

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

Yilun Liang

Columbia University

Gongbo Zhang

Columbia University

Casey Ta

Columbia University

Yilu Fang

Columbia University

Chunhua Weng

`cw2384@cumc.columbia.edu`

Columbia University

Article

Keywords:

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11210969/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

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

*Corresponding author: Chunhua Weng, Department of Biomedical Informatics, Columbia University, 622 W 168 Street, PH20 room 407, New York, NY 10032, USA. Email: cw2384@cumc.columbia.edu

Abstract

We developed a novel scalable framework that tracks the evolution of research topics across many journals and applied it to biomedical informatics, a field central to digital medicine, contributing methods for semantic harmonization across heterogeneous publication metadata. We analyzed 78,073 articles from 29 PubMed-indexed journals (2011–2025) carrying 95,872 distinct keywords: 75% author-supplied via PubMed or publishers, 24% substituted MeSH descriptors, 1% generated from abstracts. Three overlapping waves emerged: genomics and foundational informatics, machine learning, and generative AI. Attention shifted in 34 of 70 methodology and 22 of 86 health topics; rises persisted on author-supplied keywords alone, whereas most declines disappeared, implying that changes in keyword source, not waning attention, explain most apparent declines. Digital health technologies was the connective topic, clinical and computational collaboration grew, and the Journal of Biomedical Informatics bridged the journal communities. The framework can be applied to other biomedical research fields to identify emerging research priorities.

Introduction

Knowing how a scientific field is changing helps researchers, funders, and educators see which technologies are gaining ground, what skills the workforce will need, and where research and governance priorities should focus. Scientific publications provide a continuous record of research activity. Changes in the topics and terms that articles use can signal when a novel research idea emerges, spreads, or fades, and patterns of authorship, affiliation, and venue can show how collaboration and research communities evolve. Analyzing these signals at scale, and over time, can reveal not only which technologies attract attention but also how their links with clinical applications change. Yet the literature has become so vast and so interdisciplinary that no individual researcher or expert panel can keep a complete view of a scientific field.

Bibliometric and computational text-analysis methods¹, including citation analysis, co-authorship networks, co-word analysis, bibliographic coupling, and topic modeling, are widely used to describe how science develops^{2,3}, and tools such as bibliometrix⁴, VOSviewer⁵, and SciMAT⁶ make many of these analyses easy to run. Each approach, however, has limits for following a diverse, interdisciplinary field over time. Citation-based indicators accumulate slowly and mainly measure visibility rather than research content⁷. Keyword co-occurrence analyses can place synonyms or closely related terms in separate topics, which produces redundant and potentially misleading trend estimates^{8,9}. Latent topic models find statistically coherent word distributions¹, but their outputs usually need subjective, labor-intensive interpretation before they can be communicated as meaningful themes. Expert-curated topic lists are easier to interpret¹⁰ but hard to maintain as terminology and technology change.

In earlier work¹¹, we combined biomedical language representations, semantic clustering, and topic naming by a large language model to organize author keywords into named hierarchical taxonomies of methodology and health topics. Applied to the Journal of Biomedical Informatics (JBI), that approach supported analyses of topic prevalence, topic trajectories over time, keyword change within topics, and co-occurrence between methodology and health topics. A single-journal design, however, cannot tell whether a trend reflects a field-wide transition or the editorial priorities and submission patterns of one venue, and it cannot describe differences among the journal communities that make up a field.

These limitations call for methods that merge synonymous terms, produce interpretable topics with little manual curation, and allow consistent comparisons across venues. We therefore developed TRACE (Tracking Research Attention and Community Evolution), a scalable framework for field-level research surveillance, and applied it to biomedical informatics, an interdisciplinary field central to digital medicine, health care, and biomedical discovery. Expanding from one journal to multiple representative

journals for a whole field is not just a matter of scale. Author supplied keywords are available for varying fractions of articles across journals and publishers (from about 0.1% to 100% in our corpus), and 24.6% of records needed substitutes drawn from MeSH descriptors or generated by a language model (Methods). Communities describe related concepts at various levels of specificity, publication volume and topical breadth differ greatly between venues, larger corpora may need different clustering resolutions, and testing trends across many topics creates a multiple-testing burden.

TRACE innovatively addresses these open research challenges by combining keyword acquisition from four complementary sources (PubMed metadata, publisher pages, MeSH descriptors, and keywords generated from abstracts by a language model) with semantic clustering, AI-assisted topic naming, longitudinal trend analysis, and cross-journal network and specialization analyses (Figure 1). Beyond counting publications, the framework dates when topics emerge, tracks how their keyword mix changes, separates field-wide patterns from journal-specific specialization, maps similarity and community structure among journals, follows changes in authorship and institutional collaboration, and measures shifts linked to major events, with the COVID-19 pandemic as the case study. We applied it to 78,073 publications and 95,872 distinct keywords from 29 PubMed-indexed journals covering biomedical informatics, digital health, clinical informatics, and computational biology from 2011 through 2025, and we report the findings in three layers: field-level trends, the cross-journal ecosystem, and practical impact.

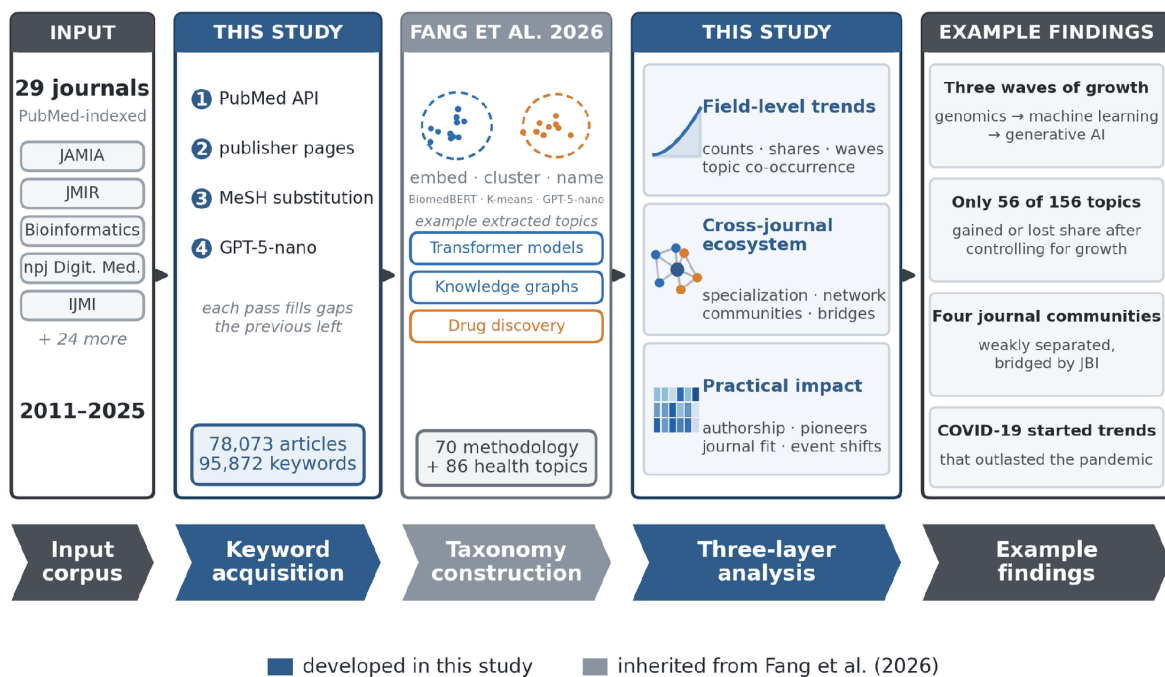


Figure 1: Overview of the TRACE framework. Module 1 (this study) assembles the corpus from 29 PubMed-indexed journals (2011–2025) and recovers publication keywords in four passes (the PubMed API, publisher pages,

MeSH-descriptor substitution, and GPT-5-nano keyword generation from abstracts), yielding 78,073 articles that carry 95,872 distinct keywords. Module 2 (inherited from Fang et al.¹¹) embeds, clusters, and names the keywords, producing 70 methodology and 86 health topics. Module 3 (this study) analyzes the topic-tagged corpus in three layers: field-level trends, the cross-journal ecosystem, and practical impact. Blue bands mark the modules developed in this study, the light gray band marks the inherited module, and dark gray bands mark the input corpus and example findings.

Results

Corpus and topics

The corpus contains 78,073 articles published between 2011 and 2025 in 29 PubMed-indexed journals that together span clinical informatics, digital health, computational biology, and bridging methods (Supplementary Table 1). The articles carry 95,872 distinct keywords, which the taxonomy pipeline grouped into 70 methodology topics and 86 health topics at the operating resolution of 100 clusters per domain ($K=100$; Supplementary Table 2). An article belongs to a topic when it carries one of the topic's keywords, so one article can belong to several topics: 4.6 on average (SD 2.0; median 4). Because this count depends on how many keywords a record carries, it varies by journal and by keyword source, from 2.6 topics per article in Nature Medicine to 7.5 in Bioinformatics Advances (Supplementary Table 3). We report the findings in three layers (field-level trends, the journal ecosystem, and practical impact) and compare them with the single-journal analysis of JBI by Fang et al.¹¹ where the comparison is informative. Supplementary Note S1 reports a sensitivity analysis on clustering resolution and keyword provenance. All figures use $K=100$ unless noted otherwise.

Publication volume grew in three overlapping waves

Annual output grew 7.6-fold, from 1,494 articles in 2011 to 11,384 in 2025 (Figure 2). The largest year-over-year increases came in 2014 (+39%), 2019 (+30%), and 2025 (+30%). Part of the early growth reflects corpus coverage rather than field expansion: journals whose PubMed indexing starts after 2011 (Supplementary Figure 1) account for 69% of the 2013 increase and 74% of the 2014 increase, when Bioinformatics entered the corpus. The 2019 and 2025 increases came from established venues. Supplementary Table 4 lists the year of the largest increase for every topic.

We therefore did not date research waves from the volume curve. Instead, we define a wave as a group of related methodology topics that broke through within a three-year span, and we date its onset with the breakthrough rule (Methods): the first year in which the wave's signature topic reached 25% of its eventual annual peak. Signature topics are the $K=100$ topics named for the wave's defining method (machine learning methods; large language models) together with the method topics that broke through alongside them, and the foundational topics are those already above the threshold in 2011. This gives

three overlapping waves; the second and third follow the deep-learning and generative-AI breakthroughs marked by AlexNet and the transformer architecture^{12,13}. Earlier themes keep producing publications throughout the window, so each wave adds to the themes already in circulation:

1. Foundational informatics and computational biology (established before 2011). Clinical decision support, statistical methods, software and usability research, and genomics and sequence bioinformatics were already established in 2011, so their onset lies before the window. At field scale this wave is carried mainly by the computational-biology journals, and it has no counterpart in the single-journal JBI analysis.
2. Machine learning (breakthrough 2019). Machine learning methods reached 25% of its eventual peak in 2019, with biomedical signal processing (2018), medical image segmentation (2019), and mixed deep learning applications (2020) alongside it. Its share of annual output held near 30% through 2020 and reached 46% by 2025.
3. Large language models and generative AI (breakthrough 2024). Transformer models (2022) and diffusion generative models (2023) came first; large language models and generative-AI methods broke through in 2024, and the large language models topic grew from 43 articles in 2022 to 578 in 2025. Part of the steep 2024–2025 climb may reflect AI-assisted writing raising output across the board, not only research about generative AI (see the limitations in the Discussion).

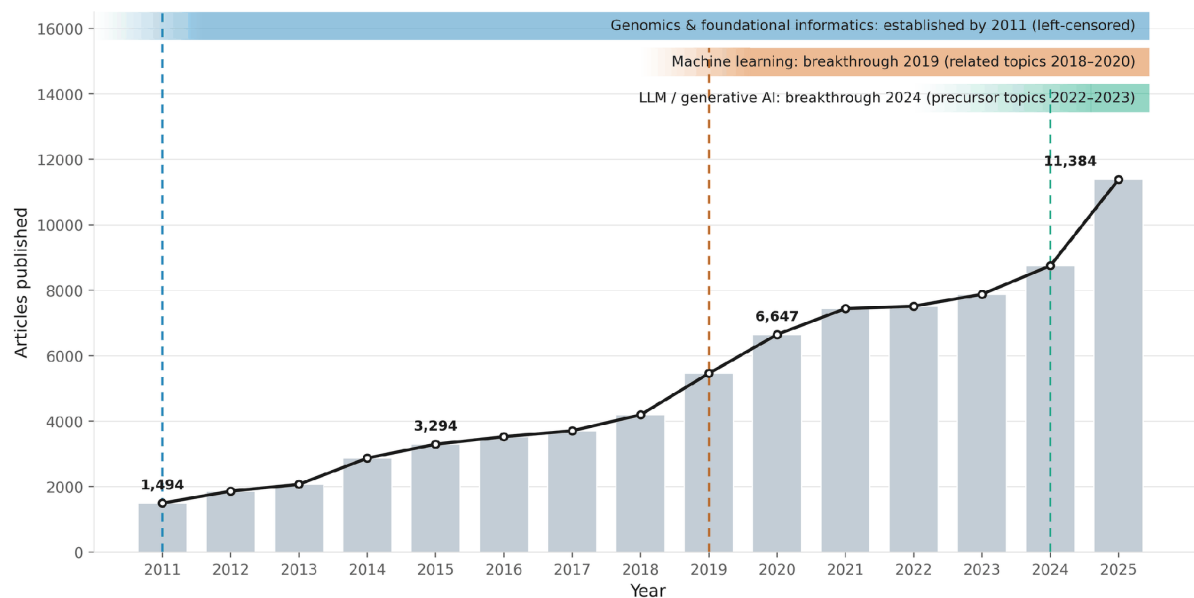


Figure 2: Annual publication volume, 2011–2025, with the three overlapping waves of research attention. Wave onsets are dated by the breakthrough rule (Methods) applied to each wave’s signature topics: the foundational topics were already established at the 2011 corpus start (left-censored), machine learning methods broke through in 2019 (related topics 2018–2020), and large language models in 2024 (transformer models 2022, diffusion models 2023). The bands mark onsets, not period boundaries; earlier themes continue to contribute publications throughout.

Figure 3 overlays the annual counts of four topics with dated milestone events that were chosen in advance (Methods; Supplementary Table 5). In each panel, output moves in step with the milestones: transformer-model research surged after the release of ChatGPT¹⁴, the large language models topic rose as medical LLMs appeared, de-identification and privacy studies grew after the GDPR took effect and the All of Us Research Program¹⁵ launched in 2018, and infectious-disease research surged during the COVID-19 pandemic, peaked in 2021 and declined from 2022, before the WHO ended the public health emergency in May 2023¹⁶, then partly rebounded in 2025 with the corpus-wide surge. Fang et al. saw the same alignment in JBI, which suggests it is a property of the field rather than of one journal. The overlay shows timing only. We do not test whether an event caused a change, and other developments, such as renewed attention to measles and Ebola, may contribute to the 2025 rebound.

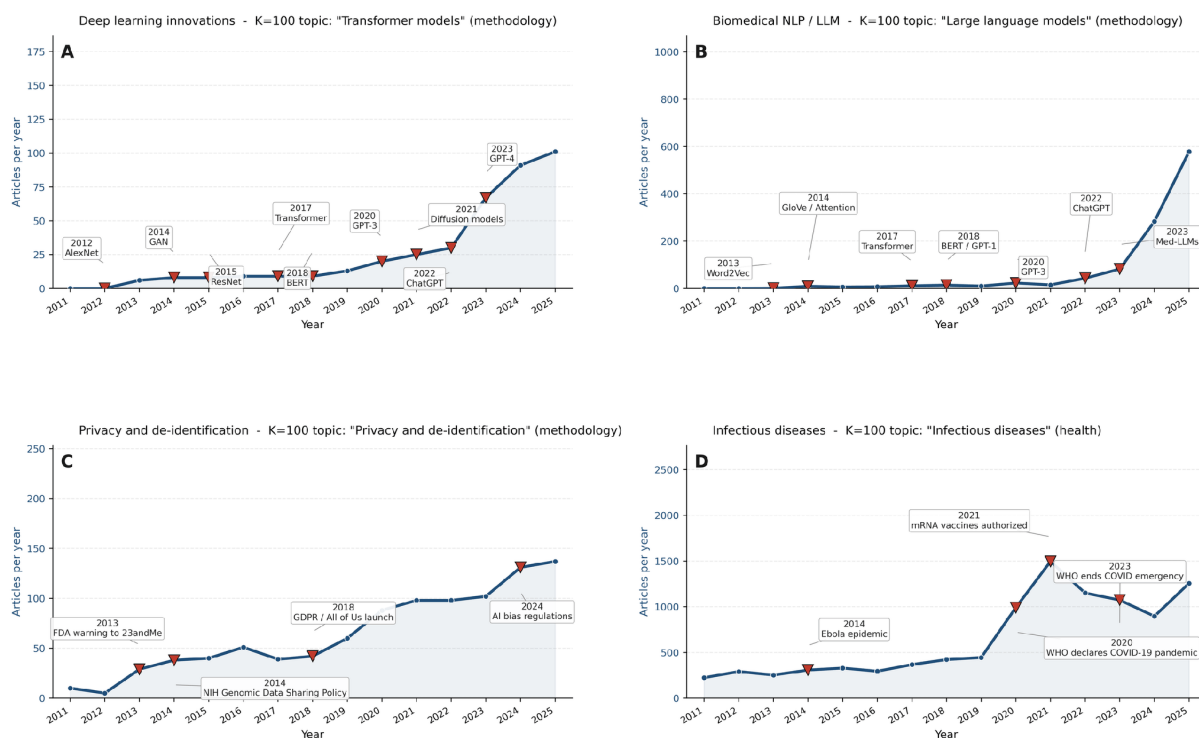


Figure 3: Milestone-annotated publication trends for four topics (three methodology, one health). Each panel overlays annual counts against dated events: (a) transformer models, from AlexNet to GPT-4; (b) large language models, from Word2Vec to medical LLMs; (c) privacy and de-identification, from the FDA warning to 23andMe to the 2024 AI bias regulations, including the GDPR and the All of Us launch (both 2018); (d) infectious diseases, from the 2014 Ebola epidemic to the end of the COVID-19 emergency. The events and their sources are listed in Supplementary Table 5.

Most topics grew, and a few grew very fast

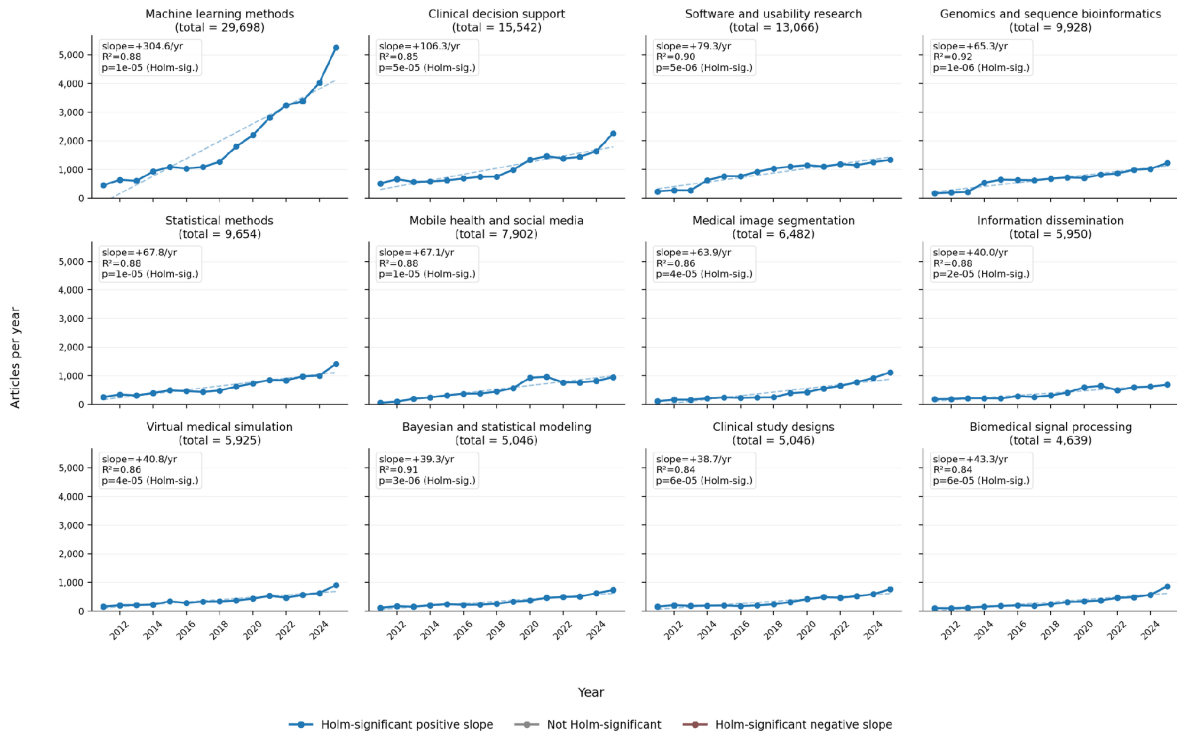
Growth was broad but uneven across topics. Figure 4 shows the 12 largest methodology and health topics with annual counts and a fitted linear trend. Most topics grew in absolute terms: 81.4% of methodology

topics (57 of 70) and 89.5% of health topics (77 of 86) have positive slopes that remain significant after Holm correction; four methodology topics had negative slopes, none significant. These count trends partly reflect the growth of the field itself, so share trends are reported below. Within JBI, Fang et al. found clear decliners (support vector machines, biomedical ontologies, and sociotechnical systems), based on author keywords only. At field scale the corpus grows so fast that even topics losing share do not shrink in article count, and most apparent declines coincide with a change in where the keywords come from rather than in research attention (see the limitations in the Discussion; Supplementary Note S1.2).

Ranking topics by growth relative to their own mean, so that topics with few articles but fast percentage growth are visible, the leading methodology risers were knowledge distillation (+30.0% per year; $n=98$; Holm-adjusted $p = 0.012$), federated learning (+29.4%; $n=154$; $p = 0.009$), transfer learning variants (+26.5%; $n=41$; $p < 0.001$), contrastive learning (+25.3%; $n=1,579$; $p = 0.001$), and data augmentation (+23.8%; $n=383$; $p < 0.001$). Large language models grew fastest of all (+31.4% per year; $n=1,082$) but not significantly ($p = 0.11$), because its late, steep rise fits a straight line poorly. The leading health risers were sequence and structure prediction (+20.6% per year; $n=198$; $p = 0.001$), health semantic similarity (+20.1%; $n=189$; $p = 0.003$), causal mediation analysis (+19.6%; $n=204$; $p = 0.002$), motor imagery (+18.0%; $n=567$; $p = 0.002$), and rare diseases (+17.7%; $n=81$; $p = 0.003$).

Growth also changes what a topic contains. Within machine learning methods, the keyword deep learning rose from almost none in 2011 to 862 articles in 2025 and overtook algorithms, the largest early keyword, whose yearly count stayed between 227 and 384 articles while its share of the topic fell (Supplementary Figure 2). The datasets that articles name changed with the waves as well: mentions of MIMIC and eICU, UK Biobank, and All of Us rose, while mentions of the i2b2/n2c2 shared-task corpora fell (Supplementary Note S3; Supplementary Table 6). A descriptive split of the corpus into three five-year windows (Supplementary Figure 3; Supplementary Note S4) confirms the ranking: machine learning methods and digital health technologies lead every window, and the third-wave topics are too new to enter the top 20 by volume.

(a) Methodology



(b) Health

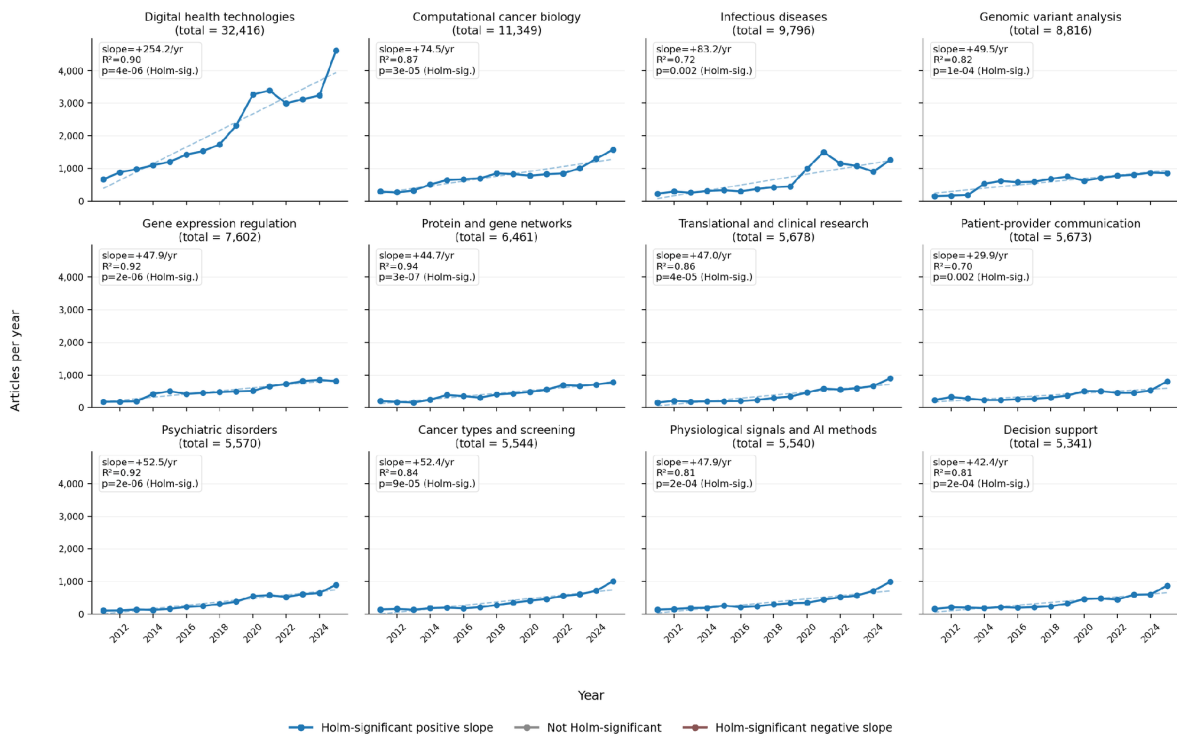


Figure 4: Top 12 methodology (a) and health (b) topics by publication volume, 2011–2025. Each panel shows annual counts with a linear-regression overlay; line color encodes Holm-corrected significance (blue = significant positive, red = significant negative, gray = not significant).

The digital health technologies topic is the connective backbone of the field

Figure 5 shows how often methodology and health topics occur together, and Supplementary Table 7 lists the 10 most frequent cross-domain pairs with their growth. Digital health technologies is the backbone of this structure: it appears in 6 of the top 10 pairs, the only topic to appear more than three times, and it forms the dark column in Figure 5. Machine learning methods appears in 3 of the top 10 pairs. The anchor of this structure depends on the corpus. Within JBI, the most frequent cross-domain pair was machine learning × patient safety; machine-learning pairs remain prominent here, but patient safety does not persist as a standalone topic at field scale. The digital-health column has no counterpart in the single-journal analysis because digital health technologies was not among the topics Fang et al. reported. This is a taxonomy effect rather than a difference in what JBI publishes: 48% of JBI’s articles carry a digital-health keyword (42% corpus-wide), but in the single-journal keyword pool those keywords (electronic health records, telemedicine, mobile health) were spread across narrower topics such as electronic medical records. The umbrella topic emerges only in the field-scale pool, to which the digital and consumer-health journals contribute most of its keywords.

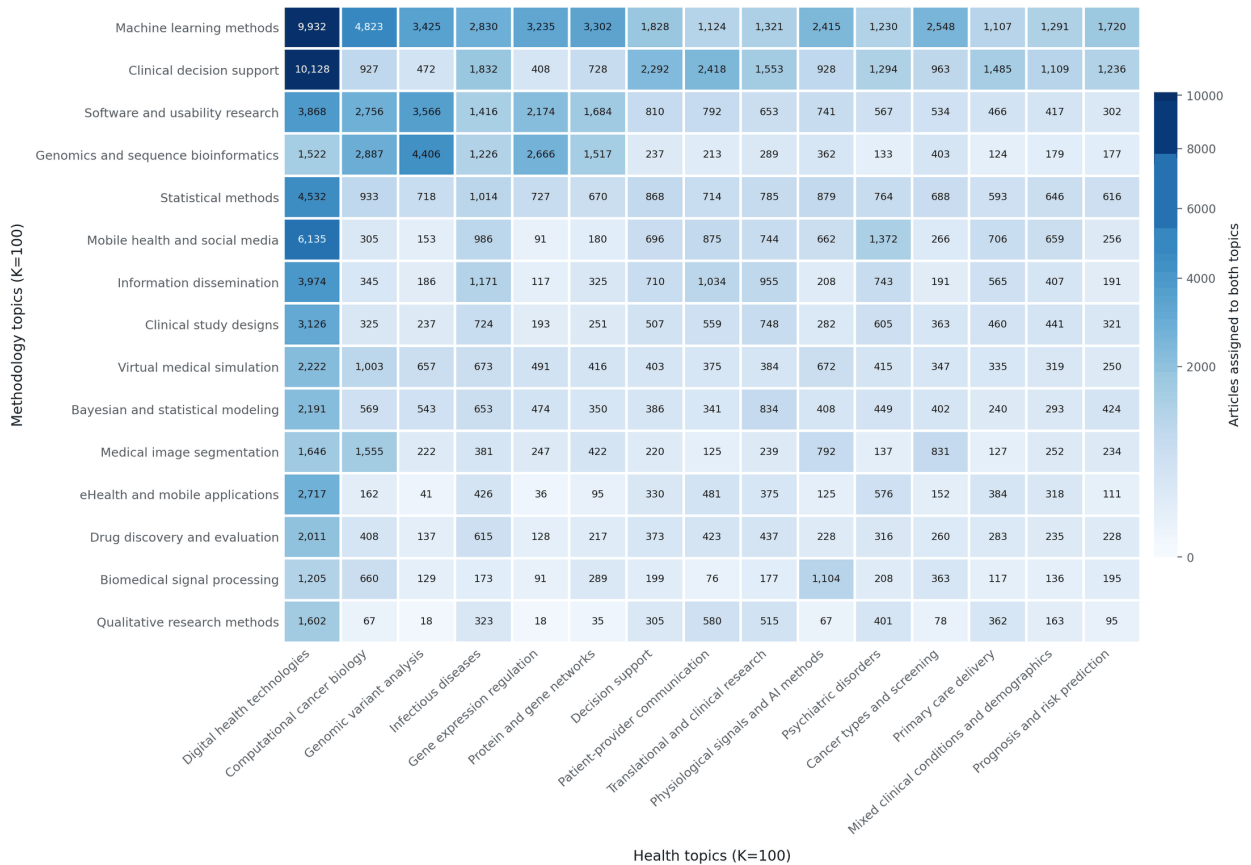


Figure 5: Methodology × health co-occurrence heat map. Cell intensity encodes the number of articles jointly assigned to a methodology topic (rows) and a health topic (columns); the digital-health-technologies column dominates the coupling structure.

Journals grew at different speeds

The next analyses look at the journals themselves. Based on their publication trajectories, we sorted the 29 journals into two criterion-defined groups and a residual group (Figure 6). High-volume veterans were indexed across the full window and published more than 4,000 articles: JMIR (9,728 articles), Computers in Biology and Medicine (8,908), and Briefings in Bioinformatics (4,579). High-growth newcomers were first indexed after 2011 and grew their annual output more than threefold by 2025: IEEE J-BHI, JMIR mHealth and uHealth, npj Digital Medicine, JMIR Medical Informatics, Frontiers in Digital Health, Journal of Clinical and Translational Science, JAMIA Open, and Bioinformatics Advances (4.5- to 12.2-fold; Supplementary Table 8). The remaining 18 journals mix long-established mid-volume venues, recently founded journals that have not yet tripled their output, and one discontinued title (Journal of Innovation in Health Informatics, indexed 2014–2019). The largest of them, Bioinformatics (9,564 articles, indexed in the corpus from 2014), peaked at 993 articles in 2019 and fell to 598 by 2025. Many of the new venues specialize in digital health, which mirrors the rise of digital health technologies as the field's largest topic.

Topic trends also differ between journals (Supplementary Table 9; Supplementary Figure 4; Supplementary Note S5). Fitting the share model within each journal, 27 of the 28 journals with at least five indexed years show at least one Holm-significant topic trend (269 of 1,419 journal-topic pairs tested; 115 rising, 154 falling). Machine learning methods gain share in 11 journals and large language models in 7, whereas the topics that most often lose share are virtual medical simulation, clinical decision support, and information dissemination. Trends inside a journal that changed its keyword source over the period (see the limitations in the Discussion) carry the same provenance caveat as the field-level declines.

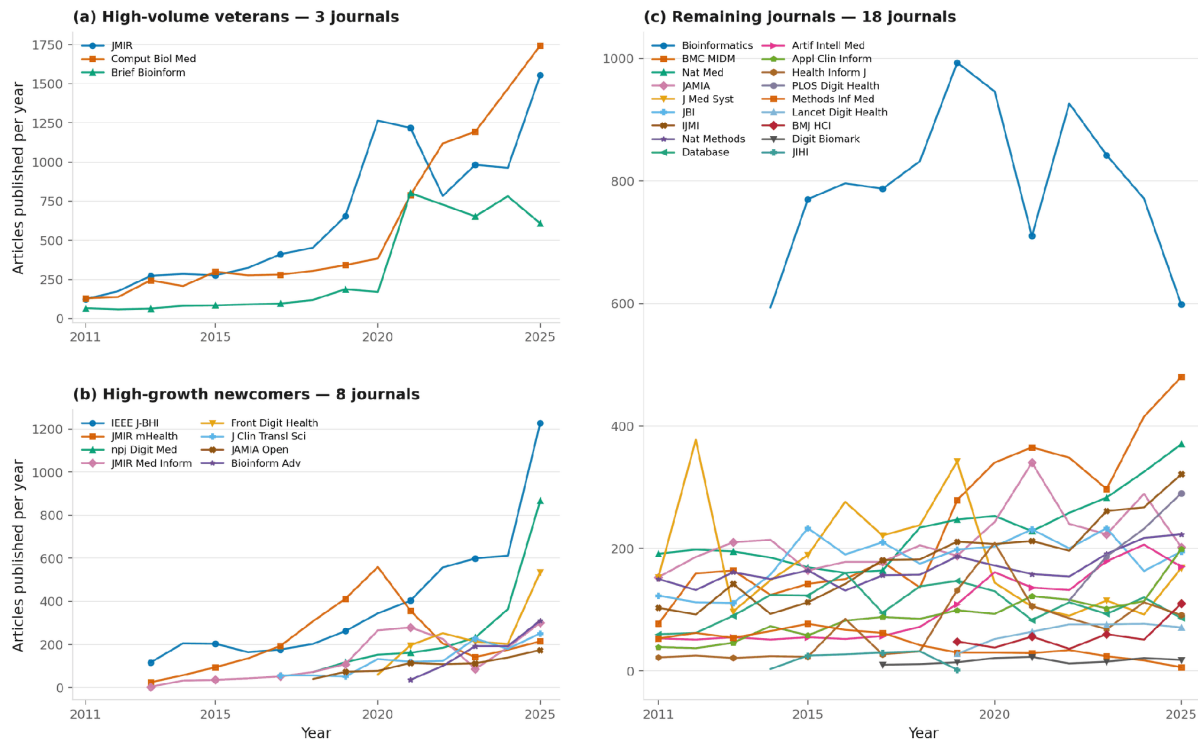


Figure 6: Per-journal publication volume over time for the 29 journals, grouped into (a) high-volume veterans, (b) high-growth newcomers, and (c) the remaining journals (right half). Each journal has its own color and marker, so every line is identifiable.

Journal specialization and topic universality

Figure 7 shows each journal's topic coverage relative to the field average, with journals and topics ordered by clustering so that similar profiles sit together. Four blocks stand out (dashed outlines). (i) Gene and variant block: Bioinformatics, Bioinformatics Advances, Briefings in Bioinformatics, Database, and Nature Methods are enriched, mostly 3–6 $\times$, for gene- and variant-centered health topics (genomic variant analysis, gene expression regulation, protein and gene networks, computational cancer biology) and depleted for health-service topics. (ii) Digital and consumer-health block: the JMIR family and related venues are enriched for mobile health and social media, eHealth and mobile applications, ecological momentary assessment, and information dissemination, and for psychiatric disorders and digital health technologies; they are modestly enriched for clinical decision support (1.2–1.7 $\times$) and depleted for computational biology and imaging topics. (iii) AI and signal- and image-processing block: Artificial Intelligence in Medicine, Computers in Biology and Medicine, and IEEE J-BHI are enriched 2–5 $\times$ for medical image segmentation, biomedical signal processing, and mixed deep learning applications. (iv) Computational biology block: the venues of block (i) are also enriched for software and usability research and for genomics and sequence bioinformatics; this block is the least clean, with mixed enrichment across the other methodology topics. Bioinformatics is the sharpest single case: 5.4 times the field average for

software and usability research (65.7% of its articles versus 12.1% corpus-wide) and 4.1 times for genomics and sequence bioinformatics, with its strongest enrichment on topics outside the plotted top 20, such as protein-peptide binding (8.8×).

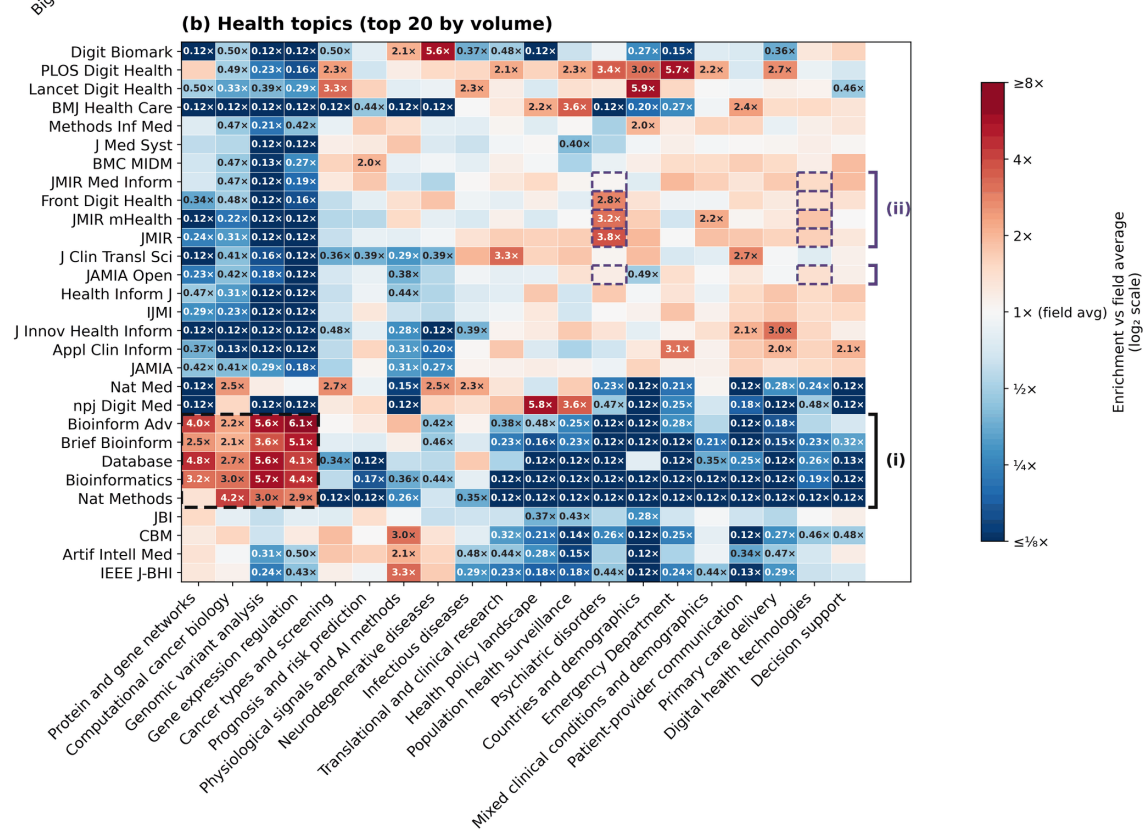
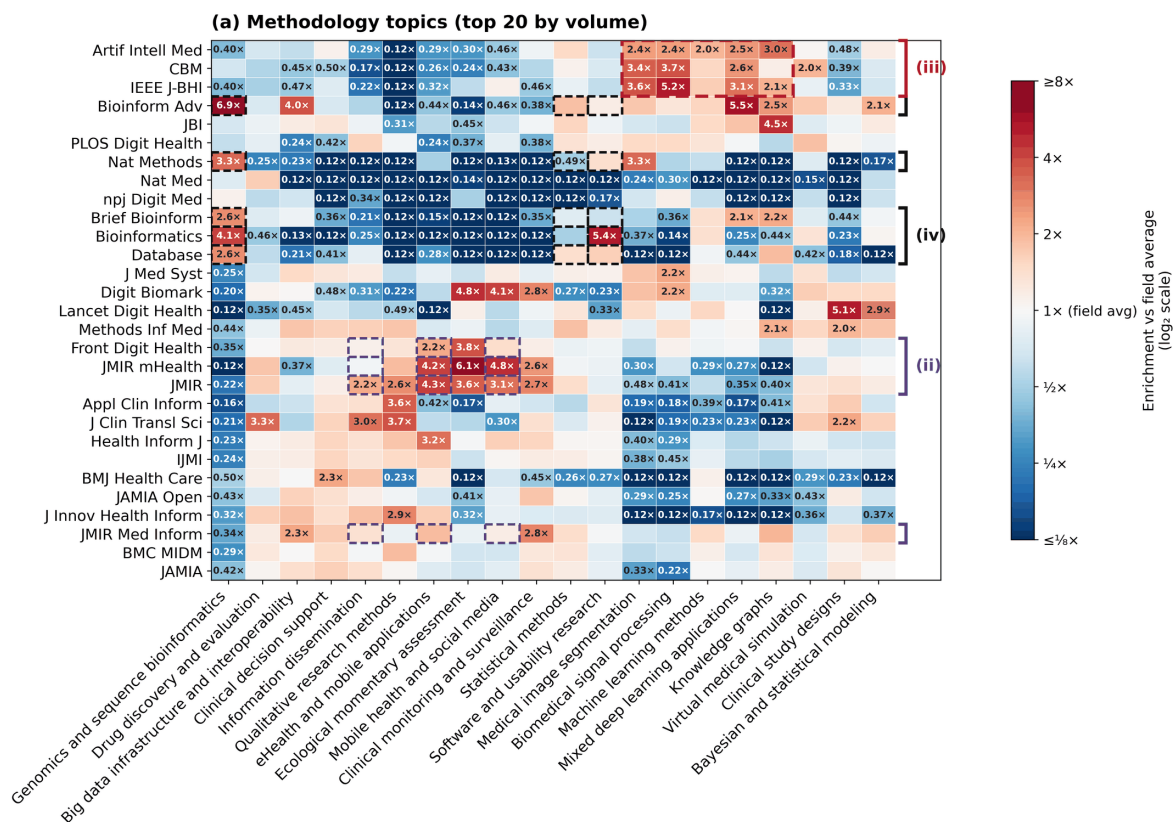


Figure 7: Journal and topic enrichment. Divergent enrichment heat maps (a: methodology topics; b: health topics) show each journal's enrichment or depletion for each topic relative to the field average (the unweighted mean of the 29 journals' shares). Rows and columns are reordered by cosine-distance average linkage. Dashed outlines mark the four blocks discussed in the text: one rectangle where a block's journals cluster together, and individually outlined member cells where the clustering interleaves other journals. The labelled brackets in the right margin mark the member rows of each block. The numerals match the blocks in the text: (i) gene and variant block, the computational-biology journals (Bioinformatics, Bioinformatics Advances, Briefings in Bioinformatics, Database, Nature Methods) for genomic variant analysis, gene expression regulation, protein and gene networks and computational cancer biology (b); (ii) digital and consumer-health block, the three JMIR journals and Frontiers in Digital Health (plus JAMIA Open in b) for mobile health and social media, eHealth and mobile applications and information dissemination (a) and for psychiatric disorders and digital health technologies (b); (iii) AI and signal- and image-processing block, Artificial Intelligence in Medicine, Computers in Biology and Medicine and IEEE J-BHI for medical image segmentation, biomedical signal processing, mixed deep learning applications and knowledge graphs (a); (iv) computational biology block, the block (i) journals for genomics and sequence bioinformatics, statistical methods and software and usability research (a).

We measured how evenly each topic is spread across the 29 journals with normalized Shannon entropy¹⁷, where 0 means a topic sits in a single venue and 1 means an even spread (Figure 8). Universal topics (entropy ≥ 0.85) are the shared core of the discipline. In methodology they are big data infrastructure (0.882), Bayesian and statistical modeling (0.862), clinical decision support (0.859), statistical methods (0.855), and drug discovery and evaluation (0.852). The health domain is more concentrated: only prognosis and risk prediction (0.874), drug safety surveillance (0.866), and infectious diseases (0.863) clear the threshold, with digital health technologies (0.844) just below it. Niche topics (entropy ≤ 0.55 , at least 80 articles) are published mainly by a few journals. Methodology niches include knowledge distillation (0.48; 44.9% in IEEE J-BHI), gene editing (0.50; 41.1% in Bioinformatics), and a heterogeneous cluster of acronym-bearing engineering keywords, such as computational fluid dynamics and finite-element methods, that the taxonomy grouped as mixed engineering and imaging acronyms (0.41; 97 articles, 48.5% in Computers in Biology and Medicine). Health niches include protein-peptide binding (0.41; 61.3% in Bioinformatics), motor imagery (0.49; 39.7% in IEEE J-BHI), and sequence and structure prediction (0.52; 43.4% in Briefings in Bioinformatics). Seven of the ten niche health topics map to Bioinformatics: within this corpus, the computational-biology venues form

the most self-contained sub-community. These topics are concentrated in a few journals here because the journal set centers on biomedical informatics; in genomics journals outside the corpus they are large fields (see the limitations in the Discussion).

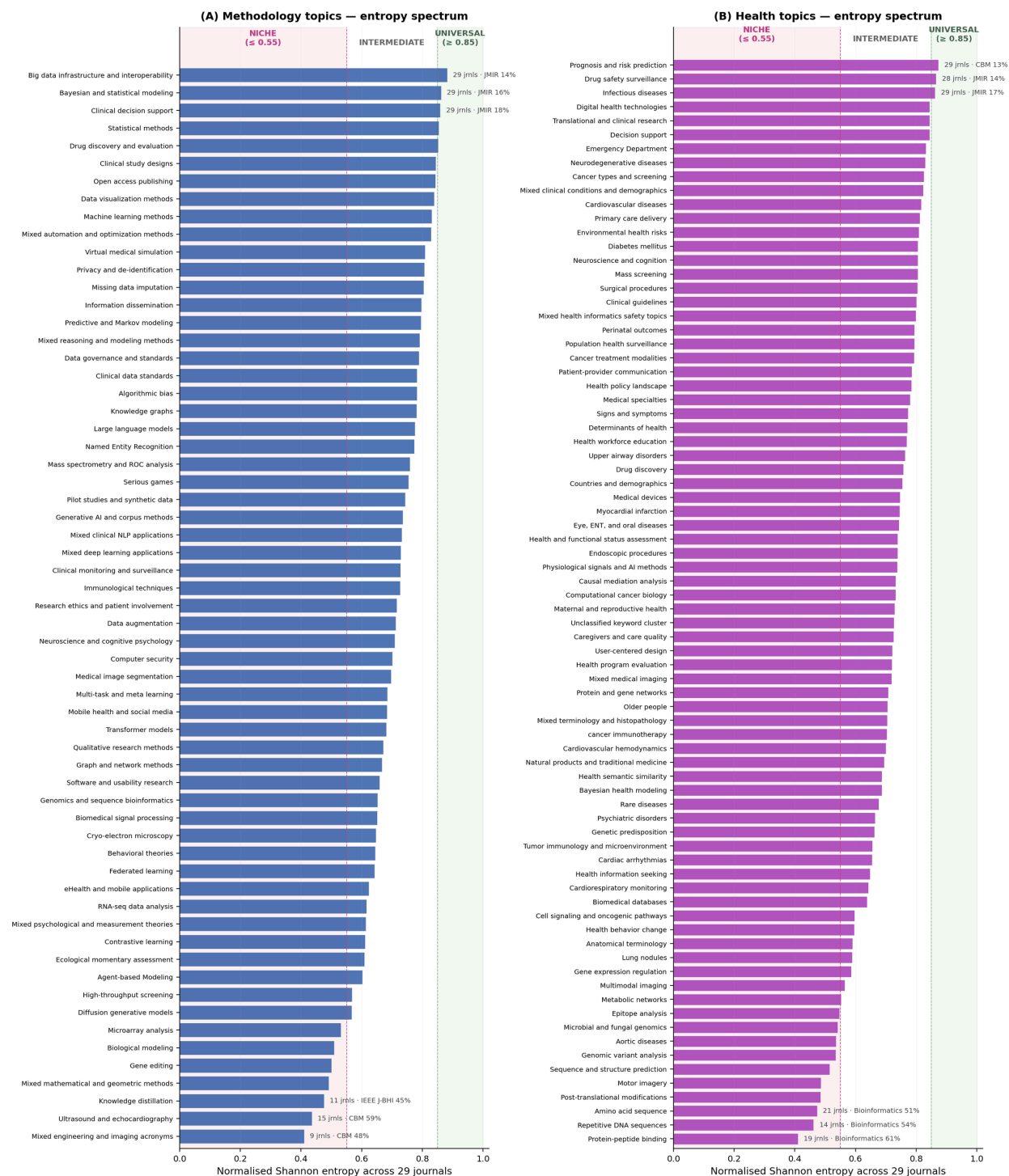


Figure 8: Cross-journal topic universality: K=100 topics ranked by normalized Shannon entropy across the 29 journals (A: methodology; B: health); shaded bands mark universal (entropy ≥ 0.85) and niche (entropy ≤ 0.55) topics.

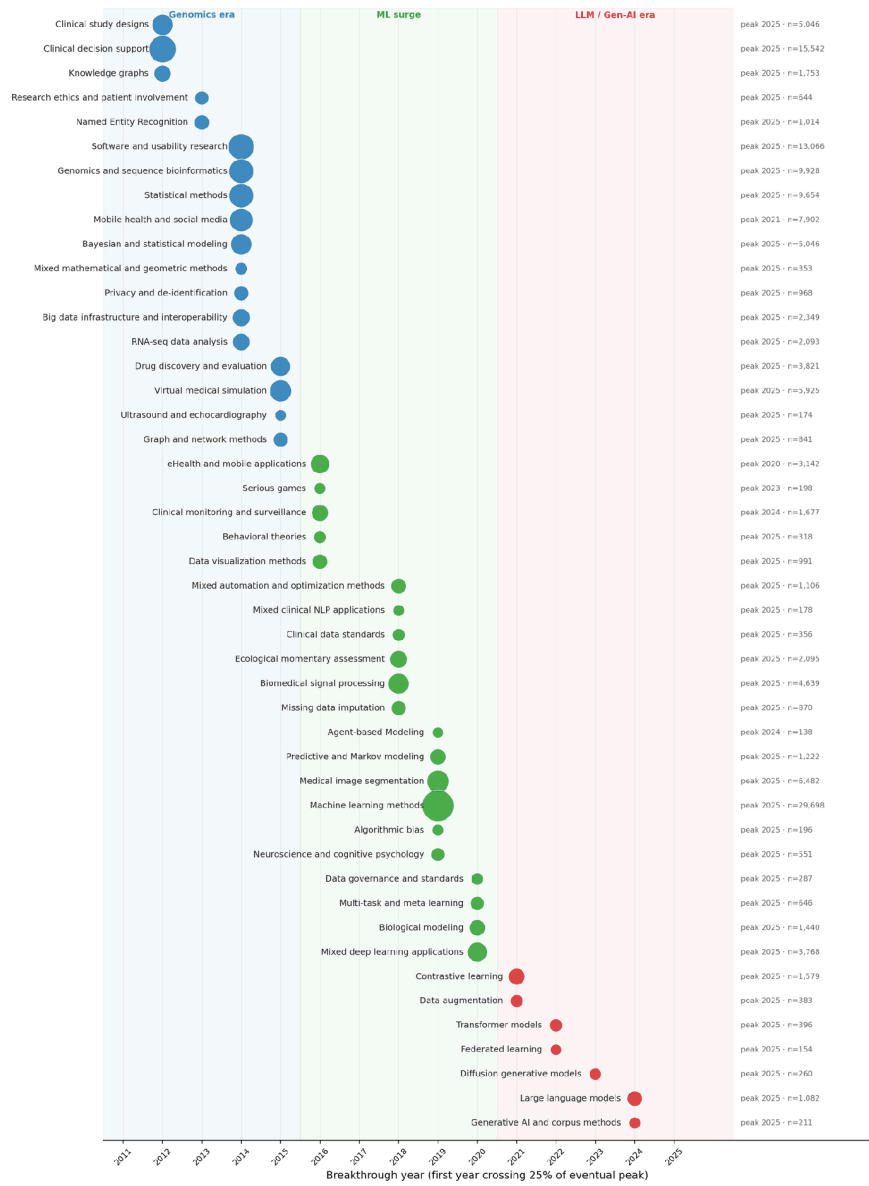
When topics broke through, and which ones rose or retreated

Figure 9 shows the breakthrough year of every K=100 topic with at least 100 articles, defined as the first year in which the topic reached 25% of its eventual peak annual volume. Methodology breakthroughs were front-loaded: 12 topics were already above the threshold when the corpus starts in 2011 (their true breakthrough is left-censored), followed by 9 in 2014, 5 in 2016, 6 in 2018, and 6 in 2019, and the median breakthrough year is 2015 (2017 without the left-censored topics). Health breakthroughs follow the same pattern (8 left-censored; median 2015). Breakthroughs become rare from 2020 onward, partly by design: a theme must accumulate enough articles to form its own K=100 topic, so the newest themes surface late. The late methodology breakthroughs are contrastive learning and data augmentation (2021), transformer models and federated learning (2022), diffusion generative models (2023), and large language models and generative-AI methods (2024). The only late health breakthroughs are sequence and structure prediction and eye, ENT, and oral diseases (2021) and health semantic similarity (2022).

(a) Methodology

True breakthrough year
if pre-2011 (unobserved)
Established at corpus start
(2011 count $\geq$ 25% of eventual peak; n=12)

- Information dissemination n=5,950
- Qualitative research methods n=2,152
- Mass spectrometry and ROC analysis n=1,302
- Mixed reasoning and modeling methods n=840
- Cryo-electron microscopy n=606
- Computer security n=535
- Open access publishing n=432
- Immunological techniques n=392
- High throughput screening n=292
- Microarray analysis n=259
- Gene editing n=248
- Pilot studies and synthetic data n=176



(b) Health

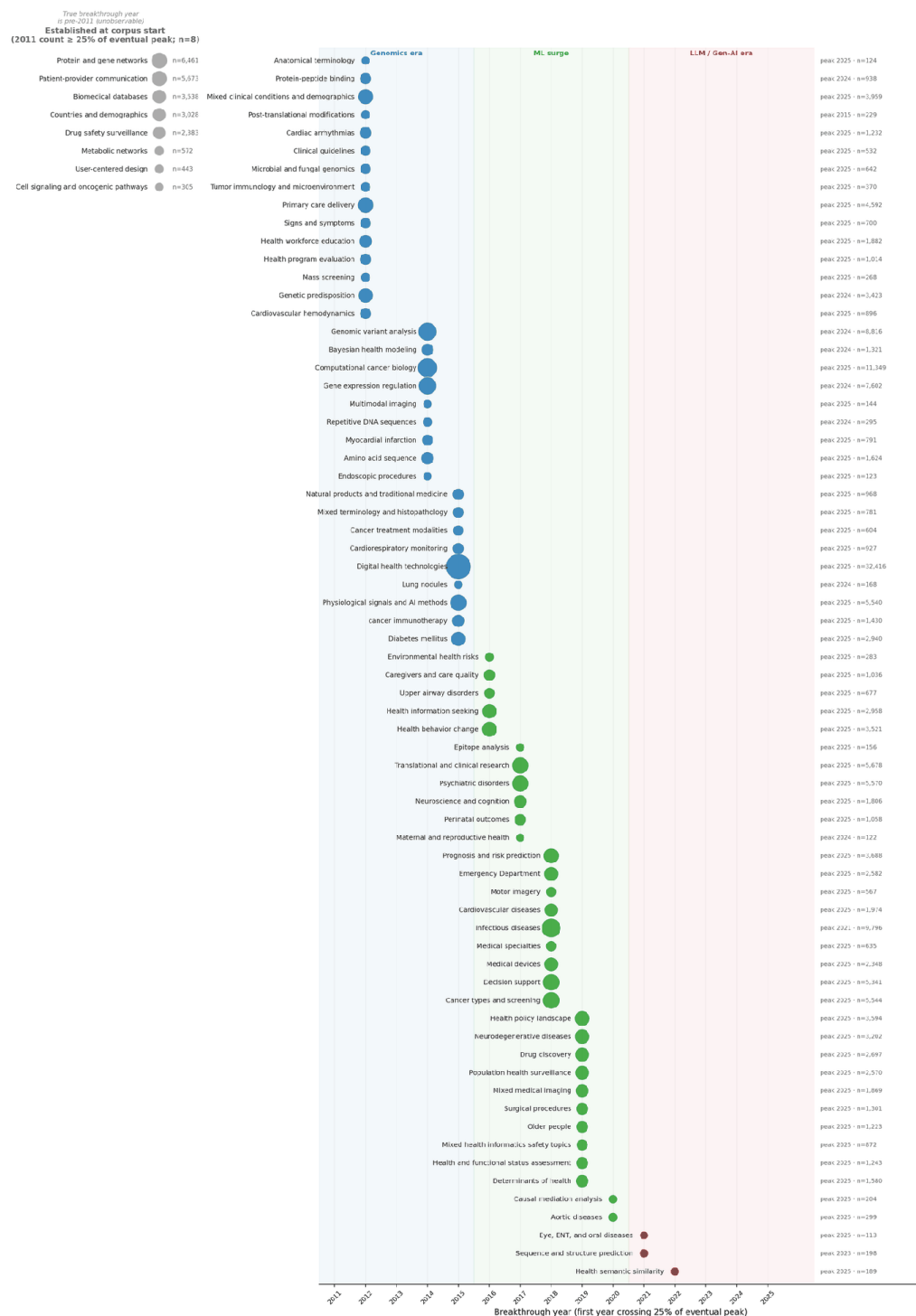


Figure 9: Topic breakthrough years for (a) methodology and (b) health topics. Each topic with at least 100 articles is plotted at the first year its annual count reached 25% of its eventual peak; the left column lists topics already at or above that bar at the 2011 corpus start (left-censored). Marker size reflects total 2011–2025 volume; marker color marks the five-year window containing the breakthrough year (2011–2015, 2016–2020, 2021–2025).

When the corpus itself grows 7.6-fold, nearly every topic's article count rises, so absolute counts are a poor guide to which topics are really gaining ground. Figure 10 therefore reports each topic's share of annual output, estimated with the overdispersion-robust Poisson model (Methods) and Holm-corrected within domain. Of the 70 methodology topics, 34 show a significant shift in share (24 gaining and 10 losing); of the 86 health topics, 12 gain share and 10 lose it. Of the 134 topics whose absolute counts rise significantly, 89 show no significant change in share.

The fastest-gaining topics are recent machine-learning methods: large language models (+65.4% per year), knowledge distillation (+51.3%), federated learning (+46.5%), transfer learning variants (+30.8%), and contrastive learning (+25.9%). The steepest losses fall on pilot studies and synthetic data (−22.2%), microarray analysis (−19.3%), computer security (−15.5%), and mixed reasoning and modeling methods (−14.5%). Health topics move within a narrower band, from sequence and structure prediction (+13.3%) to biomedical databases (−10.9%). Two findings differ between the count-based and share-based views. Digital health technologies, the largest topic at 32,416 articles, rose significantly in absolute terms, but its annual share fell from 44.3% in 2011 to 40.5% in 2025 (not significant). And seven health topics, among them drug safety surveillance, patient-provider communication, and clinical guidelines, publish more papers every year while losing ground relative to the field. These losses depend on where the keywords come from. A sensitivity analysis restricted to records with author-provided keywords (2013–2025; Supplementary Table 10; Supplementary Figure 5) keeps the sign of every gain (13 of the 24 methodology and 3 of the 12 health gains remain significant on the smaller record set), but none of the 10 methodology losses and only three of the 10 health losses (post-translational modifications, user-centered design, and health program evaluation) remain significant. The other apparent losses coincide with the shift from MeSH-substituted to author keywords over 2011–2016 (see the limitations in the Discussion), so we do not read them as declining research attention.

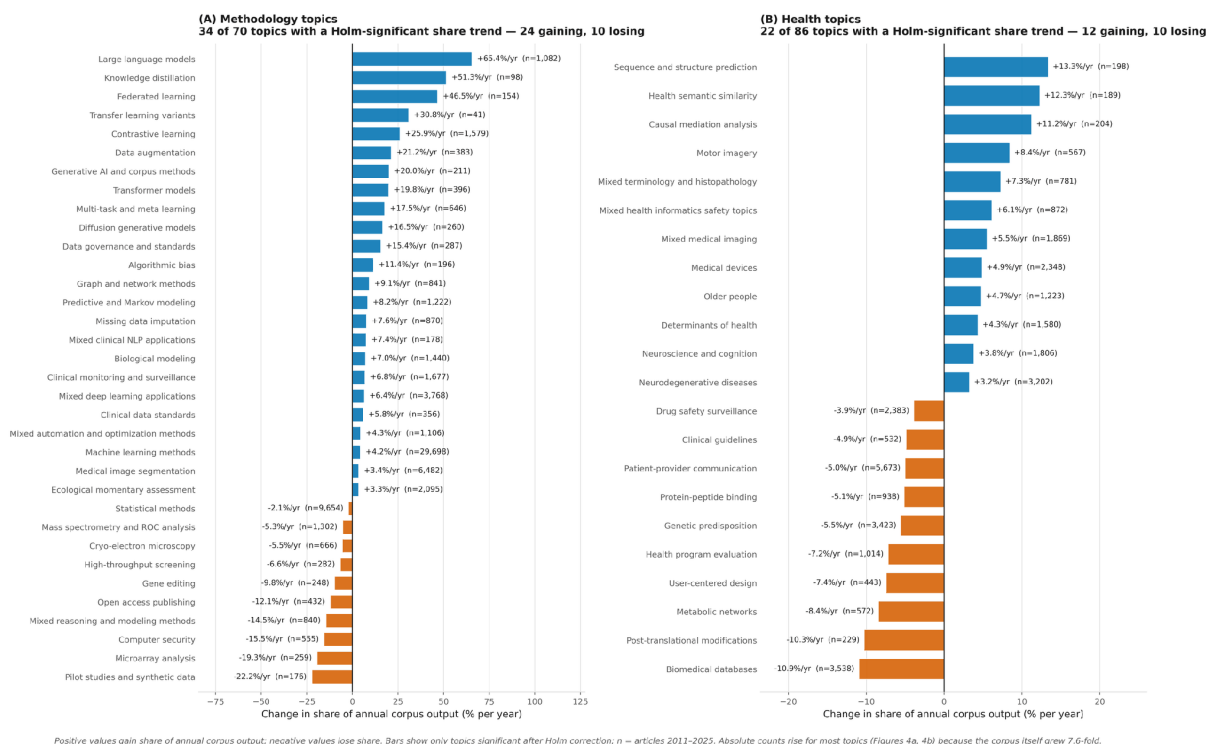


Figure 10: Topic share trends (A: methodology; B: health). Each bar is the annual change in a topic’s share of total corpus output, estimated by an overdispersion-robust Poisson model with a log-corpus-size offset and Holm-corrected within domain; only topics that remain significant after correction are shown. Blue indicates gaining share, orange losing share; n counts articles published 2011–2025.

The journal similarity network

Figure 11 places the 29 journals with the Kamada–Kawai layout, with each edge weighted by the cosine similarity between two journals’ topic-distribution vectors. JBI is the main structural bridge in the network, with a betweenness centrality of 0.339, followed by JMIR Medical Informatics (0.262), IJMI (0.243), and Briefings in Bioinformatics (0.217) (Supplementary Figure 6). JBI’s bridging role is consistent with its record as a pioneer journal: it published the first article for 24.3% of methodology topics (17 of 70); a citation-weighted version of this ranking is reported below. Louvain detection identifies four weakly separated communities (Figure 11, node color): a 15-journal clinical informatics and digital health community, a 7-journal community of AI-methods and bridging venues that includes JBI, a 5-journal computational biology community, and a pair formed by Nature Medicine and npj Digital Medicine. The modularity of 0.141^{18,19} points to a well-integrated ecosystem rather than rigid fragmentation. In line with its high betweenness centrality, the layout places JBI between the clinical-informatics and computational-biology communities. This structure matches the entropy gradient in Figure 8: within this corpus, computational biology venues publish more niche topics and sit loosely at

the edge of the graph, whereas clinical informatics journals publish more universal topics and are densely woven together.

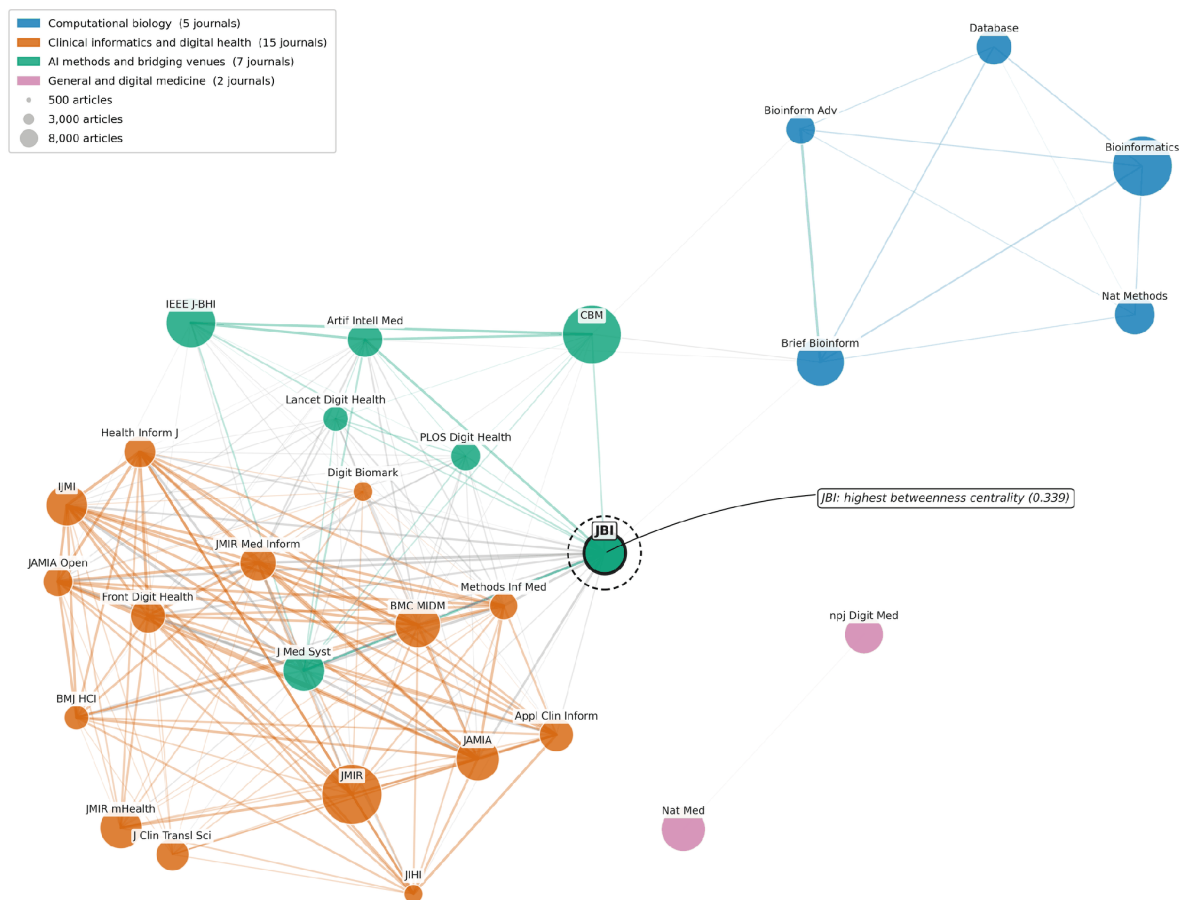


Figure 11: Journal similarity network. The 29 journals are positioned by a Kamada–Kawai layout with edges weighted by topic overlap. Node color marks the four Louvain communities (clinical informatics and digital health; AI methods and bridging venues; computational biology; general and digital medicine). The Journal of Biomedical Informatics is the structural bridge of the network (betweenness centrality = 0.339).

Teams grew larger and more mixed

The final layer looks at practical impact: collaboration, journal fit, the COVID-19 case study, and citation impact. Authorship changed in three ways (Supplementary Figure 7). Teams grew: the mean number of authors per paper rose from 5.98 in 2011 to 7.78 in 2025 (median 5 to 6), and the 95th percentile from 17 to 18. Collaboration widened: mean distinct affiliations per paper rose from 2.0 in 2014 to 4.85 in 2025 (median 1 to 4). Authorship became more mixed: the share of papers with both a clinical and a computational affiliation roughly doubled, from 16.4% in 2014 to 28.8% in 2025, while computational-only papers fell from 28.9% to 20.0%, clinical-only papers rose from 19.6% to 32.5%, and the neither/unknown category fell from 35.0% to 18.7% (we start at 2014 because earlier affiliation metadata

are sparse). Across journals (Supplementary Figure 8), mixed clinical and computational authorship is most common in JAMIA Open (51.4%), JAMIA (51.3%), and The Lancet Digital Health (51.1%), and teams are largest in Nature Medicine (mean 20.9 authors).

Journal fit and where topics have room

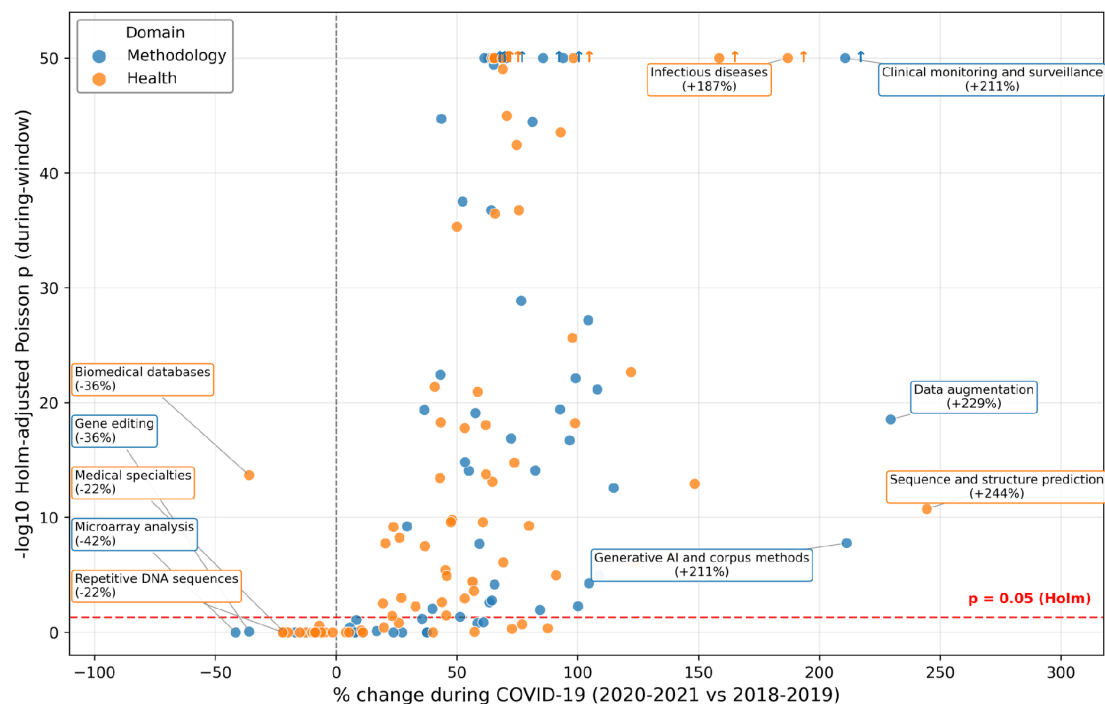
Choosing where to submit matters, because a poor topical match is a common reason for desk rejection²⁰. Two descriptive aids support that decision; neither has been validated against acceptance outcomes. Journal-fit heat maps (Supplementary Figure 9) give the share of each journal's articles on each topic: an author working on mobile health, for instance, finds JMIR mHealth and uHealth (47.6% of its articles) and JMIR (30.7%) as natural venues, and Bioinformatics, where the topic barely appears, as an unsuitable one. Submission-landscape scatter plots (Supplementary Figure 10) place each topic by the number of journals that publish it against its total article count, which separates crowded universal areas such as machine learning methods (29,698 articles across 29 journals) from topics with few articles and few venues, such as distributed ledger technology (19 articles in 6 journals).

Event-related shifts: the COVID-19 case study

The event test (Methods) can be applied to any dated event; we applied it to the COVID-19 pandemic. Figure 12 plots each $K=100$ topic by its percentage change in 2020–2021 relative to the 2018–2019 baseline against the $-\log_{10}$ Holm-adjusted Poisson p -value. The test included the 133 topics with at least 100 articles and a pre-pandemic mean of at least 3 articles per year; federated learning was excluded for insufficient pre-pandemic volume. About 68% of testable topics (91 of 133) shifted significantly in 2020–2021. Of these, 90 increased and only one decreased: biomedical databases (−36.1%), a MeSH-heavy topic that does not decline on author-keyword records (see the limitations in the Discussion). Microarray analysis (−42% raw) and gene editing (−36% raw) appeared to decline but did not stay significant after Holm correction. Part of this breadth reflects the corpus itself: total output was 46% higher in 2020–2021 than in 2018–2019, so a topic that merely kept pace with the field registers an increase of roughly that size. The leading gainers far exceeded it: sequence and structure prediction (+244%), data augmentation (+229%), generative AI and corpus methods (+211%),

clinical monitoring and surveillance (+211%), infectious diseases (+187%), population health surveillance (+158%), and mass screening (+148%).

We read the response as three patterns, as a description rather than a statistical classification: a direct surge in infectious diseases and health surveillance; a methodology pull-through, in which machine-learning and generative-AI methods grew together with pandemic-era applications; and a displacement of established database and genomics workflows, which published less during the pandemic window. The third pattern is the least robust, because those topics are dominated by MeSH-substituted records (see the limitations in the Discussion). Comparing 2020–2021 with 2022–2025 shows that post-pandemic growth outpaced the pandemic surge for most topics, for example generative AI (+211% during versus +722% after) and data augmentation (+229% versus +729%). For most affected topics the pandemic-era surge was followed by continued growth rather than a return to baseline (Supplementary Figure 11; Supplementary Table 4). We describe this association without assigning a cause.



Note: y-axis capped at $-\log_{10}(p) = 50$; 13 highly significant points marked with ↑ extend beyond.

Figure 12: COVID–19 volcano plot. Each K=100 topic is plotted by percentage change (2020–2021 versus the 2018–2019 baseline, x-axis) against $-\log_{10}$ Holm-adjusted Poisson p (y-axis); the dashed line marks Holm significance at $\alpha = 0.05$.

Citation impact and influential articles

Citation metrics were available for 78,070 articles, and 57,940 articles published in 2011–2023 have a relative citation ratio (RCR). The median RCR is 0.97 across methodology topics and 1.00 across health topics, so the corpus as a whole is cited at about the NIH benchmark rate, and 30% of articles exceed twice that rate. Attention and impact are only loosely related: across topics, the correlation between median RCR and the share trend is 0.23 for methodology and 0.24 for health topics, and topics gaining share have a modestly higher median RCR than topics losing share (1.06 versus 0.91 for methodology; 1.15 versus 0.93 for health) (Supplementary Figure 12; Supplementary Table 11). The highest-impact methodology topics are neuroscience and cognitive psychology (median RCR 2.3), cryo-electron microscopy (1.8), and federated learning (1.7); large language models sits at 1.1. The most influential individual articles are software and resource papers (Trimmomatic, Fiji, ImageJ, Bowtie 2, DADA2, and the REDCap consortium paper; Supplementary Table 12), and the articles most widely cited across topics within the corpus are BioBERT, SciPy 1.0, nnU-Net, a review of high-performance medicine, and the CONSORT-EHEALTH statement. Finally, the pioneer ranking depends on its definition: JBI leads first-to-publish (17 of 70 methodology topics), but a citation-weighted version is led by JAMIA (15 topics) and Nature Methods (14), with JBI fourth (7), and the two definitions name the same journal for 27% of methodology and 21% of health topics (Supplementary Table 13).

Discussion

We analyzed 78,073 articles published between 2011 and 2025 in 29 PubMed-indexed journals (Supplementary Table 1) to describe how the biomedical informatics field has evolved. Annual output grew 7.6-fold, from 1,494 to 11,384 articles (Figure 2). Beyond the four core analyses inherited from Fang et al.¹¹ (article volume, leading topics, keyword evolution, and topic co-occurrence), TRACE adds eleven characterizations, six of which examine the journal ecosystem. The topic breakthrough analysis (Figures 2 and 9) identifies three overlapping waves of research attention:

1. A foundational informatics and computational biology wave, already established when the corpus begins in 2011;
2. A machine learning wave whose signature topics broke through in 2018–2020, after which the share of annual output held by machine learning methods rose from about 30% to 46%;
3. A generative AI wave, with transformer models breaking through in 2022 and large language models in 2024.

Significant upward trends in annual article counts appeared in 81% of methodology topics and 90% of health topics (Figure 4). Several topics had negative slopes but did not reach corrected significance, though low statistical power and the fast-growing corpus may have hidden real declines. Finer clusterings do not change this picture: the fine-grained sub-topics that appear to decline owe their early peaks to MeSH-substituted records rather than to waning research attention (Supplementary Note S1.2).

A methodological finding of the multi-journal design is that the source of the keywords can look like a trend. Author keywords were unavailable for most 2011–2012 records, so MeSH descriptors were substituted for them. Topics that MeSH indexers tag routinely (study designs, imaging modalities, hospital information systems, support vector machines) therefore show early peaks and apparent declines that disappear when the analysis is restricted to author-keyword records (Supplementary Note S1.2; Supplementary Figure 5). The gains in share we report survive this restriction. Most of the losses do not, and we report them as provenance-sensitive.

The 29-journal ecosystem is well integrated (Figure 11). Louvain detection separates four weakly bound communities (Supplementary Figure 6), but the modularity of 0.14 is well below the 0.3 benchmark for strong community structure. JBI is the structural bridge between them (betweenness centrality 0.339) and leads the first-to-publish pioneer ranking, whereas the citation-weighted ranking is led by JAMIA and Nature Methods. Citation impact tracks attention only loosely, and the field's most influential articles are shared software and resource papers rather than the newest methods. The COVID-19 inflection reached 68% of testable topics (90 increasing, 1 decreasing) relative to the 2018–2019 baseline, and growth in the affected topics often kept accelerating after the pandemic, which suggests that the pandemic started trends rather than capping them.

These results serve five stakeholder groups in different ways.

Academic researchers. The framework offers a roadmap of the field that separates rapidly emerging areas, such as large language models, federated learning, contrastive learning, and sequence and structure prediction, from topics whose share of the field is stable or shrinking, and the provenance sensitivity analysis shows which of these trends are robust.

Journal editors. The enrichment matrix (Figure 7) shows how a journal's topic profile compares with its peers, which informs the choice between deepening an existing specialization and broadening topical reach, and the similarity network (Figure 11) shows which venues are the closest neighbors.

Prospective authors. Two decisions arise at different times. When choosing a research direction, the share trends (Figure 10) show which topics are still gaining ground, and the submission-landscape plots

(Supplementary Figure 10) show which of them are published by many journals with modest cumulative volume, that is, more venue options and fewer competing papers. Once a manuscript exists, its main topics can be matched to the journal-fit heat maps (Supplementary Figure 9) to find the venues that devote a large share of their articles to those topics.

Funders. The share trends and breakthrough years show where research attention is moving and where it has plateaued, and the COVID-19 case study shows how quickly the field re-allocated attention after a shock and that the new activity persisted. Combining share trends with citation impact gives a simple evidence map: areas gaining attention and cited above the field benchmark (for example, federated learning and contrastive learning) versus areas that grow in volume without citation traction. These are retrospective signals that can inform portfolio review; they do not forecast.

Educators. Courses on current methods can follow the breakthrough years and share risers (Figures 9 and 10). Foundational curricula can anchor on the universal topics of Figure 8, which are published across nearly all 29 journals and therefore represent the shared core of the field rather than the identity of any single venue. Provenance-robust declines (Supplementary Table 10) flag areas whose relative share is shrinking, which curricula may want to revisit, the most-cited articles (Supplementary Table 12) are natural reading-list candidates, and the milestone analysis (Figure 3) supplies context for why the field moved in particular directions.

This study has several limitations.

Truncated start year. Because the corpus starts in 2011, when journals such as Journal of Biomedical Informatics formally began the use of article keywords, we cannot observe the true breakthrough years of research areas established before then (Figure 9, left columns): 12 methodology topics (17%) and 8 health topics (9%) are left-censored. Extending the corpus further back would capture some of these milestones, but keyword metadata quality falls sharply before 2011.

Reliance on PubMed and keyword fallbacks. The method depends on PubMed to identify articles and retrieve keywords, with publisher-page scraping, MeSH substitution, and keyword generation from abstracts as fallbacks. The fallbacks create a coverage gradient: author-keyword availability ranges from about 0.1% to 100% across journals, so under-keyworded venues (notably the bioinformatics journals) may have topics systematically under-counted, which biases topic assignment and trend estimates toward the better-keyworded digital-health venues. The venues that look most topically isolated are also the least keyword-covered, so part of their apparent isolation may reflect metadata rather than content (Supplementary Note S2).

Curated journal selection. The 29 journals were chosen to span biomedical informatics but form a curated, non-exhaustive set. Field-level generalizations are conditional on this selection, informatics work published in adjacent clinical or computer-science venues is not represented, and topics that have few articles here can be large fields in journals outside the set.

Unvalidated decision tools. The journal-fit and submission-landscape tools are illustrative. They have not been validated against submission or acceptance outcomes and should not be read as predictive recommendations.

Volume confounders. Publication volume responds to influences unrelated to research content: journal launches and indexing starts (Supplementary Figure 1), changes in indexing practice, and, in 2024–2025, the mainstream adoption of generative-AI writing tools, which may raise output independently of what is being researched and not necessarily equally across topics. Wave magnitudes in Figures 2 and 3 should be read with this in mind; the share-based analyses are less exposed because they account for field-wide growth.

Keyword provenance changes over time. MeSH substitution supplied the keywords for 75% of 2011 records but 10% of 2025 records, because author keywords became available in PubMed for most clinical-informatics journals only from 2013, and two journals switched to MeSH substitution from 2018. MeSH indexers assign some descriptors, including the check tag “Humans” and headings such as “Support Vector Machine”, far more consistently than authors list the corresponding terms, so topics that absorb such descriptors show early peaks that shrink as the vocabulary shifts. We excluded the MeSH check tags from topic assignment (Methods) and treat declines as provenance-sensitive throughout; Supplementary Note S2 gives the numbers. Future applications should analyze a single keyword vocabulary or model provenance explicitly.

Metric definitions and thresholds. Several statistics rest on choices that are reasonable but not unique: the minimum article counts per analysis, the 25%-of-peak breakthrough threshold (a topic that keeps growing must clear an ever-higher bar, so digital health technologies is dated 2015 despite 677 articles in 2011), the Poisson rate test (which ignores overdispersion and field-wide growth, so we read the COVID-19 results as effect-size rankings corroborated by the share model), and the pioneer metric (which records where a topic first surfaced within the corpus, not influence). Affiliation classification ran on free-text affiliations, so the interdisciplinary shares are lower bounds. Supplementary Note S2 discusses each choice.

Citation metrics. The relative citation ratio is field- and time-normalized but immature for recent articles, so the citation analyses stop at 2023. iCite counts citations from PubMed-indexed articles only, and

within-corpus citations are 11.5% of all references, so the journal citation network describes exchange among the 29 journals rather than the field's full citation environment. Institution- and author-level influence were not analyzed because the collector retained affiliation counts and flags, not names.

Topic naming. Topic names were generated by a language model from each cluster's keywords and then audited by the authors. Clusters that the naming step could not summarize under a single theme carry a "Mixed ..." or "Unclassified" label (8 of 70 methodology and 5 of 86 health topics). We keep them for completeness, interpret them cautiously, and state their composition where the text singles one out.

Several extensions follow from these limitations.

The most immediate extension is a wider time window: a dedicated retrieval process that relaxes the keyword-availability requirement could pinpoint the breakthrough years of the 20 topics that were already established when our corpus starts. The taxonomy framework of Fang et al. is independent of any particular set of publications, so broadening the journal set would let the pipeline run on neighboring biomedical specialties, such as cardiology or neuroscience journals. Porting it to a distant discipline requires more work, because the keyword-categorization prompts and the embedding model are biomedical and would need re-validation. Two further directions address the provenance limitation directly: generating keywords for every article with a language model would put all journals and years on a single vocabulary, at the cost of departing from author-declared terms, and selecting articles individually by their fit to the target field, instead of including every article of a chosen journal set, would sharpen field boundaries, at the cost of a selection model that itself needs validation.

In summary, this study contributes a scalable framework for tracking how research trends evolve across an entire field. It combines a multi-pass pipeline that gathers publication keywords even where PubMed coverage is thin, the taxonomy construction pipeline of Fang et al. applied at field scale, and a set of analyses that read trends within and across journals. Applied to 78,073 articles from 29 journals covering biomedical informatics over 2011–2025, it showed that some single-journal findings hold at field scale (growth tied to milestone events, broad expansion of methods research), while others, such as the shape of the journal ecosystem and the sensitivity of trend estimates to keyword provenance, become visible only when many journals are analyzed together. The field grew 7.6-fold in three overlapping waves; digital health technologies became its connective backbone; the journal ecosystem is well integrated, with weakly separated communities bridged by JBI; the COVID-19 pandemic started trends rather than ending them; and the fastest-gaining topics are recent machine-learning methods, whereas apparent losses proved sensitive to keyword provenance. Code and derived data are available in the project repository (see Data availability), and the framework can be applied to any biomedical discipline with a definable set of

journals; porting it to more distant fields requires re-validating the biomedical embedding model and categorization prompts.

Methods

Study overview

TRACE runs as three modules (Figure 1). Module 1 is the multi-pass corpus and keyword acquisition pipeline. Module 2 is the taxonomy construction pipeline, which turns the harvested keywords into named methodology and health topics. Module 3 analyzes the topic-tagged corpus with trend and event analyses, cross-journal analyses, authorship analyses, and citation analyses. Modules 1 and 3 were developed in this study. Module 2 was introduced in Fang et al.¹¹ and is used here with one modification: the topic-naming model was changed from GPT-4o²¹ to GPT-5-nano²².

Corpus construction

We analyzed 29 English-language, PubMed-indexed journals selected by the study team to span the four communities of biomedical informatics: clinical informatics, digital health, computational biology, and bridging methods (Supplementary Table 1). The set covers the field's flagship, specialty, and niche venues, from high-impact journals such as *The Lancet Digital Health*, *npj Digital Medicine*, and *JAMIA* to smaller community journals. We required PubMed indexing so that bibliographic metadata were consistent across venues; Supplementary Figure 1 shows each journal's indexed span within the window. The set is curated rather than exhaustive (see the limitations in the Discussion). We chose a 15-year window (2011–2025) to ensure consistent keyword availability, reliable metadata after JMIR's 2010 adoption of XML, and alignment with the pretraining corpus of BiomedBERT²³. Articles indexed for 2026 were excluded because too few had been indexed at the time of collection (April 2026).

We built a multi-pass keyword acquisition pipeline because author keywords for many of the 29 journals were not available through PubMed. Pass 1 uses the PubMed API^{24,25} to retrieve primary metadata, including PMIDs, abstracts, and existing MeSH descriptors. Pass 2 uses a scraper keyed on each record's digital object identifier (DOI) to harvest keywords directly from publisher platforms (IEEE Xplore, Nature, Springer, and Elsevier) where PubMed coverage was incomplete. For records still lacking author keywords after passes 1 and 2, pass 3 substituted existing MeSH descriptors where available (n=18,409; 23.5%), a pattern concentrated in bioinformatics-focused journals. The MeSH check tags (15 population tags such as Humans, Female, and Middle Aged, and 9 organism tags such as Mice) and their inbred-strain qualifiers were excluded from topic assignment because they describe the study population rather than the research content (20,979 occurrences on 11,667 records). Pass 4 applied GPT-5-nano zero-shot

keyword generation from abstracts as the last resort for records with neither author keywords nor MeSH descriptors ($n=857$; 1.1%). These records make up essentially the entire Bioinformatics Advances corpus (829 of its 830 articles), because that journal had no MeSH indexing at collection time. Throughout, PMID-based deduplication and per-journal state persistence protected corpus integrity against connection interruptions. The pipeline assembled 78,073 articles published in 2011–2025 across the 29 journals, carrying 95,872 distinct publication keywords. Keyword provenance also varies over time: MeSH substitution supplied the keywords for 75% of 2011 and 83% of 2012 records but for 10% of 2025 records, because author keywords became available in PubMed for most clinical-informatics journals only from 2013 (see the limitations in the Discussion).

Taxonomy construction and topic assignment

To organize the corpus by theme, we applied the published pipeline of Fang et al.¹¹ to our 29-journal keyword set. We ran it with the cluster count set to $K=100$ per domain, producing keyword clusters²⁶ that its coherence-checking, sub-cluster-splitting, and near-duplicate-merging steps consolidate into 70 methodology and 86 health named topics (Supplementary Table 2). These named topics are the unit of analysis for all trend results; Supplementary Note S1 reports a sensitivity analysis on K . Articles are linked to topics through shared keywords, so a single paper can map to several health and methodology topics. These memberships populate three matrices: yearly article counts per topic, article counts by journal and year, and yearly topic share per journal. The number of topics per article, overall and by journal and keyword source, is reported in Supplementary Table 3.

Linear-trend testing and multiple-test correction

For every topic with at least 10 articles, we regressed annual article counts on year and recorded the slope, R^2 , and two-sided p -value against a null hypothesis of zero slope, together with the relative slope (slope divided by the mean). Because hundreds of hypotheses are tested in parallel, we applied the Holm-Bonferroni step-down procedure^{27,28} to keep the family-wise error rate at $\alpha = 0.05$ within each domain.

Because field-wide output rose over the study period, we also fitted for each topic an overdispersion-robust (quasi-Poisson) generalized linear model of annual counts with an offset for the log of that year's corpus size. The time coefficient estimates the trend in the topic's share of total output, again Holm-corrected within domain, and separates compositional change from field-wide growth. We fitted the same share model within each journal, for topics with at least 30 articles in that journal and with the journal's own yearly totals as the offset, to describe journal-specific trends (Supplementary Table 9).

Poisson rate test for event-associated shifts

To test whether an event coincides with a shift in publication output, we compare article counts in an event window and a post-event window against the rate in a short pre-event baseline, using a Poisson rate test (chi-square approximation) with Holm-Bonferroni correction across topics. Topics enter the test if they have at least 100 articles over 2011–2025 and a pre-event mean of at least 3 articles per year. For the COVID-19 case study, the baseline is 2018–2019, the event window 2020–2021, and the post-event window 2022–2025; the baseline is limited to the two preceding years because field-wide output was still expanding, so that it reflects the field's contemporary volume. Results are shown as a volcano plot of percentage change against $-\log_{10}$ adjusted p, and event-window growth is compared with post-event growth to separate transient shifts from durable ones. For every topic we also record the year of the largest year-over-year increase over 2012–2025 (Supplementary Table 4); Supplementary Figure 11 shows these as heat maps for the 20 most frequent topics per domain.

Topic breakthrough dating, waves, and share trends

To date each topic's emergence, we identified for every topic with more than 100 articles its breakthrough year, the earliest year in which its annual count first reached one quarter of its peak annual count; topics already above that bar in 2011 are left-censored. A wave is a group of related methodology topics whose breakthroughs fall within three years of each other. Its onset is the breakthrough year of the topic named for the wave's defining method, and its signature topics were fixed before the dating: machine learning methods with biomedical signal processing, medical image segmentation, and mixed deep learning applications; and large language models with transformer models, diffusion generative models, and generative-AI methods. The foundational wave comprises the methodology topics that were already above the threshold in 2011. We separately estimated each topic's share trend, the per-year change in its fraction of total corpus output from the quasi-Poisson model above, to order topics from fastest-rising to fastest-declining.

Keyword turnover within topics

To examine turnover inside individual topics, we decomposed each topic into its most frequent keywords and plotted each keyword's share of the topic year by year as a stacked-area chart. The decomposition is descriptive, with no statistical test applied to the keyword shares, and complements the corrected trend tests.

Milestone annotation

As in Fang et al., we relate topic trends to dated milestone events by overlaying the annual counts of selected topics on the event dates. The study team compiled the milestone list before looking at the curves, from three kinds of events: widely recognized technical releases (for example AlexNet, the transformer architecture, BERT, GPT-3, ChatGPT, and GPT-4), policy and regulatory changes (the FDA warning letter to 23andMe, the NIH Genomic Data Sharing Policy, the GDPR, the All of Us launch, and the 2024 AI bias regulations), and public-health events (the 2014 Ebola epidemic, the WHO pandemic declaration, the mRNA vaccine authorizations, and the end of the COVID-19 emergency). An event qualified if it had a public, datable source and a plausible link to the topic; Supplementary Table 5 lists every event with its date and source. We chose the four topics because each has such an event series and, for the three methodology topics, a counterpart in the single-journal analysis. The overlay is descriptive: it shows whether output changed at the same time as an event, and we do not test or claim that the event caused the change.

Named datasets and shared resources

To track the datasets and shared resources that articles build on, we counted case-insensitive mentions in titles and abstracts of 33 named resources curated by the study team: public clinical and biomedical datasets (for example, MIMIC, UK Biobank, TCGA, All of Us, ADNI), shared-task corpora (i2b2/n2c2), knowledge bases (KEGG, Gene Ontology, UniProt, DrugBank), imaging benchmarks (ImageNet, BraTS, ISIC, CheXpert), and data sources such as PubMed, ClinicalTrials.gov, Twitter, and Reddit. We report mention rates per 1,000 articles in each five-year window for resources named by at least 20 articles. Pattern matching undercounts resources that are used but not named in the abstract.

Cross-journal analyses

We characterize the journal ecosystem with six analyses. First, we profiled each journal's publication trajectory. From the per-journal counts of articles published each year, we sorted journals into two criterion-defined groups, high-volume veterans (full-window coverage and more than 4,000 articles) and high-growth newcomers (first indexed after 2011 and more than a threefold increase in annual output by

2025), and a residual group containing every other journal. The criteria were defined empirically for the 29-journal set; applying the framework to a broader sample may need additional profiles, such as declining or non-monotonic trajectories. Supplementary Table 8 lists objective descriptors for every journal (years indexed, growth factor, peak year, and year-to-year volatility) so that readers can apply alternative labels.

Second, we quantified each journal's relative specialization. For every journal-topic pair we computed an enrichment ratio: the fraction of that journal's articles in the topic divided by the field average for that topic, defined as the unweighted mean of the 29 journals' fractions so that each journal counts equally. Values are displayed on a base-2 logarithmic scale. Enrichment identifies the topics a journal publishes disproportionately often relative to the field, not necessarily the topics that dominate its output; a journal can be strongly enriched for a topic that accounts for a small share of its articles, and the journal-fit heat maps (sixth analysis below) report the absolute shares. Because journals differ in the mean number of keywords, and hence topic assignments, per article (Supplementary Table 3), enrichment values are also sensitive to keyword density (Supplementary Note S2).

Third, cross-journal universality measures how evenly a topic is spread across journals. For each topic we built a vector of its article counts over the 29 journals and computed the Shannon entropy of that vector normalized by $\log(29)$, which is 0 when a topic resides in a single journal and 1 under perfect spread. We labeled topics scoring at least 0.85 as universal and those at or below 0.55 as niche, restricting to topics with at least 80 articles.

Fourth, we mapped how journals relate to one another. We built a journal-similarity graph with journals as nodes and an edge between two journals when the cosine similarity between their topic-distribution vectors exceeded the median of all pairwise similarities. Communities were identified with Louvain detection²⁹, and bridge journals with NetworkX betweenness centrality³⁰, **using $1 - \text{cosine similarity}$ as the edge distance. For visualization, we positioned the nodes with the Kamada-Kawai algorithm³¹**, with node size reflecting article volume and color marking the Louvain communities.

Fifth, to identify pioneer journals, we located for every topic the article with the earliest publication year carrying that topic and recorded its journal, marking where the topic first surfaced within the corpus. Each journal was then scored by the fraction of topics on which it was first to publish.

Sixth, to describe each journal's topical focus and the submission landscape, we built two artifacts: a journal-by-topic heat map whose cells give the share of a journal's articles on each topic, and a scatter

plot placing every topic by the number of journals that have published it against its cumulative article count.

Authorship and affiliation analyses

First, to follow authorship and affiliation dynamics, we aggregated the per-article author count and the count of distinct affiliations (affiliation strings as recorded in PubMed, so two departments of one university count separately) and report the mean and median of each across all papers published in a year. We labeled each paper's affiliations as clinical, computational, both, or neither/unknown using case-insensitive regular-expression matching against two manually curated pattern lists (Supplementary Table 14). The clinical list matches care and medicine institutions and departments (for example, "school of medicine," "hospital," "medical center"); the computational list matches quantitative and engineering units (for example, "computer science," "computational," "informatics"). Papers whose affiliations match neither list, or that carry no affiliation metadata, form the neither/unknown category. We report the yearly and per-journal interdisciplinary share.

Second, treating authors per paper as the team size, we trace its yearly and per-journal mean and median, together with the 90th and 95th percentiles. We report the mean and median together deliberately: consortium-scale outlier papers inflate the mean relative to the median, so divergence between the two curves signals the growing role of very large teams, and the upper percentiles show how far that tail extends.

Citation-impact analyses

To relate research attention to citation impact, we retrieved citation counts and the relative citation ratio (RCR) of every article from the NIH iCite database (accessed September 2026)^{32,33}. The RCR is a field- and time-normalized citation rate benchmarked so that 1.0 is the rate expected for NIH-funded articles of the same age and field. We restricted RCR analyses to articles published in 2011–2023 so that each has at least two years of citations, and report for every topic with at least 30 such articles the median RCR and the share of articles with RCR of at least 2, together with the highest-RCR articles overall and within each of the 20 largest topics per domain; we relate median RCR to the share trends. We also built a citation-weighted variant of the pioneer-journal analysis: for every topic, the journal whose articles published in the topic's first three corpus years accumulated the most citations. Finally, from the iCite reference lists we counted within-corpus citations (references from one corpus article to another), aggregated them into a journal-to-journal citation matrix on which we computed betweenness centrality, and ranked articles by the number of distinct $K=100$ topics whose articles cite them.

Use of artificial intelligence

Large language models were used as computational instruments within the analysis pipeline that this study evaluates. GPT-5-nano performed few-shot classification of keywords into methodology, health, both, or neither; zero-shot keyword generation from abstracts for the 1.1% of records lacking keywords; and generation of human-readable topic names. BiomedBERT provided keyword embeddings. Claude was used to help write and review code for visualizations and pipelines. Separately, Claude and Gemini assisted with language editing and drafting during manuscript preparation. The authors reviewed and verified all AI-assisted outputs and take full responsibility for the content, accuracy, and integrity of this work.

Data availability

The derived data supporting the findings of this study are available in the project repository at <https://github.com/DanielLiangAjj/multi-journal-trend-analysis>. The repository contains the analysis tables underlying every figure and statistic reported here, including per-topic annual counts, linear-trend and share-trend statistics, journal-by-topic specialization matrices, cross-journal entropy rankings, and the journal-similarity network. The bibliographic records themselves were retrieved from PubMed through the NCBI E-utilities API and, where PubMed coverage was incomplete, from publisher platforms. PubMed metadata are freely available from the U.S. National Library of Medicine, and the PMIDs needed to reconstruct the corpus are included in the repository. Author keywords harvested from publisher platforms are redistributed only in aggregated, derived form, in accordance with publisher terms.

Code availability

All analysis code is available at <https://github.com/DanielLiangAjj/multi-journal-trend-analysis> under the MIT license and will be archived on Zenodo with a DOI upon acceptance. The repository includes the multi-pass keyword acquisition pipeline, the BiomedBERT embedding and K-means clustering steps, the GPT-assisted keyword categorization and topic-naming scripts, all statistical testing (linear trend regression, the quasi-Poisson share model, the Poisson rate test, and Holm-Bonferroni correction), and the figure build scripts that generate every figure in the main text and supplementary material. Analyses were run in Python using pandas, NumPy, scikit-learn, statsmodels, NetworkX, and Matplotlib.

References

1. Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent Dirichlet Allocation.
<http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>.
2. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. & Lim, W. M. How to conduct a bibliometric analysis: An overview and guidelines. *J. Bus. Res.* **133**, 285–296 (2021).
3. Zupic, I. & Čater, T. Bibliometric methods in management and organization. *Organ. Res. Methods* **18**, 429–472 (2015).
4. Aria, M. & Cuccurullo, C. bibliometrix : An R-tool for comprehensive science mapping analysis. *J. Informetr.* **11**, 959–975 (2017).
5. van Eck, N. J. & Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* **84**, 523–538 (2010).
6. Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E. & Herrera, F. SciMAT: A new science mapping analysis software tool. *J. Am. Soc. Inf. Sci. Technol.* **63**, 1609–1630 (2012).
7. Hirsch, J. E. An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 16569–16572 (2005).
8. Leydesdorff, L. Words and co-words as indicators of intellectual organization. *Res. Policy* **18**, 209–223 (1989).
9. He, Q. Knowledge discovery through co-word analysis. *Libr. Trends* **48**, 133–159 (1999).
10. Lipscomb, C. E. Medical subject headings (MeSH). *Bull. Med. Libr. Assoc.* **88**, 265–266 (2000).
11. Fang, Y. *et al.* A data-driven method for research trend analysis in a scientific discipline: Application to the journal of biomedical informatics. *J. Biomed. Inform.* **178**, 105013 (2026).
12. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90 (2017).

13. Vaswani, A. *et al.* Attention is all you need. *arXiv [cs.CL]* (2023)
doi:10.48550/arXiv.1706.03762.
14. Introducing ChatGPT. *OpenAI* <https://openai.com/index/chatgpt/>.
15. All of Us Research Program Investigators *et al.* The “All of Us” Research Program. *N. Engl. J. Med.* **381**, 668–676 (2019).
16. Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic. [https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-coronavirus-disease-\(covid-19\)-pandemic](https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic).
17. Shannon, C. E. A mathematical theory of communication.
<https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf>.
18. Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8577–8582 (2006).
19. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **69**, 026113 (2004).
20. Beugelsdijk, S. & Bird, A. How to avoid a desk reject: do’s and don’ts. *J. Int. Bus. Stud.* **56**, 301–310 (2025).
21. Hello GPT-4o. *OpenAI* <https://openai.com/index/hello-gpt-4o/>.
22. Introducing GPT-5. *OpenAI* <https://openai.com/index/introducing-gpt-5/>.
23. Gu, Y. *et al.* Domain-specific language model pretraining for biomedical natural language processing. *arXiv [cs.CL]* (2020) doi:10.48550/arXiv.2007.15779.
24. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **52**, D33–D43 (2024).
25. Sayers, E. *The E-Utilities in-Depth: Parameters, Syntax and More.* (National Center for Biotechnology Information, 2026).

26. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129–137 (1982).
27. Aickin, M. & Gensler, H. Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. *Am. J. Public Health* **86**, 726–728 (1996).
28. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. <https://www.ime.usp.br/~abe/lista/pdf4R8xPVzCnX.pdf> (1979).
29. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
30. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. in *Proceedings of the Python in Science Conference* 11–15 (SciPy, 2008).
31. Kamada, T. & Kawai, S. An algorithm for drawing general undirected graphs. *Inf. Process. Lett.* **31**, 7–15 (1989).
32. Hutchins, B. I. *et al.* The NIH Open Citation Collection: A public access, broad coverage resource. *PLoS Biol.* **17**, e3000385 (2019).
33. Hutchins, B. I., Yuan, X., Anderson, J. M. & Santangelo, G. M. Relative Citation Ratio (RCR): A new metric that uses citation rates to measure influence at the article level. *PLoS Biol.* **14**, e1002541 (2016).

Acknowledgements

This work was supported by the National Library of Medicine (NLM) under grant R01LM014344.

Author contributions

Y.L.: Conceptualization, Methodology, Software, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing.

G.Z.: Conceptualization, Methodology, Writing – review & editing.

C.T.: Methodology, Writing - review & editing

Y.F.: Methodology, Writing – review & editing.

C.W.: Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Competing interests

The authors declare no competing interests.

Supplementary Files

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

- [supplementarytables.csv.zip](#)
- [SupplementaryInformation.pdf](#)