SUPPLEMENTARY MATERIALS

This section provides additional methodological details regarding the data pre-processing, normalization procedures, and network construction parameters used in the bibliometric analysis.

Supplementary Method 1. Institutional standardization protocol

Institutional names were standardized using a structured manual consolidation procedure based on the exact matching of raw affiliation strings extracted from the Web of Science records (C1 field). To minimize fragmentation arising from heterogeneous address formats, candidate variants were identified from raw data and grouped after manual confirmation of institutional identity. Variations in naming conventions, including abbreviations, language differences, and hospital–university separations, were reviewed and harmonized.

University hospitals and medical schools were merged with their parent universities when they were administratively integrated, whereas major affiliated hospitals that consistently appeared as independent institutional addresses were retained as separate entities. Sub-institutes under umbrella organizations (e.g., The National Institutes of Health) were consolidated at the parent institutional level to ensure consistency in institutional attribution. No substring-based or fuzzy matching was applied, and institutional counts were based on the final merged entities used in the network analysis.

Supplementary Method 2. Author name standardization and aggregation risk assessment

Author names were standardized using a surname–initial format. Names were converted to uppercase, and punctuation marks and redundant spaces were removed to ensure consistent formatting. To evaluate the potential aggregation of distinct individuals sharing identical surname–initial combinations, institutional diversity metrics were examined. Authors exhibiting high institutional dispersion were considered at increased risk for aggregation errors. However, complete author disambiguation was not feasible because of database constraints inherent to bibliographic records.

Supplementary Method 3. Keyword normalization dictionary

To ensure semantic consistency across publications spanning three decades, author-provided keywords (DE fields) were standardized prior to network analysis using a manually curated normalization framework. Orthographic variations (e.g., British versus American spelling), plural-singular forms, and synonymous expressions were systematically harmonized. In addition, fragmented keyword tokens arising from comma-separated expressions in the original database records (e.g., “anemia, pernicious”) were reconstructed into clinically meaningful terms (e.g., “pernicious anemia”). The normalization process was conducted through a structured two-step procedure: 1) consolidation of semantically equivalent terms based on domain knowledge and 2) reconstruction of fragmented keyword tokens to preserve clinical interpretability. Nonspecific descriptors, methodological terms, and overly general biomedical expressions were excluded to enhance the thematic specificity of the network analysis. Keywords that did not require normalization and that occurred infrequently were omitted from the dictionary for clarity and parsimony.

Supplementary Method 4. Network analysis parameters

Country- and, institution-level collaboration networks were constructed and visualized using Gephi (version 0.10.1; https://gephi.org/). Author-level co-authorship networks were constructed using VOSviewer (version 1.6.20; https://www.vosviewer.com/). While bibliometric productivity indicators were calculated using a full-counting approach, fractional counting was applied to co-authorship network construction to reduce inflation effects associated with multi-author publications and large collaborative groups. Minimum publication thresholds were applied separately at each analytical level to improve the network interpretability. For the author-level collaboration analysis, only authors with at least six publications were included in the final network.

Keyword co-occurrence networks were generated using VOSviewer with association strength normalization. To focus on the representative thematic structures and reduce network fragmentation, only keywords with a minimum occurrence frequency of 30 were included in the analysis. Clustering was performed using the VOS clustering algorithm.

Reference co-citation analyses were conducted using CiteSpace (version 7.0.0; https://citespace.podia.com/). Citation bursts were identified using Kleinberg's burst detection algorithm with γ=0.4, two burst states, and a minimum burst duration of 2 years.

Supplementary Method 5. Visualization and figure construction parameters

Network visualizations were generated using VOSviewer and CiteSpace to ensure consistency and interpretability across analyses. In VOSviewer, network maps were constructed using the association strength normalization method. Node size reflected the frequency or relative weight of items, and edge thickness represented the strength of relationships (e.g., co-authorship or co-occurrence). Clustering was performed using the VOS clustering algorithm with a resolution parameter of 1.0 and a minimum cluster size of 1. Colors were used to identify distinct thematic clusters.

For the keyword co-occurrence analysis, only keywords that occurred at least 30 times were included in the network. The final keyword network comprised 57 items, 11 clusters, and 612 links.

In CiteSpace, the analyses were performed using time slicing from 1995 to 2025, with one year per slice. The keyword timeline analysis was conducted using the g-index (k=25), whereas the reference co-citation analysis used the g-index (k=12). The additional CiteSpace parameters were link retaining factor=2.5, links per node=10, and look back years=5. Pathfinder pruning and sliced network pruning were applied to simplify the network structure. Cluster labeling was performed using the log-likelihood ratio algorithm, and timeline visualizations were used to illustrate the temporal evolution of the major research clusters.

All visual outputs were reviewed and optimized for clarity, including label visibility and layout adjustment, without altering the underlying network structure.