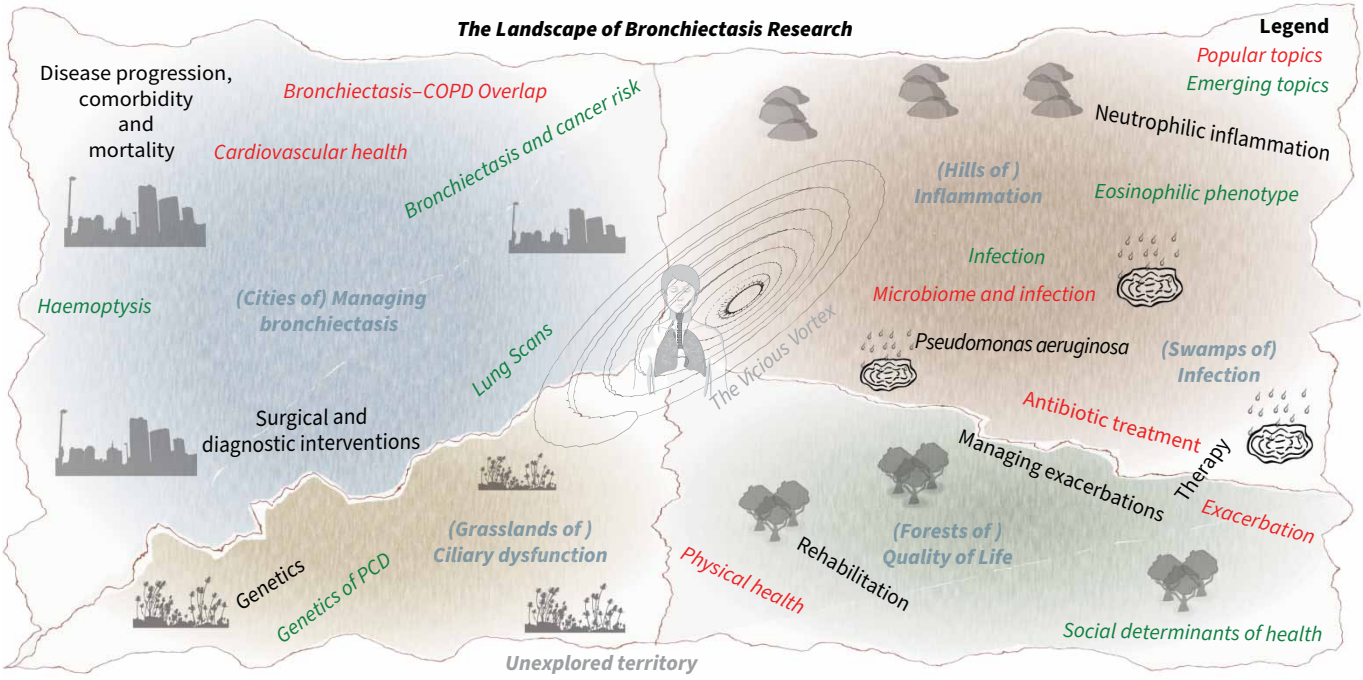
Characterising research trends in bronchiectasis through AI-powered analytics

Jayanth Kumar Narayana, Yolanda Koo Wei Ling, Micheál Mac Aogáin and Sanjay H. Chotirmall



GRAPHICAL ABSTRACT Artificial intelligence (AI)-driven topic modelling of the bronchiectasis research landscape. A summary of the current landscape of bronchiectasis research using AI. The scientific literature spans four realms: inflammation and infection, managing bronchiectasis, quality of life, and ciliary dysfunction. Each realm embeds islands of broad categories (or research themes) denoted in black text. The traditional vicious vortex framework, comprising airway dysfunction, infection, inflammation and structural disease, largely spans over two realms (infection and inflammation and primary ciliary dyskinesia (PCD)) with some overlap with surgical and diagnostic interventions. This reflects trends of bronchiectasis research evolving from the classical disease-centric approach to a patient-centric approach, including studies focused on quality of life and management of bronchiectasis. Popular topics are denoted in red, and emerging topics are denoted in green.

https://doi.org/10.1183/13993003.00894-2025 Eur Respir J 2025; 66: 2500894

Characterising research trends in bronchiectasis through AI-powered analytics

Jayanth Kumar Narayana [1], Yolanda Koo Wei Ling[1], Micheál Mac Aogáin[2,3] and Sanjay H. Chotirmall [1,4]

[1]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. [2]Department of Biochemistry, St. James's Hospital, Dublin, Ireland. [3]Clinical Biochemistry Unit, School of Medicine, Trinity College Dublin, Dublin, Ireland. [4]Department of Respiratory and Critical Care Medicine, Tan Tock Seng Hospital, Singapore, Singapore.

Corresponding author: Sanjay H. Chotirmall (schotirmall@ntu.edu.sg)

Shareable abstract (@ERSpublications)
AI-based analytics of bronchiectasis research highlights evolving trends and emerging priorities, and reveals an expanding research landscape that is shifting from disease to patient centricity, which is in early development compared to asthma and COPD research https://bit.ly/3J82wUc

Cite this article as: Narayana JK, Koo Wei Ling Y, Mac Aogáin M, *et al*. Characterising research trends in bronchiectasis through AI-powered analytics. *Eur Respir J* 2025; 66: 2500894 [DOI: 10.1183/13993003.00894-2025].

This article has an editorial commentary: https://doi.org/10.1183/13993003.01687-2025

Received: 27 April 2025
Accepted: 31 July 2025

Abstract

Background Interest in bronchiectasis is increasing and no prior study has used artificial intelligence (AI) to interrogate its rich, multidimensional literature to characterise research trends, themes and knowledge gaps.

Methods We reviewed original bronchiectasis research between 1949 and 2024 (a 75-year period) to identify, characterise and assess research trends and trajectories using two AI-powered approaches: 1) Atlas, an AI topic-modelling tool; and 2) a custom model, leveraging ChatGPT embedding and text generation models.

Results AI-powered analytics revealed a nine-fold increase in bronchiectasis research speed since 2000, typified by enhanced richness with four new research topics emerging every 5 years. Publication trends mirror clinical and technological advances, exemplified by significant rises in computed tomography, microbiome and clinical studies following the adoption of high-resolution computed tomography (1970s), next-generation sequencing (2005) and the first clinical guidelines (2008–2010), respectively. Topics with sustained growth (*i.e.* popular topics) include bronchiectasis–COPD overlap, microbiome infection, cardiovascular health and exacerbations. Those with sudden, short-term increased interest (*i.e.* trending topics) have focused on microbial pathogens and primary ciliary dyskinesia genetics. Mortality represents a nascent topic, demonstrating the highest year-on-year interest. Growth of research within the "vicious vortex" demonstrates thematic imbalance, with few studies overlapping with non-vortex components. Evolving research focus towards inflammation is evident, with increased work on comorbidities and quality of life demonstrating a shift from disease-centric to patient-centric research.

Conclusion AI captures bronchiectasis as a dynamic and interdisciplinary field in continuing growth. Emerging research topics extend beyond the vicious vortex framework, indicating a transition from disease-centric to patient-centric approaches to optimise clinical care.

Introduction

Our understanding of the pathogenesis of bronchiectasis, a chronic irreversible airways disease, has undergone a significant paradigm shift in recent times [1, 2]. This is exemplified by its "research renaissance", driven by interdisciplinary collaboration, international cooperation, advanced clinical registries, multiple clinical trials and the emergence of advanced technologies [2–5]. Once considered an orphan disease primarily due to infection, bronchiectasis is now recognised as a complex, heterogeneous condition exhibiting significant geographic heterogeneity [6–11]. Technological advances reveal distinct endophenotypes, challenging the traditional one-size-fits-all model of care with a move towards precision-based approaches [5, 10, 12–17]. This momentum underscores the emerging maturity of the bronchiectasis research field, with significant potential to offer insight into other overlapping chronic airways diseases [18, 19].

With such advances comes the challenge of assimilating increasing volumes of data, evidence and literature. Increasing awareness, strong global registries, improved clinical trial outcomes and evolving disease frameworks drive the surge in original investigations [3, 5, 20]. Our ability to organise, interpret and synthesise this literature is failing to keep pace with the volume and complexity of bronchiectasis research, challenging classical frameworks of literature review [21, 22]. As research becomes increasingly multidimensional and interdisciplinary, traditional methods of indexing, *e.g.* keyword tags and Medical Subject Heading (MeSH) terms, are proving insufficient [23]. These systems often rely on static classifications that struggle to capture dynamic trends or bridge disciplinary boundaries [23, 24].

Here, we applied large-language-model (LLM)-based topic modelling to >1900 research articles in bronchiectasis with the goal of unsupervised thematic mapping. By integrating state-of-the-art embedding models and clustering algorithms, we systematically organised existing literature into interpretable domains, assessed thematic trends over time and evaluated alignment with established disease frameworks [1, 20]. This represents the first structured application of artificial intelligence (AI)-based topic modelling in bronchiectasis, providing broadly applicable methods for other diseases and/or fields in which the relative unstructured knowledge base continues to expand. Our findings offer a reproducible, scalable roadmap to navigate the complex and varied biomedical literature in bronchiectasis, highlighting past trends, emerging areas and research gaps by applying an unbiased, data-driven approach to thematic discovery and trend analysis.

Materials and methods

Literature review

A search was conducted using the PubMed database (accessed on 24 April 2024) to identify all original research articles focused on non-cystic fibrosis bronchiectasis (NCFB). Search formulas applied in PubMed to exclude non-original research, including reviews, case reports, editorials, meta-analyses and other non-primary research articles, are shown in the supplementary methods. Following PubMed's automated filtering, titles and abstracts were screened based on predefined inclusion and exclusion criteria by two independent researchers, with discrepancies resolved through discussion, followed by a separate assessment of the full text. This study was limited to research published in the English language and focused on human subjects. Inclusion criteria were the availability of the article abstract and inclusion of patients with NCFB. Exclusion criteria included studies that were not original research articles, studies focused on cystic fibrosis, studies where NCFB was not the primary population and studies where the full text was unavailable. We employed the Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia) to facilitate title and abstract screening and full-text review. A full summary of the screening process is shown in figure 1a.

Topic modelling and analysis

The abstracts from the included studies (n=1942) were analysed using the "ChatGPT text generation model" that summarised key results to 50 words, eliminating biases from differing levels of details in the abstracts. Topic modelling, which grouped studies into differing topics (or clusters), was simultaneously performed using two models: 1) Nomic Atlas, an established topic-modelling software (Nomic Atlas Python Client v3.0.29, Nomic AI) [25]; and 2) a customised model using ChatGPT models (ChatGPT Text Embedding Model (text-embedding-3-large) and ChatGPT Text Generation Model (gpt-4o-2024-08-06), OpenAI). The Atlas API (python), with default parameters, was implemented on the study summaries to identify both broad- and narrow-level topics; for unclear topics, ChatGPT text generation models were employed to assign more descriptive names. The custom model used ChatGPT text embedding *via* OpenAI's API (python), embedding studies in a 3072-dimensional space, followed by dimensionality reduction using t-distributed stochastic neighbour embedding (t-SNE), and hierarchical clustering for grouping and topic identification (figure 1b). Assessment of cluster quality and stability revealed well-defined clusters and robust delineation. Clustered studies at various levels were labelled with topic names *via* the ChatGPT text generation model using a nested bootstrapping strategy. Statistical analysis of the labelled clusters (*i.e.* topics) was performed to assess topic evolution across the NCFB research landscape (figure 1d). To categorise studies within the "vicious vortex" framework, a ChatGPT text generation model was used, with a structured prompt-based approach that permitted overlapping classifications. Studies not fitting the vortex categories were classified as "others". Overlap analysis between vortex components was performed using Venn diagrams and UpSet plots. To explore topics within studies categorised as "others", a custom ChatGPT-based topic modelling workflow was applied to this subset (figure 1c).

Further details on comparative growth analysis, topic modelling, statistical analysis of topics, research categorisation within the vicious vortex framework, search terms and system prompts used are provided in the supplementary material.

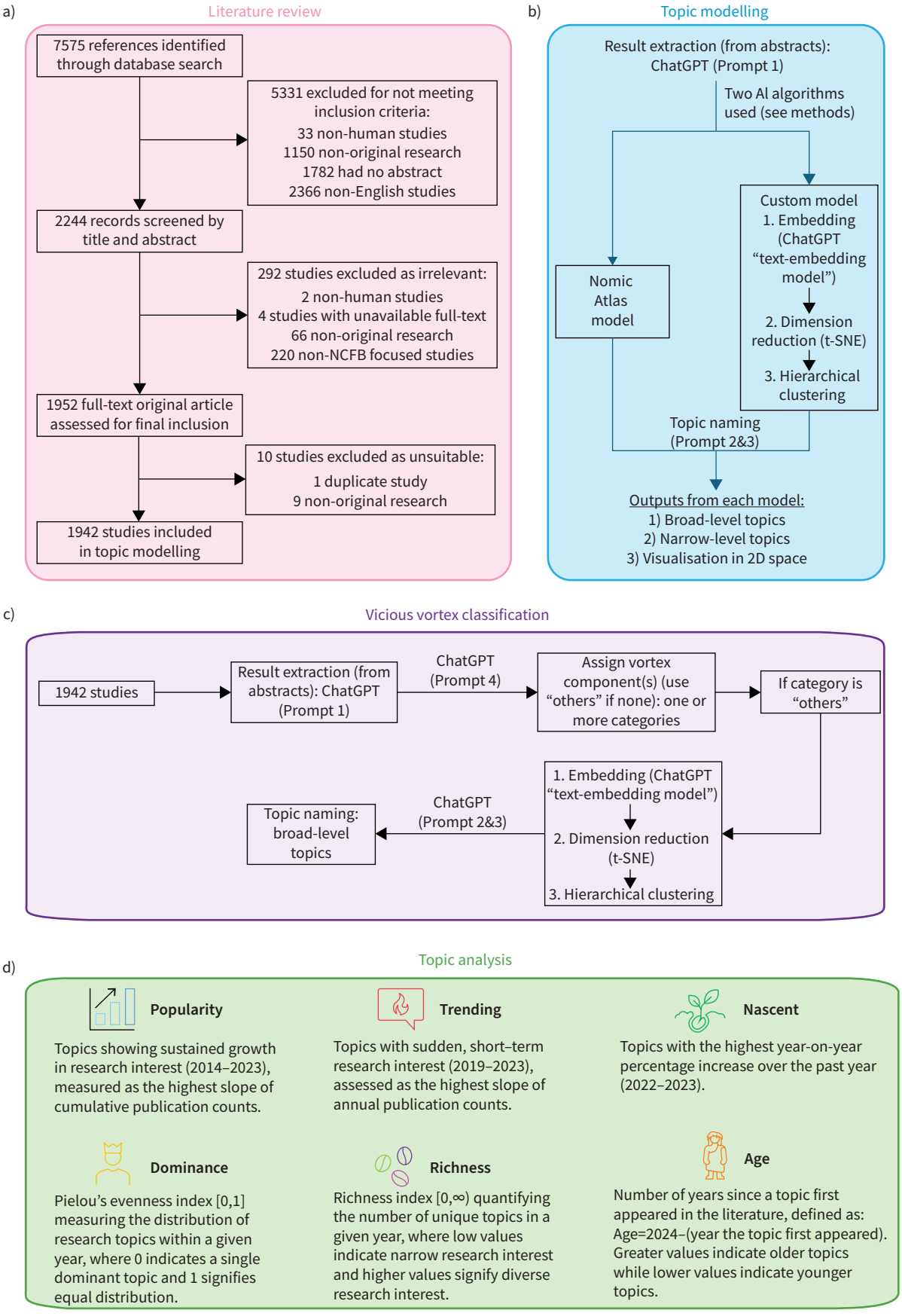


FIGURE 1 Overview of study methods including literature review, topic modelling, vicious vortex classification and topic analysis. a) A summary of the study's literature review, which included only original research involving human subjects, published in the English language and accompanied by an abstract. Records were excluded if they did not fulfil the above criteria, did not have full text available or if non-cystic fibrosis bronchiectasis (NCFB) was not the primary population of study. b) The results section from included abstracts was extracted using the ChatGPT artificial intelligence (AI) model prior to topic modelling, which grouped selected studies into their differing topics. Topic modelling was simultaneously performed using two models: Nomic Atlas, an established topic-modelling software, and a customised model involving AI-based text embedding followed by dimensionality reduction and clustering. c) Categorisation of selected studies using a ChatGPT text generation model within the vicious vortex framework. Studies were assigned to five categories, permitting overlapping classifications: "airway dysfunction", "infection", "inflammation", "structural disease" and "others" (for studies not fitting vortex categories). d) Statistical analysis of topics was performed to understand topic evolution across the NCFB research landscape. t-SNE: t-distributed stochastic neighbour embedding.

Results

The significant growth of bronchiectasis research over the past 25 years

To evaluate the evolution of NCFB research over the last three quarters of the century, our search strategy included an assessment of original research as detailed in the methods from March 1949 to 24 April 2024, yielding 1942 studies for assessment (figure 1a). The growth of NCFB research (1950 to April 2024) was evaluated using two complementary metrics: publication speed (linear slope of annual counts) and cumulative output expansion (exponential growth rate of total publications). Assessment of publication speed, defined as the annual increase in the number of new studies, revealed that, prior to 2000, publication counts increased at a rate of approximately one additional study every 2 years (slope=0.54, $p<0.001$) compared to approximately nine additional studies post-2000s (slope=4.54, $p<0.001$), reflecting an approximately nine-fold acceleration in publication speed (figure 2a) [1, 26–38]. Comparison of annual publication counts over time for bronchiectasis, asthma, COPD and life sciences research revealed that life sciences maintained the highest overall research output, followed by asthma, COPD and bronchiectasis, with COPD approaching asthma publication volumes in recent years (figure 2b). Interestingly, asthma research surged in the 1970s (with the introduction of inhaled corticosteroids), and COPD research in the early 2000s (with the introduction of Global Initiative for Chronic Obstructive Lung Disease guidelines). An equivalent pronounced rise has yet to occur in bronchiectasis, suggesting its future acceleration potential (figure 2b).

Growth analysis *via* exponential modelling of cumulative counts revealed that bronchiectasis research is closely following an exponential trajectory, whereas asthma outpaced the theoretical exponential growth rate in the 1970s but subsequently fell below exponential rates (supplementary figure S1b, c). To quantify bronchiectasis research growth dynamics, piecewise growth-curve analysis [39] was performed to objectively identify break-points (1973, 1985, 1996) in its exponential trajectory and compute growth rates (rate of cumulative expansion) across distinct time periods (figure 2c). An assessment of cumulative output expansion, reflecting the percentage increase in publications each year relative to the previous year's cumulative total, revealed four distinct phases. The highest growth rate (19.16%) occurred during 1973–1985 following the introduction of high-resolution computed tomography (HRCT) scanning [40–43]. The second phase (1985–1996) demonstrated 9.98% annual growth, partially driven by post-computed tomography (CT) developments and Cole's vicious-cycle hypothesis description [28]. The modern era (post-1996) [5] is largely driven by the development of patient registries (supplementary figure S1a), establishment of management guidelines and technological advancements (figure 2a), exhibiting 8.34% annual growth, 18.4% higher than the pre-CT era and 73.1% higher than a mature field such as life sciences research (figure 2c).

To probe beyond research growth, we next employed topic modelling to categorise similar studies into distinct clusters, which we refer to as "topics". This approach allowed the unsupervised identification of "focus areas" and an exploration of trends across the research landscape. To facilitate topic modelling, we used an AI model: the ChatGPT text generation model to extract key results from each individual study (see methods for details) [44]. Given that AI-based topic modelling techniques are actively being developed across many research fields [45], we chose to leverage two independent models: the Nomic Atlas, a proprietary software [25], and our own customised model involving text embedding, dimension reduction, clustering and subsequent descriptors of these clusters (figure 1b). By leveraging more than a single approach, we aimed to increase confidence in the results by validating common trends and patterns across models. The custom model's clustering approach was fine-tuned to align with the Atlas model results, with the aim to allow equitable comparisons and accurately capture topics across the bronchiectasis research landscape (supplementary figure S2c).

Multi-level topic modelling highlights the shifting focus of bronchiectasis research from disease-centric to patient-centric approaches

Our Atlas-based analysis uncovered eight broad-level topics, while the customised model revealed seven categories with notable overlap between approaches (figure 3a, b). Overlapping topics include "neutrophilic

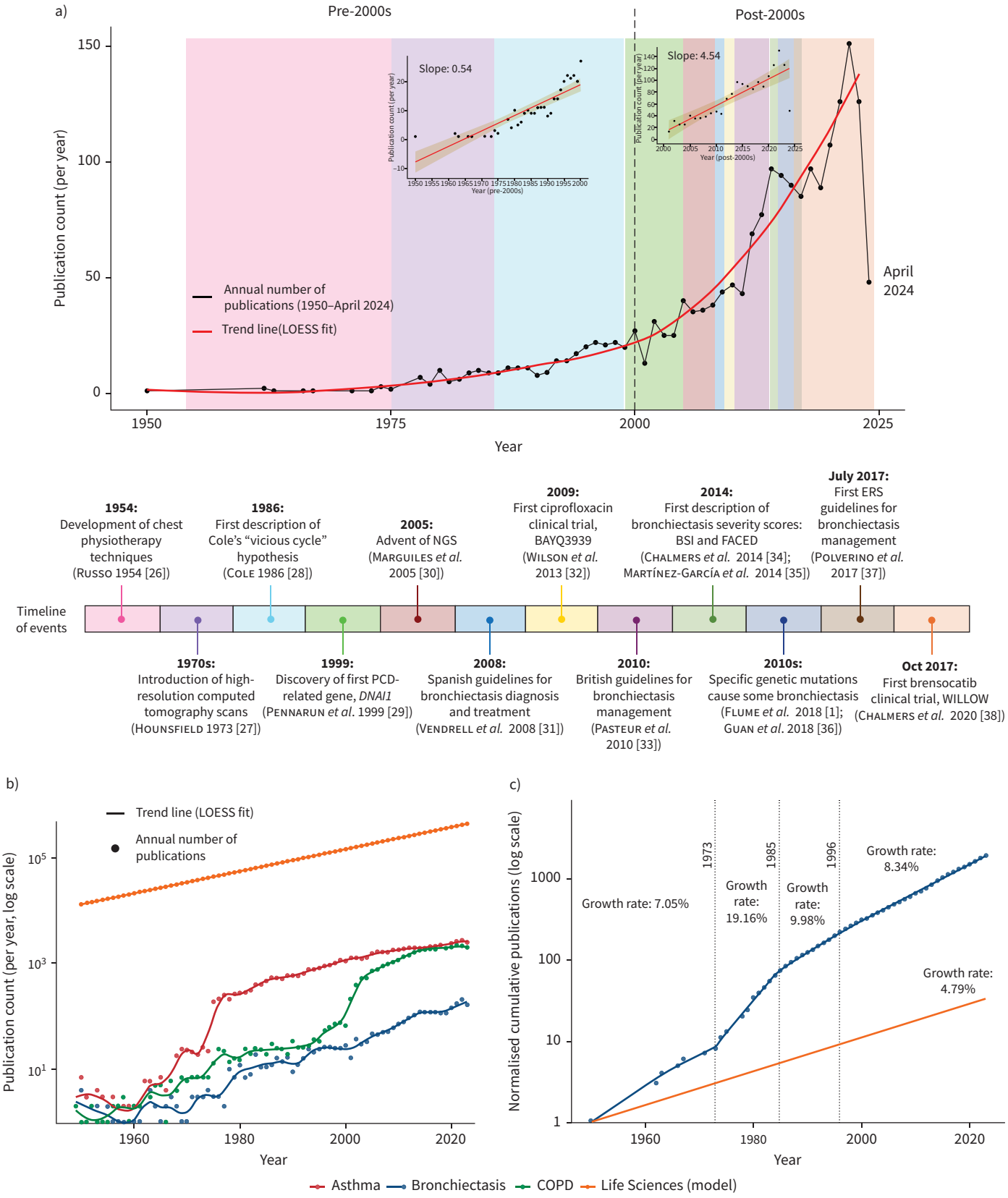


FIGURE 2 Overview of non-cystic fibrosis bronchiectasis (NCFB) research growth across key historic events between January 1950 and April 2024. a) Line plots illustrating the annual count of original research articles on the subject NCFB (y-axis) from the year 1950 through to April 2024 (x-axis). Red line indicates the trend of increasing publications (LOESS-fit line) over time with the timeline overlaid with background colours

denoting human-curated key historical events in the field of bronchiectasis (detailed below graph). Inset plots show regression analysis illustrating the research speed: linear-growth rate (indicated as red lines) of new NCFB publications by timeframe (left) pre-2000s and (right) post-2000s. b) Annual publication counts (dots) for bronchiectasis (blue), COPD (green) and asthma (red) with LOESS trend lines (lines), compared with simulated Life Sciences counts (orange). The y-axis displays publication counts (log-scale); the x-axis shows years. c) Piecewise exponential modelling of normalised cumulative bronchiectasis publications (blue) compared against simulated Life Sciences growth (orange) [39]. The y-axis shows normalised cumulative publication counts (log-scale); the x-axis shows years. Segmented regression analysis identified three break-points (1973, 1985, 1996; dotted lines) defining four growth periods (exponential growth rate %): 1950–1973 (7.05%), 1973–1985 (19.16%), 1985–1996 (9.98%) and 1996–present (8.34%), compared to Life Sciences overall exponential growth rate of 4.79%. PCD: primary ciliary dyskinesia; DNAI1: dynein axonemal intermediate chain 1; NGS: next-generation sequencing; BSI: Bronchiectasis Severity Index; ERS: European Respiratory Society.

inflammation", "rehabilitation", "genetic ciliary dysfunction or PCD [primary ciliary dyskinesia]", "surgical intervention (or lung transplantation outcomes)" and "therapy or treatment". Additionally, our customised model identified important distinct topics not captured by the Atlas-based approach, including "*Pseudomonas aeruginosa*" and "disease progression, comorbidities and mortality", while Atlas analysis yielded the additional broad topic "bronchiectasis: management and diagnosis". Assessment of the temporal evolution of these broad-level topics reveals that, in the Atlas-based model, "bronchiectasis: management and diagnosis" demonstrated a pronounced recent surge in publications, followed by "neutrophilic inflammation" and "therapy". In contrast, the custom model highlighted "disease progression, comorbidities and mortality" followed by "*Pseudomonas aeruginosa*" and "management and treatment: exacerbation" (supplementary figure S2a, b).

Unlike the Atlas model, which permits assessment at only two levels (*i.e.* broad and narrow), our customised approach enabled topic analysis at varying levels of granularity, akin to examining the same scene through differently powered lenses. At higher levels of abstraction and going beyond broad topics, we stratified the bronchiectasis research landscape into four central themes: 1) primary ciliary dysfunction; 2) understanding bronchiectasis: impact of risk factors on progression and mortality including the role of surgical intervention in patient outcomes; 3) infection and inflammation influencing disease severity; and 4) enhancing quality of life through improved treatment and management practices (figure 3c). When assessed at the narrow topic level, Atlas uncovered 64 topics, informing the granularity setting for our customised model.

To examine shifts in focus areas of bronchiectasis research over time, we next assessed the relative abundance of narrow topics from both models pre- and post-2000. We found that key research areas pre-2000 focused on CT scans (the highest studied topic), alongside ciliary dysfunction, surgical interventions, neutrophilic inflammation, antibiotic therapies, association with rheumatoid arthritis and genetic factors (figure 3d, f). In contrast, post-2000 key focus areas (common to both methods) included infections, mortality, aetiology and management (figure 3e, g). While the top 15 topics were largely consistent between methods pre-2000s, this was less prominent post-2000, likely due to an increase in topic diversity in the recent era, resulting in a reduced relative abundance of the individual topics.

Temporal research trends reflect key clinical, research and technological advances in bronchiectasis

We next assessed temporal trends in bronchiectasis (AI-identified) against the backdrop of key human-curated historical events. This ranged from early work focused on physiotherapy and imaging techniques to more recent work on evidence-based management and treatment. Our assessment of temporal trends of narrow topics using both approaches reflects the impact of key clinical, research and technological advances in bronchiectasis, including the introduction of next-generation sequencing (NGS) in 2005 and the publication of clinical guidelines for the management of bronchiectasis in 2010 (supplementary figure S3a, b). Following the introduction of HRCT in the 1970s, a significant rise in publications related to the topic of CT emerged (*i.e.* Atlas: "bronchiectasis: computed tomography"; customised model: "CT diagnostic accuracy improvement") [40–43]. With the emergence of NGS in 2005, publications investigating the microbiome in bronchiectasis soared (Atlas: "neutrophil inflammation: microbiome"; customised model: "pathogen detection and analysis") [30, 46]. Similarly, after the publication of the Spanish and British guidelines for the management of bronchiectasis in 2008 and 2010, a clear spike in interest and publication surrounding bronchiectasis management emerged (Atlas: "bronchiectasis: evidence-based management"; customised model: "bronchiectasis: aetiology and management") [31, 33, 47]. In 2014, the initial description of the bronchiectasis severity scores Bronchiectasis Severity Index (BSI) [34] and FACED [35], a widely used multidimensional clinical and research tool, saw a surge in research with respect to mortality (Atlas: "bronchiectasis: mortality";

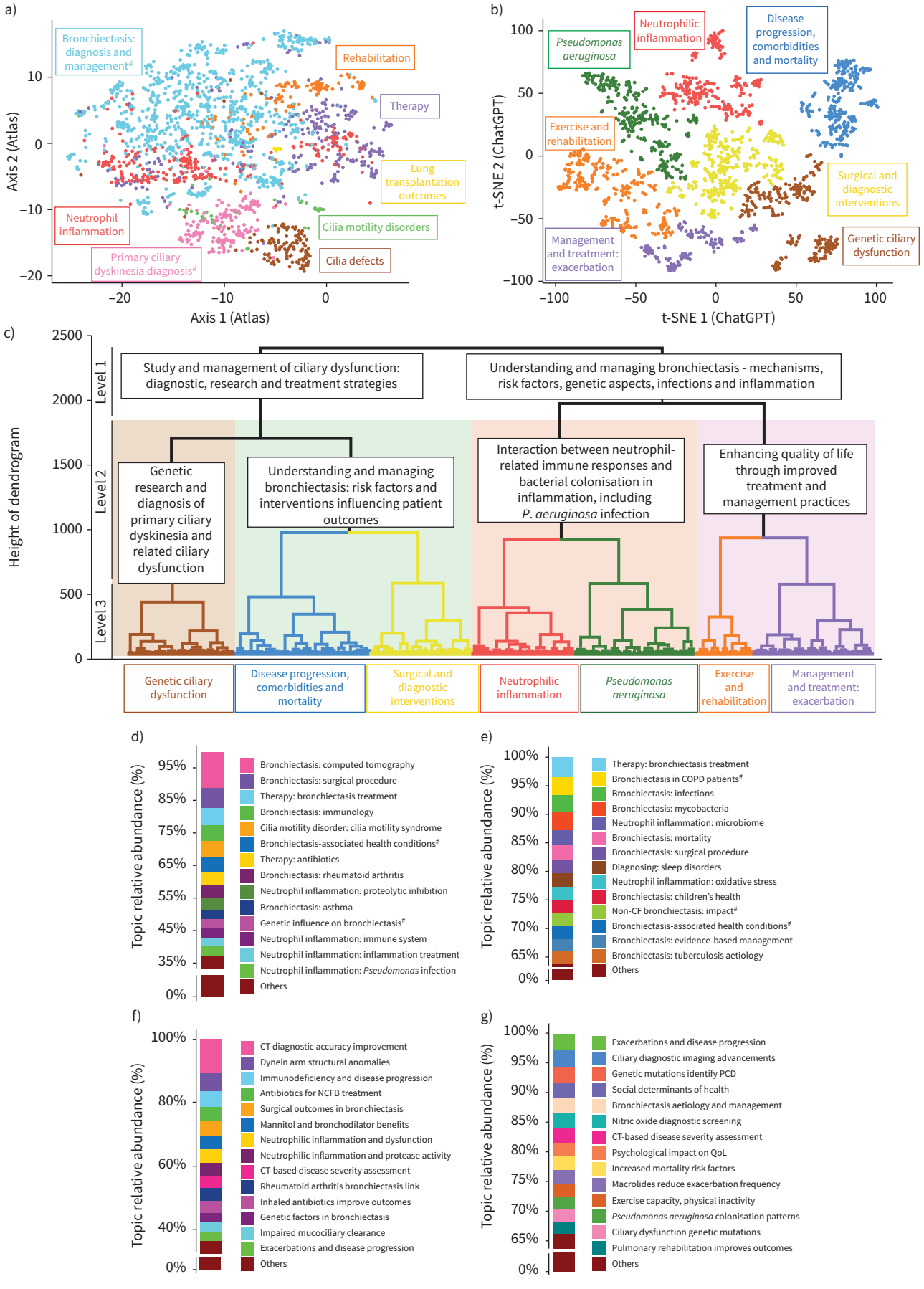


FIGURE 3 Topic modelling in bronchiectasis research between January 1950 and April 2024. Two-dimensional (2D) visualisation of included studies based on semantic similarity using a) the Atlas model and b) the customised model is illustrated. Each respective point represents a study, with proximity within the 2D space indicating higher semantic similarity between studies. Different colours indicate studies grouped by distinct broad-level topics as determined by topic modelling. In the customised model, selected studies were embedded in a 3072-dimensional space using the ChatGPT text embedding model before dimensionality reduction (2D) using t-distributed stochastic neighbour embedding (t-SNE) and for visualisation. Hierarchical clustering was applied to the dimension reduced data (t-SNE) to allow grouping of studies and to identify distinct topics of bronchiectasis research. c) A dendrogram illustrating clustering of included studies, where branching indicates the hierarchical relationships between clusters at various levels of granularity (*i.e.* levels 1 to 3, where broad-level topics are defined as level 3 topics). Studies within each generated cluster at different levels were summarised by a topic name using the ChatGPT text generation model and are indicated as text on the dendrogram. d–g) Stacked bar plots illustrating the relative abundance of the top 20 most abundant narrow-level topics identified in bronchiectasis research d, f) pre-2000 and e, g) post-2000 are shown, based on the d, e) Atlas and the f, g) customised model approach. The Atlas model narrow-level topics are minimally reformatted to enhance clarity, presented in the format "broad-level topic: narrow-level topic". #: for topics that were not clear, we used the ChatGPT text generation model to rename them based on their original titles.

customised model: "increased mortality risk factors"). Taken together, these data reveal a tight relationship between key clinical, research and technological advances and the temporal trends in bronchiectasis research that follow.

Popular, trending and nascent topics in bronchiectasis research

To provide an unbiased evaluation of current and emerging research interests in bronchiectasis, we assessed narrow-level topics to identify popular, trending and nascent topics. Popular topics are defined as those with a sustained growth in research interest over the past decade (2014–2023), determined by their highest slope value, calculated from cumulative topic growth. This analysis revealed that both models identify similar "popular" topics, including bronchiectasis COPD overlap, microbiome and infection and cardiovascular health. Our customised model further identified exacerbations and physical health, while the Atlas model highlighted antibiotic treatments (figure 4a, b). To identify trending research, we assessed topics experiencing sudden, short-term increases in interest over the past 5 years (2019–2023), determined by their highest slope value, calculated from annual counts of new original research publications. Both models identified infection (pathogen analysis) and PCD (genetics) as trending topics. The customised model further identified bronchiectasis and cancer risk, haemoptysis, eosinophilic phenotype and social determinants, while the Atlas model highlighted lung scans as trending topics (figure 4c, d). Nascent research areas are those demonstrating the highest year-on-year percentage increase in research interest over a recent 12-month period (2022–2023). This analysis detected a single topic, mortality in bronchiectasis, as common to both models. The customised model identified symptom assessment tools, *Pseudomonas aeruginosa*, eosinophilic phenotypes and ciliary dysfunction as additional nascent topics, while the Atlas model identified corticosteroids, hypertonic saline, imaging and HIV-related bronchiectasis risk as nascent topics (figure 4e, f).

Evolution of NCFB research diversification

We next examined bronchiectasis research topic evenness and richness over time (figure 5a–d). Using narrow-level topics determined from both models, we calculated Pielou's evenness index (range 0–1), where 0 indicates dominance by a single topic and 1 indicates an equal distribution of all research topics. This analysis revealed that the models exhibited comparable trends in the evolution of topic evenness, with the lowest value of 0.85, generally indicative of an overall balanced research focus (topics) over time. Notably, however, there were clear instances of decreased evenness, signalling key periods in which specific areas gained major attention. Both models identified 1980, 1987 and 1994 as such periods, primarily centred on ciliary structural anomalies (1980) and CT scans (1987 and 1994). The custom model further highlighted two key periods of leading focus: antibiotics in 1994 and genetic mutations related to ciliary dysfunction, macrolides and nitric oxide testing in 2013 (figure 5a, b). We next assessed for temporal evolution of research topic diversification by analysing trends in topic richness, defined as the number of distinct narrow-level topics in both models. This revealed similar upward trajectories in both models (slope approximately equal to 1), indicating an annual rate of diversification of one new distinct topic each year (figure 5c, d). This sustained growth in thematic diversity in bronchiectasis suggests a vibrant, dynamically evolving research landscape, reflective of the emerging and ongoing innovation and expansion of bronchiectasis research. Finally, we conducted an age analysis of research topics to identify the most prominent "youngest" narrow-level topics, and both models showed cardiovascular health, rehabilitation and genetic mutations in PCD as among the youngest topics. The customised model further identified microbiome and bronchiectasis–COPD overlap while Atlas identified haemoptysis and severity scoring systems as additional young topics (figure 5e, f). To further assess the distribution of research

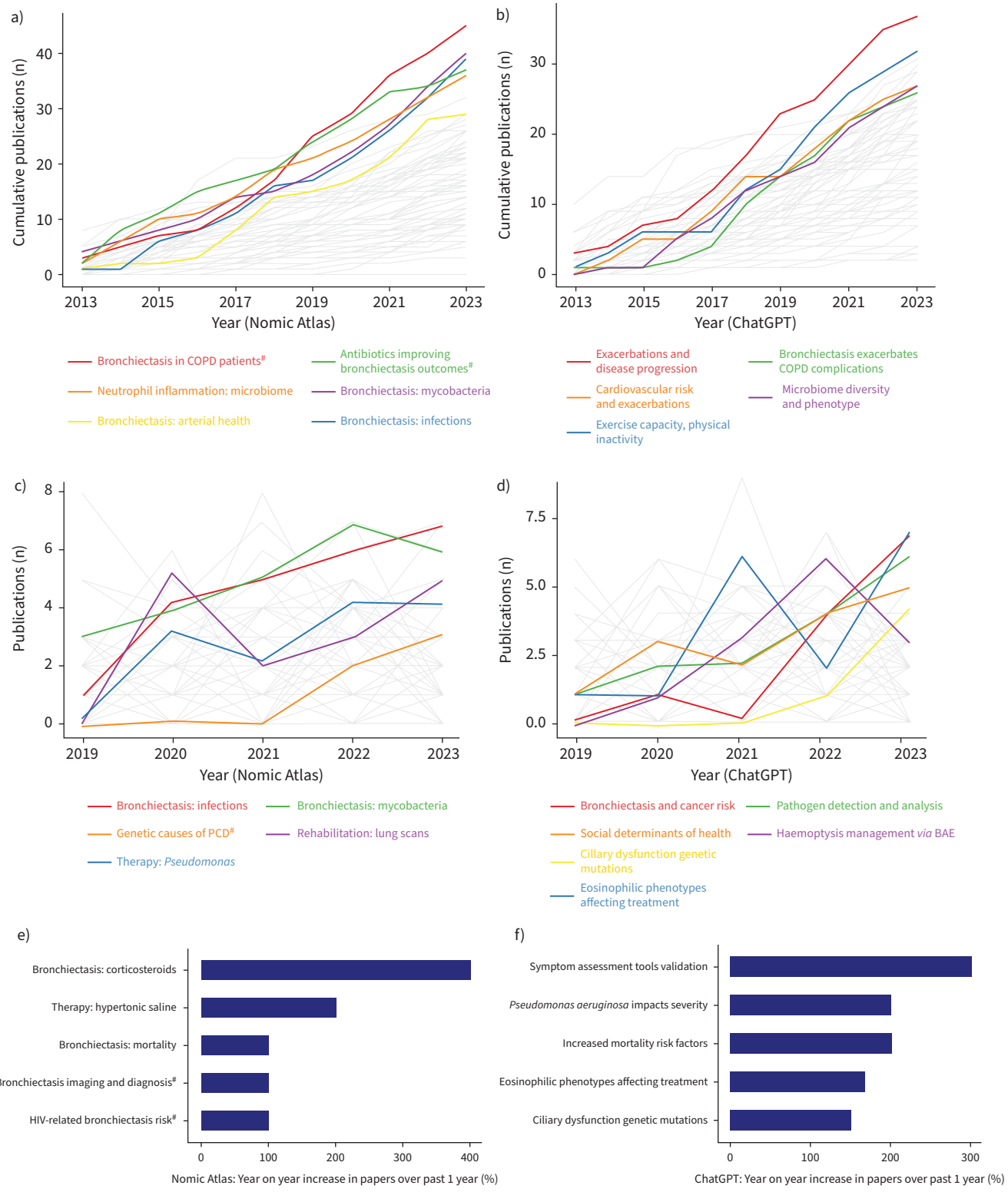


FIGURE 4 Popular and trending topics in bronchiectasis research over the past decade (2014–2023). Line plots illustrating the a, b) popular and c, d) trending topics in bronchiectasis research determined by the a, c) Atlas and b, d) customised models. Popular topics were defined as the top five topics demonstrating sustained increase in research interest, determined by highest slope values from the cumulative growth of the research topic over the past decade (2014–2023). Trending topics were defined as the top five topics which have experienced sudden, short-term increases in research interest, determined by highest slope values from the annual count of new research articles for each respective topic (y-axis) over the past 5 years (2019–2023; x-axis). e, f) Bar plots illustrating the top five topics with highest year-on-year per cent increase in the recent past (2022–2023)

determined by the e) Atlas and f) customised models. The Atlas models narrow-level topics were reformatted to enhance clarity, with each topic presented in the format "broad-level topic: narrow-level topic". PCD: primary ciliary dyskinesia; BAE: bronchial artery embolisation. #: for topics that were not clear, we used the ChatGPT text generation model to rename them based on their original titles.

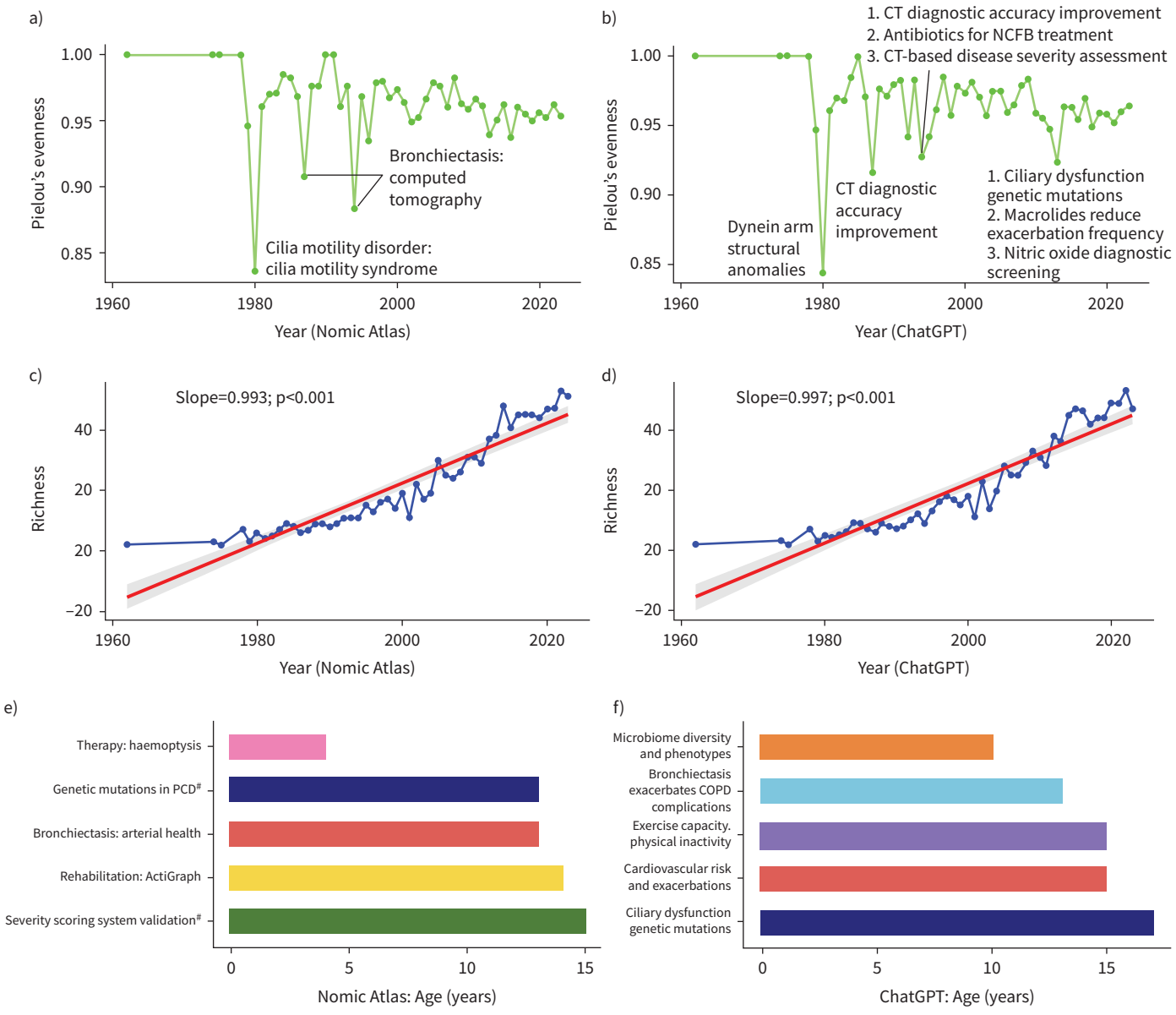


FIGURE 5 Bronchiectasis research topic evenness and richness between January 1950 and April 2024. a, b) Line plots illustrating the evolution of topic evenness determined by a) Atlas and b) customised models. The y-axis represents Pielou's evenness index and the x-axis represents the year of assessment. A value of 0 signals dominance, indicative that a single topic was predominantly studied, while a value of 1 indicates that all topics were on average equally studied. Predominantly studied topics are highlighted when decreased evenness is observed. c, d) Line plots illustrating topic richness determined by c) Atlas and d) customised models. The y-axis indicates the number of distinct topics and the x-axis represents the year of assessment. The calculated slope reflects the rate of change in topic richness, providing insight into the rate of diversification of research focus within the non-cystic fibrosis bronchiectasis (NCFB) field. e, f) Bar plots indicating the top five "youngest" research topics determined by e) Atlas and f) customised models using: 2023–"when topic first appeared". The Atlas model's narrow-level topics were reformatted to enhance clarity, with each topic presented in the format "broad-level topic: narrow-level topic". CT: computed tomography; PCD: primary ciliary dyskinesia. #: for topics that were not clear, we used the ChatGPT text generation model to rename them based on their original titles.

attention among young topics, we repeated the evenness analysis restricted to the five youngest topics. This showed generally even attention, except in 2013, when a sharp drop in evenness was driven by a surge in studies on genetic mutations in PCD (supplementary figure S4a, b).

Growth of bronchiectasis research within the vicious vortex framework

Our understanding of the pathogenesis of bronchiectasis has evolved from Cole's vicious cycle hypothesis in 1986 [28], in which infection, inflammation and airway destruction are proposed to occur sequentially, to the vicious vortex framework proposed in 2018 [1], in which all aspects occur concurrently, to differing extents and severity, and may vary and change over time at the individual level [5]. Here, we sought to assess the growth of bronchiectasis research over time within the conceptual framework of the vicious vortex hypothesis [1]. To evaluate if the bronchiectasis literature reflected this compartmentalisation, we used the ChatGPT text generation model to classify studies into one or more of four categories, aligned to the vortex's core axes: "airway dysfunction", "inflammation", "infection" and "structural disease". Where a publication did not fit clearly into one of these axes, it was classified as "others" (figure 1c). The assessment of study frequency addressing one or more aspects of the vortex revealed thematic imbalance. Interestingly, the "others" category emerged as the most frequently studied single domain (*i.e.* 22.2% of all studies) followed by "airway dysfunction" (14.9%), "infection" (11.7%), "structural disease" (8.1%) and "inflammation" (5.2%). Mapping the broad-level topics identified by our custom unsupervised model onto the supervised vortex categories revealed strong concordance between the two approaches. Specifically, the "infection" axis of the vortex was predominantly associated with the broad topic of "*Pseudomonas aeruginosa*" while the "inflammation" axis was primarily linked to "neutrophil inflammation". Additionally, we observed that the majority of bronchiectasis literature has focused on a single domain, with ~62.2% of studies addressing only one axis of the vortex, while 37.8% encompassed multiple domains (figure 6a). Assessment of the evolution of single-domain *versus* multi-domain studies revealed that, since the 2000s, multi-domain research has consistently lagged behind single-domain studies (supplementary figure S5a).

Temporal analysis of the various components of the vicious vortex revealed chronological trends. Specifically, studies focused on "airway dysfunction" started to emerge in the early 1990s, gaining momentum in the 2000s. By contrast, research on "inflammation" (the least-studied single domain) demonstrated a more recent surge, primarily occurring in the 2020s in line with an emerging paradigm shift toward viewing bronchiectasis as an inflammatory disease and the arrival of neutrophil serine proteases as novel therapeutics [5, 38, 48, 49]. Meanwhile, studies focused on "structural disease", "infection" and "others" gained momentum starting in 2010 and progressing over the following decade, reflective of a consistent, gradual increase in attention to all these areas over time, in contrast to a more acute focus on "inflammation" since 2020 (supplementary figure S6b). While several studies assessed more than a single aspect of the vortex, close examination of the "others" category (the most-studied single domain) demonstrated a notable lack of overlapping studies, suggesting that the "others" category is a relatively independent research focus (figure 6a). Further, temporal analysis of overlap studies, *i.e.* those spanning both vortex components and non-vortex components ("others"), showed a persistent scarcity of integrative work. Most studies focused either solely on vortex components (including overlap within the vortex) or exclusively on non-vortex topics, with minimal research bridging these domains (supplementary figure 6b). To further interrogate the "others" category, we assessed broad-level topics within this category using the customised model, revealing four themes: 1) risk factors and comorbidities; 2) phenotypes, clinical variability and socioeconomic factors; 3) PCD: genetics basis and diagnosing; and 4) enhancing quality of life through physical and mental health (figure 6c), each importantly distinct from the key aspects of the vicious vortex framework. Assessment of these broad topics over time revealed an increasing trend in quality-of-life studies, which started to peak around 2010, followed by studies on phenotypes, clinical variability and socioeconomic factors, and a recent surge in studies exploring risk factors and comorbidities (supplementary figure S5c).

Discussion

This study presents the first comprehensive analysis of bronchiectasis research using AI-powered topic modelling. Our analysis reveals increases in both the speed of publication (nine-fold surge post-2000) and the total research corpus (8.34% annual growth, current). Using a dual modelling approach, Nomic Atlas [25] and a customised pipeline built on OpenAI-ChatGPT (text generation and embedding) models, we have identified thematic pillars of bronchiectasis research and temporal shifts in study focus. Our findings confirm growing thematic richness and diversification, reflective of the increasingly mature, multidimensional and interdisciplinary field. Situating this trend within the broader lifecycle of scientific disciplines, our findings suggest bronchiectasis remains in an early developmental phase, in contrast to asthma and COPD, which show signs of maturity and/or slowed growth. This indicates that bronchiectasis

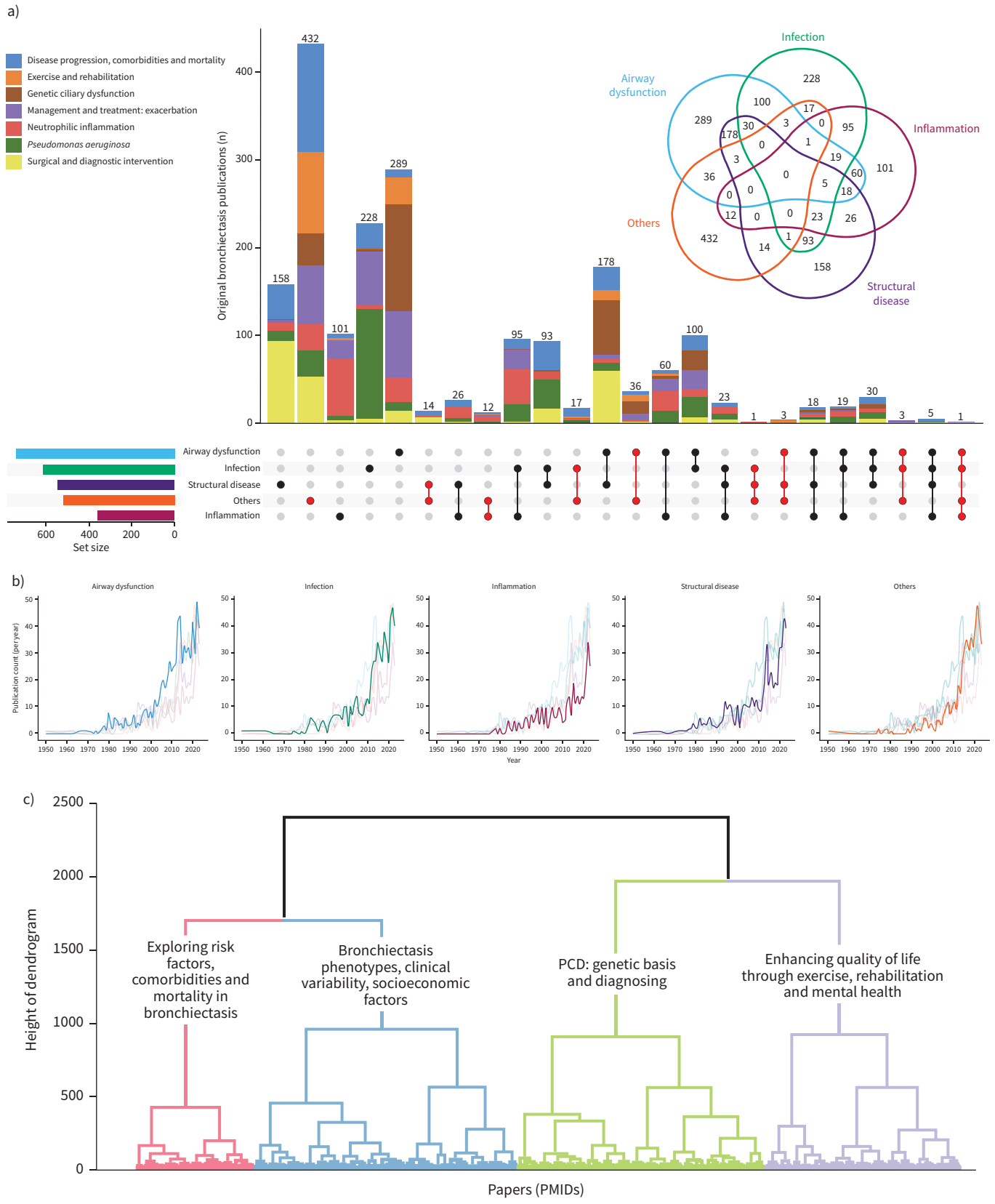


FIGURE 6 Growth of bronchiectasis research within the vicious vortex framework of disease pathogenesis. a) An UpSet plot and corresponding Venn diagram (inset) representing categorisation of bronchiectasis studies (and their respective overlaps) between the key compartments of the vicious vortex framework of disease pathogenesis. An "others" category was included for studies not traditionally fitting within any of major areas

of the vicious vortex framework. Studies were grouped into one or more of the following compartments using the ChatGPT text generation model: “airway dysfunction”, “inflammation”, “infection”, “structural disease” and “others”. Stacked bar plots (within UpSet plot) illustrate the composition of NCFB research focus coloured by broad-level topics (mapped by custom model) across various aspects of the vicious vortex framework, including their intersections. Set size (*i.e.* the number of studies within each compartment) is indicated by horizontal bar plots (airway dysfunction>infection>structural disease>others>inflammation; located bottom-left). Individual compartments and their intersections are indicated as an intersection-matrix (below the stacked bars) with black dots indicating sets, and connecting lines indicating set intersections related to each stacked bar chart. Sets that intersect with the “others” category are highlighted in red (dot and lines). Text within the Venn diagram and above the stacked bars reflect the number of respective studies within each category (and their intersections). b) Line plots illustrating the evolution of various aspects of the vicious vortex framework, including the “others” category, evaluating the number of studies falling within each aspect (y-axis), as categorised by the ChatGPT model, over time in years (x-axis). c) Dendrogram illustrating clustering of all studies within the “others” category, based on our customised model, in which branching indicates hierarchical relationships between topics within the “others” category. Topics were summarised using the ChatGPT text generation model and are indicated as text on the dendrogram. PCD: primary ciliary dyskinesia; PMID: PubMed identifier.

research is expanding, and critically entering its primary exponential phase, suggesting substantial future momentum. However, critical areas, including integration of comorbidities, quality of life and socioeconomic factors, remain only partially aligned with prevailing pathological frameworks such as the vicious vortex, revealing persistent conceptual, research and clinical gaps.

Current methods to organise and interpret scientific literature resemble outdated library systems that filed books alphabetically, functional but largely inefficient to facilitate deeper inquiry. Modern library systems now classify texts by subject and subfield, enabling researchers to locate related materials and explore adjacent domains [50, 51]. By analogy, AI, particularly LLMs, offers a powerful tool to impose structure on unstructured data, including scientific literature [52, 53]. LLMs now embed semantic relationships while preserving latent themes, enabling coherent and accessible representation of the scientific research landscape [54, 55]. Leveraging this approach for bronchiectasis, we have demonstrated “key” thematic areas, highlighting ciliary dysfunction, inflammation, infection, therapeutic interventions and quality of life. Moreover, we observed profound transformation in the bronchiectasis research landscape, shifting from disease-centric to patient-centric approaches.

Prior to 2000, research predominantly focused on diagnostic technologies, including chest radiography or HRCT, surgical interventions, ciliary dysfunction and disease mechanisms, including neutrophilic inflammation. By contrast, the post-2000 era reflects a more holistic research approach, emphasising infection, mortality, aetiology and endophenotypes, resulting in the consideration of more comprehensive and personalised management. This shift is further substantiated by our analysis of broad-level topic evolution, which shows that patient-centric themes, *e.g.* “bronchiectasis: diagnosis and management”, “disease progression, comorbidities and mortality”, and “management and treatment: exacerbation”, have experienced a surge in recent publication activity. While “neutrophil inflammation” appeared less prominent in the post-2000 abundance analysis compared to pre-2000, this does not necessarily imply reduced scientific interest, but rather reflects the relatively rapid growth of other patient-centric themes and diversification of the research landscape.

Analysis of AI-derived topic evolution in bronchiectasis closely tracks human-curated events, including major clinical, diagnostic, patient-registry and technological developments. In this sense, it provides a consistent and interpretable framework that faithfully captures the underlying structure and thematic meaning of the research corpus. For instance, the introduction of HRCT (1970s) is associated with surges in publications focused on imaging and diagnosis [40–42]. Similarly, the emergence of NGS (2005) corresponded with a rapid rise in microbiome research [6], while publication of the first bronchiectasis guidelines by the Spanish Society of Pneumology and Thoracic Surgery (SEPAR) [31] (2008) and the development of the BSI [34] and FACED score [35] (both 2014) were each followed by notable increases in research on evidence-based management and disease risk stratification.

Our findings reveal a dynamic research landscape that is increasingly aligned with evolving clinical priorities and emerging tools. Popular topics over the past decade have centred on bronchiectasis in the context of comorbidities such as COPD and cardiovascular disease. Meanwhile, trending topics highlight infection, the microbiome and PCD genetics, pointing to shifting priorities, with emergent nascent themes like mortality and eosinophilic phenotypes signalling a move towards precision medicine and treatable traits. The convergence of findings between both AI approaches lends further strength to our observations, with only subtle differences in topics highlighting the multifaceted nature of bronchiectasis.

At a macro-level, bronchiectasis research demonstrates remarkable dynamism and continuous expansion. The research focus is remarkably balanced over time, even among the field's youngest topics, with only minor fluctuations. This indicates periods of concentrated scientific interest in specific topics. In contrast, periods of decreased evenness (research concentration) signal key technological and clinical advancements, such as the emergence of CT (1980s), investigations into ciliary structural anomalies and antibiotic treatments (1994) and macrolide treatment (2013). While Pielou's evenness index reflects the distribution of research topics rather than their necessary clinical importance, these periods highlight marked research concentration within topics. Noticeably, the consistent annual diversification of research topics, with one new distinct topic emerging each year, underscores the field's innovative nature and reflects an increasingly complex and emerging interdisciplinary approach to bronchiectasis.

The vicious vortex framework has emerged as a transformative conceptual model to understand disease pathogenesis [1]. It conceptualises bronchiectasis as a self-perpetuating cycle of interacting factors, including infection, inflammation, impaired mucociliary clearance and structural damage, each reinforcing the others and sustaining disease progression. Our sub-analysis based on this framework highlights significant thematic imbalance between its four axes, with the "others" (non-vortex related) category emerging as most frequently studied. Such research distribution partially reflects the field's expanding appreciation of bronchiectasis beyond traditional conceptual frameworks. Interestingly, our comprehensive assessment of the "others" category (non-vortex related) highlights recently rising patient-centric broad-level topics, including risk factors, comorbidities, quality of life and socioeconomic factors. These findings further underscore the ongoing shift from traditional disease-centric (vortex-component) models toward more holistic, patient-centred clinical approaches. Notably, this category also encompasses research focused on clinical variability and phenotypic heterogeneity. Together, these align well with the treatable traits and precision medicine paradigms gaining traction in bronchiectasis care and management [5, 56]. Additionally, the strong concordance between unsupervised broad-level topics and the supervised vicious vortex axes, specifically, the dominance of "*Pseudomonas aeruginosa*" in "infection" and "neutrophil inflammation" in "inflammation", validates our analytical approach.

Our temporal analysis reveals distinct research trajectories, with "airway dysfunction" emerging in the 1990s, "structural disease" and "infection" gaining momentum in the 2010s and "inflammation" experiencing a notable surge from 2020. This shift aligns with emerging disease paradigms [10, 11, 13, 57–59]. These trends track with the development of anti-inflammatory therapies, culminating in landmark positive clinical trials including those of dipeptidyl peptidase 1 inhibition, demonstrating the ability of LLMs to track emerging therapeutic paradigms in near real-time [38, 48, 60, 61]. While ∼37% of studies addressed multiple vortex components, highlighting the intricate and interconnected nature of bronchiectasis, our analysis reveals that truly holistic, multi-domain research remains limited. Despite a recent interdisciplinary push, the evolution of overlap studies since the 2000s has consistently lagged behind single-domain investigations. Furthermore, studies overlapping both vortex and non-vortex ("others") components demonstrate persistent scarcity. Most research continues to focus exclusively on either disease pathogenesis (*i.e.* vortex components including overlaps within the vortex itself) or non-vortex topics (*i.e.* patient-centric), with minimal studies bridging these domains. This ongoing gap underscores the need for more holistic and integrative approaches to bronchiectasis research that can better reflect the multifaceted nature of the disease and inform comprehensive patient care.

To our knowledge, this is the first structured implementation of LLM-based AI for a comprehensive literature assessment in the field of respiratory medicine. This approach demonstrates the potential of LLMs to extract, synthesise and classify research at scale, providing an unbiased and high-throughput approach to complement conventional reviews and meta-analyses. Human-led reviews often suffer from author bias, citation self-selection and disproportionate self-citation [62, 63]. By contrast, LLM-driven assessment offers an impartial, reproducible overview of a field, balancing literature appraisal and uncovering "blind spots". As the volume of published literature expands, these AI tools will be critical to enable clinicians and researchers to maintain situational awareness of evolving research, offering timely, accurate and important clinical insight [62]. Notwithstanding the novelty of the approaches, our customised analytical pipeline is grounded in conventional statistical rigour. Our customised model transcends standard Atlas capability by offering more granular stratification of research themes, including assessment of cluster quality and robustness, going beyond Atlas's two-level topic modelling. However, variability between AI approaches exist. For instance, our customised model identified "*Pseudomonas aeruginosa*" as a broad topic, unlike the Atlas model. Although we minimised bias through unsupervised methods, choices made during preprocessing (*e.g.* selection of abstracts, embedding model, number of clusters) inevitably shape results [62]. Nevertheless, this does not reflect a fundamental weakness, but rather an opportunity to leverage multiple models to achieve richer and comprehensive insights. For instance, the Atlas model embedded

"*Pseudomonas aeruginosa*" within the closely related broader concept of "inflammation" (as evidenced by narrow-level topics such as "neutrophil inflammation: *Pseudomonas* infection"), highlighting the impact of hierarchical granularity and clustering thresholds inherent to each AI methodology. This inherent variability between AI methods will be a critical challenge in emerging future applications of AI to clinical medicine, underscoring the need for explicit reporting, use of multi-method validation or ensemble-based approaches to ensure transparency, reproducibility and, ultimately, trustworthy clinical implementation.

While novel, our study has several limitations. First, we assume equal weight and quality across all included research articles, while in reality, study quality varies [64]. Second, while use of LLMs introduces novelty and scalability, these models remain in dynamic development and should be considered as "black boxes" whose outputs will evolve as new versions are released [65]. Third, thematic clustering inherently involves some degree of abstraction, and subtle but important nuances at the level of individual studies may be lost [55, 66, 67]. Fourth, our analysis was restricted to abstracts from original research articles, potentially omitting critical contextual details and other study types such as meta-analyses. Future iterations should consider incorporating multiple study types with appropriate weighting informed by established hierarchies of evidence synthesis. Lastly, the absence of "ground truth" due to the subjective nature of thematic interpretation hinders formal validation of our identified research trends. While our dual AI approach partly mitigates this, future research should pursue formal validation, using structured expert panels or Delphi surveys. Despite these caveats, our approach provides high-resolution mapping of the bronchiectasis research landscape, offering a data-driven view of the field past, present and future. Future studies should incorporate full-text mining, integrate multimodal data, and explore geographic trends and predictive modelling of topic evolution. Additionally, building on the existing pipelines developed here, future work may evolve into a scalable, automated analytic framework that allows for dynamic "dashboard-like" interfaces for near-real-time evidence synthesis, monitoring and interpretation of literature in bronchiectasis and beyond.

The research landscape in bronchiectasis has expanded significantly. AI-powered analytics offers an unbiased perspective on research evolution, identifying imbalances and guiding future investigation without the tedium of traditional literature review. As the field of bronchiectasis and others continue to evolve, embracing AI and these tools will be key to accelerating timely insight and improving patient outcomes.

Acknowledgements: The authors thank The Academic Respiratory Initiative for Pulmonary Health (TARIPH) and the Lee Kong Chian School of Medicine Centre for Microbiome Medicine for collaboration support.

Conflict of interest: S.H. Chotirmall has served on advisory boards for CSL Behring, Pneumagen Ltd., Zaccha Pte Ltd, Boehringer Ingelheim, GSK and Sanofi; on data monitoring boards for Inovio Pharmaceuticals and Imam Abdulrahman Bin Faisal University; and has received personal fees from AstraZeneca and Chiesi Farmaceutici, all unrelated to this work. The remaining authors have no potential conflicts of interest to disclose.

Support statement: This research is supported by the National Research Foundation Singapore under its Open Fund-Large Collaborative Grant (MOH-001636) and administered by the Singapore Ministry of Health's National Medical Research Council, the Singapore Ministry of Health's National Medical Research Council under its Clinician-Scientist Individual Research Grant (MOH-001356) (S.H. Chotirmall), and the Singapore Ministry of Education under its AcRF Tier 1 Grant (RT1/22) (S.H. Chotirmall). Funding information for this article has been deposited with the Open Funder Registry.

References

1 Flume PA, Chalmers JD, Olivier KN. Advances in bronchiectasis: endotyping, genetics, microbiome, and disease heterogeneity. *Lancet* 2018; 392: 880–890.
2 Long MB, Chotirmall SH, Shteinberg M, *et al.* Rethinking bronchiectasis as an inflammatory disease. *Lancet Respir Med* 2024; 12: 901–914.
3 Chalmers JD, Goeminne PC, Ringshausen FC. EMBARCing on a new era for bronchiectasis: a review series for the Seventh World Bronchiectasis Conference. *Eur Respir Rev* 2024; 33: 240124.
4 Chotirmall SH, Chalmers JD. Bronchiectasis: an emerging global epidemic. *BMC Pulm Med* 2018; 18: 76.
5 Chotirmall SH, Chalmers JD. The precision medicine era of bronchiectasis. *Am J Respir Crit Care Med* 2024; 210: 24–34.
6 Mac Aogain M, Dicker AJ, Mertsch P, *et al.* Infection and the microbiome in bronchiectasis. *Eur Respir Rev* 2024; 33: 240038.
7 Perea L, Faner R, Chalmers JD, *et al.* Pathophysiology and genomics of bronchiectasis. *Eur Respir Rev* 2024; 33: 240055.

8 Chalmers JD, Polverino E, Crichton ML, *et al.* Bronchiectasis in Europe: data on disease characteristics from the European Bronchiectasis registry (EMBARC). *Lancet Respir Med* 2023; 11: 637–649.
9 Chalmers JD, Chang AB, Chotirmall SH, *et al.* Bronchiectasis. *Nat Rev Dis Primers* 2018; 4: 45.
10 Mac Aogain M, Tiew PY, Lim AYH, *et al.* Distinct "immunoallertypes" of disease and high frequencies of sensitization in non-cystic fibrosis bronchiectasis. *Am J Respir Crit Care Med* 2019; 199: 842–853.
11 Mac Aogain M, Chandrasekaran R, Lim AYH, *et al.* Immunological corollary of the pulmonary mycobiome in bronchiectasis: the CAMEB study. *Eur Respir J* 2018; 52: 1800766.
12 Narayana JK, Tsaneva-Atanasova K, Chotirmall SH. Microbiomics-focused data integration: a fresh solve for the Rubik's Cube of endophenotyping? *Am J Respir Crit Care Med* 2022; 206: 365–368.
13 Narayana JK, Mac Aogain M, Hansbro PM, *et al.* The bronchiectasis microbiome: current understanding and treatment implications. *Curr Opin Pulm Med* 2025; 31: 135–144.
14 Mac Aogain M, Narayana JK, Tiew PY, *et al.* Integrative microbiomics in bronchiectasis exacerbations. *Nat Med* 2021; 27: 688–699.
15 Mac Aogain M, Ivan FX, Jaggi TK, *et al.* Airway "resistotypes" and clinical outcomes in bronchiectasis. *Am J Respir Crit Care Med* 2024; 210: 47–62.
16 Huang JT, Cant E, Keir HR, *et al.* Endotyping chronic obstructive pulmonary disease, bronchiectasis, and the "chronic obstructive pulmonary disease-bronchiectasis association". *Am J Respir Crit Care Med* 2022; 206: 417–426.
17 Chalmers JD, Chotirmall SH. Bronchiectasis: new therapies and new perspectives. *Lancet Respir Med* 2018; 6: 715–726.
18 Poh TY, Mac Aogain M, Chan AK, *et al.* Understanding COPD-overlap syndromes. *Expert Rev Respir Med* 2017; 11: 285–298.
19 Tiew PY, Lim AYH, Keir HR, *et al.* High frequency of allergic bronchopulmonary aspergillosis in bronchiectasis-COPD overlap. *Chest* 2022; 161: 40–53.
20 Shteinberg M, Waterer G, Chotirmall SH. A global effort to stop the vicious vortex: a special *American Journal of Respiratory and Critical Care Medicine* issue for world bronchiectasis day 2024. *Am J Respir Crit Care Med* 2024; 210: 1–3.
21 Blaizot A, Veettil SK, Saidoung P, *et al.* Using artificial intelligence methods for systematic review in health sciences: a systematic review. *Res Synth Methods* 2022; 13: 353–362.
22 Hasan B, Saadi S, Rajjoub NS, *et al.* Integrating large language models in systematic reviews: a framework and case study using ROBINS-I for risk of bias assessment. *BMJ Evid Based Med* 2024; 29: 394–398.
23 Jin Q, Leaman R, Lu Z. PubMed and beyond: biomedical literature search in the age of artificial intelligence. *EBioMedicine* 2024; 100: 104988.
24 Porter AL, Rafols I. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics* 2009; 81: 719–745.
25 Duderstadt B, Mulyar A, Schmidt B, *et al.* Atlas Whitepaper: Scalable Information Cartography. 2023. Available from: https://static.nomic.ai/atlas_tech_report.pdf
26 Russo L. Pulmonary postural drainage for bronchiectasis. *Dis Chest* 1954; 26: 81–91.
27 Hounsfield L. Computerized transverse axial scanning (tomography). 1. Description of system. *Br J Radiol* 1973; 46: 1016–1022.
28 Cole PJ. Inflammation: a two-edged sword: the model of bronchiectasis. *Eur J Respir Dis Suppl* 1986; 147: 6–15.
29 Pennarun G, Escudier E, Chapelin C, *et al.* Loss-of-function mutations in a human gene related to *Chlamydomonas reinhardtii* dynein IC78 result in primary ciliary dyskinesia. *Am J Hum Genet* 1999; 65: 1508–1519.
30 Margulies M, Egholm M, Altman WE, *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005; 437: 376–380.
31 Vendrell M, de Gracia M, Olveira C, *et al.* Diagnóstico y tratamiento de las bronquiectasias [Diagnosis and treatment of bronchiectasis]. *Arch Bronconuemol* 2008; 44: 629–640.
32 Wilson M, Welte T, Polverino E, *et al.* Ciprofloxacin dry powder for inhalation in non-cystic fibrosis bronchiectasis: a phase II randomised study. *Eur Respir J* 2013; 41: 1107–1115.
33 Pasteur MC, Bilton D, Hill AT, *et al.* British Thoracic Society guideline for non-CF bronchiectasis. *Thorax* 2010; 65: Suppl. 1, i1–i58.
34 Chalmers JD, Goeminne P, Aliberti S, *et al.* The Bronchiectasis Severity Index. An international derivation and validation study. *Am J Respir Crit Care Med* 2014; 189: 576–585.
35 Martínez-García MA, de Gracia J, Vendrell Relat M, *et al.* Multidimensional approach to non-cystic fibrosis bronchiectasis: the FACED score. *Eur Respir J* 2014; 43: 1357–1367.
36 Guan M, Li J-C, Liu F, *et al.* Next-generation sequencing for identifying genetic mutations in adults with bronchiectasis. *J Thorac Dis* 2018; 10: 2618–2630.
37 Polverino E, Goeminne PC, McDonnell MJ, *et al.* European Respiratory Society guidelines for the management of adult bronchiectasis. *Eur Respir J* 2017; 50: 1700629.
38 Chalmers JD, Haworth CS, Metersky ML, *et al.* Phase 2 trial of the DPP-1 inhibitor brensocatib in bronchiectasis. *N Engl J Med* 2020; 383: 2127–2137.

39 Bornmann L, Haunschild R, Mutz R. Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. *Humanit Soc Sci Commun* 2021; 8: 224.

40 Engeler CE, Tashjian JH, Engeler CM, *et al.* Volumetric high-resolution CT in the diagnosis of interstitial lung disease and bronchiectasis: diagnostic accuracy and radiation dose. *AJR Am J Roentgenol* 1994; 163: 31–35.

41 Grenier P, Maurice F, Musset D, *et al.* Bronchiectasis: assessment by thin-section CT. *Radiology* 1986; 161: 95–99.

42 Phillips MS, Williams MP, Flower CD. How useful is computed tomography in the diagnosis and assessment of bronchiectasis? *Clin Radiol* 1986; 37: 321–325.

43 Schulz RA, Stein JA, Pelc NJ. How CT happened: the early development of medical computed tomography. *J Med Imaging (Bellingham)* 2021; 8: 052110.

44 Hurst A, Lerer A, Goucher AP, *et al.* Gpt-4o system card. *arXiv* 2024, preprint [https://doi.org/10.48550/arXiv.2410.21276].

45 Churchill R, Singh L. The evolution of topic modeling. *ACM Comput Surveys* 2022; 54: 1–35.

46 van Dijk EL, Auger H, Jaszczyszyn Y, *et al.* Ten years of next-generation sequencing technology. *Trends Genet* 2014; 30: 418–426.

47 Martinez-Garcia MA, Maiz L, Olveira C, *et al.* Spanish guidelines on treatment of bronchiectasis in adults. *Arch Bronconeumol (Engl Ed)* 2018; 54: 88–98.

48 Chalmers JD, Shteinberg M, Mall MA, *et al.* Cathepsin C (dipeptidyl peptidase 1) inhibition in adults with bronchiectasis: AIRLEAF, a phase II randomised, double-blind, placebo-controlled, dose-finding study. *Eur Respir J* 2025; 65: 2401551.

49 Chalmers JD, Mall MA, Chotirmall SH, *et al.* Targeting neutrophil serine proteases in bronchiectasis. *Eur Respir J* 2025; 65: 2401050.

50 Taylor AG, Joudrey DN. The Organization of Information. New York, NY, Bloomsbury Publishing, 2008.

51 Bliss HE. The Organization of Knowledge in Libraries and the Subject-approach to Books. New York, The H. W. Wilson Company, 1933.

52 Dagdelen J, Dunn A, Lee S, *et al.* Structured information extraction from scientific text with large language models. *Nat Commun* 2024; 15: 1418.

53 Dunn A, Dagdelen J, Walker N, *et al.* Structured information extraction from complex scientific text with fine-tuned large language models. *arXiv* 2022, preprint [https://doi.org/10.48550/arXiv.2212.05238].

54 Wang L, Yang N, Huang X, *et al.* Improving text embeddings with large language models. *arXiv* 2023, preprint [https://doi.org/10.48550/arXiv.2401.00368].

55 Mars M. From word embeddings to pre-trained language models: a state-of-the-art walkthrough. *Applied Sciences* 2022; 12: 8805.

56 Boaventura R, Sibila O, Agusti A, *et al.* Treatable traits in bronchiectasis. *Eur Respir J* 2018; 52: 1801269.

57 Shoemark A, Shteinberg M, De Soyza A, *et al.* Characterization of eosinophilic bronchiectasis: a European multicohort study. *Am J Respir Crit Care Med* 2022; 205: 894–902.

58 Choi H, Ryu S, Keir HR, *et al.* Inflammatory molecular endotypes in bronchiectasis: a European multicenter cohort study. *Am J Respir Crit Care Med* 2023; 208: 1166–1176.

59 Perea L, Bottier M, Cant E, *et al.* Airway IL-1β is related to disease severity and mucociliary function in bronchiectasis. *Eur Respir J* 2024; 64: 2301966.

60 Chalmers JD, Burgel PR, Daley CL, *et al.* Brensocatib in non-cystic fibrosis bronchiectasis: ASPEN protocol and baseline characteristics. *ERJ Open Res* 2024; 10: 00151-2024.

61 Chalmers JD, Burgel PR, Daley CL, *et al.* Phase 3 trial of the DPP-1 inhibitor brensocatib in bronchiectasis. *N Engl J Med* 2025; 392: 1569–1581.

62 van Dijk SHB, Brusse-Keizer MGJ, Bucsan CC, *et al.* Artificial intelligence in systematic reviews: promising when appropriately used. *BMJ Open* 2023; 13: e072254.

63 Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *Milbank Q* 2016; 94: 485–514.

64 Mays N, Pope C. Qualitative research in health care. Assessing quality in qualitative research. *BMJ* 2000; 320: 50–52.

65 Zhao H, Chen H, Yang F, *et al.* Explainability for large language models: a survey. *ACM Trans Intell Syst Technol* 2024; 15: 1–38.

66 Nie Z, Feng Z, Li M, *et al.* When text embedding meets large language model: a comprehensive survey. *arXiv* 2024, preprint [https://doi.org/10.48550/arXiv.2412.09165].

67 Keraghel I, Morbieu S, Nadif M. Beyond words: a comparative analysis of LLM embeddings for effective clustering. In: International Symposium on Intelligent Data Analysis. Princeton, NJ, Springer, 2024, p. 205–216.