numerical summaries and frequency counts. The analytical procedure first extracts the retained records, then calculates overall summaries, constructs descriptive groups, summarizes citation outcomes within each group, and calculates concentration shares. Arithmetic means and sample standard deviations follow Equations 2 and 3.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Here, $\bar{x}$ denotes the sample mean, $x_i$is the observed value for publication i, and n is the number of publications included in the calculation.

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}[x_i - \bar{x}]^2}{n-1}}$$

Here, s_x denotes the sample standard deviation. The denominator uses one fewer observation than the sample size. Medians are calculated from ordered observations, using the average of the two central values when the number of observations is even. Minimum and maximum values describe the observed range. Exploratory comparisons divide reference counts into 1 to 10, 11 to 25, 26 to 50, and at least 51. Author counts are divided into one, two to three, four to six, and at least seven. Publication ages are divided into 1 to 3, 4 to 6, 7 to 9, and 10 to 13 calendar years. These boundaries were introduced for descriptive comparison and were not prespecified theoretical thresholds. Each publication belongs to exactly one band for each characteristic. Within every band, the analysis calculates the number of publications, mean citations, median citations, and percentage with zero indexed citations. Equation 4 defines the latter percentage.

$$Z_g = 100\frac{n_{0g}}{n_g}$$

Here, Z_g denotes the percentage with zero indexed citations in group g, n_{0g} is the number of publications with a constructed citation count of zero in that group, and $n_g$is the total number of publications in the group. Differences between group percentages are expressed in percentage points. Every publication receives equal weight within its group. Citation concentration is calculated by sorting annual citation counts in descending order and dividing the citations received by the highest ranked publications by total sample citations. Equation 5 defines this calculation.

$$C_k = 100\frac{\sum_{j=1}^{k} y_{[j]}}{\sum_{i=1}^{n} y_i}$$

Here, C_k denotes the percentage of total citations attributable to the khighest citation counts, $y_{[j]}$ denotes the citation count at descending rank j, and y_i denotes citations received by publication i. Shares are calculated for the highest one, five, ten, and twenty counts. Ties at a boundary do not change the corresponding citation sum because tied observations contribute equal

counts. All calculations use unrounded values. Reported means, standard deviations, and percentages are rounded to two decimal places. Group comparisons remain descriptive; differences are not classified through significance thresholds.

4.4 Verification, Reproducibility, and Interpretation:

Verification checks identifier uniqueness, analytical completeness, numerical boundaries, and consistency between source fields and constructed measures. The retained dataset contains 199 distinct work identifiers. Publication years fall within the eligibility period, indexed reference and author counts are positive integers, and annual citation counts are nonnegative integers. Calculated publication ages reconcile with Equation 1. The grouping procedure preserves all retained observations. Publication counts across the bands of each characteristic reconcile with the complete analytical sample. Group citation totals reconcile with the overall citation total. These checks assess calculation consistency without treating it as evidence of complete bibliographic coverage.

Valid extreme observations remain in the calculations. Means are interpreted alongside medians, indexed zero percentages, and concentration shares so that unusually large citation counts remain visible without determining the interpretation alone. Comparisons across these summaries provide complementary descriptions rather than robustness tests of an estimated model. The workbook preserves the extraction query, sample seed, persistent identifiers, raw bibliographic lists, annual citation entries, variable dictionary, and coding rules. The calculation sequence and group boundaries specify how to reproduce the reported summaries from that saved extraction. Reproduction should use the preserved records because later database revisions can alter returned metadata and citation counts. The method addresses the research questions by distinguishing average citation volume, typical citation outcomes, absence of indexed annual attention, and concentration. Its contribution lies in the coordinated interpretation of established descriptive measures. Observed differences remain specific to the accessible frame, the retained records, and the common citation year.

5. Findings:

This analysis examines how citation attention varies across reference counts, author counts, and publication ages. The findings combine distributional summaries with exploratory comparisons addressing the three research questions. Interpretation concerns observed differences within the sample rather than causal effects or model validation.

5.1 Sample Composition and Citation Concentration:

The analytical sample comprises 199 distinct publication records issued between 2010 and 2022, with citation attention measured during 2023. One of the 200 selected records was excluded because indexed authorship was unavailable. All retained publications have complete values for reference count, author count, publication age, and the constructed citation outcome. Consequently, every comparison uses the same underlying sample. The publications received 822 indexed citations during 2023. The mean was 4.13 citations, compared with a median of 1 and a standard deviation of 12.48. Citation counts ranged from 0 to 147. These differences indicate that the average reflects a concentration of attention among a relatively small number of publications. The publication with the largest citation count accounted for 17.88 percent of all recorded citations. The ten publications with the largest counts accounted for 51.82 percent, while the corresponding twenty publications accounted for 66.42 percent.

Table 4: Descriptive characteristics of the analytical sample

Measure	Publications	Mean	Standard Deviation	Median	Minimum	Maximum
Citations received in 2023	199	4.13	12.48	1	0	147
Indexed reference count	199	33.00	29.97	26	1	280
Indexed author count	199	4.67	5.33	3	1	55
Publication age in calendar years	199	5.53	3.74	5	1	13

Eighty-one publications, representing 40.70 percent of the sample, had zero indexed citations during the observation year. Another 85 received between 1 and 5 citations, 15 received between 6 and 10, and 18 received at least 11. The annual citation series explicitly reported a positive count for 118 publications. For the remaining 81, the absence of a 2023 entry was coded as an indexed zero. These values describe the source database rather than the complete universe of scholarly citations. The distinction between annual and cumulative attention is central to interpretation. Every publication contributes citations from the same calendar year, although publications entered the scientific record at different times. Growth in the scientific literature changes the environment in which publications receive attention, as documented by Bornmann and Mutz, 2015. The reported counts therefore characterize attention within a common observation window rather than equivalent lifetime citation opportunities.

5.2 Reference Count and Subsequent Citation Attention:

To address the first research question, the analysis compares citation outcomes across four reference count bands. The bands are exploratory descriptive groupings rather than estimated thresholds. Publications with 1 to 10 indexed references received a mean of 0.42 citations, compared with 8.54 among publications with at least 51 references. Their respective median citation counts were 0 and 4. The difference therefore extends beyond the arithmetic means. The proportion with zero indexed citations declined across all four reference bands, from 78.79 percent among publications with 1 to 10 references to 16.22 percent among those with at least 51. This represents a difference of 62.57 percentage points. Within this sample, broader indexed referencing corresponded to a lower frequency of receiving no indexed attention during 2023.

Table 5: Citation attention across bibliographic characteristics

Characteristic	Observed Band	Publications	Mean Citations in 2023	Median Citations in 2023	Zero Indexed Citations, Percent
Reference count	1 to 10	33	0.42	0	78.79
Reference count	11 to 25	65	3.92	1	49.23
Reference count	26 to 50	64	3.70	1.5	26.56

Characteristic	Observed Band	Publications	Mean Citations in 2023	Median Citations in 2023	Zero Indexed Citations, Percent
Reference count	At least 51	37	8.54	4	16.22
Author count	1	19	0.37	0	73.68
Author count	2 to 3	81	4.02	0	50.62
Author count	4 to 6	69	3.99	1	30.43
Author count	At least 7	30	7.13	4	16.67
Publication age	1 to 3 years	76	5.55	2.5	22.37
Publication age	4 to 6 years	47	3.43	1	40.43
Publication age	7 to 9 years	36	1.39	0	55.56
Publication age	10 to 13 years	40	4.73	0	62.50

Mean citations did not increase consistently across the intermediate reference bands. Publications with 11 to 25 references averaged 3.92 citations, slightly exceeding the 3.70 recorded for publications with 26 to 50 references. However, the median increased from 1 to 1.5, while the proportion with zero indexed citations declined. Reference count therefore shows a clearer descriptive gradient in the incidence of receiving citation attention than in average citation volume. These results answer the first research question through observed distributional differences: publications in the largest reference band received more annual citation attention and were less frequently uncited than those in the smallest band. Reference quantity nevertheless remains distinct from the intellectual composition of references. Uzzi et al., 2013, demonstrate the relevance of knowledge combinations to scientific impact, whereas the present measure captures only the number of indexed references.

5.3 Author Count and Subsequent Citation Attention:

The second research question concerns variation across indexed team sizes. Publications with one author received a mean of 0.37 citations and had a median of 0. Publications with at least seven authors received a mean of 7.13 citations and had a median of 4. The proportion with zero indexed citations fell from 73.68 percent in the single author group to 16.67 percent in the largest team band. The intermediate bands reveal an important distinction between average attention and the distribution of attention. Publications with two to three authors averaged 4.02 citations, while those with four to six averaged 3.99. Their means were therefore similar. Nevertheless, the median increased from 0 to 1, and the proportion with zero indexed citations declined from 50.62 percent to 30.43 percent. Comparable averages conceal differences in how broadly citation attention was distributed across publications.

Across the four author bands, the proportion with zero indexed citations declined consistently as indexed team size increased. Mean citation counts showed a less regular pattern, with little difference between the two intermediate groups. The descriptive answer to the second research question is therefore strongest for the contrast between single author publications and the largest team band, and for the declining frequency of indexed zeros across bands. These comparisons do not establish that adding authors increases citation attention. Author count records participation scale without measuring individual contributions, disciplinary composition, or research organization. Wu et al., 2019, distinguish the scientific contributions associated with smaller and larger teams. The present findings concern annual citation distributions across team sizes rather than those underlying contribution mechanisms.

5.4 Publication Age and the Distribution of Annual Attention:

The third research question examines citation attention across publication ages. Publications aged 1 to 3 years received a mean of 5.55 citations and a median of 2.5. Mean citations declined to 3.43 among publications aged 4 to 6 years and to 1.39 among those aged 7 to 9 years. The corresponding medians declined to 1 and 0. The oldest band interrupts this decline in average citation volume. Publications aged 10 to 13 years received a mean of 4.73 citations, exceeding the averages of both intermediate age bands. However, their median remained 0, and 62.50 percent had zero indexed citations. The higher mean in the oldest band therefore coexists with a majority receiving no indexed attention during the observation year. The proportion with zero indexed citations increased across every age band, from 22.37 percent among publications aged 1 to 3 years to 62.50 percent among those aged 10 to 13 years. Mean citations followed a different pattern because the oldest band retained publications with substantial annual attention. This combination distinguishes the persistence of attention among some older publications from the typical outcome within their age group. Citation dynamics involving aging and publication specific trajectories, examined by Wang et al., 2013, provide a relevant interpretation of this heterogeneity without establishing its cause in the present sample.

The descriptive answer to the third research question is that older publication bands contained progressively larger proportions of indexed zeros, while average annual citations did not decline consistently across all bands. Publication age therefore cannot be summarized adequately by a uniformly increasing or decreasing pattern. Across the three research questions, distributional comparisons reveal patterns that averages alone obscure. Larger reference and author bands contained fewer publications with zero indexed citations, while older publication bands contained more. Citation concentration and differences between means and medians qualify these comparisons. As emphasized by the range of citation determinants reviewed by Tahamtan et al., 2016, these observed characteristics provide a partial account of scholarly attention. The evidence establishes descriptive variation within the retained sample without identifying independent effects or causal mechanisms.

6. Discussion:

Average citation counts can obscure whether scholarly attention reaches many publications or concentrates among a few. Tables 4 and 5 reveal this distinction across reference counts, team sizes, and publication ages. Larger reference and author bands contained progressively smaller proportions of publications with zero indexed citations, although their mean citation counts did not increase consistently across intermediate bands. The contribution of this analysis lies in separating the distribution of annual attention from its average volume within scientific benchmarking publications. This distinction refines the interpretation of

bibliographic characteristics: a characteristic can correspond to broader citation participation without producing a consistent gradient in mean citations. Tahamtan et al., 2016, identify multiple determinants of citation counts; the present evidence adds a focused descriptive comparison showing why means, medians, and indexed zeros require joint interpretation. Reference count captures connections to existing literature, but Table 5 shows that these connections correspond to different dimensions of citation attention. The proportion with zero indexed citations declined across every reference band, whereas average citations differed little between the two intermediate bands. The analytical implication is that reference quantity distinguishes the frequency of receiving attention more consistently than its volume across the selected groups. Uzzi et al., 2013, associate scientific impact with the composition of knowledge combinations. The present findings complement that distinction by showing that a simple reference count can describe differences in citation participation while remaining insufficient to explain intellectual contribution. The evidence therefore supports treating reference quantity as an observable bibliographic characteristic rather than a direct measure of novelty or research quality.

Team size exhibits a comparable separation between typical outcomes and averages. In Table 5, publications with two to three authors and those with four to six authors had similar mean citation counts, yet the latter group had a higher median and fewer indexed zeros. The largest team band also combined a higher median with a smaller proportion receiving no indexed citations. These patterns matter because an average alone would conceal differences in how attention is distributed across publications. Wu et al., 2019, distinguish the developmental and disruptive contributions associated with different team sizes. This analysis identifies a separate descriptive dimension: team size corresponds to variation in the distribution of annual citation attention. The resulting contribution concerns citation patterns across observed teams, without attributing those patterns to coordination, dissemination, or individual contributions that the dataset does not measure. Publication age exposes a different asymmetry. Using the age measure defined in Equation 1, Table 5 shows progressively larger proportions of indexed zeros across older publication bands, alongside a rebound in mean citations within the oldest band. Older publications therefore combine widespread absence of indexed annual attention with continuing recognition among a smaller subset. The concentration reported in Table 4 strengthens the need to distinguish typical outcomes from persistent attention at the upper end of the distribution. Wang et al., 2013, explain citation trajectories through interacting processes involving aging and publication fitness. The present evidence contributes a descriptive qualification: an older cohort can retain a substantial mean citation count even when its median is zero. Age comparisons consequently require distributional interpretation rather than a single assumption of continuing accumulation or uniform decline.

The international relevance of these findings rests on a measurement problem shared across research evaluation systems: citation averages can conceal differences in both participation and concentration. Bornmann and Mutz, 2015, document expansion in scientific publications and cited references, placing these comparisons within a growing system of competition for attention. Although the sample has no geographical restriction, it does not support comparisons between national research systems. Its transferable contribution is an evaluation approach that reports annual citation means alongside medians, indexed zeros, and concentration. Research evaluators should use these summaries together when comparing publication groups and should avoid interpreting longer reference lists or larger teams as demonstrated routes to greater impact. The evidence redirects attention toward how recognition is distributed, while keeping the interpretation proportionate to the observed bibliographic measures.

7. Conclusion and Implications:

Research evaluation can misread scholarly recognition when averages conceal how unevenly attention is distributed. This analysis shows that broader participation in citation attention and greater average attention represent distinct empirical patterns. The contribution is a distributional perspective that separates typical recognition from concentration among a smaller set of publications. This perspective qualifies interpretations that equate stronger aggregate performance with broadly shared visibility. Its theoretical value lies in showing how apparently favorable averages can coexist with limited recognition across much of a publication group. Research managers should therefore assess distributions before using aggregate indicators to compare outputs or allocate support. Evaluation policies should require transparent observation windows and complementary summaries of typical outcomes, absent indexed attention, and concentration. Operationally, institutions can incorporate these distinctions into reporting routines and review the underlying records before interpreting differences as evidence of research quality. Such practices can support more accountable judgments across international research settings.

The descriptive cross section provides a foundation for longitudinal inquiry into how recognition develops and persists. Broader samples and alternative bibliographic databases could assess whether the observed patterns extend beyond the selected records. Comparisons accounting for disciplinary composition, publication venues, and document types could distinguish compositional differences from more general relationships. Repeated annual observations could examine whether the contrast between typical attention and concentrated recognition remains stable. These extensions would strengthen generalizability and enable direct assessment of mechanisms that descriptive comparisons cannot identify.

References:

1. Bornmann, L., & Mutz, R. 2015. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Journal of the Association for Information Science and Technology, 66, issue 11, 2215-2222. https://doi.org/10.1002/asi.23329
2. Tahamtan, I., Safipour Afshar, A., &Ahamdzadeh, K. 2016. Factors affecting number of citations: A comprehensive review of the literature. Scientometrics, 107, issue 3, 1195-1225. https://doi.org/10.1007/s11192-016-1889-2
3. Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. 2013. Atypical combinations and scientific impact. Science, 342, issue 6157, 468-472. https://doi.org/10.1126/science.1240474
4. Wang, D., Song, C., &Barabási, A. L. 2013. Quantifying long term scientific impact. Science, 342, issue 6154, 127-132. https://doi.org/10.1126/science.1237825
5. Wu, L., Wang, D., & Evans, J. A. 2019. Large teams develop and small teams disrupt science and technology. Nature, 566, issue 7744, 378-382. https://doi.org/10.1038/s41586-019-0941-9