

Supplementary Information

A Scalable Framework for Longitudinal Surveillance of Research Evolution in Digital Medicine

Yilun Liang, Gongbo Zhang, Casey Ta, Yilu Fang, Chunhua Weng*

Department of Biomedical Informatics, Columbia University, New York, NY, USA

Supplementary Information

Supplementary Note S1: Granularity and provenance sensitivity

The K parameter in K-means clustering sets the target number of clusters and therefore the granularity of the topic taxonomy. To check that the field-level findings are not artifacts of one granularity, we re-ran the keyword clustering at finer settings: K=750, and the per-domain values of the original JBI study (K=1,050 for methodology and K=1,100 for health). Note S1.1 reports which headline results hold across these runs. Note S1.2 asks whether the finer runs reveal declining sub-topics hidden inside rising K=100 umbrella topics, and shows that the apparent declines are keyword-provenance artifacts.

S1.1 Robustness across granularities

The 7.6-fold growth is a corpus-level count and does not depend on the taxonomy. The wave onsets, dated from signature-topic breakthroughs, hold at the keyword level as well: the keywords machine learning, deep learning, and large language models break through in 2019, 2020, and 2024. The widespread COVID-19 inflections also hold in the finer-grained runs. The journal-network findings do not carry over unchanged. Rebuilding the journal-similarity network from finer-grained topic vectors changes the community partition (four communities at K=100, three at K=750, four at K=1,050/1,100, all with modularity below 0.15). JBI's betweenness lead holds at K=750 (rank 1) but not at the finest granularity (rank 4). We therefore read the partition as weak at every resolution and the JBI bridge as a K=100 and K=750 result. We did not re-derive every downstream analysis at finer K: topic identity, the enrichment blocks, and the entropy rankings are reported at the K=100 operating resolution.

S1.2 Do finer granularities reveal hidden declines?

The absolute-count trend table at K=100 reports no topic with a Holm-significant negative slope. We asked whether finer clusterings expose declining sub-topics hidden inside rising umbrellas. Every keyword carries both a fine-grained and a K=100 cluster assignment, and a sub-topic is “inside” the K=100 topic that holds the plurality of its keywords. A descriptive screen (peak at least three years ago, mean of the last three years below 70% of the peak, negative slope, at least 100 articles) flags 42 methodology and 36 health fine-grained clusters at K=1,050/1,100, all inside Holm-significantly rising K=100 umbrellas. Only one of them (a cluster of MeSH “... as topic” headings) reaches Holm significance under the larger fine-K test family.

These candidates do not survive a provenance check. Their early peaks sit in MeSH-substituted records, which dominate 2011–2012: MeSH indexers assign headings such as “Support Vector Machine”, “Hospital Information Systems”, or “Information Storage and Retrieval” consistently, whereas authors list the

corresponding terms less often, so the counts fall as author keywords replace MeSH descriptors from 2013 onward. Repeating the screen on author-keyword records (2013–2025) leaves 3 methodology and 2 health clusters, and none of the four cases we had previously highlighted survives. The K=750 human–computer interaction topic (3,861 articles) is flat only because a rising author-keyword series (103 to 241 articles per year over 2013–2025) is offset by a falling MeSH series (263 to 34). The support vector machine cluster (544 articles) rises from 10 to 22–35 articles per year on author keywords while its MeSH series falls from 58 to 3. The hospital information systems cluster (631 articles) rises from 8 to 21 on author keywords while its MeSH series falls from 101 to 3. And the interoperability cluster (1,021 articles), whose most frequent keyword is the MeSH heading “information storage and retrieval”, shows a V-shape that is entirely the MeSH series, while keywords containing “interoperability” rise steadily from 1 to 60 articles per year (Supplementary Figure 5); its apparent early peak and recovery are an artifact of the MeSH series, not a change in research attention.

An earlier version of this note grouped these cases into mechanisms (algorithm displacement, paradigm replacement, and an infrastructure assumption followed by policy-driven recovery). We withdraw those labels because the declines they were meant to explain are not real: in every case the “decline” is the fading of MeSH-substituted records while the author-keyword series is flat or rising, so there is no displacement or replacement to explain. We therefore report no robust fine-grained declines. The finer runs instead show that decline claims in a multi-source keyword corpus are sensitive to keyword provenance, which is why the main text treats share losses as provenance-sensitive and reports the author-keyword sensitivity analysis alongside the full-corpus results (Supplementary Table 10). K=100 remains the operating resolution for the field-scale narrative.

Supplementary Note S2: Extended limitations

Keyword provenance. MeSH substitution supplied the keywords for 75% of 2011 records and 83% of 2012 records, 44% in 2013, and 10% by 2025. Two journals switched to MeSH substitution from 2018 (Applied Clinical Informatics, Methods of Information in Medicine), and three are MeSH-based throughout (Bioinformatics, Database, The Lancet Digital Health). MeSH check tags had entered the topic taxonomy: the tag “Humans” alone sat inside the clinical study designs and mixed medical imaging topics on 10,609 records. Excluding the check tags from topic assignment reduced those two topics from 14,528 to 5,046 and from 12,385 to 1,869 articles and removed their spurious pairing from the leading co-occurrence pairs. Share losses concentrated among MeSH-flavored topics largely do not replicate when the share model is refitted on author-keyword records (Supplementary Table 10), whereas the gains do.

Coverage gradient and network structure. For Bioinformatics and Database, over 99% of records fall back to MeSH descriptors, and these are also the venues that look most topically isolated in the similarity network. Part of their apparent isolation may therefore reflect metadata provenance rather than content. Cross-referencing with Web of Science or Scopus was beyond the scope of this study.

Thresholds. The minimum article counts differ by analysis (80 articles for universality, 100 for breakthrough dating and the event test) because each statistic needs a different amount of data to be stable: dating a breakthrough from yearly counts needs a denser series than computing an entropy. These thresholds were tuned to this corpus, and other fields or smaller corpora may need different values. The breakthrough threshold is a fraction of each topic’s own peak, so topics with sustained growth must clear an ever-higher bar. The Poisson rate test assumes Poisson counts against a fixed pre-event rate and accounts for neither overdispersion

nor field-wide growth, so an elevated event-window rate partly reflects the corpus expanding over 2020–2021, and the two-year baseline limits the power of any yearly comparison. The pioneer-journal metric records where a topic first surfaced within the corpus; an early paper may be rarely cited, a topic may have appeared earlier in venues outside the 29 journals, and the citation-weighted variant ranks journals differently. Affiliation classification was performed on free-text affiliations, so journals with sparse affiliation metadata (notably IEEE J-BHI and most pre-2014 records) skew toward the neither/unknown category.

Supplementary Note S3: Named datasets and shared resources

The resources that articles name changed with the waves (Supplementary Table 6). Among the 29 resources mentioned by at least 20 articles, mentions of public clinical datasets rose fastest: MIMIC and eICU from 3.8 to 12.8 articles per 1,000 between 2011–2015 and 2021–2025, UK Biobank from none to 3.8, All of Us from none to 2.1, and OMOP/OHDSI from 0.4 to 2.7. The imaging benchmarks ImageNet, BraTS, and ISIC went from none to 1.4–2.0 per 1,000. Mentions of the i2b2/n2c2 shared-task corpora fell from 9.5 to 1.8 per 1,000 as clinical natural language processing moved to other data, and mentions of the bioinformatics knowledge bases (KEGG, Gene Ontology, UniProt) stayed flat or declined. PubMed itself is the most-mentioned resource (3,540 articles), followed by ClinicalTrials.gov (1,715). Overall, 13% of articles (10,217) name at least one of the 33 resources.

Supplementary Note S4: Topic volumes across three five-year windows

To compare topic volumes over the study period, we split the corpus into three equal five-year windows and ranked the top 20 topics in each domain by cumulative volume (Supplementary Figure 3). The windows are a fixed descriptive partition; they are not the waves of the main text, which are dated from topic breakthroughs. In 2011–2015, machine learning methods (3,672 articles) and clinical decision support (2,908) led the methodology ranking, with genomics and sequence bioinformatics prominent alongside them (1,721 articles). In 2016–2020, machine learning methods doubled to 7,352 articles, and software and usability research grew fastest among the leaders (+131%, to 4,940). In 2021–2025, machine learning methods surged again (18,674), and digital health technologies, the largest health topic in every window, reached 17,349. The signature topics of the third wave, large language models and generative AI, are too new to enter these top-20 rankings: roughly 90% of their articles appeared after 2020, and large language models ranks 24th by 2021–2025 volume. The union of the per-window top-10 lists lies inside the cumulative top 20, so a per-window ranking would not add topics.

Supplementary Note S5: Per-journal topic trends

For every journal with at least five indexed years (28 of 29; PLOS Digital Health has four) and every topic with at least 30 of that journal's articles, we fitted the quasi-Poisson share model of the main text with the journal's own yearly totals as the offset and applied the Holm correction within the journal. Supplementary Table 9 summarizes the largest rises and falls per journal, and Supplementary Figure 4 shows the share trends of the 20 largest topics per domain across journals. Of 1,419 journal-topic pairs, 269 have a Holm-significant trend (115 rising, 154 falling), and 27 journals show at least one. Machine learning methods gains share in 11 journals and large language models in 7 (BMC Medical Informatics and Decision Making, Computers in

Biology and Medicine, Frontiers in Digital Health, JMIR Medical Informatics, JBI, Journal of Medical Systems, and JAMIA). The most frequent losers are virtual medical simulation (9 journals), clinical decision support (8), information dissemination (8), and patient-provider communication (7). Within-journal trends share the provenance caveat of the main text: for journals whose keyword source changed over the period (Applied Clinical Informatics and Methods of Information in Medicine switched to MeSH substitution from 2018), a trend can reflect the vocabulary change rather than the journal's content.

Supplementary tables

Supplementary Tables 2–14 are also provided as CSV files.

Supplementary Table 1: The 29 PubMed-indexed biomedical-informatics journals in the corpus: sub-domain, founding year, and article count (78,073 articles published 2011–2025; a further 352 records indexed for 2026 were retrieved but excluded from analysis).

#	Journal	Sub-domain	Founded	n articles in our corpus
1	Journal of Medical Internet Research (JMIR)	digital/consumer health	1999	9,728
2	Bioinformatics	comp. biology	1985	9,564
3	Computers in Biology and Medicine	bridging methods	1970	8,908
4	IEEE Journal of Biomedical and Health Informatics (J-BHI)	bridging methods	2013	5,071
5	Briefings in Bioinformatics	comp. biology	2000	4,579
6	BMC Medical Informatics and Decision Making	clinical informatics	2001	3,654
7	Nature Medicine	clinical informatics	1995	3,460
8	Journal of the American Medical Informatics Association (JAMIA)	clinical informatics	1994	3,213
9	JMIR mHealth and uHealth	digital/consumer health	2013	2,872
10	Journal of Medical Systems	clinical informatics	1977	2,755
11	Journal of Biomedical Informatics (JBI)	clinical informatics	1967	2,734
12	International Journal of Medical Informatics (IJMI)	clinical informatics	1970	2,722
13	Nature Methods	comp. biology	2004	2,503
14	npj Digital Medicine	digital/consumer health	2018	2,147
15	JMIR Medical Informatics (JMI)	digital/consumer health	2013	1,684
16	Database (J. Biological Databases & Curation)	comp. biology	2009	1,623
17	Artificial Intelligence in Medicine	bridging methods	1989	1,539
18	Frontiers in Digital Health	digital/consumer health	2019	1,458
19	Applied Clinical Informatics	clinical informatics	2009	1,353
20	Journal of Clinical and Translational Science	clinical informatics	2017	1,194
21	Health Informatics Journal	clinical informatics	1995	1,063
22	JAMIA Open	clinical informatics	2018	836

23	Bioinformatics Advances	comp. biology	2021	830
24	PLOS Digital Health	digital/consumer health	2022	824
25	Methods of Information in Medicine	clinical informatics	1962	652
26	The Lancet Digital Health	digital/consumer health	2019	444
27	BMJ Health & Care Informatics	clinical informatics	2019	399
28	Digital Biomarkers	digital/consumer health	2017	145
29	Journal of Innovation in Health Informatics	clinical informatics	2015	119

Supplementary Table 2: The 70 methodology and 86 health K=100 topics, with the number of distinct keywords, the number of 2011–2025 articles assigned to each after excluding MeSH check tags, and the five most frequent keywords.

Supplementary Table 3: Keyword provenance and topic density by journal (2011–2025, after excluding MeSH check tags from MeSH-substituted records). Topic density scales with keyword density ($r = 0.98$ across journals); by source, author and publisher keywords give 4.5 topics per article, MeSH substitutes 5.0, and GPT-generated keywords 7.5.

Supplementary Table 4: COVID-19 event test for the 133 eligible topics: average annual articles in the 2018–2019 baseline, the 2020–2021 event window, and the 2022–2025 post-event window; percentage change and Holm-adjusted Poisson p for each window; and the year of the largest year-over-year increase (2012–2025). Federated learning is ineligible (pre-event mean below 3 articles per year).

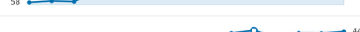
Supplementary Table 5: Milestone events used in Figure 3: panel, year, event, and the public source that dates it. The list was compiled by the study team before the curves were examined.

Supplementary Table 6: Named datasets and shared resources mentioned in titles or abstracts (case-insensitive pattern matching; resources with at least 20 mentioning articles), with mention rates per 1,000 articles in each five-year window (Supplementary Note S3).

Supplementary Table 7: Top-10 methodology–health co-occurring topic pairs with publication-share sparklines (2011–2025).

Top 10 co-occurring methodology × health topic pairs (K=100)

Sparklines auto-scale per row so the within-pair trend is readable; Δ 2011–25 = percent change between 2011–13 mean and 2023–25 mean.

Methodology topic	Health topic	2011 → 2025 trend (per-row scale)	Δ 2011–25	Total
Clinical decision support	Digital health technologies		▲ +155%	10,128 
Machine learning methods	Digital health technologies		▲ +487%	9,932 
Mobile health and social media	Digital health technologies		▲ +578%	6,135 
Machine learning methods	Computational cancer biology		▲ +605%	4,823 
Statistical methods	Digital health technologies		▲ +246%	4,532 
Genomics and sequence bioinformatics	Genomic variant analysis		▲ +404%	4,406 
Information dissemination	Digital health technologies		▲ +205%	3,974 
Software and usability research	Digital health technologies		▲ +211%	3,868 
Software and usability research	Genomic variant analysis		▲ +1026%	3,566 
Machine learning methods	Genomic variant analysis		▲ +596%	3,425 

Co-occurrence: an article is counted once per pair if its author keywords map to both topics in that pair. Open circle = peak year for that pair. Per-row sparkline scaling reveals the trend shape; the magnitude bar (right) and Total column preserve cross-row magnitude comparison.

Supplementary Table 8: Per-journal trajectory descriptors (years indexed, first and last year, total and 2025 volume, peak year, growth factor, and year-to-year volatility), from which labels such as long-lived, high-growth, volatile, stable, or discontinued can be derived; growth factors use 2025 volume relative to the first (or, for journals with a partial first year, the second) indexed year.

Supplementary Table 9: Per-journal topic share trends (Supplementary Note S5): for each journal, the number of topics tested (at least 30 articles in the journal), the number of Holm-significant trends, and the three largest rises and falls in the topic's share of the journal's output (% per year, with the number of articles). The full journal-by-topic table is provided as a CSV file in the project repository.

Supplementary Table 10: Provenance sensitivity of the share-trend model. For every K=100 topic: the annual change in share of corpus output (quasi-Poisson model with log-corpus offset, Holm-corrected within domain) on the full corpus (2011–2025) and on records with author, publisher, or GPT keywords only (2013–2025), both after excluding MeSH check tags.

Supplementary Table 11: Citation impact by K=100 topic (NIH iCite, accessed September 2026): median and mean relative citation ratio (RCR) and the share of articles with $RCR \geq 2$ among articles published 2011–2023 (topics with at least 30 such articles), total citations and citations per article, and the share trend from Figure 10.

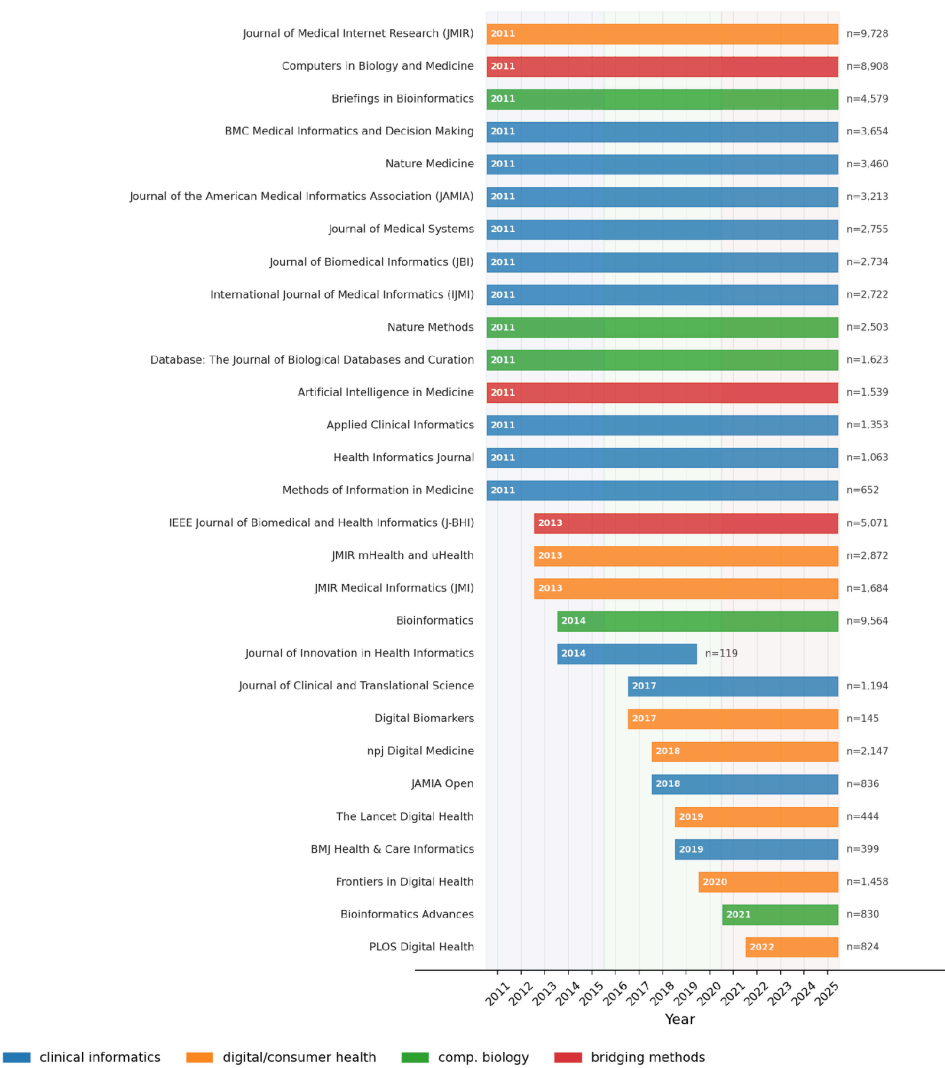
Supplementary Table 12: Influential articles. (A) The 25 articles with the highest relative citation ratio; (B) the three highest-RCR articles of each of the 20 largest methodology and health topics; (C) the 25 articles cited by the largest number of distinct K=100 topics within the corpus (within-corpus citations only).

Supplementary Table 13: Pioneer journals under two definitions for every topic: the journal of the first article in the corpus and the journal whose articles published in the topic's first three corpus years accumulated the most citations (iCite).

Supplementary Table 14: Regular-expression lists used to classify affiliations as clinical or computational (case-insensitive, matched against the concatenated affiliation strings of each article; a match on both lists = interdisciplinary; no match on either, or no affiliation metadata = neither/unknown).

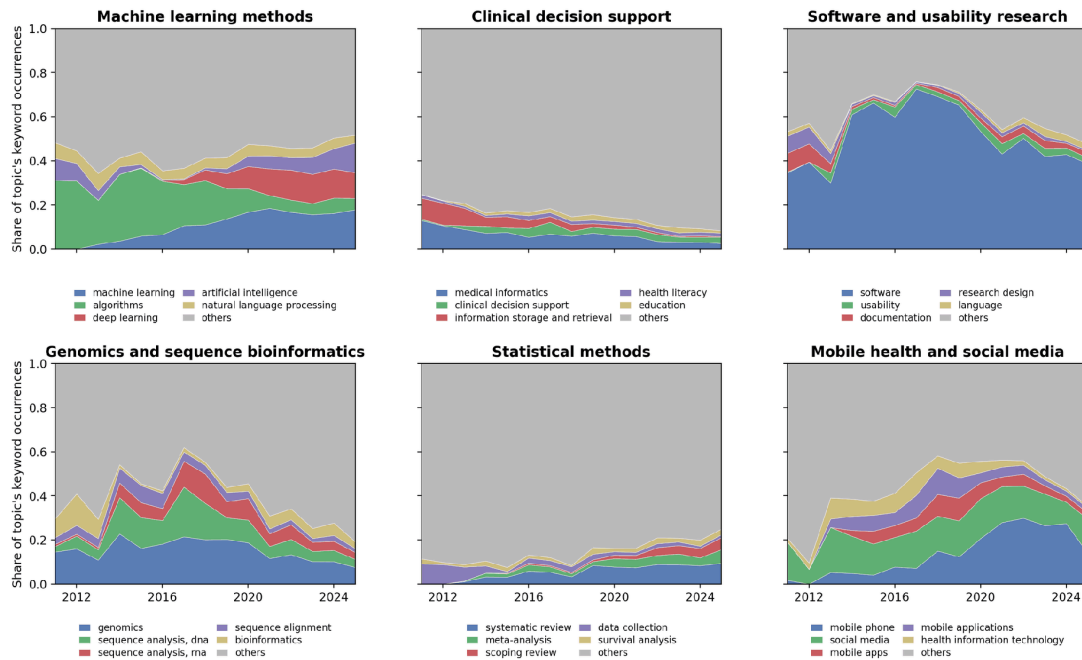
Supplementary figures

Bars show each journal's first → last indexed year in our corpus; right-side n = total articles. Colour = sub-domain.

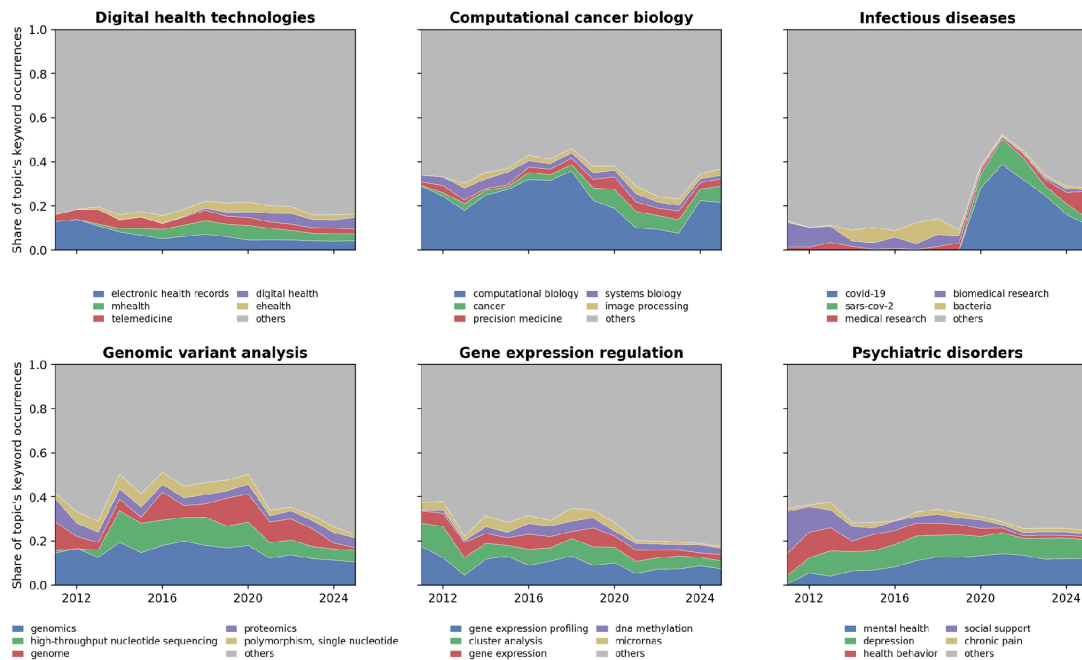


Supplementary Figure 1: Per-journal coverage timeline showing the indexed publication span of each of the 29 journals; the label at the left of each bar is the first indexed year and n is the number of articles.

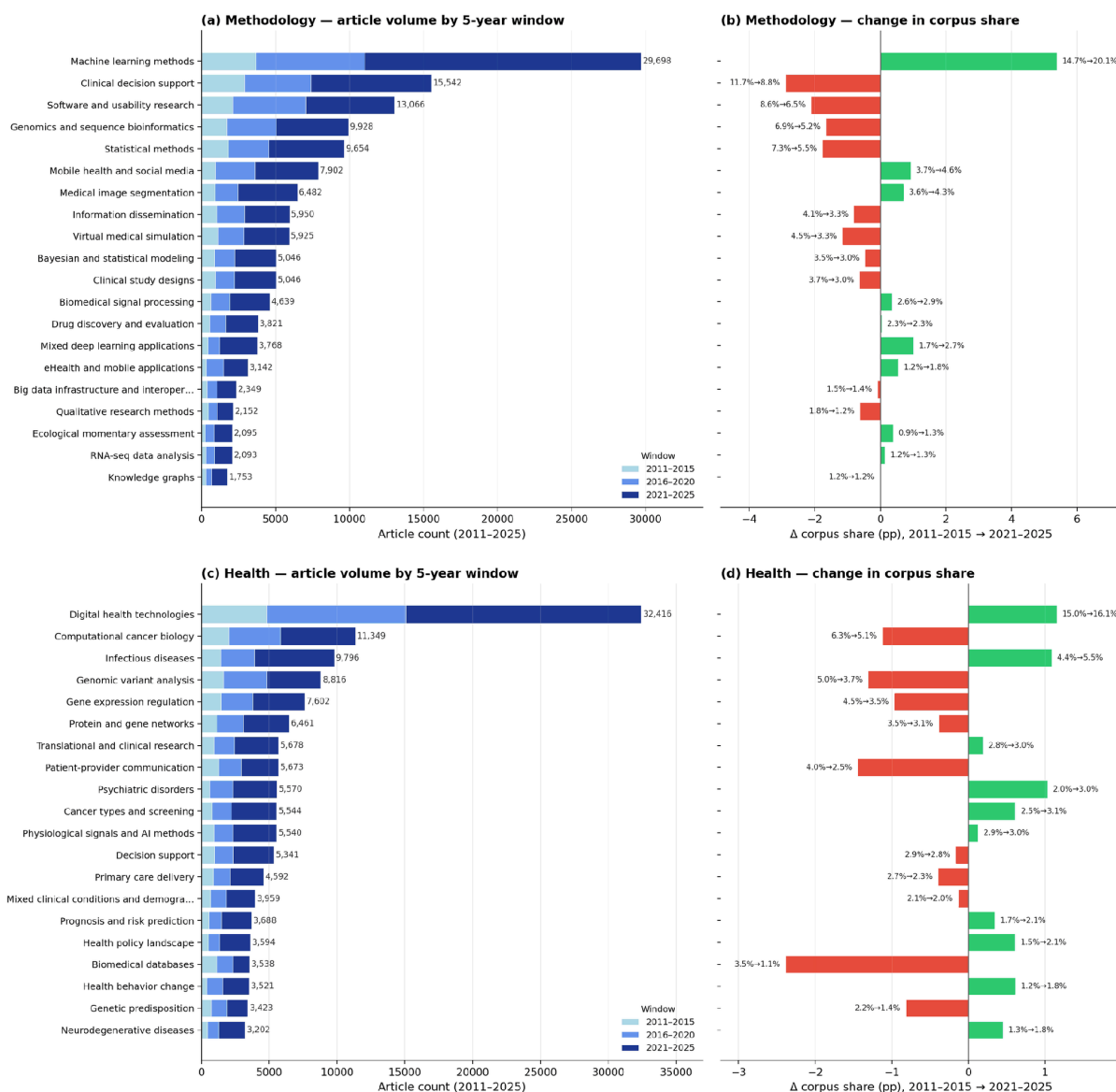
(a) Methodology



(b) Health

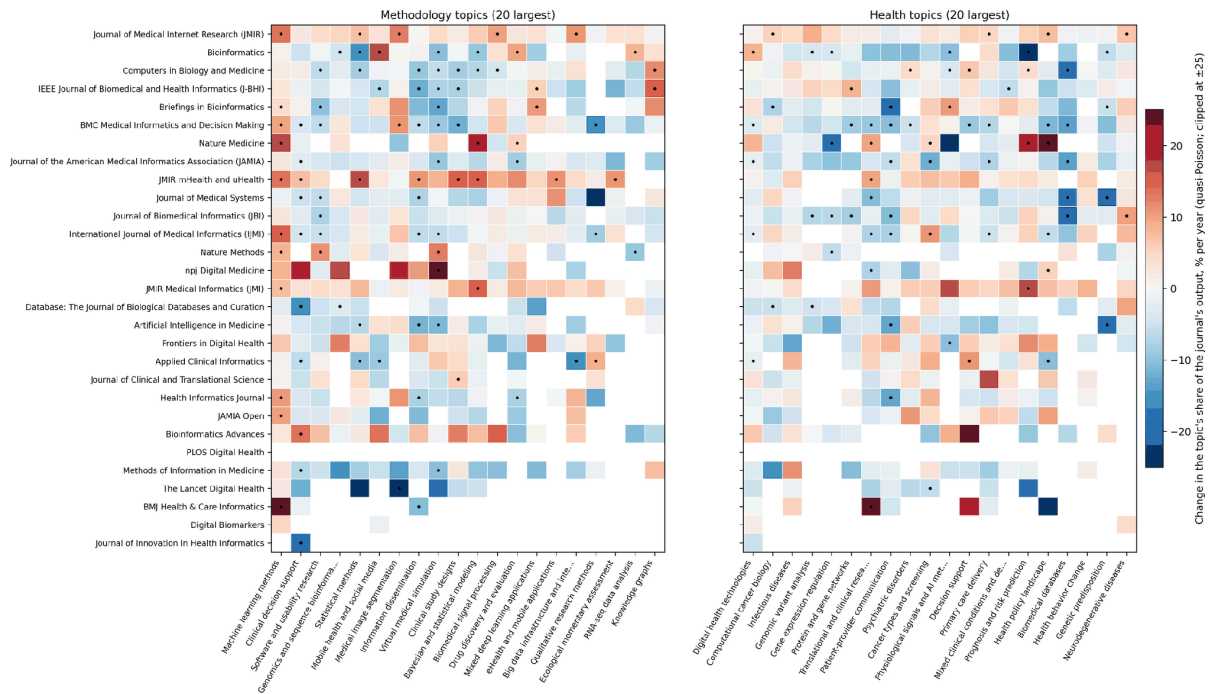


Supplementary Figure 2: Within-topic keyword composition over time. For each domain's six largest topics, the share of the topic's keyword occurrences held by its five most frequent keywords ("others" pools the rest): (a) methodology; (b) health. Within machine learning methods, for example, deep learning rises from zero to overtake algorithms, the largest early-window keyword.

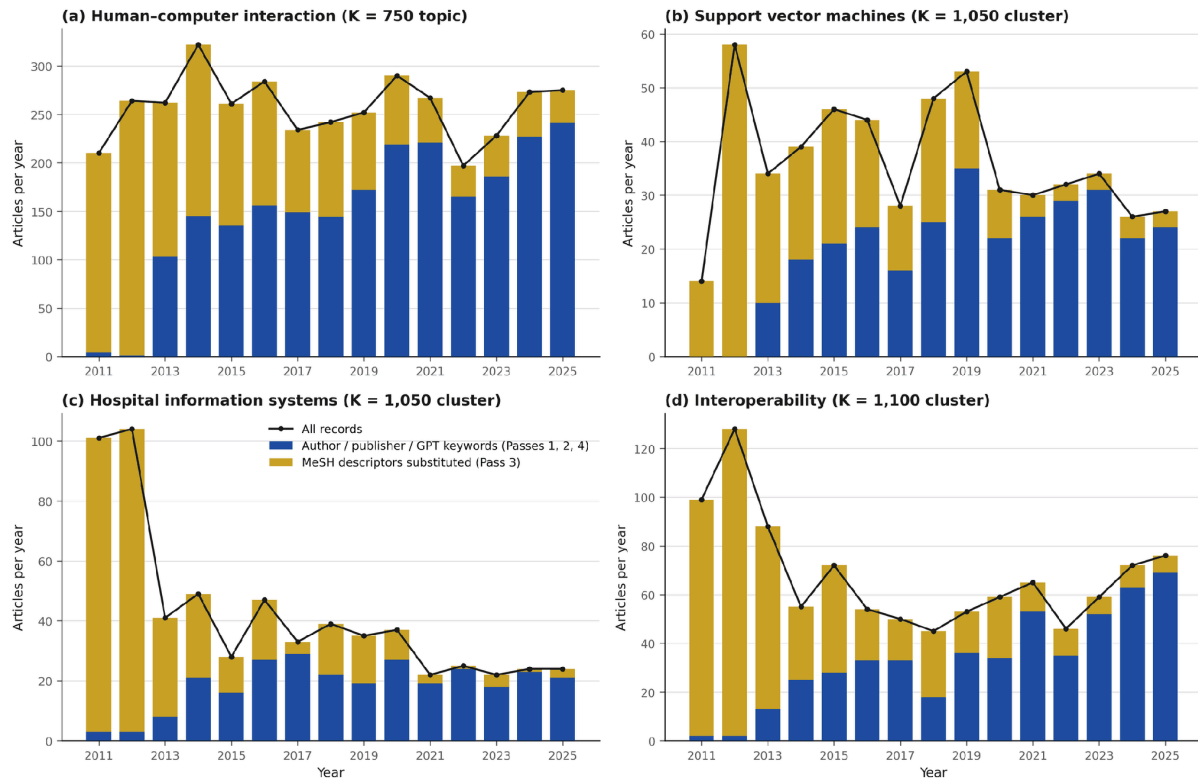


Supplementary Figure 3: Methodology and health topics across three five-year windows (top 20 by 2011–2025 volume): (a) methodology article volume per window, (b) methodology change in corpus share from 2011–2015 to 2021–2025, (c) health article volume, (d) health share change. Bar labels give 2011–2025 totals (Supplementary Note S4).

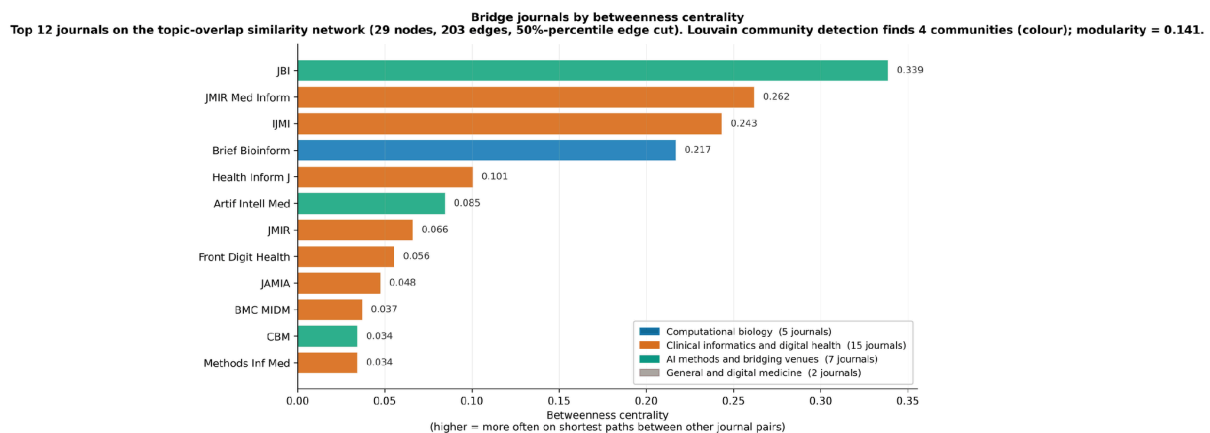
Per-journal topic share trends, 2011–2025 (dot = Holm-significant within the journal; blank = fewer than 30 articles)



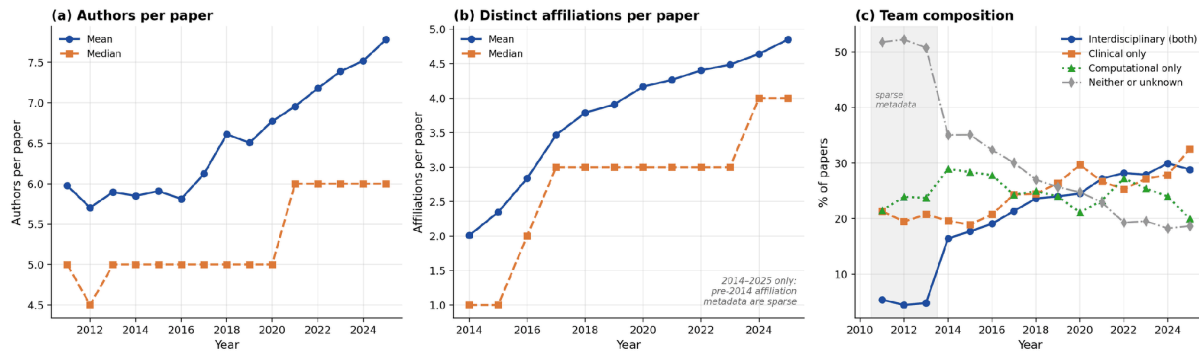
Supplementary Figure 4: Per-journal topic share trends, 2011–2025. Cells give the annual change in a topic's share of the journal's output from a quasi-Poisson model with the journal's yearly total as the offset (clipped at $\pm 25\%$ per year); dots mark Holm-significant trends within the journal; blank cells have fewer than 30 articles. Columns are the 20 largest methodology and health topics; rows are journals ordered by volume (Supplementary Note S5; Supplementary Table 9).



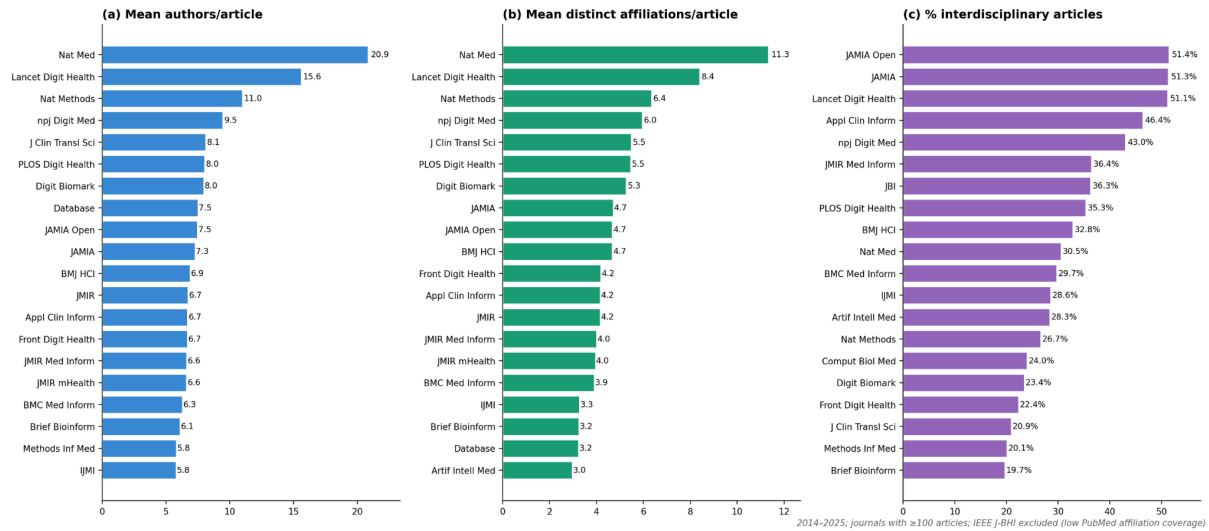
Supplementary Figure 5: Keyword-provenance decomposition of the four fine-grained decline candidates discussed in Supplementary Note S1.2. Each panel stacks the yearly article count of the cluster or topic into records whose keywords come from authors, publishers, or GPT (passes 1, 2, and 4) and records whose keywords are MeSH descriptors substituted in pass 3: (a) human–computer interaction (K=750 topic); (b) support vector machines (K=1,050 cluster); (c) hospital information systems (K=1,050 cluster); (d) interoperability (K=1,100 cluster). The early peaks and apparent declines sit entirely in the MeSH-substituted stack.



Supplementary Figure 6: Journal bridge centrality: betweenness-centrality ranking of the 12 most central journals in the journal-similarity network, colored by Louvain community (four communities, modularity = 0.141).



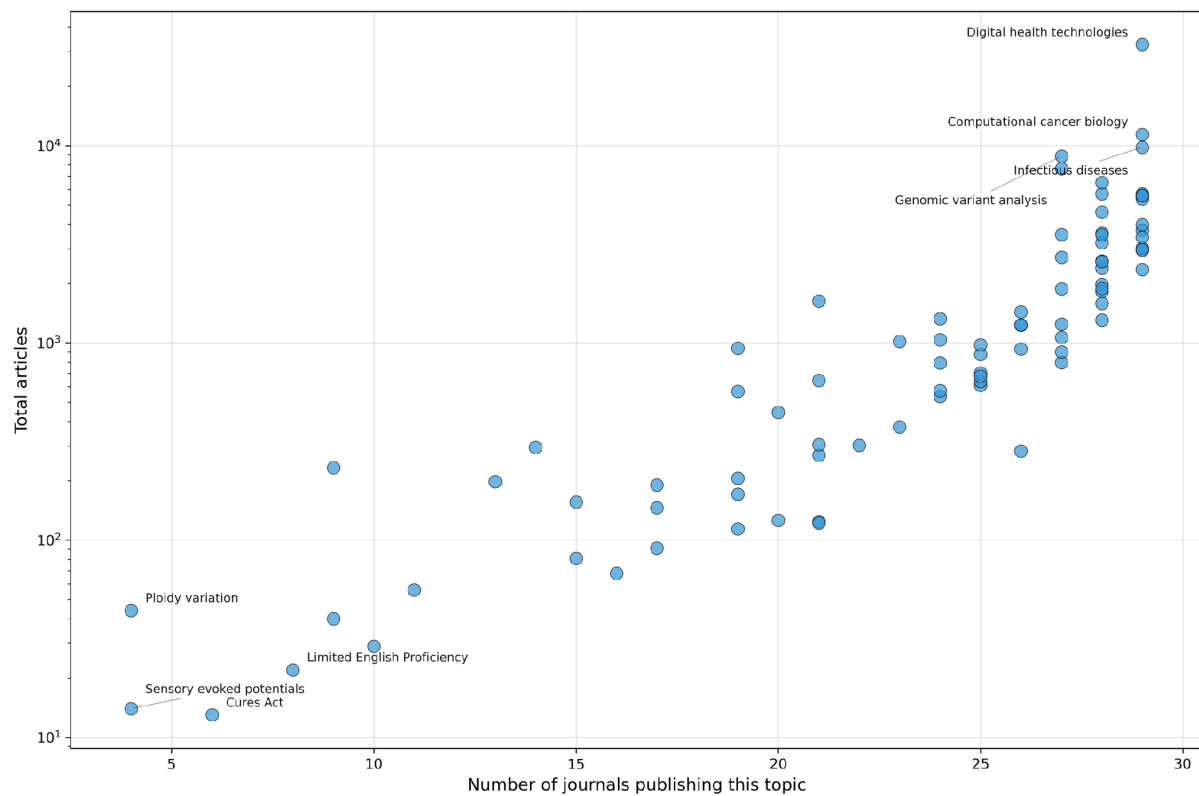
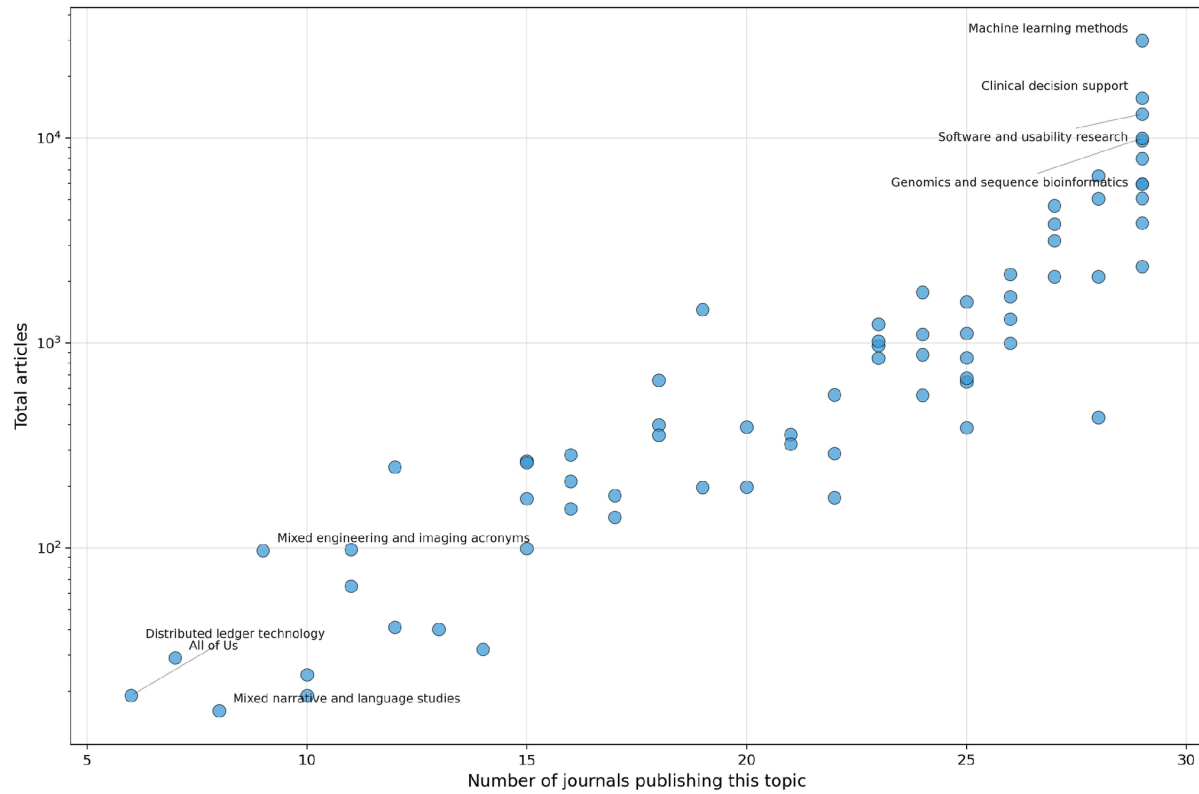
Supplementary Figure 7: Authorship dynamics, 2011–2025: (a) authors per paper (mean and median); (b) distinct affiliations per paper (2014–2025; earlier affiliation metadata are sparse); (c) team composition, the share of papers with clinical-only, computational-only, both (interdisciplinary), or neither/unknown affiliations (the pre-2014 band is shaded to indicate unreliable affiliation metadata).



Supplementary Figure 8: Per-journal team-science indicators, computed on 2014–2025 (pre-2014 affiliation metadata are sparse): (a) mean authors per article; (b) mean distinct affiliations per article; (c) percentage of articles with both a clinical and a computational affiliation. Journals with at least 100 articles in the window; IEEE J-BHI is excluded for low PubMed affiliation coverage.



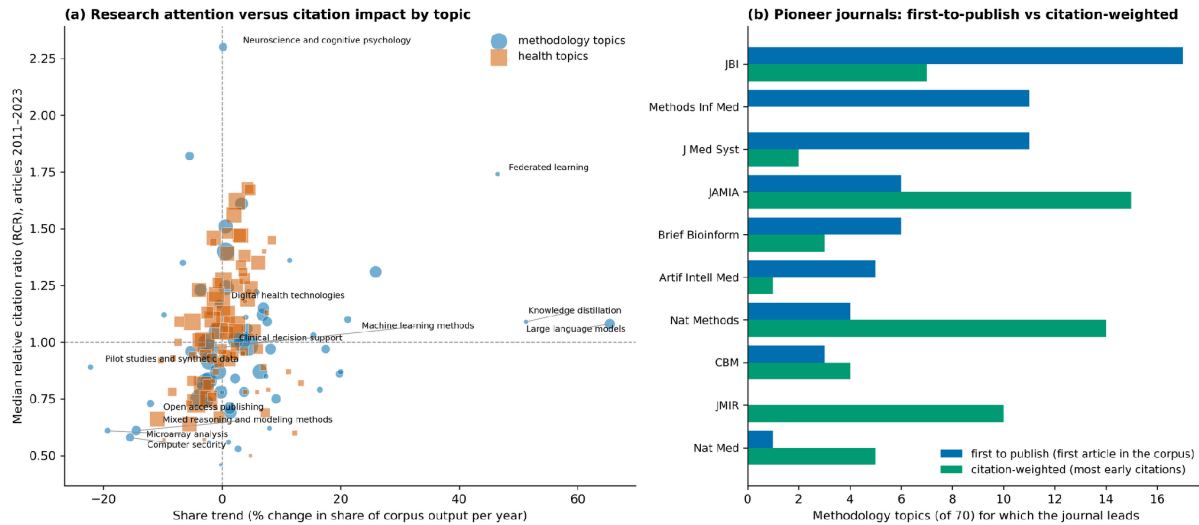
Supplementary Figure 9: Journal-fit heat maps (A: methodology; B: health): the share of each journal's articles tagged with each topic, for the 15 highest-volume journals; the full 29-journal matrices are in the project repository.



Supplementary Figure 10: Submission-landscape scatter plots (A: methodology; B: health): each topic plotted by the number of publishing journals (submission options) against total articles (competition).



Supplementary Figure 11: Year-over-year change in article count for the 20 most frequent topics per domain (methodology and health); the year of the largest year-over-year increase for every topic is listed in Supplementary Table 4.



Supplementary Figure 12: Citation impact. (a) Research attention versus citation impact: each K=100 topic is plotted by its share trend (Figure 10) against the median relative citation ratio (RCR) of its articles published 2011–2023 (marker size scales with article count; dashed lines mark zero share change and the RCR benchmark of 1). (b) Pioneer journals under two definitions: the number of methodology topics for which a journal published the first article in the corpus versus the number for which its articles of the topic’s first three years accumulated the most citations.