Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

When One Paper Counts Many Times: A Bibliometric Study of Authorship Expansion, Publication-Credit Inflation, and Research Integrity Across Ten Medical Specialties, 2000–2025

Michael Shalman

michael.shalman@gmail.com

Research Article

Keywords: Authorship, Team size, Fractional counting, Publication counts, Research integrity, Research assessment, Academic promotion, Academic leadership, Generative artificial intelligence, Bibliometrics

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11413212/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Publication counts influence selection for medical training, academic promotion and administrative leadership positions even as the number of authors per paper rises. Awarding a full publication credit to every listed name can distort comparisons of research records and raises questions of fairness and accountability. I examined how 26 years of authorship expansion affected full versus equal-fraction counts for individual bibliographic author names.

Methods

Bibliometric census of PubMed-indexed research articles published 2000–2025 in 32 journals of ten specialties, with two PubMed-wide random samples. I modeled named authors per article (negative binomial models with journal-clustered standard errors; interrupted time series) and, for three 3-year windows, compared author names by full count, equal-fraction count and lead (first, last or sole) positions under four name-disambiguation rules.

Results

Among 177,486 research articles, mean authors rose from 6.4 to 14.2 (annual increase 3.5%, 95% CI 3.1–3.9), in all ten specialties and all 32 journals. At the author-name level (names with ≥ 2 initials), names with the same equal-fraction credit (0.9–1.1) listed 3–8 papers (10th–90th percentile) in 2000–2002 and 4–16 in 2023–2025 (increase in the 90th percentile 8 papers; 95% bootstrap CI 6–9.7), and the share of lead positions among the papers of the 100 names with the highest full count fell from 39% to 23% (decrease 15.1 percentage points; 95% CI 9.0–21.5). In 2023–2025, the top 100 names by full count shared 44–55 names with the top 100 by fractional count under the three stricter name rules (89 when all names were included). Of the 100 most prolific names, 35 probably merged different people. No clear discontinuity in author counts was observed after 2022.

Conclusions

Authorship expansion has made publication rankings of individual names strongly dependent on how credit is counted. Neither full nor equal-fraction counts measure actual contribution, and bibliographic records cannot identify honorary authorship. Nevertheless, using unadjusted publication volume to allocate competitive training, academic or administrative positions risks unfairly equating accumulated bylines with scientific achievement. Evaluation should prioritize the quality and originality of selected works and documented individual contributions.

Background

The list of authors on a medical scientific article, the peer-reviewed publication indexed in MEDLINE and PubMed that is the basic unit of the biomedical literature, serves two purposes that are increasingly in tension. It is a record of who did the work and who is accountable for it, and it is also the main currency of academic careers. The International Committee of Medical Journal Editors (ICMJE) defines authorship by four cumulative criteria, including accountability for all aspects of the work [1], and the Singapore Statement on Research Integrity states that author lists “should include all those and only those who meet applicable authorship criteria” [2]. Yet the same author list is counted, almost without adjustment, when candidates are compared for residency and fellowship positions, academic appointments, promotion, research funding and administrative leadership posts, for which academic rank is often a prerequisite. In an international sample of 92 biomedical faculties, the number of publications was mentioned in 95% of promotion and tenure documents, whereas data sharing was mentioned in 1% [3].

The number of authors per article in medicine has risen for decades [4–7]. Much of this growth reflects genuine changes in the way clinical research is done: multicenter trials, consortia, multidisciplinary teams and methodological specialization [8]. Some of it may not. Papatheodorou and colleagues showed that author numbers grew even where study complexity did not [6], and surveys of authors consistently report honorary (gift or guest) authorship on roughly one in five articles [9–11]. The most comprehensive recent analysis of PubMed, covering about 17 million records, found that mean authors per publication rose from 3.99 in 2000 to 6.25 in 2020, while single-author papers fell from 17.0% to 5.7% [12]. Comparable growth has been documented in neuroscience and psychology [13], in anesthesiology [14–16] and in cardiothoracic surgery [17]. Two phenomena must therefore be distinguished: authorship expansion, an increase in the number of people listed, which can be measured from bibliographic records; and inappropriate authorship, the listing of people who do not meet authorship criteria, which cannot. The question I address lies between them: how far a full publication count, the unit used on curricula vitae, still corresponds to an individual’s share of the published work.

Two developments make the question timely. First, generative artificial intelligence (AI) has lowered the cost of producing text. At least 13.5% of PubMed abstracts published in 2024 show linguistic traces of large language model processing, rising to about 40% in some subgroups [18]. Formulaic single-factor analyses of one freely available dataset, the US National Health and Nutrition Examination Survey, rose from about four per year in 2014–2021 to 190 in 2024 (to 9 October); the authors attributed this growth to AI-assisted workflows and paper mills, although AI use was inferred rather than measured [19]. The total number of indexed articles has also grown faster than the number of active scientists [20]. If AI enables many more manuscripts while careers still depend on publication counts, incentives to accumulate byline credit may intensify. Second, competition for clinical training positions has become openly quantitative. Mean self-reported numbers of abstracts, presentations and publications of matched applicants, as recorded by the US National Resident Matching Program, have nearly doubled

within four years in anesthesiology and plastic surgery [21, 22]; verified peer-reviewed output is much smaller than self-reported output, and some listed publications are misrepresented [23–25].

Most previous studies described a single specialty, ended before 2022, or did not examine how growth in team size changes comparisons between individual researchers [12–17]. Growth in mean team size itself is well documented, and its aggregate consequences follow arithmetically: across all authorships of a set of papers, total full credit divided by total equal fractional credit equals the mean number of authors per paper. What is not determined by the mean is how credit is distributed among individual names, which depends on the variation of team sizes across each person's papers. I therefore performed a census of research articles published over 26 years in 32 journals of ten medical specialties, together with two random samples of the whole of PubMed. My principal aim was to describe how authorship expansion changes publication-based comparisons between individual author names under full, fractional and lead-position counting. Secondary aims were to quantify authorship expansion by specialty, to estimate the growth of authorship positions across PubMed, and to describe the first years after generative AI became widely available, as a baseline for future monitoring.

Methods

Design and data source

This was a retrospective bibliometric census of PubMed records, reported as original bibliometric meta-research. PubMed was chosen because it is the principal index of the medical scientific literature and the database against which applicants' reported publications have been verified in studies of residency selection [23, 25]. Records were retrieved through the National Center for Biotechnology Information Entrez Programming Utilities [26] on 14–15 October 2026, at least nine months after the end of the last study year. No human participants were involved and ethical approval was not required. Reporting follows the BIBLIO guideline for bibliometric studies of the biomedical literature [27] as far as it applies to a primary bibliometric study; the completed checklist is provided in Additional file 1. All queries, data and code are available (see Availability of data and materials).

Specialties and journals

I selected ten specialties that together cover medicine, surgery, diagnostics and mental health: anesthesiology, general medicine, cardiology, oncology, neurology, general surgery, orthopedics, pediatrics, radiology and psychiatry. Journals were selected purposively before data retrieval, using four criteria: general rather than subspecialty scope within the specialty; publication of original clinical research; MEDLINE indexing; and status as a flagship journal of a major specialty society or a widely recognized high-impact general journal of the field. Three or four journals were chosen per specialty (Table 1). Selection was not based on a formal ranking, and the sample is not intended to represent the whole medical literature; the journals are widely recognized and influential in their fields, but I did not

measure their comparative weight in academic evaluation. Several journals were launched or renamed during the study period; the years in which they entered the sample are given in Table 1, and their influence was examined in sensitivity analyses.

Table 1
Specialties, journals and numbers of articles, 2000–2025

Specialty	Journals (PubMed abbreviation)	Eligible records	Research articles analyzed
Anesthesiology	Anesthesiology; Br J Anaesth; Anesth Analg; Anaesthesia	20,823	20,446
General medicine	N Engl J Med; Lancet; JAMA; BMJ	22,887	19,503
Cardiology	Circulation; Eur Heart J; J Am Coll Cardiol	25,661	25,512
Oncology	J Clin Oncol; Lancet Oncol (from 2001); JAMA Oncol (from 2015)	14,702	14,258
Neurology	Neurology; Lancet Neurol (from 2002); Ann Neurol	18,007	16,018
General surgery	Ann Surg; JAMA Surg (from 2013); Br J Surg	13,142	12,602
Orthopedics	J Bone Joint Surg Am; Bone Joint J (from 2013); Am J Sports Med	16,195	16,064
Pediatrics	Pediatrics; JAMA Pediatr (from 2013); J Pediatr	21,934	20,417
Radiology	Radiology; Eur Radiol; AJR Am J Roentgenol	28,354	27,749
Psychiatry	Am J Psychiatry; JAMA Psychiatry (from 2013); Lancet Psychiatry (from 2014)	5,134	4,917
Total	32 journals	186,839	177,486

Eligible records: journal articles with an abstract after exclusion of non-research publication types. Research articles analyzed: records with a structured abstract or an abstract of at least 100 words and at least one named author. Years in parentheses indicate the first year in which the journal was indexed under its current title; earlier volumes of renamed journals (for example Archives of Surgery) were not included. Journals were selected purposively (see Methods).

Eligibility

All records with the publication type “journal article” and an abstract, published from 2000 to 2025, were retrieved for each journal. I excluded editorials, letters, comments, reviews, systematic reviews and meta-analyses, guidelines, consensus statements, case reports, news, biographies, historical articles, interviews, lectures, errata, retractions and retracted publications. During quality control I found that one

journal began to deposit abstracts for news features, patient pages and opinion pieces from 2022 onwards, which are not distinguished by PubMed publication type. Because some journals deposit research abstracts in PubMed without section labels, I applied a single objective criterion to all journals and years: the primary analysis included articles with either a structured abstract (at least two labeled sections, such as Background, Methods, Results and Conclusions) or an unstructured abstract of at least 100 words. In the excluded items the abstract field was empty or contained a one- or two-sentence summary, typical of news, opinion, patient-information and research-letter formats. Articles without any named individual author were excluded from the analysis of author counts. The PubMed publication-date field ([dp]) matches both the electronic and the print publication date. Records were retrieved year by year, and an article whose electronic and print dates fell in different years was retrieved under both; the earlier year (usually the electronic publication) was used as the publication year in all main analyses. This applied to 17,443 articles (9.8%). As an alternative definition, I assigned these articles to the later year, which agreed with the year of the print issue recorded in PubMed for 99.3% of 28,800 articles checked, and repeated the main models.

Variables

For each article I extracted the number of named individual authors, the presence of a collective (group or consortium) author, the publication types and the publication year. The main outcome was the number of named authors per article. Secondary outcomes were the median number of authors, the proportion of articles with ten or more authors, the proportion with a single author and the proportion with more than 100 authors.

PubMed-wide samples and estimated authorship positions

To place the specialty data in context, I drew two random samples of PubMed. The primary sample was designed as a two-stage sample: for each year from 2000 to 2025, 20 calendar days were drawn at random (Python random module, seed 7), and for each day up to 100 records were drawn at random from all records with the publication type “journal article”, the given publication year and a PubMed record-creation date ([crdt]) on that day. Days without eligible records were skipped (438 days contributed records). Because the number of eligible records differs between days, the total number of eligible records was obtained for every sampled day and annual means were weighted by these totals; for 33 days with more than 5,000 records the draw was made from the first 5,000 identifiers returned. Uncertainty was estimated by a bootstrap that resampled days within years (2,000 resamples). Because restricting these days to the first 5,000 identifiers could introduce bias that weighting does not remove, I repeated the estimates after excluding them. As a validation, I drew a second, independent random numerical-PMID sample by selecting 140,000 PubMed identifiers (PMIDs) uniformly at random between 10,500,000 and 41,800,000 (seed 2026) and retaining records with the publication type “journal article” and a publication year from 2000 to 2025 (107,123 records). Each record with a PMID in this range had the same probability of selection; PMIDs are assigned sequentially as records are created, so the range

corresponds to records created between 29 September 1999 and 9 March 2026, which included at least 99.06% of eligible journal articles in each publication year examined (99.06% for 2025, because some 2025 articles were indexed later). Coverage was calculated as the proportion of eligible records whose record-creation date ([crdt]) fell within the dates of the first and last PMIDs of the range (code coverage.py). I regard this sample as a validation sample rather than a demonstrably unbiased sample of all PubMed records. The annual number of PubMed records with the publication type “journal article” was obtained from PubMed counts. Estimated authorship positions per year were calculated as the number of journal articles multiplied by the mean number of named authors. This quantity is the number of byline positions across indexed journal articles; it approximates the number of publication entries that would appear on curricula vitae if every author counted every paper in full, but it is not a count of curriculum-vitae entries, of researchers or of research performed. Estimates were repeated after exclusion of articles with more than 100 authors.

Name-level analysis

I linked every research article to its author list and analyzed credit at the level of author names (surname plus initials, as given in PubMed). For three 3-year windows (2000–2002, 2011–2013 and 2023–2025) I calculated, for each name, the full count (number of papers on which the name appears, the method used in curricula vitae), the equal-fraction count (sum of 1/number of authors over the name’s papers) and the number of papers in a lead position (first, last or sole author). Equal fractional counting is an accounting convention that divides each paper equally among its authors [28]; it does not measure how much work an author performed, and first and last positions are only imperfect proxies of leadership. None of the three counts is a validated measure of contribution. Two aggregate quantities are arithmetic consequences of team size and were not treated as findings: the share of middle positions, which for a paper with n authors is (n − 2)/n, and the ratio of total full to total fractional credit, which across all names equals the mean number of authors per paper. The name-level outcomes were the distribution of full counts among names with the same equal-fraction credit (0.9–1.1; names with at least two papers), the Spearman correlations between full count and fractional count and between full count and lead positions (names with at least five papers), the overlap of the 100 names with the highest full count with the top 100 by fractional count and by lead positions, and the share of lead positions among the papers of the top 100 names by full count. These outcomes depend on how team sizes vary across the papers of each name, not only on the mean. Because the three windows contain different cross-sectional populations of names, papers and team sizes, the comparisons between windows are descriptive. To describe sampling uncertainty, I calculated 95% percentile bootstrap confidence intervals for the 10th, 50th and 90th percentiles of full counts at equal fractional credit and for the lead-position share among the top 100 names, by resampling names with replacement within each window (2,000 resamples; numpy default_rng, seed 303; the top 100 were re-selected in each resample), and for the differences between 2023–2025 and 2000–2002 from independent resamples of the two windows. These intervals do not capture uncertainty arising from name disambiguation, which was examined with the four name rules.

Name-based linkage merges different people who share a surname and initials and may split one person across name variants. I therefore compared four rules: all names; names with at least two initials; names with at least two initials after exclusion of the 50 most frequent surnames; and names with at least two initials after exclusion of probable merged identities among the 100 most prolific names. For this last rule, full author records (forename, ORCID identifier and affiliation) were retrieved for all 2023–2025 papers of the 100 names with the highest full count, and a name was classified as a probable merge of different people if fewer than 80% of its papers carried the same first forename or if it was linked to two or more ORCID identifiers with none accounting for at least 80% of ORCID-linked papers. The primary rule was names with at least two initials. The fourth rule is an operational approximation: missing ORCID identifiers, incomplete metadata, transliteration and name changes can affect classification, and none of the rules yields a fully disambiguated set of individual researchers.

Publication volume after 2022

To describe publication volume after the release of ChatGPT, I obtained annual PubMed counts of research articles (journal articles with an abstract, after exclusion of the same non-research publication types; the abstract-length criterion was not applied to these counts) and compared counts for 2023–2025 with those projected from a log-linear trend fitted to 2010–2022. This comparison is descriptive and was not designed to estimate an effect of AI.

Statistical analysis

Means are reported with 95% bootstrap confidence intervals (2,000 resamples) and medians with interquartile ranges. Because author counts are overdispersed counts, annual change was estimated with negative binomial generalized linear models (log link) with publication year as a continuous predictor; the dispersion parameter was estimated by maximum likelihood on a random subsample of 30,000 articles ($\alpha = 0.18$) and then fixed, and the exponentiated year coefficient gives the average annual percentage change. The pooled model included specialty as a fixed effect. Because articles are clustered within journals, the primary pooled estimates use standard errors clustered by journal (32 clusters); I also fitted models with journal fixed effects, estimated journal-specific trends and pooled them with a DerSimonian–Laird random-effects model, and repeated the pooled model leaving out one journal at a time. Specialty-specific estimates use heteroscedasticity-robust standard errors, because three or four journals per specialty are too few for reliable clustering; these ten estimates are descriptive, and I report whether they remain significant after Bonferroni correction.

To describe the period after the release of ChatGPT (November 2022), I fitted an interrupted time-series model with the 2000–2022 trend, a level change for 2023–2025 and a slope change after 2022, at article level with journal fixed effects and journal-clustered standard errors. Because the release of ChatGPT was not an intervention applied simultaneously to all researchers, adoption has been gradual and publication lags behind research, this model can detect only a change in trend; it cannot attribute a change to AI, and there is no unexposed comparison group. I repeated the model after excluding 2025,

whose indexing may be incomplete, and with the post-period starting in 2024 to allow for publication lag. To assess serial correlation, I also fitted the model to the 26 annual log-means, reporting the Durbin–Watson statistic and the Ljung–Box test, and estimated it with Newey–West (lag 2) standard errors and with first-order autoregressive errors. The pooled and interrupted time-series models were also repeated with the alternative (later) publication-year definition. Sensitivity analyses repeated the pooled estimates (i) for all eligible records regardless of abstract type, (ii) for journals indexed under their current title throughout 2000–2025, (iii) after exclusion of articles with a collective author or more than 100 authors, and (iv) for randomized and other clinical trials only. Analyses were performed in Python 3 with pandas, statsmodels and SciPy. Two-sided P values below 0.05 were considered significant.

Use of AI-assisted tools and validation of code

AI-assisted tools (Elicit and Perplexity, used 2025–2026) were used to support literature identification, to write the data-retrieval, analysis and figure-plotting code, and to assist with drafting and language editing under the author's direction. The study question, design, interpretation and conclusions are the author's own, and the author checked all AI output, including every reference. Code was validated in four ways. First, all scripts were re-run end to end from the raw downloads. Second, for 500 randomly selected analyzed articles the author lists were downloaded again in a different format (XML rather than the summary format used in the main retrieval) and parsed with independently written code; the number of named authors agreed for 500 of 500 articles, and the publication year for 449, the remainder differing by one year (50 articles) or two years (1 article) because of online-ahead-of-print dates. Third, the PubMed-wide estimates were compared with published values [12] and between two independent sampling designs. Fourth, internal consistency checks were performed (unique identifiers, totals across tables and figures). The author takes full responsibility for the code, the analyses and the text. All figures were drawn from the data by this code; no figure was produced with a generative AI image tool. No AI tool meets the criteria for authorship and none is listed as an author.

Reporting of the bibliometric analyses is mapped to the BIBLIO checklist (Additional file 1) [27].

Results

Study sample

The search identified 186,839 eligible journal articles in the 32 journals. After exclusion of 8,790 items without a research-type abstract (mainly news, opinion and research-letter items in JAMA-network journals after 2017) and 563 articles without a named individual author, 177,486 research articles were analyzed (Table 1, Fig. 1). The random-day PubMed sample comprised 42,331 journal articles and the random numerical-PMID validation sample 107,123. Weighted mean authors in the random-day sample were 4.24 in 2000 and 6.34 in 2020, and in the random-PMID sample 4.02 and 6.31, close to the values reported for the full PubMed database by Jakab and colleagues (3.99 and 6.25) [12].

Authorship expansion across specialties

The mean number of authors per research article increased in every specialty (Table 2, Fig. 2). Across all ten specialties the mean rose from 6.4 in 2000 to 14.2 in 2025, and the median from 6 to 11. The pooled average annual increase was 3.52% (95% CI 3.11–3.93, journal-clustered standard errors), and 3.42% (95% CI 3.00–3.85) with journal fixed effects. Every one of the 32 journals showed a statistically significant increase (journal-specific annual change 1.1% to 6.7%; median 3.7%, IQR 2.7–4.4; random-effects pooled estimate 3.62% (95% CI 3.22–4.01), I^2 = 99%), and leaving out any single journal gave pooled estimates between 3.42% and 3.59%. The fastest growth occurred in general medicine (5.4% per year; 6.7 to 24.4), oncology (4.0% per year) and cardiology (4.0% per year). The slowest growth was in radiology (2.4% per year), pediatrics and orthopedics. All ten specialty-specific trends remained significant after Bonferroni correction (all $P < 0.001$). In anesthesiology the mean rose from 5.0 to 9.6 (2.7% per year).

The distribution shifted accordingly. Articles with ten or more authors increased from 14% to 57% of all research articles, and from 2% to 41% in anesthesiology; in oncology, general medicine and cardiology more than four in five articles had at least ten authors in 2025 (Fig. 3). Single-author research articles fell from 1.2% to 0.15%. Articles with a collective (consortium or group) author increased from 0.7% to 15.3%, whereas articles with more than 100 named authors remained rare (0.11% in 2025).

Table 2
Authors per research article by specialty, 2000 and 2025

Specialty	Mean (95% CI), 2000	Mean (95% CI), 2025	Median (IQR), 2000 → 2025	≥ 10 authors, % 2000 → 2025	Annual change, % (95% CI)
General medicine	6.7 (6.5–7.0)	24.4 (23.0–25.7)	6 (4–9) → 21 (13–31)	20 → 87	5.4 (5.2–5.6)
Oncology	9.7 (9.3–10.1)	23.1 (22.1–24.2)	10 (7–12) → 21 (17–28)	50 → 93	4.0 (3.9–4.2)
Cardiology	7.3 (7.2–7.5)	18.8 (17.8–19.9)	7 (5–9) → 17 (11–23)	21 → 83	4.0 (3.9–4.1)
General surgery	5.6 (5.4–5.9)	14.2 (13.3–15.2)	5 (4–7) → 11 (8–17)	6 → 60	3.8 (3.7–4.0)
Neurology	7.0 (6.7–7.3)	15.5 (14.5–16.7)	6 (5–9) → 12 (8–18)	17 → 65	3.6 (3.4–3.7)
Psychiatry	5.9 (5.6–6.2)	12.2 (10.8–13.9)	6 (4–7) → 10 (7–15)	11 → 51	3.5 (3.2–3.8)
Anesthesiology	5.0 (4.8–5.1)	9.6 (9.2–10.1)	5 (3–6) → 9 (6–12)	2 → 41	2.7 (2.6–2.8)
Orthopedics	4.3 (4.1–4.5)	8.1 (7.7–8.5)	4 (3–5) → 7 (6–9)	0 → 23	2.7 (2.4–2.9)
Pediatrics	5.2 (5.0–5.4)	9.7 (9.1–10.3)	5 (4–6) → 8 (5–12)	7 → 36	2.6 (2.5–2.8)
Radiology	5.9 (5.7–6.1)	10.6 (10.2–11.1)	6 (4–7) → 9 (7–13)	7 → 49	2.4 (2.3–2.5)
All ten specialties	6.4 (6.3–6.4)	14.2 (13.9–14.5)	6 (4–8) → 11 (7–18)	14 → 57	3.5 (3.1–3.9)*
PubMed random-day sample (weighted)	4.2 (4.1–4.4)	7.0 (6.7–7.2)	4 → 6	—	—

Specialties are ordered by annual change. CI, confidence interval (bootstrap with 2,000 resamples for means; day-clustered for the PubMed sample); IQR, interquartile range. Annual change is the exponentiated year coefficient of a negative binomial regression for 2000–2025, with heteroscedasticity-robust standard errors for specialties. *Pooled estimate adjusted for specialty, with standard errors clustered by journal.

The period after 2022

No clear discontinuity in authorship counts was observed after 2022 (Fig. 4). In the original article-level model with heteroscedasticity-robust standard errors, the level change in 2023 was 0.98 (95% CI 0.95–1.01) (P = 0.15) and the slope change after 2022 was 1.013 (95% CI 1.001–1.026) per year (P = 0.04), a small nominal increase. With journal fixed effects and journal-clustered standard errors, neither the level change (0.98 (95% CI 0.92–1.03); P = 0.37) nor the slope change (1.014 (95% CI 0.995–1.033); P = 0.14) was significant, and the same was true after exclusion of 2025 (slope change 1.04 (95% CI 0.99–1.10); P = 0.11) and with the post-period starting in 2024 (slope change 0.99 (95% CI 0.95–1.02); P = 0.50). The annual series showed positive serial correlation (Durbin–Watson 1.04; Ljung–Box P = 0.008); with first-order autoregressive errors the slope change was 1.01 (95% CI 0.98–1.04) (P = 0.44), and with Newey–West standard errors 1.010 (95% CI 0.998–1.022) (P = 0.12). The observed mean in 2025 (14.2) was slightly above the value projected from the pre-2023 trend (13.8).

Author-name level: full, fractional and lead-position counts

The 177,486 research articles contained 1,663,368 authorship positions. Middle-author positions increased from 69% in 2000 to 86% in 2025 and first-author positions fell from 15.5% to 7.1% (Fig. 5A); the shares expected from the distribution of team sizes alone (68.8% and 85.9%) were identical to the observed shares. Likewise, the ratio of total full to total fractional credit across all names (6.54, 8.89 and 13.74 in the three windows) equaled the mean number of authors per paper (6.54, 8.90 and 13.75). Both quantities restate the growth of team size and are reported only for completeness.

At the level of individual names, the results under the primary rule (names with at least two initials) are shown in Table 3. Among names with the same equal-fraction credit, the full count ranged from 3 to 8 papers (10th–90th percentile; median 5) in 2000–2002, from 4 to 11 (median 6) in 2011–2013 and from 4 to 16 (median 8) in 2023–2025 (Fig. 5B and C). The 90th percentile was 8 (95% CI 8–8.4), 11 (95% CI 10–12) and 16 (95% CI 14.7–17.7) papers in the three windows; between 2000–2002 and 2023–2025 it increased by 8 papers (95% bootstrap CI 6–9.7) and the median by 3 papers (95% CI 3–4). Of the papers of the 100 names with the highest full count, 39% (95% CI 34–44), 30% (95% CI 27–34) and 23% (95% CI 20–27) were in a lead position (decrease between the first and last windows 15.1 percentage points, 95% CI 9.0–21.5). The correlation between full and fractional counts was 0.66, 0.61 and 0.56, and between full count and lead positions 0.37, 0.36 and 0.30. The top 100 names by full count shared 61, 64 and 55 names with the top 100 by fractional count, and 44, 53 and 46 with the top 100 by lead positions; these overlaps show that rankings depend on the counting method but did not change consistently over time. When all names were included, the widening spread of full counts at equal fractional credit (3–8 to 5–16) and the falling lead-position share among the top 100 (41% to 15%) were reproduced, but the correlation between full and fractional counts showed no consistent trend (0.68, 0.63, 0.66). These are descriptive comparisons of different cross-sectional populations; the bootstrap intervals describe sampling variation among names, not uncertainty in name disambiguation.

Table 3
Author-name level comparisons of full, fractional and lead-position counts (names with at least two initials)

Window	Names	Mean authors per paper	Papers listed by names with equal fractional credit, P10–P90 (median)	Spearman ρ: full vs fractional / full vs lead	Top-100 overlap: fractional / lead	Lead-position share, top 100 by full count, % (95% CI)
2000–2002	36,329	6.5	3–8 (5; n = 487); P90 95% CI 8–8.4	0.66 / 0.37 (n = 1,861)	61 / 44	39 (34–44)
2011–2013	45,164	8.9	4–11 (6; n = 353); P90 95% CI 10–12	0.61 / 0.36 (n = 2,526)	64 / 53	30 (27–34)
2023–2025	47,521	13.7	4–16 (8; n = 184); P90 95% CI 14.7–17.7	0.56 / 0.30 (n = 2,598)	55 / 46	23 (20–27)
Difference, 2023–2025 minus 2000–2002	–	–	P10 + 1 (0 to + 2); median + 3 (+ 3 to + 4); P90 + 8 (+ 6 to + 9.7)	–	–	−15.1 (− 21.5 to − 9.0)

Full count: papers on which the name appears. Fractional count: sum of 1/number of authors. Lead: first, last or sole author. Equal fractional credit: 0.9–1.1, names with at least two papers; correlations: names with at least five papers. 95% CIs: percentile bootstrap over names within each window (2,000 resamples); differences from independent resamples of the two windows. The windows are different cross-sectional populations, and the comparisons are descriptive; the intervals do not include uncertainty from name disambiguation. None of these counts is a validated measure of contribution. Results under four name-disambiguation rules are given in Additional file 2.

Of the 125,754 names in 2023–2025, 78,233 (62%) had a single initial, and a further 4,040 carried one of the 50 most frequent surnames. Among the 100 most prolific names, 35 were classified as probable merges of different people on the basis of forenames and ORCID identifiers, which shows that the most prolific bibliographic identities are strongly affected by name collisions. The rankings depended on the rule (Additional file 2): the overlap between the top 100 by full and by fractional count was 89 when all names were included and 44–55 under the three stricter rules, and the overlap between the top 100 by full count and by lead positions was 66 and 34–46. The spread of full counts at equal fractional credit (90th percentile about 16 papers under every rule) and the lead-position share among the top 100 under the stricter rules (23–25%) were more stable. The corrected results still do not represent a fully disambiguated set of individual researchers.

PubMed-wide output and estimated authorship positions

The number of PubMed records with the publication type “journal article” rose from 488,062 in 2000 to 1,764,212 in 2025 (×3.6); research articles, a narrower category that excludes records without an abstract and non-research types, rose from 328,850 to 1,348,347 (Fig. 6A). Estimated authorship positions use the broader category, because the random samples were drawn from all journal articles. In the weighted random-day sample, mean authors increased from 4.24 to 6.96, and estimated authorship positions from 2.07 million (95% CI 2.01–2.14) in 2000 to 12.28 million (11.90–12.72) in 2025, a 5.9-fold increase (Fig. 6B). The random numerical-PMID validation sample gave 1.96 and 12.26 million (×6.2), exclusion of articles with more than 100 authors changed the 2025 estimate by less than 1%, and exclusion of the 33 days restricted to the first 5,000 identifiers changed annual means by at most 0.46 authors and the 2000–2025 increase from 5.93-fold to 5.94-fold. The peaks in 2017 and 2021 were caused by single sampled articles with several hundred or more authors and disappeared when such articles were excluded; medians rose from 4 to 6 and 5% trimmed means from 4.1 to 6.5.

Research articles grew by 5.2% per year from 2010 to 2022. Counts in 2023 (1,212,795), 2024 (1,246,541) and 2025 (1,348,347) were below the values projected from this trend (1,322,286, 1,391,434 and 1,464,199). This descriptive comparison shows no surge in volume in the first three years after the release of ChatGPT; it does not test for an effect of AI.

Sensitivity analyses

The pooled annual increase was 3.19% (95% CI 3.14–3.24) when all eligible records were included regardless of abstract type, 3.01% (95% CI 2.97–3.05) after exclusion of articles with a collective author or more than 100 authors, 3.46% (95% CI 3.41–3.51) in journals indexed throughout the study period, and 3.87% (95% CI 3.79–3.95) for randomized and other clinical trials (n = 40,462); these intervals use robust standard errors and are therefore narrower than the journal-clustered primary interval. Medians showed the same pattern as means in every specialty (Table 2). With the alternative publication-year definition, the pooled annual increase was 3.41% (95% CI 2.98–3.83, journal fixed effects and clustered standard errors; 3.42% with the main definition), the 2025 mean was 14.1 (14.2), and the post-2022 slope change was 1.011 (95% CI 0.992–1.029; P = 0.26).

Discussion

Principal findings

In this census of 177,486 research articles from 32 journals of ten specialties, the number of authors per article more than doubled between 2000 and 2025, in every specialty, every journal and every sensitivity analysis, and estimated authorship positions across PubMed increased about six-fold. This growth in team size is the descriptive background; its aggregate consequences, a larger share of middle positions and a higher ratio of full to fractional credit, follow arithmetically from it. The principal additional finding concerns individual names. Names with the same equal-fraction credit now list between 4 and 16 papers (10th–90th percentile), compared with 3 to 8 two decades earlier, a widening of the distribution of full counts at equal fractional credit; the share of lead positions among the most prolific names fell from

39% to 23%. Both findings were also observed when all names were included, and the bootstrap confidence intervals for the changes in the median, the 90th percentile and the lead-position share excluded zero. Which names appear at the top of a publication ranking depends on the counting method, but the overlap between the top 100 by full and by fractional count did not decline consistently (61, 64 and 55 names), and the decline in the correlation between full and fractional counts depended on the name rule. When such rankings shape careers, the problem is ethical as well as statistical. These findings extend the PubMed-wide analysis of Jakab and colleagues [12] beyond 2020 and agree with earlier evidence [13–16].

What the data show and what they do not

Three levels of evidence should be distinguished. First, the study directly demonstrates authorship expansion: more names per article in every specialty and journal studied. Second, it documents publication-credit inflation in an accounting sense: each paper contributes one unit of fractional credit in total but one full credit to every listed name, so names with comparable fractional credit now hold more dispersed full counts, and rankings depend on the metric. Third, it does not measure inappropriate authorship. The largest increases occurred in specialties dominated by multicenter trials and consortia, and legitimate collaboration, methodological specialization and research networks can explain an unknown part of the growth [8]. Evidence that part of the growth is inappropriate comes from outside this study: surveys in which roughly one in five articles is reported to include honorary authors [9–11] and the finding that author numbers grew where study complexity did not [6]. Over a century of science, individual productivity measured fractionally has not increased, and the apparent growth in output per researcher is largely an effect of crediting each of more names with the whole paper [29].

Fractional credit has its own limitations. Equal fractional counting assumes equal contributions, which is rarely true: a first author who designed the study, collected the data and wrote the manuscript has usually contributed more than a middle author, yet both receive the same fraction. The meaning of the last position differs between fields and teams, and author order and contribution statements are often not aligned [30, 31]. Fractional counting therefore cannot distinguish honorary authorship from legitimate collaboration. Its value in this study is narrower: it is a transparent accounting reference that shows how much of the growth of individual publication lists is due to the multiplication of names rather than of papers [28]. The same applies to rankings. A researcher who is a middle author on many multicenter trials may make substantial legitimate contributions without a first or last position, and a high full count with few lead positions is not evidence of honorary authorship. My results show that rankings are metric-dependent; they do not show which ranking is scientifically or ethically correct, which is why I argue for assessing verified contributions rather than for replacing one count with another.

Ethical interpretation: competition for academic and administrative positions

The following is an ethical interpretation supported by existing evidence, not a finding of the present data. An author list is a public statement of fact. The Singapore Statement requires that it include “all

those and only those" who meet authorship criteria, and the European Code of Conduct for Research Integrity names manipulating authorship, misusing seniority to advance one's career and misrepresenting research achievements as unacceptable practices [2, 32]. In surveys of health-science researchers, 18–26% perceived at least one co-author of the publication at issue as honorary, and 51% when declared contributions were compared with ICMJE criteria [11]. Honorary authorship is strongly associated with the automatic listing of department heads (adjusted odds ratio 3.80) [33], and its reporting varies with academic rank, with many researchers feeling pressured to add names [34].

Residency places, fellowships, academic appointments, promotion, grants and administrative leadership posts, such as head of department, unit director or dean, are allocated competitively, so an advantage gained by one candidate is a disadvantage to another. Publication counts appear in 95% of promotion criteria [3] and are associated with academic rank in anesthesiology [35]. Where unearned authorship occurs and evaluation relies on full counts, it is plausible that it confers an unfair advantage. My data show how strongly full counts and fractional credit now diverge for individual names, but they do not establish how often this mechanism operates or how much of the divergence it explains. The hypothetical model in Fig. 7 illustrates the mechanism under the explicit assumption that the added names made no qualifying contribution; five researchers who genuinely collaborated on all ten papers would produce the same bibliometric pattern. Such advantages may be amplified through collective self-citation [36], the h-index [37] and the cumulative advantage described by Merton as the Matthew effect [38]. The same pressure operates at the start of careers: the mean self-reported number of abstracts, presentations and publications of matched US MD anesthesiology applicants rose from 5.2 to 9 between 2020 and 2024 [21]. In orthopedics, PubMed-verified publications of matched applicants amounted to 23% of mean self-reported research outputs, most of which were non-peer-reviewed activities [25]. Publication misrepresentation has been found in about 5% of residency and fellowship applicants overall and in about 20% of orthopedic applications [23, 24]. The advantage may extend to administrative power. In many countries academic rank is a formal or informal prerequisite for leadership: the h-index rises with academic rank and is highest among department chairs [39], and 73% of US deans of medicine had previously served as department chairs [40]. In a national analysis of Lithuanian medical researchers, prolific leaders had more publications (mean 78 vs 49) and a higher h-index (19.7 vs 12.6) than prolific non-leaders, yet only 10.8% of their papers were first-authored, compared with 36.3% [41]. The ethical risk is structural as well as individual: when raw publication counts decide selection, systems may reward the multiplication of names and disadvantage those who keep honest author lists.

Generative AI: hypotheses for prospective testing

The present data cannot establish whether generative AI has influenced authorship or publication volume: no clear discontinuity was observed, the post-2022 period is short, adoption has been gradual, publication lags behind research, and there is no unexposed comparison group. The prospect of AI-enabled publication growth nevertheless makes this an urgent problem for research assessment. Use of AI is increasing among physicians, 81% of whom reported professional use in 2026, more than double the share in 2023 [42], and among researchers for research or publication tasks (62% in 2025 vs 45% in

2024) [43]. A substantial share of recent PubMed abstracts show traces of large language model processing [18]. A recent preprint estimated that by the end of 2025, 89% of open-access biomedical full-text papers in PubMed Central showed an excess of vocabulary associated with large language models [44], and formulaic analyses of public databases have increased sharply [19]. I propose, as hypotheses for prospective testing, that if AI raises output per researcher while evaluation relies on full counts, future data will show more papers per author name, initially outside the journals studied here; a rebound of small-team papers [45]; and further widening of the distribution of full counts at equal fractional credit. The present data provide a baseline for these tests.

Implications for research assessment

Research assessment should stop treating raw publication totals as proxies for merit, especially in residency selection, academic promotion and administrative leadership. Decisions should prioritize the quality, originality and verified contribution of selected work, not replace one numerical indicator with another. Table 4 distinguishes established assessment principles, proposed implementation measures and their evidence status, together with possible unintended effects. Established principles, such as those of DORA, the Leiden Manifesto, the Hong Kong Principles and CoARA [46–49], call for qualitative assessment of contribution and integrity. Implementation measures include assessment of a limited number of selected works, as practiced by the German Research Foundation and in China's representative-works evaluation [50, 51]; contribution-weighted credit based on CRediT [31, 52–54]; verification of listed publications and roles [23–25]; and verifiable human accountability for each contribution. I did not identify convincing prospective evaluations of the effectiveness of these measures in preventing inappropriate authorship, although I did not search systematically, and each can have unintended effects. Fractional counting applied without field-specific adjustment may discourage legitimate large-team research. Portfolios of selected works may favor researchers with time and institutional support to prepare them. Contribution statements are usually self-reported and do not by themselves verify that contributions occurred [30]. Disclosure of AI use is required by journals [1, 55], and concealing it is an integrity violation [32], but as AI use becomes near-universal a generic disclosure will no longer distinguish one paper or candidate from another. What remains informative is each author's attestation of the parts personally performed and verified, together with data and code that can be checked and selected works that candidates can explain and defend.

Table 4

Assessment principles and proposed measures: evidence status and possible unintended effects

Measure	What it changes	Precedent and evidence status	Possible unintended effects and safeguards
Principle: qualitative assessment of contribution and integrity	Assessment of quality, originality, integrity and clinical impact replaces counting	Established principles: DORA, Leiden Manifesto, Hong Kong Principles, CoARA [46–49]; no convincing prospective evaluation of effect on authorship identified	Greater reliance on reviewers' judgment and potential bias; requires trained panels and transparent criteria
1. Limited portfolio of selected works	Only a small number of self-selected works are assessed; additional names on additional papers bring no advantage	Implemented by the German Research Foundation (up to 10 peer-reviewed publications in the CV) [50] and in China's representative-works evaluation [51]; no convincing prospective evaluation identified	May favor researchers with time and support to prepare portfolios; provide templates and allow narrative context
2. Contribution-weighted credit	Fractional or role-weighted counting replaces full counting	CRediT taxonomy (ANSI/NISO Z39.104) [31, 52, 53]; proposals for graded credit [54]; fractional counting standard in some bibliometric contexts [28]	May discourage legitimate large-team research; equal fractions misrepresent unequal contributions; use field-specific adjustment and never as a sole metric
3. Verified contributions and publications	Candidates state their roles for each listed paper; listed publications are checked against PubMed or DOI records	Misrepresentation documented in 5–20% of applicants [23, 24]; contribution statements are usually self-reported [30]	Administrative burden; self-report bias; confirmation by corresponding authors and random audits
4. Verifiable human accountability	Each author attests to the parts personally performed and verified; data and code available; candidates explain and defend selected works	ICMJE 2026 recommendations [1]; EC living guidelines [55]; ALLEA code [32]; widespread AI-assisted writing [44]; proposed; no convincing prospective evaluation identified	Interviews favor confident speakers; attestation can become a formality; use structured questions and verify against data and code

CRediT, Contributor Roles Taxonomy; DORA, San Francisco Declaration on Research Assessment; CoARA, Coalition for Advancing Research Assessment; ICMJE, International Committee of Medical Journal Editors; EC, European Commission; ALLEA, All European Academies. The principal recommendation is to assess the quality, originality and verified contribution of researchers, not to replace one numerical metric with another.

In medicine, authorship is an ethical claim of responsibility, not just a career asset. Accepting authorship of work one did not do, or demanding it from a junior colleague, misrepresents responsibility for the work. Teaching the meaning of authorship should be part of residency education, and declining an undeserved authorship should be recognized as an act of professional integrity.

Limitations

The main strengths relevant to interpretation are a complete census of research articles in the selected journals, coverage of ten specialties over 26 years, two independent PubMed-wide samples, linkage of 1.66 million authorship positions with validation of the most prolific name-based identities, and consistency across journal-level, clustering, autocorrelation, publication-year and other sensitivity analyses.

This study has several limitations. First, PubMed records do not reveal who contributed what; I measured the number of authors, not the prevalence of honorary authorship, and part of the growth reflects genuine collaboration. Second, journals were selected purposively rather than by a formal ranking; the 32 journals publish more multicenter trials and consortium studies than the average journal, and the findings cannot be generalized to the whole medical literature, although every journal showed the same direction of change and the PubMed-wide samples showed growth of the same kind. Third, the abstract-based criterion for research articles may misclassify some short research reports and some non-research items, and journal formats changed over time; the sensitivity analysis without this criterion gave consistent results. Some journals entered the sample only after a launch or change of title, but results were unchanged when the analysis was restricted to journals indexed throughout. Fourth, the post-2022 period is short, publication lags behind research, and team size in leading journals is not the outcome most likely to be affected by AI; the interrupted time-series analysis cannot attribute changes to AI. Fifth, estimated authorship positions are byline positions, not curriculum-vitae entries or researchers. The random-day sample drew from the first 5,000 records on the busiest days, and the numerical-PMID sample covers records created before March 2026; agreement between the two designs is reassuring but does not establish that either is unbiased. Sixth, name-level analyses were based on surnames and initials, which merge different people with the same name and split people across name variants; 35 of the 100 most prolific names were probable merges, top-100 overlaps depended on the rule, and the analyses describe the distribution of credit among names rather than the productivity of identifiable individuals. Seventh, full counts, equal fractional counts and lead positions are all accounting conventions, not measures of work, and aggregate ratios between them are determined by team size. Eighth, about 10% of articles had electronic and print dates in different years; results were unchanged with the alternative year definition. Ninth, PubMed indexing of 2025 may be slightly incomplete; results were unchanged when 2025 was excluded. Finally, the model in Fig. 7 is hypothetical and illustrates a mechanism; it is not an estimate.

Conclusions

Between 2000 and 2025, the number of authors per research article increased substantially across all ten medical specialties examined. At the level of individual author names, this expansion made publication-based comparisons strongly dependent on the counting method: names with the same equal fractional credit listed increasingly different numbers of papers, the most prolific names held a declining share of lead positions, and top rankings by full, fractional and lead-position counts overlapped only partly. None of these counts measures actual contribution or identifies honorary authorship, yet using raw publication totals as evidence of individual merit in competitive decisions risks inequitable evaluation.

No post-2022 discontinuity in authorship counts was established, and the study cannot identify effects of generative AI; a possible AI-driven acceleration of publication output makes prospective monitoring and contribution-sensitive assessment increasingly urgent.

Academic and clinical institutions should evaluate selected works, documented individual contributions and accountability instead of relying primarily on publication volume, including when selecting candidates for academic and administrative positions. Assessment reform need not wait for a further increase in AI-assisted output.

Abbreviations

AI
Artificial intelligence
ALLEA
All European Academies
CI
Confidence interval
CoARA
Coalition for Advancing Research Assessment
CRediT
Contributor Roles Taxonomy
DFG
Deutsche Forschungsgemeinschaft (German Research Foundation)
DOI
Digital object identifier
DORA
San Francisco Declaration on Research Assessment
EC
European Commission
ICMJE
International Committee of Medical Journal Editors
IQR

Interquartile range
ORCID
Open Researcher and Contributor ID
PMID
PubMed identifier

Declarations

Ethics approval and consent to participate

Not applicable. The study analyzed publicly available bibliographic records and involved no human participants. No individual researcher is identified in the article, and author names in the deposited data are replaced by codes.

Consent for publication

Not applicable.

Availability of data and materials

Source bibliographic records were retrieved from PubMed. A public repository containing pseudonymized bibliometric analysis inputs, Python code, query methods, aggregated results, and figures is openly available in Zenodo (https://doi.org/10.5281/zenodo.23435572) [56]. The public package omits the original author-name lists, detailed identity-validation records, the name-to-code linkage, and publication titles. These verification materials are retained separately and are available from the corresponding author for justified scholarly verification requests, subject to responsible handling and any applicable redistribution limitations. Public PMIDs permit linkage to source records; pseudonymization is not a guarantee of anonymity. The dataset represents a historical PubMed extraction and may differ from current PubMed records.

Competing interests

The author declares that he has no competing interests.

Funding

Not applicable. This study received no funding.

Authors' contributions

Michael Shalman conceived and designed the study, acquired and analyzed the data, interpreted the findings, drafted and revised the manuscript, and approved the final version. The sole author accepts full responsibility for the work.

Acknowledgements

Not applicable.

References

1. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. Updated January 2026. Available from: https://www.icmje.org/recommendations/ (accessed 14 October 2026).
2. World Conference on Research Integrity. Singapore Statement on Research Integrity. 2010. Published in: Resnik DB, Shamoo AE. The Singapore Statement on Research Integrity. Account Res. 2011;18(2):71 – 5. 10.1080/08989621.2011.557296. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC3954607/
3. Rice DB, Raffoul H, Ioannidis JPA, Moher D. Academic criteria for promotion and tenure in biomedical sciences faculties: cross sectional analysis of international sample of universities. BMJ. 2020;369:m2081. 10.1136/bmj.m2081.
4. Weeks WB, Wallace AE, Kimberly BC. Changes in authorship patterns in prestigious US medical journals. Soc Sci Med. 2004;59(9):1949–54. 10.1016/j.socscimed.2004.02.029.
5. Tilak G, Prasad V, Jena AB. Authorship inflation in medical publications. Inquiry. 2015;52:0046958015598311. 10.1177/0046958015598311.
6. Papatheodorou SI, Trikalinos TA, Ioannidis JP. Inflated numbers of authors over time have not been just due to increasing research complexity. J Clin Epidemiol. 2008;61(6):546–51. 10.1016/j.jclinepi.2007.07.017.
7. Chau HD, Gan ZJ, Bin Abd Razak HR, Allen JC, Koh SJ, Howe TS. A review of authorship inflation and multicenter collaboration trends in orthopedic, medical, and surgical journals over the last 60 years. Cureus. 2024;16(8):e66107. 10.7759/cureus.66107.
8. Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. Science. 2007;316(5827):1036–9. 10.1126/science.1136099.
9. Flanagin A, Carey LA, Fontanarosa PB, Phillips SG, Pace BP, Lundberg GD, et al. Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. JAMA. 1998;280(3):222–4. 10.1001/jama.280.3.222.
10. Wislar JS, Flanagin A, Fontanarosa PB, DeAngelis CD. Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. BMJ. 2011;343:d6128. 10.1136/bmj.d6128.
11. Meursinge Reynders RA, ter Riet G, Di Girolamo N, Cavagnetto D, Malički M. Honorary authorship is highly prevalent in health sciences: systematic review and meta-analysis of surveys. Sci Rep. 2024;14(1):4385. 10.1038/s41598-024-54909-w.
12. Jakab M, Kittl E, Kiesslich T. How many authors are (too) many? A retrospective, descriptive analysis of authorship in biomedical publications. Scientometrics. 2024;129(3):1299–328. 10.1007/s11192-024-04928-1.

13. Lin Z, Lu S. Exponential authorship inflation in neuroscience and psychology from the 1950s to the 2020s. Am Psychol. 2025;80(2):264–78. 10.1037/amp0001216.
14. An JY, Marchalik RJ, Sherrer RL, Baiocco JA, Rais-Bahrami S. Authorship growth in contemporary medical literature. SAGE Open Med. 2020;8:2050312120915399. 10.1177/2050312120915399.
15. De Cassai A, Dost B, Boscolo A, Navalesi P. Authorship trends in systematic reviews: a 30-year analysis in anaesthesia and pain medicine. Br J Anaesth. 2025;135(4):1119–21. 10.1016/j.bja.2025.06.035.
16. Miller C, Rosenberger J, Sablewski A, Nemeth M, Paal P, Dejaco A. Global evolution of female authorships in anesthesiology articles: an affiliation-based, longitudinal, scientometric analysis. BMC Anesthesiol. 2025;25(1):489. 10.1186/s12871-025-03407-1.
17. Massey J, Zahid Z, Shaw M, Modi P. Continuing authorship proliferation in the cardiothoracic surgical literature. Interdiscip Cardiovasc Thorac Surg. 2026;41(7):ivag171. 10.1093/icvts/ivag171.
18. Kobak D, González-Márquez R, Horvát EÁ, Lause J. Delving into LLM-assisted writing in biomedical publications through excess vocabulary. Sci Adv. 2025;11(27):eadt3813. 10.1126/sciadv.adt3813.
19. Suchak T, Aliu AE, Harrison C, Zwiggelaar R, Geifman N, Spick M. Explosion of formulaic research articles, including inappropriate study designs and false discoveries, based on the NHANES US national health database. PLoS Biol. 2025;23(5):e3003152. 10.1371/journal.pbio.3003152.
20. Hanson MA, Barreiro PG, Crosetto P, Brockington D. The strain on scientific publishing. Quant Sci Stud. 2024;5(4):823–43. 10.1162/qss_a_00327.
21. Bhardwaj A, Bajaj P, Rifai K, Patel A, Bhatt A, Abolhassani A, et al. Association of the change in USMLE Step 1 scoring with research output among applicants to anesthesiology residency programs: a retrospective observational study. J Educ Perioper Med. 2026;28(1):E760. 10.46374/VolXXVIII_Issue1_Bhardwaj.
22. Elemosho A, Sarac BA, Janis JE. The law of diminishing returns in the integrated plastic surgery residency match: a deeper look at the numbers. Plast Reconstr Surg Glob Open. 2024;12(7):e5937. 10.1097/GOX.0000000000005937.
23. Wiggins MN. A meta-analysis of studies of publication misrepresentation by applicants to residency and fellowship programs. Acad Med. 2010;85(9):1470–4. 10.1097/ACM.0b013e3181e2cf2b.
24. Burkhart RJ, Hecht CJ, Karimi AH, Acuña AJ, Kamath AF. What are the trends in research publication misrepresentation among orthopaedic residency and fellowship applicants from 1996 to 2019? A systematic review. Clin Orthop Relat Res. 2023;481(7):1292–303. 10.1097/CORR.0000000000002549.
25. Bozocea A, Michael BR, Zhao AY, Garcia D, Siegel N, Agarwal AR, et al. Verified research productivity among matched orthopaedic surgery residency applicants: establishing a national baseline and comprehensive analysis. Cureus. 2025;17(12):e98930. 10.7759/cureus.98930.
26. National Library of Medicine. PubMed and Entrez Programming Utilities (E-utilities). Bethesda (MD): National Center for Biotechnology Information. Available from: https://www.ncbi.nlm.nih.gov/books/NBK25501/ (accessed 14 October 2026).

27. Montazeri A, Mohammadi S, Hesari M, Ghaemi P, Riazi M, Sheikhi-Mobarakeh H. Preliminary guideline for reporting bibliometric reviews of the biomedical literature (BIBLIO): a minimum requirements. Syst Rev. 2023;12:239. 10.1186/s13643-023-02410-2.
28. Waltman L, van Eck NJ. Field-normalized citation impact indicators and the choice of an appropriate counting method. J Informetr. 2015;9(4):872–94. 10.1016/j.joi.2015.08.001.
29. Fanelli D, Larivière V. Researchers' individual publication rate has not increased in a century. PLoS ONE. 2016;11(3):e0149504. 10.1371/journal.pone.0149504.
30. Sauermann H, Haeussler C. Authorship and contribution disclosures. Sci Adv. 2017;3(11):e1700404. 10.1126/sciadv.1700404.
31. Larivière V, Pontille D, Sugimoto CR. Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). Quant Sci Stud. 2021;2(1):111–28. 10.1162/qss_a_00097.
32. ALLEA – All European Academies. The European Code of Conduct for Research Integrity. Revised edition 2023. Berlin: ALLEA. 2023. doi:10.26356/ECOC. Available from: https://allea.org/code-of-conduct/
33. Eisenberg RL, Ngo L, Boiselle PM, Bankier AA. Honorary authorship in radiologic research articles: assessment of frequency and associated factors. Radiology. 2011;259(2):479–86. 10.1148/radiol.11101500.
34. Fong EA, Wilhite AW. Authorship and citation manipulation in academic research. PLoS ONE. 2017;12(12):e0187394. 10.1371/journal.pone.0187394.
35. Pagel PS, Hudetz JA. An analysis of scholarly productivity in United States academic anaesthesiologists by citation bibliometrics. Anaesthesia. 2011;66(10):873–8. 10.1111/j.1365-2044.2011.06860.x.
36. Jaksic C, Gayet-Ageron A, Perneger T. In health research publications, the number of authors is strongly associated with collective self-citations but less so with citations by others. BMC Med Res Methodol. 2023;23:230. 10.1186/s12874-023-02037-w. Correction in: BMC Med Res Methodol. 2023;23:260. doi:10.1186/s12874-023-02084-3.
37. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A. 2005;102(46):16569–72. 10.1073/pnas.0507655102.
38. Merton RK. The Matthew effect in science. Science. 1968;159(3810):56–63. 10.1126/science.159.3810.56.
39. Zaorsky NG, O'Brien E, Mardini J, Lehrer EJ, Holliday E, Weisman CS. Publication productivity and academic rank in medicine: a systematic review and meta-analysis. Acad Med. 2020;95(8):1274–82. 10.1097/ACM.0000000000003185.
40. Jacobson CE, Beeler WH, Griffith KA, Flotte TR, Byington CL, Jagsi R. Common pathways to Dean of Medicine at U.S. medical schools. PLoS ONE. 2021;16(3):e0249078. 10.1371/journal.pone.0249078.
41. Aliukonis V, Gefenas E. Authority and authorship: examining leadership's role in medical publications. J Empir Res Hum Res Ethics. 2026;21(1–2):49–59. 10.1177/15562646251392340.

42. 2026 Physician Survey on Augmented Intelligence. Chicago: AMA; American Medical Association, Center for Digital Health and AI, March. 2026. Available from: https://www.ama-assn.org/sites/ama-assn.org/files/2026-03/physician-ai-sentiment-report_0.pdf (accessed 17 October 2026).
43. Wiley. ExplanAItions. 2025: the evolution of AI in research. Hoboken (NJ): John Wiley & Sons; 2025. Available from: https://www.wiley.com/en-us/about-us/ai/study/key-findings (accessed 14 October 2026).
44. Holzwarth L, González-Márquez R, Kobak D. Most biomedical publications show signs of LLM-assisted writing. arXiv [Preprint]. 2026 Aug 11. arXiv:2608.10715. Available from: https://arxiv.org/abs/2608.10715
45. Matsui A. Return of the solo author: the changing division of labor in science in the age of generative AI. arXiv [Preprint]. 2026 Jul 12. arXiv:2607.10780. Available from: https://arxiv.org/abs/2607.10780
46. San Francisco Declaration on Research Assessment (DORA). 2012. Available from: https://sfdora.org/read/ (accessed 14 October 2026).
47. Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I. Bibliometrics: the Leiden Manifesto for research metrics. Nature. 2015;520(7548):429–31. 10.1038/520429a.
48. Moher D, Bouter L, Kleinert S, Glasziou P, Sham MH, Barbour V, et al. The Hong Kong Principles for assessing researchers: fostering research integrity. PLoS Biol. 2020;18(7):e3000737. 10.1371/journal.pbio.3000737.
49. Coalition for Advancing Research Assessment (CoARA). Agreement on Reforming Research Assessment. 20 July 2022. Available from: https://www.coara.org/agreement/the-agreement-full-text/ (accessed 14 October 2026).
50. Deutsche Forschungsgemeinschaft. Package of measures to support a shift in the culture of research assessment. Information for Researchers No. 61, 2022; and FAQ on CVs. Available from: https://www.dfg.de/en/research-funding/proposal-funding-process/faq/cv (accessed 14 October 2026).
51. Ministry of Education and Ministry of Science and Technology of the People's Republic of China. Opinions on standardising the use of SCI paper indicators in higher education institutions and establishing a correct evaluation orientation. February 2020. Available from: http://en.moe.gov.cn/news/media_highlights/202002/t20200224_423535.html (accessed 14 October 2026).
52. National Information Standards Organization. ANSI/NISO Z39.104-2022, CRediT, Contributor Roles Taxonomy. Baltimore (MD): NISO. 2022. Available from: https://credit.niso.org/ (accessed 14 October 2026).
53. Brand A, Allen L, Altman M, Hlava M, Scott J. Beyond authorship: attribution, contribution, collaboration, and credit. Learn Publ. 2015;28(2):151–5. 10.1087/20150211.
54. Lin Z. Modernizing authorship criteria and transparency practices to facilitate open and equitable team science. Acc Res. 2025;32(8):1564–87. 10.1080/08989621.2024.2405041.

55. European Commission, Directorate-General for Research and Innovation. ERA living guidelines on the responsible use of generative AI in research. Brussels: European Commission. 2024 (updated 16 October 2026). Available from: https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/era-living-guidelines-responsible-use-generative-ai-research-updated-2026-10-16_en (accessed 17 October 2026).
56. Shalman M. Data and code for: When One Paper Counts Many Times: A Bibliometric Study of Authorship Expansion, Publication-Credit Inflation, and Research Integrity Across Ten Medical Specialties, 2000–2025 [dataset]. Version 1.1. Zenodo; 2026. https://doi.org/10.5281/zenodo.23435572.

Figures

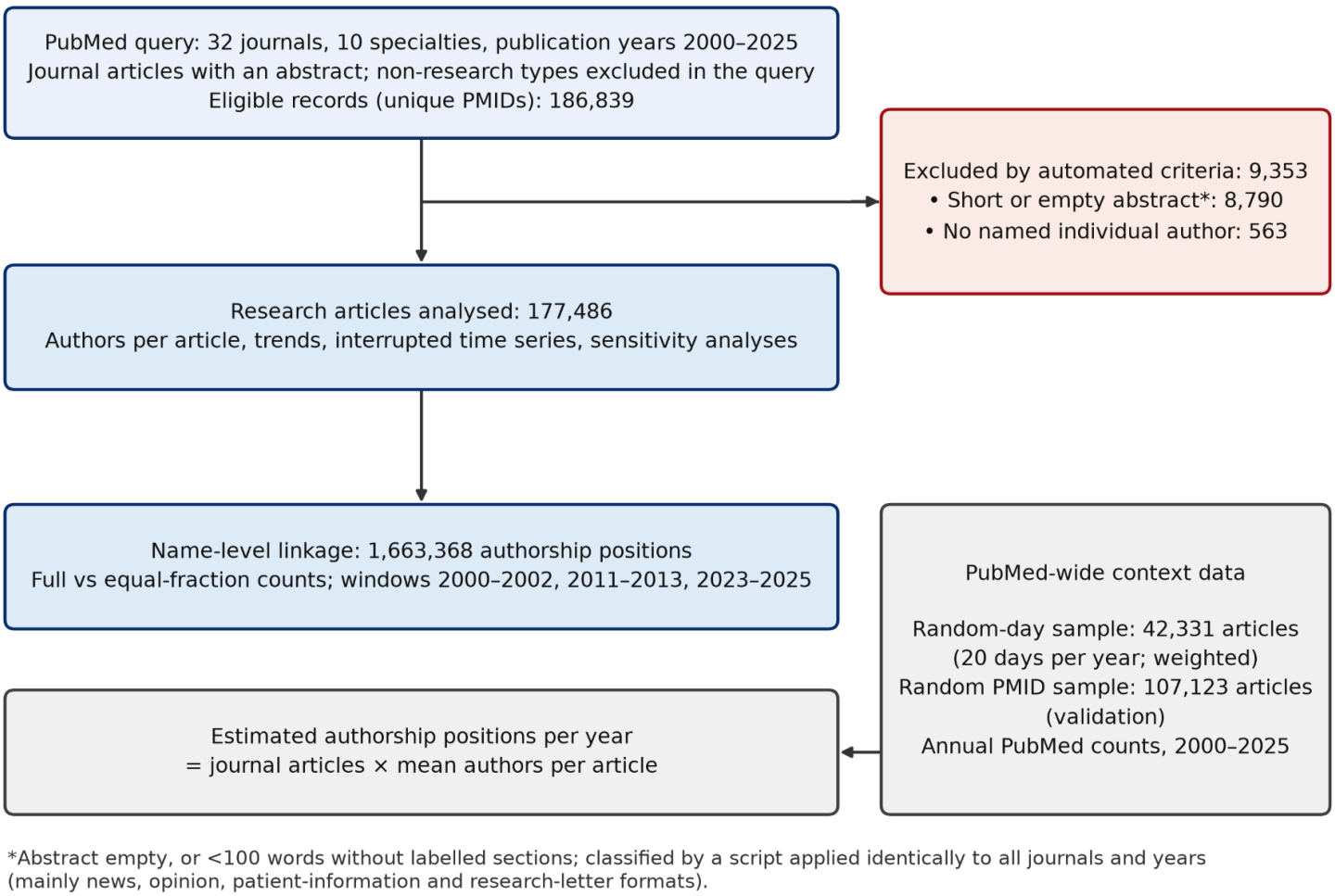


Figure 1

Study flow diagram. Selection of research articles and data sources. Exclusions after retrieval were made by automated criteria applied identically to all journals and years. PubMed-wide context data were used for the estimates of authorship positions.

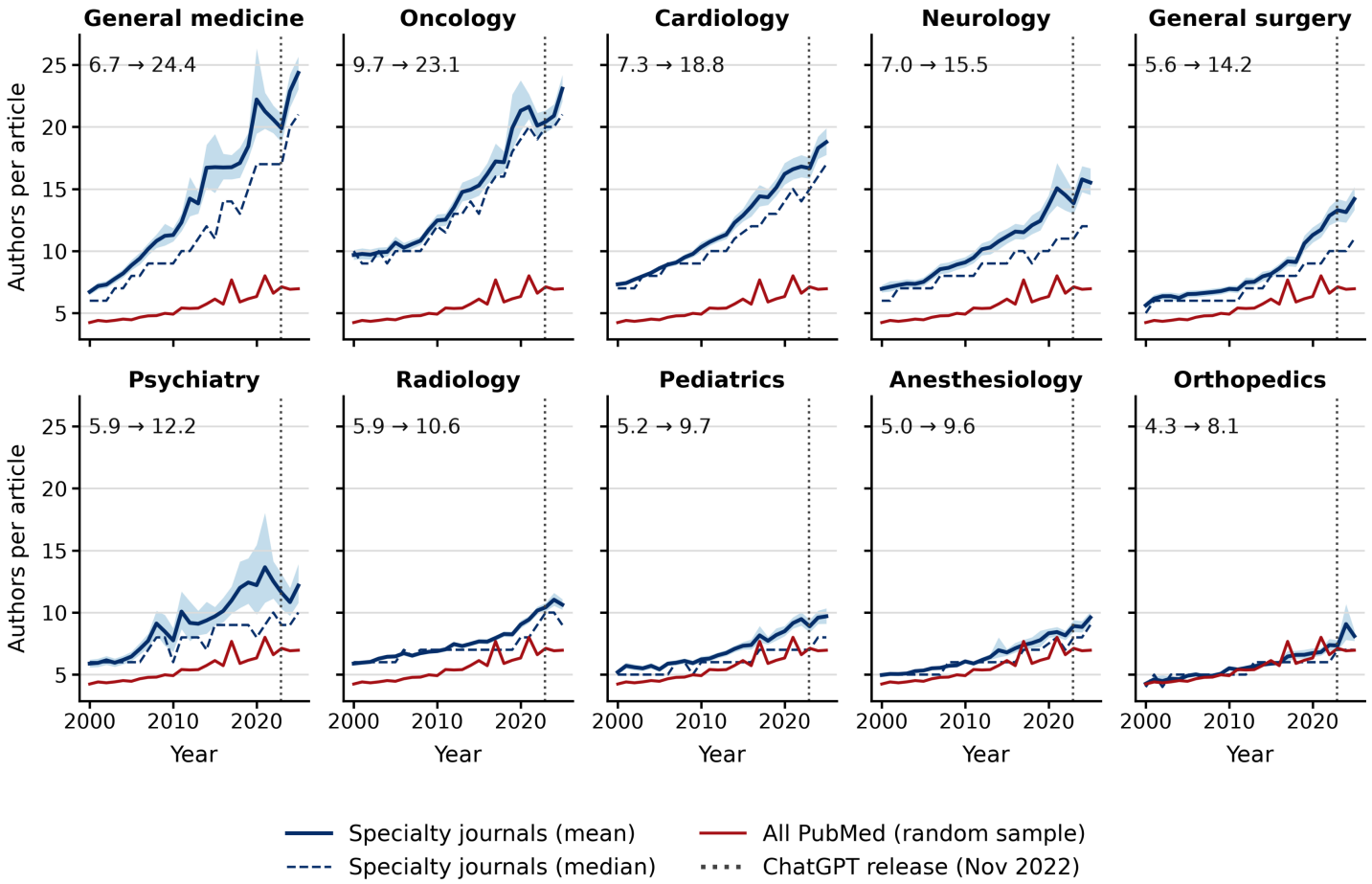


Figure 2

Authors per research article by specialty, 2000–2025. Mean (solid line, with 95% bootstrap confidence band, 2,000 resamples) and median (dashed line) number of authors per research article in the selected journals of each specialty, compared with the weighted random-day sample of all PubMed journal articles (red line). Numbers give the mean in 2000 and 2025. The dotted vertical line marks the public release of ChatGPT (November 2022). Journals that were launched or renamed during the study period enter the series in the year given in Table 1; results were unchanged when the analysis was restricted to journals indexed throughout.

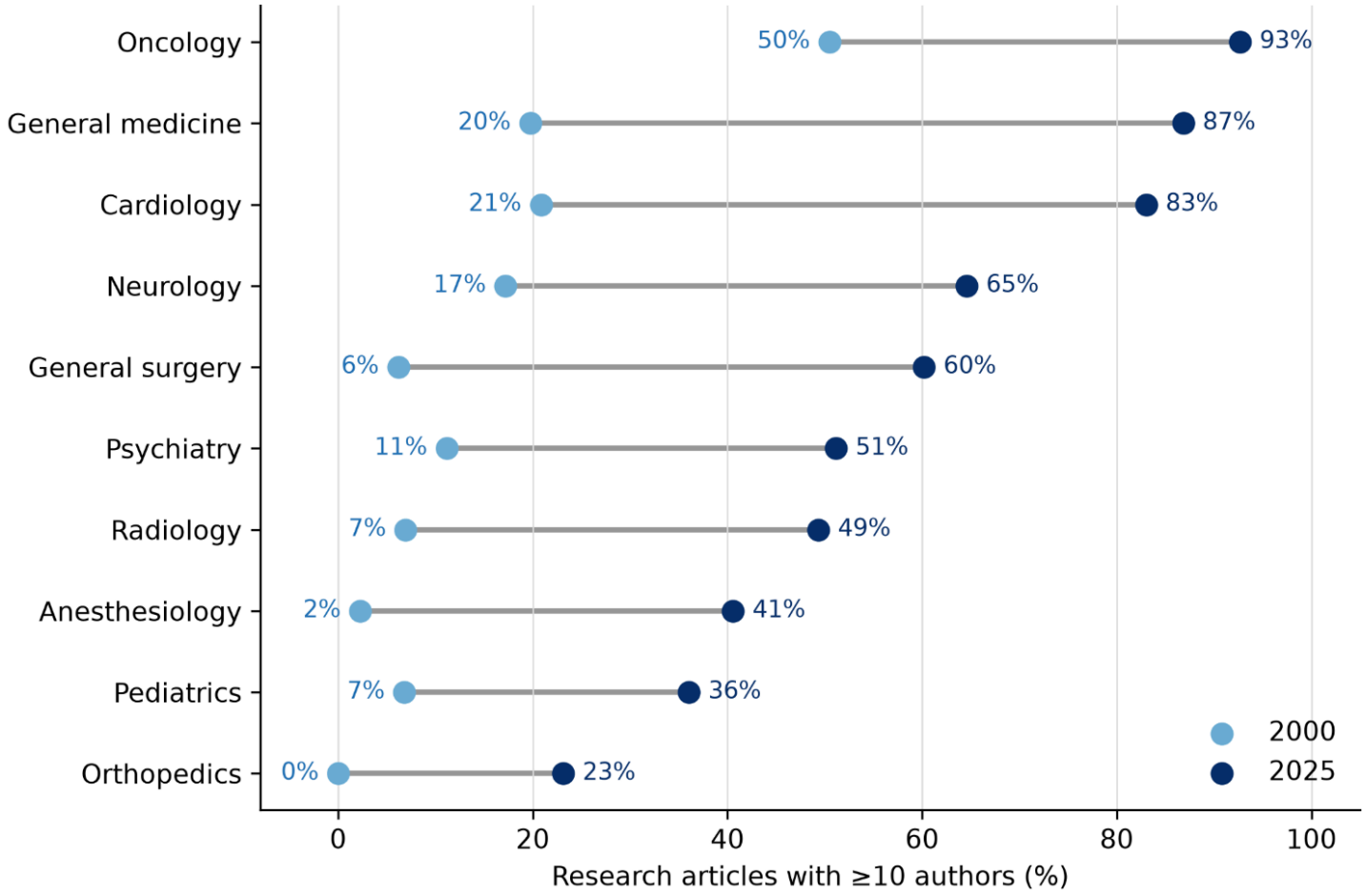


Figure 3

Research articles with ten or more authors, 2000 and 2025. Proportion of research articles with ten or more authors, by specialty. The threshold is descriptive and does not indicate inappropriate authorship.

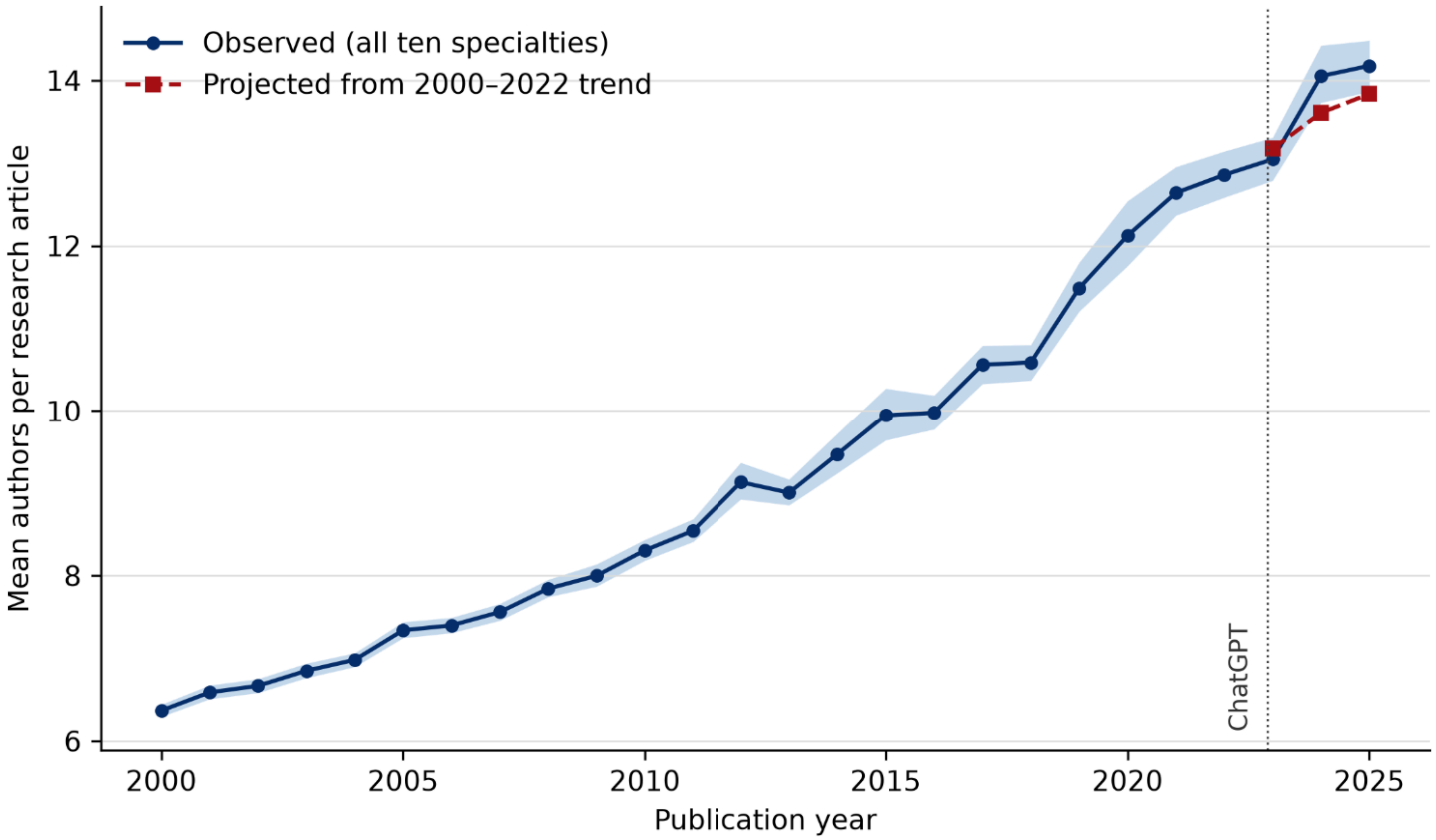


Figure 4

Observed and projected authors per research article. Mean authors per research article across the ten specialties (observed, with 95% confidence band) and values for 2023–2025 projected from the 2000–2022 trend of a negative binomial model adjusted for specialty. No clear discontinuity was observed; a small nominal slope increase in the robust-error model was not significant with journal-clustered standard errors or autoregressive errors, and the design cannot attribute changes to generative AI.

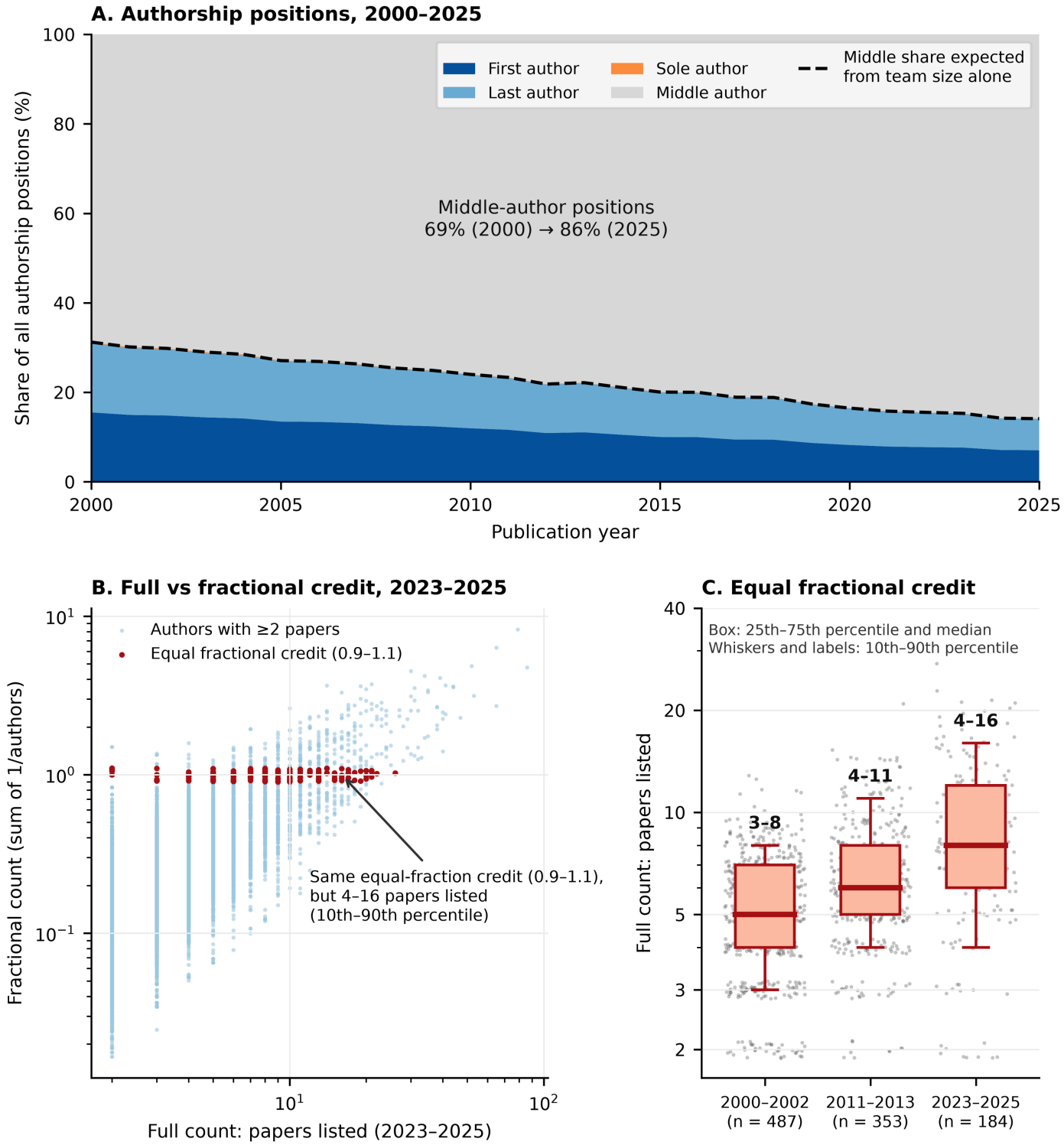


Figure 5

Authorship positions and full versus fractional counts. (A) Distribution of authorship positions in research articles, 2000–2025; the dashed line shows the middle-author share expected from the empirical distribution of team sizes, which coincides with the observed share. (B) Full publication count versus equal-fraction count for author names with at least two initials and at least two papers in 2023–2025 (log scales). Red points identify names with fractional credit of 0.9–1.1 (10th–90th percentile of

full counts, 4–16 papers). (C) Distribution of full publication counts for author names with at least two initials, at least two papers, and equal-fraction credit of 0.9–1.1 across three 3-year windows; boxes show median and interquartile range, and whiskers the 10th–90th percentiles. Name-level comparisons do not identify unique researchers or measure actual scientific contribution.

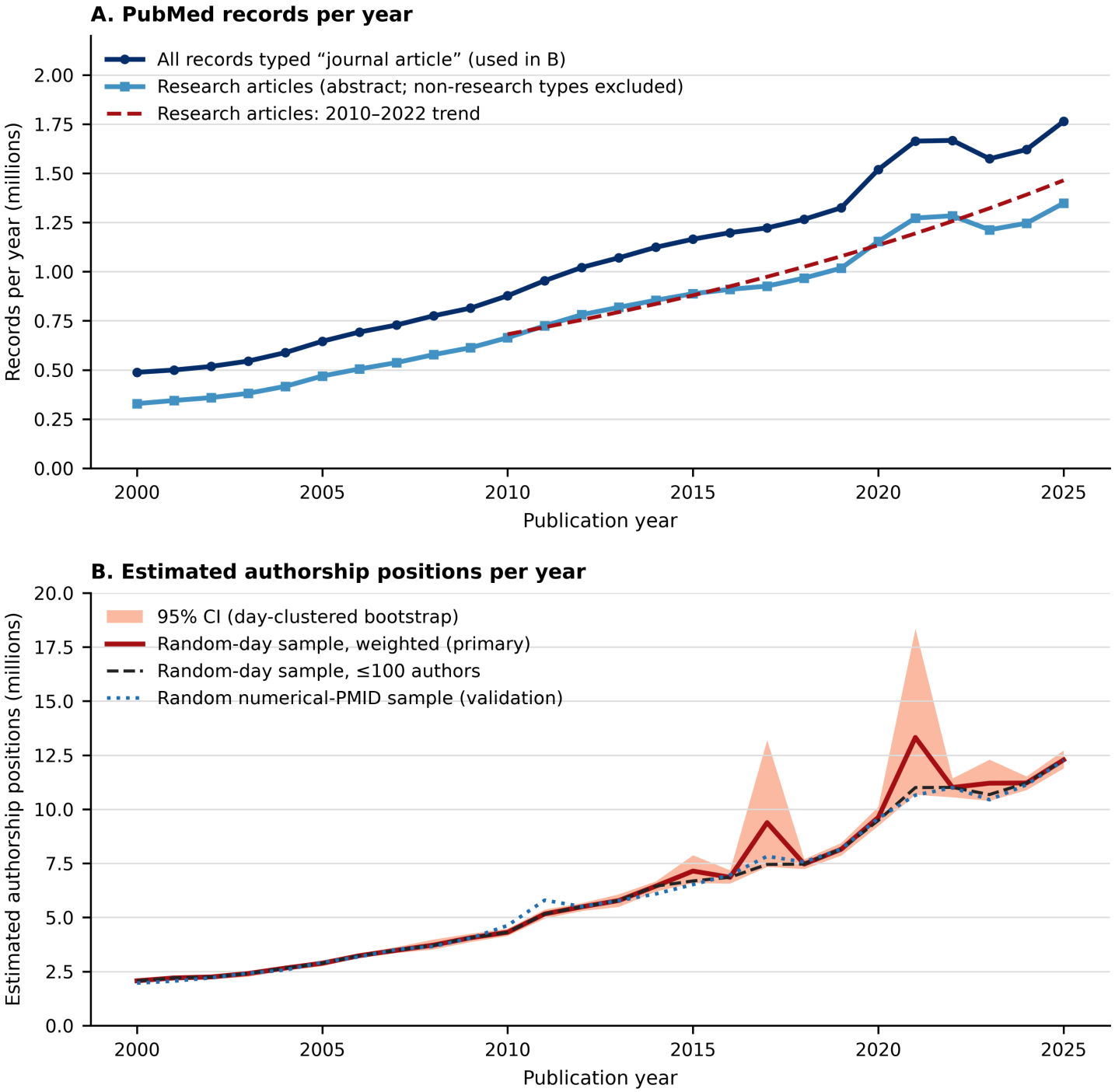


Figure 6

PubMed output and estimated authorship positions, 2000–2025. (A) PubMed records per year with the publication type “journal article” (used for panel B) and the narrower category of research articles

(journal articles with an abstract, excluding reviews, editorials, letters, case reports and other non-research types), with the log-linear trend for research articles fitted to 2010–2022. (B) Estimated authorship positions per year (journal articles multiplied by mean authors): weighted random-day sample with 95% confidence band from a day-clustered bootstrap (primary), the same sample after exclusion of articles with more than 100 authors, and the independent random numerical-PMID validation sample. Peaks in 2017 and 2021 reflect single sampled articles with several hundred or more authors.

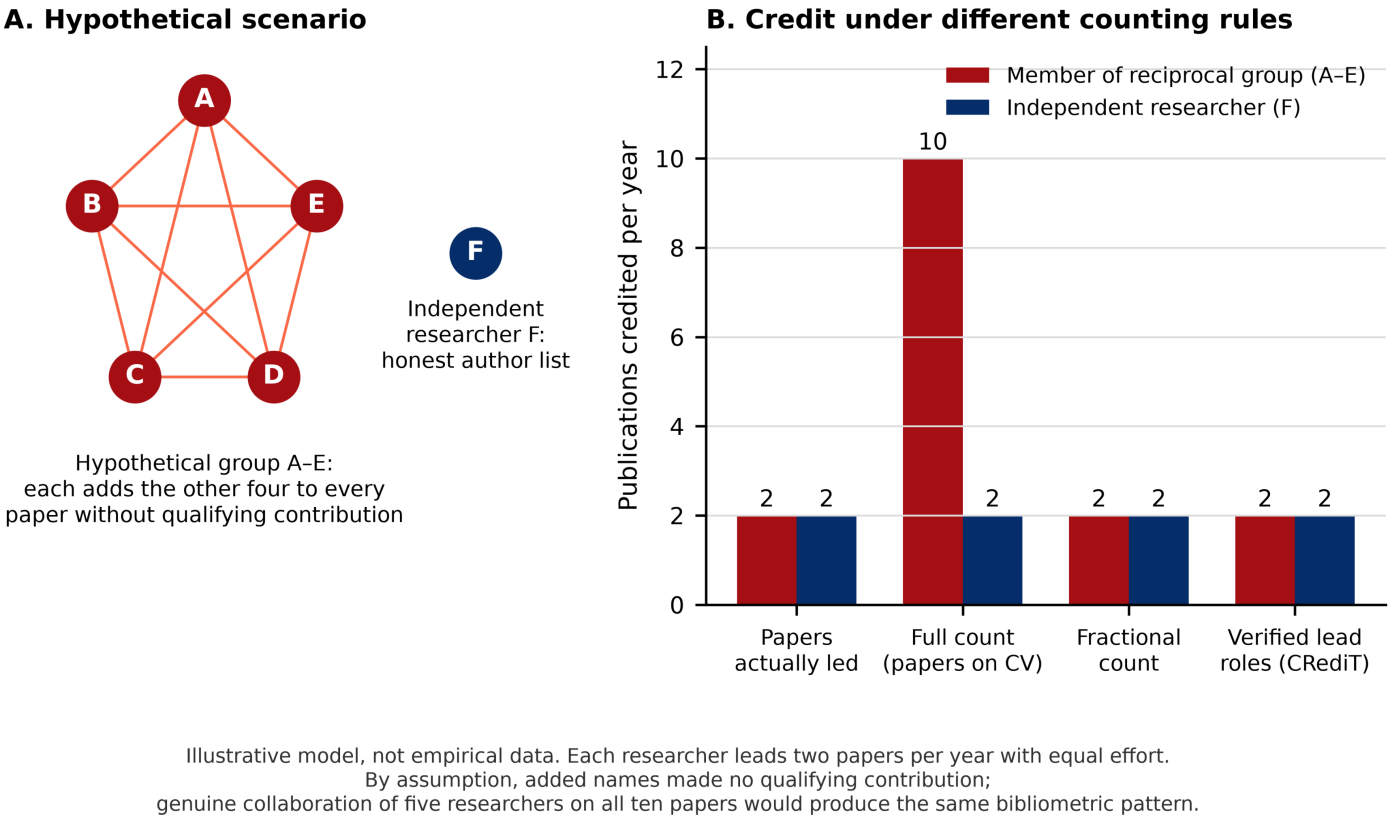


Figure 7

Hypothetical illustration of reciprocal honorary authorship. Hypothetical model, not empirical data. Five researchers (A–E) each lead two papers per year and, by assumption, add the other four to every paper without any qualifying contribution; each therefore appears on ten papers per year, whereas an equally productive independent researcher (F) appears on two. Under full counting the members of the group appear five times more productive; under fractional counting or verified lead roles the difference disappears. Five researchers who genuinely collaborated on all ten papers would produce the same bibliometric pattern; the scenario represents misconduct only because of the stated assumption.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1BIBLIOchecklist.docx

- Additionalfile2TableS1.docx