The Changing Geography of Medical Research

Hongyu Zhou Prashant Garg Thiemo Fetzer*

September 29, 2025

Abstract

Medical research remains concentrated in high-income settings, risking misalignment with global health needs. We build a geography-aware knowledge graph linking articles in the 524 leading medical journals to the diseases they study, the countries or territories whose data they analyse, author institutions and funders. We use large-language-model extraction to compare research output with disease burden across 204 countries and 15 major disease groups from 1990 to 2021. Research output has become twice as responsive to domestic disease burden since 1990, yet lower-income regions remain underrepresented in authorship despite serving as frequent research contexts. Maternal–neonatal, nutritional, and many infectious diseases are still under-studied relative to their burden. Philanthropic funders targets neglected burdens, corporations focus on profitable chronic diseases, and governments fall in between. Analyzing WHO disease outbreak news alerts in an event-study design, we show that health shocks trigger rapid, durable increases in both domestic and global research attention, strongest for high-lethality threats. The system is becoming more needs-driven yet remains uneven. Our scalable framework enables near-real-time tracking of convergence.

Keywords: Research Responsiveness, Geographic Disparities, Disease Burden, LLM-enabled knowledge graph

*Zhou is based at the University of Cambridge, supported by funding from the Accelerate Programme for Scientific Discovery, made possible by a donation from Schmidt Sciences, and is also affiliated with the University of Antwerp. Garg is based at Imperial College London (Email: prashant.garg@imperial.ac.uk). Fetzer is based at University of Warwick & Bonn and affiliated with CEPR, CAGE, NIESR, and ECONtribute and acknowledges ME[GP]E?O[A-Z]{0,4}.

1 Introduction

Why do some diseases that impose a heavy toll on human life attract scant scientific attention, while others enjoy a relative surplus of studies, trials, and funding? Although 80% of the world's population lives in low- and middle-income countries (LMICs), fewer than half of registered clinical trials take place there and only a fragment of the peer-reviewed medical literature is authored by researchers based in LMIC institutions (World Health Organization, 2024b; Lewis, Schneegans, Straza, et al., 2021).[1] Such spatial disparities in knowledge production have direct consequences for global health equity, the external validity of clinical evidence (Bareinboim & Pearl, 2016), and the speed with which new treatments reach the populations that need them most. These imbalances also reflect cumulative advantage (Merton, 1968) even as international collaboration networks globalize (Adams, 2012; Hoekman, Frenken, & Tijssen, 2010).

Yet diagnosing "misallocation" is harder than it appears. Both disease burden and research effort are endogenous: a country's epidemiological profile shapes its scientific agenda, but new knowledge can in turn bend the path of morbidity and mortality. Data constraints compound the problem (Røttingen et al., 2013) and known coverage/language biases in bibliographic data persist (Mongeon & Paul-Hus, 2016; Van Leeuwen, Moed, Tijssen, Visser, & Van Raan, 2001). Large-scale scholarly data rarely record where study participants lived; health metrics such as Disability-Adjusted Life Years (DALYs) are reported in epidemiological silos; and the two worlds (scholarly publishing and disease surveillance) are seldom linked by a common ontology. As a result, most existing studies rely on narrow disease categories, coarse regional aggregates, or hand-curated case studies that cannot scale (Evans, Shim, & Ioannidis, 2014; Yegros-Yegros, Van de Klippe, Abad-Garcia, & Rafols, 2020; Access to Medicine Foundation, 2024).

In this paper, we offer a fine-grained global view of how medical research is dis-

[1] UNESCO's 2023 Science Report estimates that Africa, with roughly 12.5% of the world's population, accounts for under 1% of global research output and invests barely 0.5 % of GDP in R&D—ten times below high-income averages (Lewis et al., 2021).

tributed across space, time, and disease, and how closely that distribution tracks the geography of health needs. To achieve this, we begin by constructing a geography-aware medical knowledge graph that links 308,135 research articles in 524 leading general medical journals (see Methods and Appendix A) to (i) the diseases they study, (ii) the countries or territories whose data or patients they analyse, and (iii) the institutional homes of their authors. The underlying pipeline combines two large-language-model (LLM) prompts: one that extracts biomedical relations, another that parses study context, and then maps resulting free-text strings to controlled vocabularies via neural embeddings. A new crosswalk maps Medical Subject Headings (MeSH) to the Global Burden of Disease (GBD) taxonomy (Global Burden of Disease Collaborative Network, 2024), aligning each paper with the relevant GBD causes. A companion module classifies 32,438 funders into public, corporate, philanthropic, and hybrid categories.

Overlaying this graph on Disability-Adjusted Life Year (DALY) statistics for 204 countries or territories (1990–2021) reveals three striking patterns. First, endemic research responsiveness (i.e., the elasticity of publication output with respect to domestic disease burden) has more than doubled over the past three decades, signalling a gradual move toward needs-driven science. Second, aggregate progress masks stark heterogeneity: High-income countries remain the most responsive; yet low- and middle-income regions such as Africa account for 32% of study contexts for neglected tropical diseases and malaria but contribute only 14% of authorship. This pattern echoes systematic evidence of authorship imbalances and "parachute" research in global health collaborations (Hedt-Gauthier et al., 2019; Naidoo, Hodkinson, King, & Wallis, 2021; Abimbola et al., 2021; Dimitris, Gittings, & King, 2021), for instance roughly half of National Cancer Institute–funded publications include no LMIC-affiliated co-authors (Eldridge, Garton, Duncan, & Gopal, 2024). Third, funding priorities diverge sharply: philanthropic foundations disproportionately support research into diseases endemic to lower income regions, whereas corporate R&D tilts toward diseases associated with higher stages of economic development, such as diabetes and kidney disorders, with public agencies

occupying an intermediate position.

To probe whether these patterns reflect causal responses to emerging health risks, we exploit 3,134 WHO Disease Outbreak News alerts as quasi-random shocks to perceived burden (see Methods). An event-study difference-in-differences (DiD) design (saturated with country-disease, disease-year, and country-year fixed effects, together with estimators robust to staggered timing (Sun & Abraham, 2021; Callaway & Sant'Anna, 2021; Goodman-Bacon, 2021) show that alerts trigger rapid, statistically significant jumps in both domestic and global research output. The surge peaks for emerging and high-profile pathogens such as COVID-19, Ebola, and Zika, is smaller and shorter-lived for endemic or persistent conditions such as malaria, diphtheria, and diarrheal diseases, and has grown markedly stronger in the 2010s relative to the 1990s.

Taken together, our results paint a cautiously optimistic picture: the world's scientific apparatus is more geographically inclusive and more attuned to local disease landscapes than it was thirty years ago, yet significant misalignments endure.[2] Diseases such as cardiovascular conditions, neoplasms, and "profitable" chronic disorders still attract outsized attention, while maternal–neonatal disorders, nutritional deficiencies, and many infectious diseases remain under-studied relative to their DALY shares. Funding composition and research capacity (rather than burden alone) emerge as pivotal drivers of these gaps.

Our work speaks to four strands. First, metascience on (mis)alignment and shock-driven reallocation (Gillum et al., 2011; Evans et al., 2014; Yegros-Yegros et al., 2020; Furuse, 2019; Ioannidis, Bendavid, Salholz-Hillel, Boyack, & Baas, 2022; Wagner, Cai, Zhang, & Fry, 2022; Zhao, Wang, & Zhang, 2022) and the broader science studies on geography, collaboration, and cumulative advantage (Fortunato et al., 2018; Merton, 1968; Adams, 2012; Hoekman et al., 2010); we move beyond cross-sections by estimating time-varying

[2]Recent syntheses highlight the gap: the Access to Medicine Foundation's 2024 Index finds that most large pharmaceutical firms "are still not maximising their potential to reach patients in LMICs" (Access to Medicine Foundation, 2024); a Lancet Viewpoint on global R&D priorities highlights the same tension for non-communicable diseases (Vollset et al., 2024); and WHO's 2024 NTD report stresses that R&D bottlenecks remain a core barrier to disease elimination (World Health Organization, 2024a).

elasticities and causal alert responses at the country–disease level. Second, incentives in R&D and portfolio choice shape where research goes. Market size and time-to-payoff (Acemoglu & Linn, 2004; Budish, Roin, & Williams, 2015), shocks from public funding (Azoulay, Graff Zivin, Li, & Sampat, 2019; Chu & Evans, 2021), and the switching costs implied by the elasticity of science to targeted funding (Myers, 2020) all play a role. We add to this a funder-type decomposition, linking sponsor portfolios to the observed growth in responsiveness. Third, external validity and transportability (Bareinboim & Pearl, 2016), motivating our focus on *where* evidence is produced relative to *where* burden lies. Fourth, in natural-language processing, a rapidly expanding corpus demonstrates how large language models can be harnessed to construct (biomedical) knowledge graphs (Kandibedala, Pyayt, Piraino, Caballero, & Gubanov, 2023; Jiralerspong, Chen, More, Shah, & Bengio, 2024; Sudhahar et al., 2024; Gao et al., 2025).[3] Closest to our approach, Schmallenbach et al. map disease-coded publications to GBD causes and attribute recent alignment mainly to burden shifts rather than reallocation Schmallenbach2025NatMed; we complement this by estimating elasticities, decomposing funder portfolios, and leveraging WHO alerts in a staggered DiD to identify causal responsiveness.

[3]See also (Barakati et al., 2025), (Fetzer, Lambert, Feld, & Garg, 2024) or (Garg & Fetzer, 2025b) for examples where LLMs are used to recover edges from materials-science, value chains, or causal claims in economics papers, illustrating the broader potential of LLM-assisted graph construction.

2 Results

2.1 Disparities and shifts in global disease research

We begin by charting the global landscape and evolution of disease research in relation to disease burden. Across our dataset of research papers, the majority of work focuses on cardiovascular diseases and neoplasms, while comparatively less attention is directed toward neglected tropical diseases, malaria, and nutritional deficiencies (Figure 1a). However, the volume of research efforts does not correspond closely to patterns of disease burden (disability-adjusted life years, DALYs) globally. For instance, respiratory infections and tuberculosis, as well as maternal and neonatal conditions, exhibit a much higher disease burden relative to their paper count ranking, whereas neoplasms, digestive diseases, and diabetes and kidney diseases are overrepresented in research output relative to their burden. On the other hand, emergency disease outbreaks, measured by WHO alerts, are most frequently associated with infectious diseases, particularly respiratory infections, tuberculosis, and neglected tropical diseases and malaria.

A geographic lens further highlights disparities. As expected, the Americas, Asia, and Europe account for the largest shares of research authorship and in-text mentions, while Africa lags behind (Figure 1c). Yet the imbalance is nuanced. For neglected tropical diseases and malaria, Africa accounts for 32% of research context but only 14% of authorship. This imbalance means that regions often feature prominently as objects of study but less as sources of authorship. The pattern recurs across disease groups in Africa.

Encouragingly, over time we observe decreasing geographic concentration of research efforts, as measured by the Herfindahl–Hirschman Index (HHI), suggesting that research across most disease groups has become more geographically diverse. An exception emerges in substance use disorders, where research concentration has increased since 2015, largely reflecting the growing dominance of contributions from the United States (Appendix Figure D3).

But does such geographic diversification also mean that research efforts are increasingly directed toward the regions bearing the greatest burden, thereby leading to greater responsiveness globally? Figure 1d provides an illustrative example in the case of HIV/AIDS and STIs. In 2006, the burden fell disproportionately on the Global South, particularly sub-Saharan Africa, yet research efforts from the region were limited. By 2021, while the disease burden remained highest, several countries in sub-Saharan Africa had emerged as leading contributors. This example hints at increasing responsiveness, yet unevenly, prompting us to examine the pattern more formally and empirically across countries and diseases.

2.2 Research responsiveness to endemic disease burden

We conceptualize endemic research responsiveness as the extent to which countries adjust their research output in line with changes in their endemic disease burden, reflecting how they structure their research portfolios and set priorities amid shifting health challenges. To capture this empirically, we estimate fixed-effects regressions of national publication counts on DALYs, controlling for country, disease, and year. The resulting coefficient, which we term elasticity, measures the percentage change in publications associated with a 1% change in DALYs (see Methods). This framework allows us to examine how responsiveness varies across countries, diseases, and over time.

Similar to other areas of national capacity, marked disparities exist in endemic research responsiveness. Responsiveness tends to rise with GDP per capita, but high-income countries vary widely: countries such as Japan, Saudi Arabia, and the United States exceeding the trend, while Luxembourg, Switzerland, and Iceland fall below it (Figure 2a). Comparable disparities also emerge across diseases. Estimated as baseline deviations from burden-expected levels, we show that conditions more prevalent in high-income settings, such as cardiovascular and digestive diseases, receive about 18% more publications than the average disease with comparable burden, while nutritional deficiencies and maternal and neonatal disorders attract 11–13% fewer (Figure 2c).

Building on our earlier observation of HIV/STI and AIDS, we observe a clear increase in responsiveness over time, though the pace of growth differs sharply across country income groups. Year–income-group specific elasticities show that responsiveness roughly doubled in high-income countries between 1990 and 2021, nearly tripled in upper-middle-income countries, and rose more modestly in lower-middle- and low-income groups. The rise offers grounds for cautious optimism, yet gaps persist: high- and upper-middle-income countries consistently lead, while lower-income groups remain well behind (Figure 2b).

These gains are not driven by research on all diseases equally. To further unpack which diseases contribute most to rising responsiveness within each income group, we construct a counterfactual that holds post-2017 publication counts constant at 2017 levels and then re-estimate the change in elasticity. This approach allows us to isolate the contribution of specific diseases to overall growth in responsiveness. The decomposition shows that the increase in responsiveness among low-income countries is driven largely by more research on neglected tropical diseases and malaria, respiratory infections and tuberculosis, and maternal and neonatal conditions (Figure 2d). By contrast, in high-income countries, the gains come primarily from expanded research on cardiovascular diseases and neoplasms, a pattern also observed in upper-middle-income countries. Overall, this shows that increasing endemic research responsiveness stems from more research on the conditions most responsible for disease burden in each income setting.

2.3 The portfolio preference and impact of research funders

Funding agencies play a central role in shaping the direction and focus of scientific research, an influence that is especially consequential for the Global South, where external resources often determine whether local health challenges receive adequate attention. To examine how funders shape disease research and contribute to responsiveness growth, we classified all 32,437 funders listed in the 2024 OpenAlex snapshot into six categories: direct government, independent public, private corporate, private philanthropic, aca-

demic institution, and hybrid, and study their differences in both funding tilt across diseases and their impact on responsiveness growth.

To examine how funders allocate resources across diseases, we calculate each type's funding tilt, defined as the log-odds ratio of a disease being funded by that type relative to the overall average. Funding tilt varies sharply across funder types, highlighting their divergent funded disease research portfolios (Figure 3a). Private philanthropic funders tilt strongly toward nutritional deficiencies as well as HIV/AIDS and neglected tropical diseases, which disproportionately affect low-income countries. By contrast, corporate funders allocate relatively little to these conditions and instead show a strong orientation toward non-communicable diseases, particularly cardiovascular diseases, neoplasms, and diabetes and kidney diseases, which dominate in high-income settings.

The consequences of these portfolio choices are also visible in their contributions to endemic responsiveness growth (Figure 3b). Similar to our earlier counterfactual analysis, we estimate each funder type's impact by assuming that papers it supported were absent and re-calculating how elasticity changed between 2014–2017 and 2018–2021. We observe that, in low-income countries, responsiveness gains rely heavily on philanthropic and government support: without papers funded by them, the estimated responsiveness growth would shrink by 38% and 32%, respectively. Similar patterns are observed in lower-middle-income countries. In high-income countries, however, the pattern reverses: removing papers funded by governments or philanthropic organizations significantly reduces responsiveness growth, i.e., funding research on conditions not most prevalent in their own settings but aligned with a broader global health agenda. Put differently, these two funder types play a central role in sustaining research that extends beyond conditions prevalent in high-income settings, by directing attention toward diseases that impose heavier burdens in low-income contexts.

2.4 Research responsiveness in emergency: event-study evidence from WHO disease outbreak news

Outbreaks create another form of disease burden, producing sudden shocks that often leave countries unprepared. To assess how research systems respond under such conditions, we focus on emergency responsiveness, which is the capacity to scale up research activity in the wake of acute health shocks. We study this using nearly three thousand outbreaks announced by the WHO. For each outbreak, we estimate a difference-in-differences event–study specification that compares changes in research activity on the focal disease in the affected country before and after the outbreak to contemporaneous changes in other diseases and other countries. The model includes country–disease fixed effects to absorb persistent heterogeneity in research portfolios, disease–year effects to capture global shocks to individual disease areas, and country–year effects to account for overall shifts in national output. The resulting estimates capture the increase in the likelihood of publication on the focal disease after an outbreak, reflecting the capacity of national research systems to mobilize and deliver science in response to health emergencies.

Tracing evolution over time, Figure 4a shows once again increasing emergency research responsiveness: Outbreak-triggered research surges were minimal in the 1990s and modest in the 2000s, but became pronounced in the 2010s, with outbreaks generated substantial and near immediate increases in research output activity. This pattern suggests that research effort allocation may increasingly anticipate disease risks. We show the robustness of this finding to a broad menu of modeling choices in the Appendix by documenting a generally increasing trend in before–after shifts in event–study coefficients across years and specifications.

Not all disease alerts, however, trigger substantial research responses. Figure 4b reports the mean pooled estimate across alerts and highlights considerable variation. Emerging and high-profile pathogens such as COVID-19, Ebola, Zika, and measles gen-

erated the largest increases in related research activity, whereas endemic or historically persistent conditions such as malaria, diphtheria, and diarrheal diseases elicited weaker responses. These patterns suggest that research systems tend to mobilize disproportionately around novel or highly salient threats, where existing knowledge is limited and rapid scientific exploration is required.

Which factors explain variation in emergency research responsiveness? To answer this, we utilized Random Forests to predict the size of the estimated research response (pooled estimate) using a broad set of country-, disease-, and alert-level features. Figure 4c shows that four predictors stand out: population structure (share of senior citizens and urban residents), internet penetration, research capacity (average citations per publication), and the length of outbreak alerts. These results indicate that demographic structure, digital infrastructure, research preparedness, and the nature of alerts are important factors in explaining a nation's ability to mobilize research during health crises.

3 Discussion

Using a large-scale, LLM-constructed geographical knowledge graph linking 308,135 medical research articles to the diseases they study and the countries whose data they analyse, we distill three main insights. First, research responsiveness to domestic health needs has more than doubled since 1990, reflecting a global shift toward burden-aligned agendas. Second, participation remains uneven and nuanced: Africa, for instance, provides nearly one-third of study contexts for neglected tropical diseases and malaria but only 14% of authorship, underscoring persistent asymmetries between where research is relevant and who conducts it. Third, funding priorities diverge, with philanthropic foundations emphasizing diseases prevalent in lower-income regions, corporate RD concentrating on conditions of advanced economies such as diabetes and kidney disorders, and public agencies occupying an intermediate position. Finally, quasi-experimental evidence from 3,134 WHO outbreak alerts shows that sudden health shocks trigger rapid and lasting surges of scientific attention, with responses becoming markedly stronger in the 2010s than in the 1990s.

Implications for global health policy Our burden-responsiveness estimates provide a quantitative complement to prior calls to align research portfolios with health need (Røttingen et al., 2013) and to address structural inequities in who leads and benefits from global health science (Hedt-Gauthier et al., 2019; Abimbola et al., 2021). The upward trend in responsiveness is encouraging. It suggests that efforts to expand research capacity through open-access platforms, regional training programmes and cross-border collaborations are bearing fruit. Nonetheless, the persistent gap between where disease burden is highest and where research is produced underscores the need for targeted investments. Philanthropic donors play an outsized role in plugging critical gaps for neglected diseases.[4] Sustaining and scaling those investments will be vital if the global

[4]This pattern was also noted in the 2024 WHO Global Report on Neglected tropical diseases (World Health Organization, 2024a).

community is to meet the 2030 Sustainable Development Goals for health. Public agencies and multilateral funders can use burden-adjusted publication metrics to identify under-served diseases and geographies, directing new grants or fellowships to the places of greatest need, though care must be taken not to reinforce "safe-topic" incentives in evaluation (Wang, Veugelers, & Stephan, 2017).

Heterogeneity and plausible mechanisms Why has research become more burden-responsive? Two complementary channels are likely at work. On the one hand, "capacity deepening" in emerging economies (exemplified by rapid growth in China's, India's and Brazil's biomedical infrastructure) has raised the global baseline of papers that both originate in and address high-burden settings. On the other hand, "capacity borrowing" by international consortia and "parachute" research teams during major outbreaks (e.g. Ebola 2014, Zika 2016) has temporarily amplified study output in smaller low-income countries. Disentangling these channels is an important task for future work, for example by combining our graph with author-affiliation trajectories, grant-award data, or the emerging North–South collaboration metrics documented by the Nature Index (Nature, 2023). The contrast between corporate and philanthropic tilt is consistent with market-size–driven innovation in pharmaceuticals (Acemoglu & Linn, 2004) and with time-to-payoff considerations in clinical development (Budish et al., 2015), helping explain why non-communicable diseases endemic to high-income settings attract outsized attention

Limitations and future directions Several caveats apply. First, although we began with 524 leading medical journals from JCR 2023, successive filters for English-language research articles with abstracts and at least one citation reduced the sample to about 308,135 papers. Access to full texts (rather than abstracts alone) would likely yield richer geolocation of study cohorts and finer detail on data subjects. Second, our MeSH–ICD–GBD cross-walk covers roughly 86% of articles; certain high-profile pathogens (e.g. MERS coronavirus) lack direct GBD codes and are therefore excluded from burden-adjusted

regressions, but not from the analysis of WHO alerts. Third, we rely on OpenAlex's coverage of English-language journals, which may undercount regional titles and grey literature, particularly in parts of Africa and South America. Fourth, our focus is on the geography and volume of research; we do not assess methodological quality or synthesize treatment effects as a systematic review would. Finally, while LLMs enable scalable extraction of diseases and contexts, further validation (especially for very rare conditions) would strengthen confidence in tagging accuracy.

In conclusion, we stress that the geography of medical knowledge is shifting: global research systems are becoming more attuned to local health needs, yet significant misalignments remain, highlighting the need for transnational cooperation and coordination. By fusing bibliographic data with disease-burden data through modern language models, we provide both a diagnostic and a dynamic platform. As language models and open-data initiatives evolve, this framework can be updated in near real time, guiding funders, policymakers and researchers toward a more equitable and effective global health agenda.

4 Methods

4.1 Paper Sample

To construct a bibliographic dataset that reflects global research output in medical research, we sourced data from OpenAlex (Priem, Piwowar, & Orr, 2022), which curates the largest scholarly documents and various information such as abstracts, authorship, and funding source.

Journal selections. We focused on 524 journals listed in the Journal Citation Reports (JCR) by Clarivate in 2023, selecting journals published in English and classified under three clinical medicine categories: *Medicine, General & Internal*, *Medicine, Research & Experimental*, and *Pharmacology & Pharmacy*. This selection includes leading journals such as *The Lancet*, *The New England Journal of Medicine* (NEJM), and *JAMA*.

Paper filters. We further filtered this subset using the paper-level classification of broad domains provided by OpenAlex, retaining only papers in the domain of Health Sciences. We also included only research articles (not reviews or other publication types), so that only papers with the publication type "article" were kept. In addition, we restricted our dataset to papers that have at least one citation to focus on citable items. Since our analysis focuses on the geographical attribution of research efforts in authorships, we included only papers for which OpenAlex provides author country information. This filtering process resulted in a dataset of about 603k papers.

Geographical mention filters. To identify papers that potentially mention geographic locations, we used two methods. The first is named entity recognition, applying the `dslim/distilbert-NER` model to detect location (LOC) and organization (ORG) entities. The second is a dictionary-based approach using the GeoText Python package, which matches country and city names from GeoNames.org. We kept papers where at least one geographic entity was identified by either method. After this step, we obtained a

final set of about 308,135 papers. We show the complete dataset construction steps and filtering procedure in Appendix A.

4.2 LLM-based extraction of paper metadata and diseases

We transform every article's title and abstract into a structured record with a *single* call to `gpt-4o-mini-2024-07-18`. The model receives the raw text together with a JSON schema that enforces strict field names, default values, and type checking. This large-scale retrieval of information from academic papers follows the methodology of Garg and Fetzer (2025b) and is reproduced detailed in Appendix B.

Paper–data metadata. For each paper the LLM reports whether data were used, the unit of analysis (with optional details), the temporal , start year, end year, an ISO-3 list of geographical context countries, and the ownership of the dataset (e.g. open source, private company). Missing information is filled with schema defaults (e.g. `NA` or empty arrays), guaranteeing parseability at scale.

Disease tagging and normalisation. The same response contains an array of free-text disease strings extracted from the abstract. Because disease names vary in spelling, abbreviation, and phrasing, we use text embeddings to enable semantic matching rather than relying on exact text overlap. We convert each string into a 1,024-dimensional embedding using a leading embeddings model (OpenAI's `text-embedding-3-large`)[5] and assign it to the MeSH descriptor with the highest cosine similarity. The resulting MeSH codes are later mapped to ICD-10 and Global Burden of Disease (GBD) level-2 causes, supplying the bridge between research effort and national DALYs.

All JSON outputs pass automatic schema validation before entering our analysis pipeline. This end-to-end retrieval yields a reproducible, disease-coded, geography-aware knowledge graph for the full corpus of 308,135 articles. Our approach is informed

[5]More details about the model at (https://platform.openai.com/docs/guides/embeddings)

by recent work showing that large language models encode clinically relevant knowledge (Singhal et al., 2023), yet require careful domain evaluation and guardrails (Wornow et al., 2023; Garg & Fetzer, 2025a), motivating our strict-schema, embeddings-to-MeSH normalization.

4.3 Embeddings to match to medical data nomenclature

While our LLM-based retrieval pipeline provides structured representations of biomedical relationships and disease information, these outputs are initially in free-text form. However, medical data is typically denominated in controlled vocabularies. To bridge this gap and ensure compatibility with external datasets, we match our free-text representations (both the nodes in our knowledge graph and the diseases extracted at the paper level) to established medical nomenclatures, specifically MeSH codes.

Controlled vocabularies. MeSH (Medical Subject Headings) is a hierarchical vocabulary developed by the United States National Library of Medicine (NLM) for indexing biomedical literature. It assigns an alphanumeric code to each medical term, organizing descriptors into 16 broad categories such as Anatomy, Organisms, Diseases, and Chemicals and Drugs. This hierarchical arrangement, spanning up to thirteen levels, enables both broad and precise literature searches. The use of MeSH terms has been shown to enhance search precision and recall compared to simple text-word searches (National Library of Medicine, 2023; Sampson et al., 2009). In our study, MeSH serves as a comprehensive discrete node space that standardizes the free-text outputs from our LLM and facilitates the mapping of additional medical data.

For disease-specific information, we also employ the International Classification of Diseases, 10th Revision (ICD-10). Established by the World Health Organization (WHO), ICD-10 provides a standardized system for coding diseases and clinical conditions. Its widespread adoption in mortality and morbidity statistics, healthcare reimbursement, and decision-making makes it essential for aligning our extracted disease information

with global health metrics, such as Disability-Adjusted Life Years (DALYs). In our framework, ICD-10 codes are mapped from disease-specific MeSH nodes, ensuring compatibility with global disease burden datasets.

The Unified Medical Language System (UMLS) plays a critical bridging role by integrating multiple biomedical vocabularies, including MeSH and ICD-10. UMLS facilitates seamless translation between these systems, ensuring that our free-text extractions can be reliably matched to standardized terminologies.

Semantic matching. To perform the matching, we leverage text embeddings. Using OpenAI's `text-embedding-3-large` model, we convert our free-text fields into 1024-dimensional embeddings. These embeddings capture the semantic nuances of biomedical language, allowing us to compare our extracted text with precomputed embeddings of MeSH descriptors. We compute cosine similarity between the generated embeddings and the MeSH embeddings, assigning each free-text entry the MeSH descriptor with the highest similarity. This process is applied both to the nodes derived from our knowledge graph extraction and to the disease information extracted at the paper level. In some instances, we generate ensemble representations by combining raw text with its standardized form to further refine the matching process.

Key decisions in this approach include the choice of a high-dimensional embedding model and the use of cosine similarity as a robust metric for semantic comparison. These choices ensure that our mappings are both precise and consistent. By standardizing our dataset through these embeddings, we enable reliable cross-paper comparisons and facilitate the integration of our research outputs with global health datasets, ultimately supporting a nuanced analysis of the relationship between medical research and disease burden.

4.4 Global Burden of Disease

The Global Burden of Disease (GBD) is a comprehensive research program that systematically quantifies the health impact of diseases, injuries, and risk factors worldwide by analyzing mortality and disability across different demographic groups (Vos et al., 2020; Global Burden of Disease Collaborative Network, 2024). The GBD uses metrics such as prevalence rates and Disability-Adjusted Life Years (DALYs) to assess the overall burden of disease, enabling comparisons across populations and over time. This approach helps identify health priorities and inform policy decisions to improve global health outcomes. In our study, we focus on the DALYs metric as an indicator of disease burden.

The GBD dataset quantifies the magnitude and impact of major diseases for 204 countries and territories from 1990 to 2021. The identified disease burdens are structured in a four-level hierarchy, starting from the broadest category with nine major burden groups and progressing to the most detailed category with 175 distinct causes. In some of our descriptive analyses, we focus on the second level, which includes 22 causes such as HIV/AIDS and sexually transmitted infections, cardiovascular diseases, and neglected tropical diseases and malaria.

4.5 Constructing the MeSH-ICD-GBD crosswalk

An essential task in this study is to connect a country's research effort on specific diseases with the actual disease burden it faces. To achieve this, we constructed a crosswalk that links MeSH codes assigned to individual papers to ICD codes, and then connects these to disease burdens as identified in the GBD dataset. The overall crosswalk is built upon these two linkages.

MeSH and ICD linkage. The first linkage between MeSH and ICD codes is based on the ontology provided by the Unified Medical Language System (UMLS). We employed two methods to create the linkage: direct mapping and indirect mapping. In the direct mapping approach, we used two sources of information. The first source is the UMLS

Metathesaurus, where we identified MeSH and ICD-10-CM entries that share the same Concept Unique Identifier (CUI) in the UMLS Metathesaurus (MRCONSO.RRF), resulting in a straightforward mapping. The second source is the SNOMED CT to ICD-10-CM Map provided in UMLS[6]. We first linked MeSH to SNOMED CT via shared CUIs, then mapped SNOMED CT codes to ICD-10-CM codes using the SNOMED CT to ICD-10-CM Map. The SNOMED-to-ICD mapping includes a `mapRule` column, indicating whether the mapping is unconditional (TRUE) or subject to conditions such as age or gender. In our analysis, we only used unconditional mappings.

For indirect matching, we focused on MeSH codes not directly linked to ICD codes in the previous steps. Here, we relied on UMLS inter-concept relationships, which connect concepts through different semantic associations. We primarily used four types of relationships to link ICD codes to related MeSH codes: RB ("broader than"), which connects to more general concepts; RN ("narrower than"), linking to more specific concepts; PAR (Parent), identifying the immediate "is-a" parent in the MeSH hierarchy (where a concept is a subtype of its parent); and CHD (Child), indicating immediate "is-a" children (where a concept serves as a broader type for its child). In the last step, we manually identified the corresponding ICD codes for 23 MeSH codes that were not linked in the previous steps but are both frequent and important in our dataset. This was achieved by searching online documentation and reference sources for each MeSH code, such as D053118 (Influenza A Virus, H1N1 Subtype) and D045473 (Severe acute respiratory syndrome-related coronavirus).

ICD and GBD linkage. For the second linkage between ICD codes and GBD disease categories, we used the mappings provided by the Global Burden of Disease dataset[7] which specified the corresponding ICD ranges for each disease category. This allowed us to further establish the MeSH-ICD-GBD crosswalk.

Using this crosswalk, we were able to link 86% of papers to level-2 disease causes

[6] https://www.nlm.nih.gov/research/umls/mapping_projects/snomedct_to_icd10cm.html
[7] https://ghdx.healthdata.org/record/ihme-data/gbd-2021-cause-icd-code-mappings

defined in the GBD dataset.

4.6 WHO Disease Outbreak Alerts

To capture real-time public health events, we assembled a comprehensive dataset of WHO Disease Outbreak News alerts by downloading the full content from each page on the WHO website (https://www.who.int/emergencies/disease-outbreak-news). The resulting dataset comprises 3,134 alerts reported on 2,481 unique dates, covering a period from 1996 up to the most recent alert dated 2025-01-20. Each alert record includes key metadata, such as the alert title, publication date, a unique link identifier, and the full text content describing the outbreak.

LLM-based information retrieval. Each alert is passed (one call per page) to *gpt-4o-mini* together with a JSON schema that enforces strict field names to guarantee parseability. The model returns an array of tuples, one for every disease–geography–time mention, with slots for *disease*, *day*, *month*, *year*, *cases*, *deaths*, calculated case-fatality ratio, and the free-text *geography*. This flexible extraction accommodates alerts that report multiple disease-geography-time tuples, ensuring that each distinct mention is preserved as an individual record. Full prompts and associated deliberations available in Appendix Section C.

Normalising diseases. All free-text disease strings are embedded with OpenAI's *text-embedding-3-large* model (1,024 dimensions) and matched to the nearest MeSH descriptor by cosine similarity. Matched MeSH codes are later cross-walked to ICD-10 and to level-2 Global Burden of Disease (GBD) causes, providing a common ontology for event-study regressions.

Normalising geographies. Alert prose often mixes labels such as "Central Africa" with fine-grained places ("Kasese District"). We therefore run a second LLM call, once for

every distinct geography string (approximately 1,500), that decomposes the text into *city*, *country* (ISO-3), *continent*, *region* and an *other* field. Each unique geography receives a deterministic `geography_id`, allowing many-to-one joins back to the alert table.

Temporal refinement. To further ensure data completeness, we implemented a procedure to fix missing temporal details. In cases where the 'year' field was absent from a disease-geography-time entry, one option would have been to impute the year from another entry within the same alert. However, internal time mentions can be inconsistently captured or vary across different parts of the alert. Instead, we uniformly extracted the year from the publication date of the alert. This approach leverages the reliability of the alert date. By design, these alerts reflect developing events in real time, ensuring consistency and providing a robust approximation for missing temporal information.

Limitations. While this approach is scalable, it has some limitations: (i) Case/death counts follow no uniform template and occasionally conflate suspected and confirmed numbers. (ii) Alerts covering several diseases inherit any ambiguity present in the source text. (iii) When geography strings are continental ("Europe") finer locations remain empty, which may under-state sub-national heterogeneity. Despite these caveats, the pipeline delivers a high-frequency, geography-resolved panel of 3,000+ outbreak shocks that powers the event-study design.

Integration with epidemiological data. To integrate our outbreak alerts with broader epidemiological datasets, we mapped the standardized MeSH codes to ICD-10 codes via the Unified Medical Language System (UMLS) and subsequently linked these codes to global disease burden metrics such as Disability-Adjusted Life Years (DALYs).

Integration with MeSH tree nomenclatures. In addition, we merged our alerts data with a MeSH tree mapping file to convert unique MeSH identifiers into hierarchical levels, thereby adding further granularity for our analyses. Each MeSH unique ID (e.g.,

D00028384) is potentially associated with multiple tree structures. Our algorithm selects the modal coarse tree at the most general (level 1) category; in the event of a tie, a random selection is made. Within this modal category, we then select the modal level 2 hierarchy, and subsequently the modal level 3 hierarchy, again resolving ties randomly. The final dataset thus includes one column for the MeSH unique ID and separate columns for the level 1, level 2, and level 3 tree codes. These hierarchical mappings are particularly valuable for incorporating fixed effects and investigating heterogeneity across different disease categories in our econometric analyses.

Uncertainty measurement. We also quantified uncertainty within the alert texts using a dictionary-based approach with a curated corpus from the Vermont Uncertainty Corpus (https://vermontconversationlab.com/uncertainty-corpus/). The corpus includes an overall uncertainty dictionary along with three subgroups of uncertainty terms: hedging, possibility, and confusion. These terms were converted into regular-expression patterns and applied to the alert content to compute counts and derive shares relative to the total word count. For example, the average share of uncertainty terms was 0.0101, hedging terms 0.0030, possibility terms 0.0041, and confusion terms 0.0002. These metrics provide a quantitative measure of uncertainty expression within the alerts, allowing for comparisons across different outbreak reports.

In short, we undergo following steps: data collection, LLM-based extraction, embedding-based matching, geographic refinement, temporal correction, integration with epidemiological datasets and hierarchical mapping, and uncertainty quantification. Together, these steps ensure that our WHO Alerts dataset is robust, standardized, and fully prepared for our event study analysis.

4.7 Descriptive analysis

We begin our study with a set of descriptive analyses designed to provide a broad perspective on the data. First, we map the geographical distribution of research output,

public health alerts, and disease burden across different diseases and regions, highlighting global patterns and disparities. Second, we investigate how the evenness of country contributions in research authorship and contextual evidence has evolved over time, revealing shifts in the diversity of participation. Third, we examine the allocation of funding across disease areas, with particular attention to funding tilt, i.e., how different types of funders prioritize or under-support specific diseases. These analyses collectively establish the empirical foundation for the more detailed modeling and event study analyses that follow.

The distribution of research, disease burden, and alerts Figure 1a presents the distributions of research papers, WHO alerts, and disease burden across 15 level-2 GBD disease categories. The left panel displays the count of unique papers associated with each disease category, while the middle panel shows the number of unique alerts. The right panel reports the mean DALYs rate for each disease category, averaged across all countries, territories, and years in the GBD dataset.

To understand the distribution of research efforts and contexts across continents, Figure 1c displays the composition of papers by continents. We present the percentage of papers by continents across disease categories, for both authorship and research context. Since a single paper may be classified under multiple disease categories, we employed fractional counting: for each paper, we assigned a weight of 1/N to each relevant category, where N is the number of categories associated with that paper. This approach ensures that the total contribution of each paper across categories sums to one.

The concentration of disease research over time To trace the evolution of global research output across diseases (Figure 1b), we employed the Herfindahl-Hirschman Index (HHI) to quantify the concentration of disease research output and the concentration of countries and territories mentioned as contextual references. The HHI for a given disease d in year t is defined as:

$$HHI_{d,t} = \sum_{i \in C} s_{i,d,t}^2$$

where $s_{i,d,t}$ measures the share of research output produced by author countries or by paper context countries i on disease d in year t. A higher HHI value indicates greater concentration, meaning a few countries account for most of the output or contextual mentions, while a lower value indicates increased evenness across countries. The trend lines are smoothed using LOWESS with a fractional parameter of 0.2. To further understand the observed rise in evenness for substance use disorders since 2015, we present in Figure D3 the percentage of articles on this disease contributed by the ten countries with the highest publication counts.

Quantifying the funding tilt on disease research To analyze the allocation of funding across different disease categories (Figure 3a), we examined the funding tilt in disease research by various funding sources, which captures whether a particular disease receives disproportionately more or less funding from a specific type of funder compared to the overall average. Our approach consisted of two main steps. First, we classified 32,438 funding institutions curated in OpenAlex into six categories using a large language model (see Appendix D for details). The categories are defined as follows:

- **Direct Government:** Government agencies fully funded by public resources.
- **Independent Public:** Organizations that receive public funds but operate with independent decision-making.
- **Private Corporate:** For-profit corporate entities, often funding research as part of strategic or corporate social responsibility initiatives.
- **Private Philanthropic:** Philanthropic or non-profit organizations that fund research through endowed or donated resources.

- **Academic Institution:** Academic or research institutions funding research internally or through affiliated bodies.
- **Hybrid:** Entities that combine multiple funding sources (public, private, etc.) or do not clearly fit into the above categories.

For each institution, the classification prompt was submitted five times, and the modal result was selected. The overall agreement across prompts was high, with 96–99% consistency, indicating robust and stable category assignments. Full prompt text, JSON schema and validation statistics are documented in Appendix Section D.

The funding tilt for a disease d toward a specific funder type f is quantified by the log-odds ratio (LOR):

$$LOR(d, f) = \ln\left(\frac{p_{d,f}(1 - p_f)}{(1 - p_{d,f})p_f}\right)$$

where $p_{d,f}$ denotes the proportion of papers about disease d that are funded by type f, and p_f is the proportion of all papers (across all diseases) funded by type f.

A higher (positive) value of $LOR(d, f)$ indicates that disease d is more likely than average to be funded by type f. Conversely, a lower (negative) value suggests that disease d is less likely to receive funding from that funder type.

4.8 Regression analysis

To further explore the alignment between research output and health needs, we employ fixed-effects regression models structured around three main research questions: (i) Which disease burdens receive disproportionate research attention relative to their health impact; (ii) How responsive is a country's research output to its own pressing disease burden; and (iii) How has this responsiveness evolved over time. Below, we introduce each regression model.

Relative research effort on disease. To identify which diseases receive more or less research attention than expected, we focus on the disease fixed effects from the regression:

$$\log(\text{Publications}_{ict} + 1) = \beta_1 \log(\text{DALYs}_{ict}) + \gamma_c + \delta_i + \lambda_t + \epsilon_{ict}.$$

Here, γ_c are country fixed effects, δ_i are disease fixed effects, and λ_t are year fixed effects. The disease coefficients δ_i are estimated using effect coding (sum-to-zero), so each coefficient represents the deviation in research attention for disease i relative to the average across all diseases, conditional on disease burden, country, and year. To aid interpretation, we convert these coefficients into percentage differences using

$$\text{pct_diff} = 100 \times \left(e^{\delta_i} - 1\right).$$

In Figure 2c, the x-axis shows these percentage differences: for example, a value of $+10$ means a disease receives about 10% more publications than the cross-disease average, while a value of -10 means about 10% fewer publications under the same conditions.

Temporal and Income-Group Variation in Research Responsiveness. This model examines how the responsiveness of research output to disease burden varies over time and across countries at different income levels. Specifically, we estimate a fixed-effects regression of the form:

$$\log(\text{Publications}_{ict} + 1) = \sum_t \sum_g \beta_{1tg} \log(\text{DALYs}_{ict}) \cdot D_t \cdot G_g + \gamma_c + \delta_i + \lambda_t + \epsilon_{ict}.$$

The interaction terms between $\log(\text{DALYs}_{ict})$, year indicators D_t, and income group indicators G_g allows the elasticity of research output with respect to disease burden to vary across both time and income level. The coefficients β_{1tg} capture these year–income group–specific elasticities. This specification enables us to assess whether countries at

different income levels have followed divergent trajectories in aligning research activity with health needs over time.

In the main article (Figure 2b), we assign countries and territories to World Bank income groups based on their 1990 classification and hold these groups fixed over the observed period. This approach ensures that comparisons are not affected by countries moving across income categories over time. As a robustness check, we re-estimated the model using 2021 classifications and obtained similar observations (Figure D4).

Country-Level Responsiveness to Disease Burden. To assess whether countries or territories differ in how their research output responds to disease burden, we estimate a fixed-effects regression model that allows for country-specific elasticity. The model is specified as:

$$\begin{aligned}\log(\text{Publications}_{ict} + 1) = \beta_1 \log(\text{DALYs}_{ict}) \\ + \sum_c \beta_{2c} \left[\log(\text{DALYs}_{ict}) \times C_c\right] \\ + \gamma_c + \delta_i + \lambda_t + \epsilon_{ict}\end{aligned}$$

where C_c is an indicator variable for country c. The interaction coefficients β_{2c} capture the responsiveness (elasticity) of each country's research output to changes in disease burden, relative to the global average. As before, fixed effects are included for country, disease category, and publication year. This approach enables us to quantify and compare how different countries align their research efforts with health needs.

4.9 Event study analysis: research response to WHO alerts

We employ an event study difference-in-differences (DID) framework to identify the causal effect of World Health Organization (WHO) public health alerts on subsequent scientific research activity. The core intuition behind this design is to treat a WHO alert

on a specific disease in a particular country and year as an exogenous shock, namely, an unexpected event that raises the salience and perceived urgency of the disease in that context. By comparing research activity before and after the alert, as well as between alerted and non-alerted diseases and countries, we can estimate the extent to which scientific attention responds to sudden public health emergencies. The approach is broadly following FetzerGarg2025$_{m}ediacoverage$.

Using this framework, we can answer several important questions: Do WHO alerts prompt a measurable increase in research activity on the targeted disease in the affected country? How quickly does the scientific community respond to emerging public health threats? This design allows us to disentangle the dynamic and heterogeneous effects of public health alerts on research, providing insight into the responsiveness and allocation of global scientific effort.

Counterfactual construction and identification strategy. A central challenge in estimating the impact of disease shocks on research responsiveness is constructing a credible counterfactual, i.e., which research activity would have been observed in the absence of a WHO alert. In our analysis, the "treatment" is defined as a WHO alert issued for a particular disease in a specific country and year. To address this, we exploit variation in the timing and targeting of alerts across both diseases and countries. Our approach is grounded in the logic of comparative interrupted time series: by comparing changes in research output not only before and after a treatment (alert), but also between treated (alerted) and control (non-alerted) country–disease pairs, we can distinguish the impact of alerts from general trends or shocks.

We operationalize this strategy through several distinct comparison groups:

- **Between diseases, between countries:** We compare alerted disease–country pairs to all other disease–country combinations, providing a comprehensive control group for benchmarking the treatment effect. We use this design as our primary analysis in the main text.

- **Within disease, between countries:** We compare research output for a given disease in the country experiencing the WHO alert (the "treated" unit) to research on the same disease in countries not subject to the alert (the "control" units). This controls for global trends in research on that disease.

- **Between diseases, within country:** We compare research output for the alerted (treated) disease in the alerted country to research on other diseases within the same country. This controls for country-specific changes in research funding, capacity, or policy that may affect all diseases.

By leveraging these multiple dimensions of variation, our identification strategy aims to mitigate confounding factors that may operate at the country, disease, or global level. In essence, we combine temporal contrasts (before vs. after the treatment) with cross-sectional contrasts (treated vs. control diseases/countries) to construct a counterfactual approximation of what research activity might have looked like in the absence of the alert.

This approach allows us to isolate the impact of WHO alerts on research output, distinguishing genuine responses to heightened disease salience from underlying secular trends or shared shocks across the research ecosystem. In the main article, we present results using the between-disease within-country specification as our primary design. Results from the alternative specifications (within-disease between-countries and between-diseases within-countries) are reported as robustness checks.

Centering and fixed effects. A central concern in estimating the research response to WHO alerts is the presence of confounding factors that vary systematically across countries, diseases, and years. To address this, we incorporate fixed effects that absorb these sources of variation and prevent spurious correlations from biasing the estimated alert effect.

In the main article, we rely on a specification with three-way fixed effects: country-by-disease (`iso_disease`), disease-by-year (`disease_year`), and country-by-year (`iso_-`

`year`). This combination removes time-invariant differences between country–disease pairs, accounts for global shifts in attention to particular diseases, and captures temporal shocks affecting specific countries. With this implementation, we aim to isolate the variation associated with WHO alerts as much as possible and to generate estimates that can be interpreted as shifts in research activity following alerts, rather than artifacts of pre-existing trends.

As robustness checks, we also estimate models with alternative fixed-effect structures:

- Between diseases, between countries
  - Country-by-disease and event-time fixed effects (`iso_disease + time_to_event`)
  - Treatment and event-time fixed effects (`treat + time_to_event`)
- Between diseases, within country
  - Treatment and event-time fixed effects (`treat + time_to_event`)
  - Country-by-disease and event-time fixed effects (`iso_disease + time_to_event`)
- Within disease, between countries
  - Treatment and event-time fixed effects (`treat + time_to_event`)
  - Country-by-disease and event-time fixed effects (`iso_disease + time_to_event`)

Attribution methods. Research output is attributed to countries or territories using two complementary approaches: (1) author affiliations, mapping institutional addresses in publications to country codes, and (2) contextual mentions, identifying country names referenced in article titles or abstracts. This dual attribution provides robust assignment of research activity to national contexts and provide two distinct connotations of research response.

Data resolution. Our analysis examines whether research output increases when a disease outbreak occurs in a given country. Defining "disease," however, is not straightforward: medical conditions are typically organized hierarchically, belonging to broad categories while also subdividing into more specific conditions. To account for this structure, we rely on two complementary taxonomies: the MeSH tree, which reflects how biomedical literature is indexed, and the Global Burden of Disease (GBD) hierarchy, which organizes diseases from broad categories down to specific causes.

As discussed in previous sections, we integrated the alert-level MeSH descriptors into the MeSH tree nomenclature and constructed a crosswalk linking MeSH to the Global Burden of Disease (GBD) hierarchy. Both systems are organized as trees, with broad categories branching into more specific conditions. This structure allows us to define disease fixed effects and heterogeneity at different levels of aggregation, while ensuring robustness across alternative classification schemes.

As an illustration, the disease *Type 2 Diabetes Mellitus* (D003924) is represented in both systems:

MeSH tree:
L1 (C18) Endocrine System Diseases
└─ **L2** (C18.452) Metabolic Diseases
  └─ **L3** (C18.452.394) Glucose Metabolism Disorders
    └─ **L4** (C18.452.394.750.149) Diabetes Mellitus, Type 2

GBD hierarchy:
L1 Non-communicable diseases
└─ **L2** Diabetes and kidney diseases
  └─ **L3** Diabetes mellitus
    └─ **L4** Diabetes mellitus type 2

We also use a flat identifier, denoted *MeSH dis*, which refers directly to the MeSH code itself without hierarchical aggregation. In this specification, the comparison group consists of publications tagged with all other MeSH codes.

Using both taxonomies, and considering different levels of aggregation, provides robustness to alternative ways of organizing diseases and ensures that our findings are

not an artifact of a particular classification scheme.

Extensive vs. intensive margin. Finally, we assess both the extensive margin (the probability that any research is produced on a country-disease pair in a given year) and the intensive margin (the volume of research output), providing a comprehensive view of how WHO alerts influence not only the likelihood but also the scale of subsequent research activity.

Pooled and dynamic estimates. The primary output of our event study analysis is the estimated effect of WHO alerts on research activity, which quantifies how alerts shift publication patterns relative to control groups defined in previous sections. We take two complementary forms of estimation:

(i) **Pooled estimate** Collapses all post–alert periods into a single coefficient, providing a parsimonious summary of the average increase in research following an alert which quantifies the overall magnitude of responses.

(ii) **Dynamic estimate** Estimates a separate coefficient β_τ for each relative time period five yeras before and after the alert. These coefficients can be interpreted as dynamic treatment effects: they represent the additional research attention causally attributable to the alert at each point in time, beyond what would have been expected from historical patterns and broader trends. In other words, the β_τ trace out the temporal profile of research responsiveness to public health shocks, with significant positive values indicating periods of detectable increases in research output and values near zero suggesting no effect above baseline.

4.10 Random Forest analysis of predictors of responsiveness

To complement the event–study analysis, we investigate which factors explain variation in research responsiveness across countries and diseases. For this purpose, we employ Random Forest regression models where the outcome variable is the pooled event–study

estimate, and the predictors span research capacity, socioeconomic conditions, governance, health system preparedness, disease burden, and textual features of outbreak alerts. This machine–learning approach allows us to assess the relative importance of a wide set of potential drivers while accommodating non-linear relationships and interactions. We present results from one representative specification in the main article and provide the full set of results across specifications in the Appendix.

Outcome variable. The dependent variable is the pooled event–study estimate for each country–disease–alert pair, which represents the causal effect of a WHO alert on research output as defined in our event–study design.

Predictor variables. To identify the drivers of research responsiveness, we compiled a broad set of predictors spanning research capacity, socioeconomic conditions, governance, health system preparedness, disease burden, and outbreak text features. The complete list of variables and their data sources is described below.

- **Research capacity (Scimago Journal & Country Rank).** We included three bibliometric indicators at the country level: *H index*, *AVG Citations*, and *Citable Documents*. These variables were downloaded from the Scimago Journal & Country Rank portal (https://www.scimagojr.com/countryrank.php), which compiles publication and citation data from Elsevier's Scopus database. They provide complementary perspectives on the quantity and impact of scientific research output.
- **Demographics (World Development Indicators).** We use World Bank WDI indicators at the country–year level:
  - Total population: from the corresponding "Population" dataset (https://github.com/datasets/population).
  - Urban population (% of total): `SP.URB.TOTL.IN.ZS` (`urban_pct`)

- Population density (people per sq. km of land area): `EN.POP.DNST` (`pop_density`)
- Population ages 65 and above (% of total): `SP.POP.65UP.TO.ZS` (`pct_age_65_plus`)

- **Socioeconomic context.** We used multiple indicators of socioeconomic capacity:
  - Internet use (`internet_pct`): share of the population using the Internet, obtained from Our World in Data (https://ourworldindata.org/grapher/share-of-individuals-using-the-internet, downloaded May 3, 2025).
  - GDP per capita: from the "GDP" dataset maintained by `datasets` on GitHub (https://github.com/datasets/gdp).

- **Healthcare quality.** We included the 2019 Healthcare Access and Quality (HAQ) Index (`haq_val_2019`), published by the Global Burden of Disease (GBD) Study (https://ghdx.healthdata.org/record/ihme-data/gbd-2019-healthcare-access-and-quality-index-1990-2019).

- **Governance and resilience (Worldwide Governance Indicators, 2024 update).** To capture institutional and political capacity, we used indicators on *Political and security risk*, *Government effectiveness*, *Socio-economic resilience*, and *Public confidence in government*. These variables come from the Worldwide Governance Indicators project (https://info.worldbank.org/governance/wgi/), with the November 2024 update (https://www.govindicators.org).

- **Global Health Security (GHS) preparedness pillars.** We incorporated the six core domains of the Global Health Security Index (https://ghsindex.org/report-model/), which measure national preparedness across: (1) Prevention of the emergence or release of pathogens; (2) Early detection and reporting for epidemics of potential international concern; (3) Rapid response to and mitigation of the spread

of an epidemic; (4) Sufficient and robust health sector to treat the sick and protect health workers; (5) Commitments to improving national capacity, financing, and adherence to norms; (6) Overall risk environment and country vulnerability to biological threats.

- **Disease burden.** We used the mean DALY rate for each disease in the alerted country, drawn from the Global Burden of Disease (GBD) Study (https://ghdx.healthdata.org/gbd-results-tool). For diseases mapping to multiple GBD causes, we computed the mean across those causes.

- **Disease prevention / immunization (World Development Indicators).** We include DPT3 immunization coverage among children ages 12–23 months: `SH.IMM.IDPT` (`dpt3_coverage`).

- **Alert text features (WHO outbreak reports).** We also derived features directly from the text of WHO outbreak alerts as introduced in previous sections. To capture uncertainty, we applied a dictionary-based approach using the Vermont Uncertainty Corpus (https://vermontconversationlab.com/uncertainty-corpus/), which contains an overall uncertainty lexicon and three subcategories of terms: *hedging* (e.g., "anticipate," "suppose," "hypothesis"), *possibility* (e.g., "could," "maybe," "potential"), and *confusion* (e.g., "confound," "dubious," "unlikely"). These dictionaries were converted into regular-expression patterns and matched against tokenized alert texts.

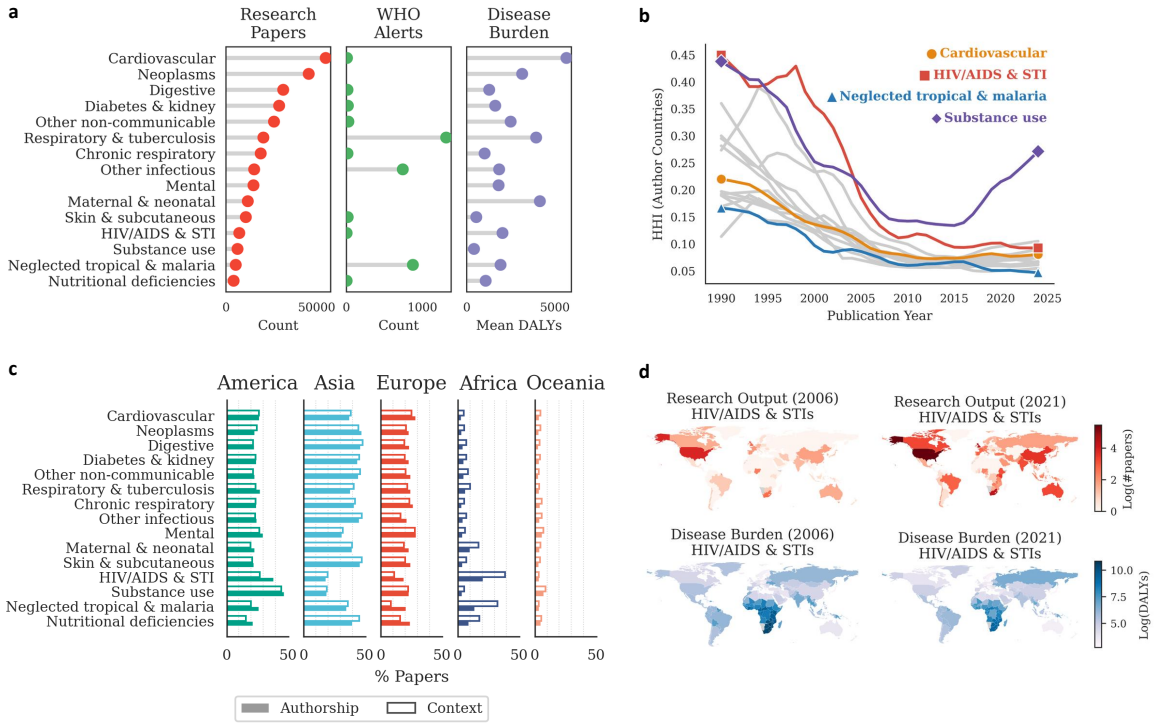
Model implementation. We employed Random Forest regression models, implemented separately for each unique specification of the event–study framework, defined by the combination of fixed effects, disease aggregation level, research attribution type, and dependent variable setting. Running models across all specifications allowed us to assess the robustness of the predictors identified.

For each specification, the data were randomly split into training (80%) and test (20%)

sets. Missing values were imputed using medians, with corresponding missing–value indicators included as additional features. The missingness share was negligible (median 0.3%, max 1.5%), indicating imputations did not affect results. Model performance was evaluated using five–fold cross–validation on the training set (reporting RMSE) and by both RMSE and R^2 on the held–out test set. Across specifications, the random forest models reported a cross-validated RMSE of 0.188 ± 0.063, a test RMSE of 0.186 ± 0.065, and a test R^2 of 0.775 ± 0.056.

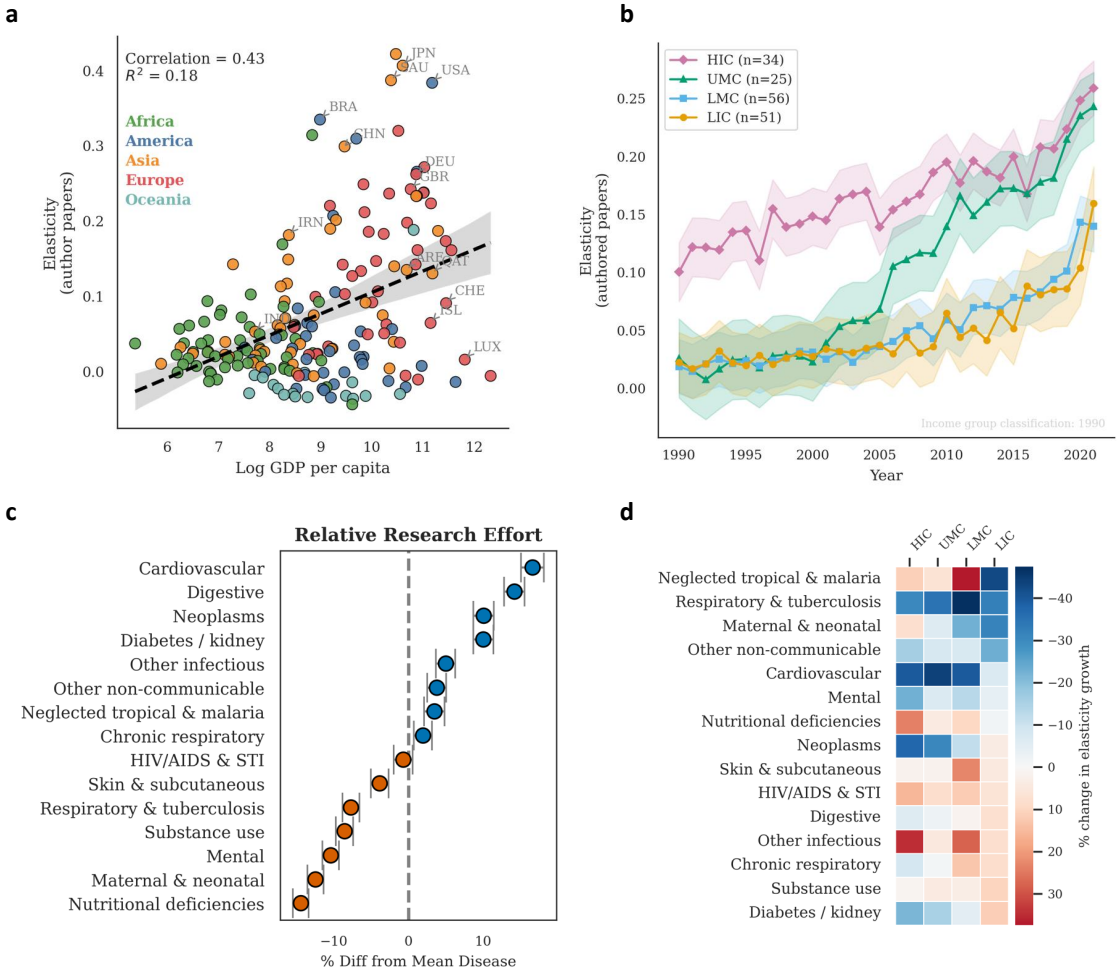
5 Main Figures and Tables

Figure 1: Global representation and evolution of disease research and burden.



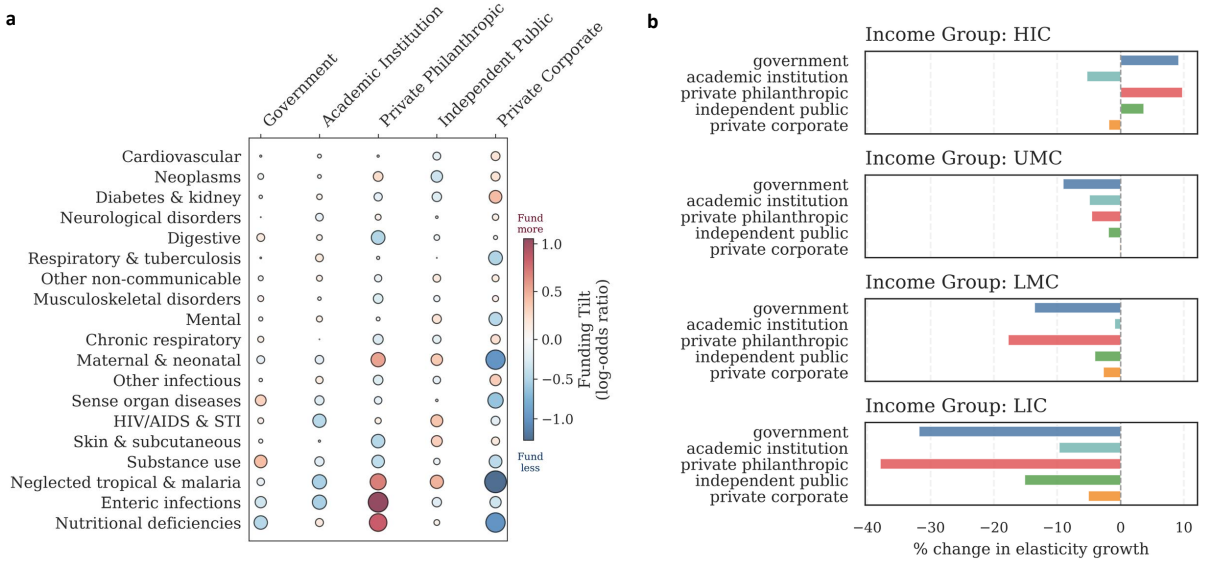
a, Cross-comparison of fifteen disease groups by total number of research papers, number of WHO alerts, and mean disability-adjusted life years (DALYs). **b,** Temporal trends in the geographic concentration of research, measured by the Herfindahl–Hirschman Index (HHI) of author countries. A general decline indicates increasing diversity of national contributions across most disease groups. Lines show LOWESS-smoothed trajectories. **c,** Continental share of research output (authored articles) and research context (in-text mentions) by disease group, calculated using fractional counting. **d,** Illustration of increasing alignment between research effort and disease burden for HIV/AIDS and STIs in 2006 and 2021. Research output (upper) and disease burden (lower) are shown on log-transformed scales.

Figure 2: Endemic responsiveness of national research efforts to disease burden.



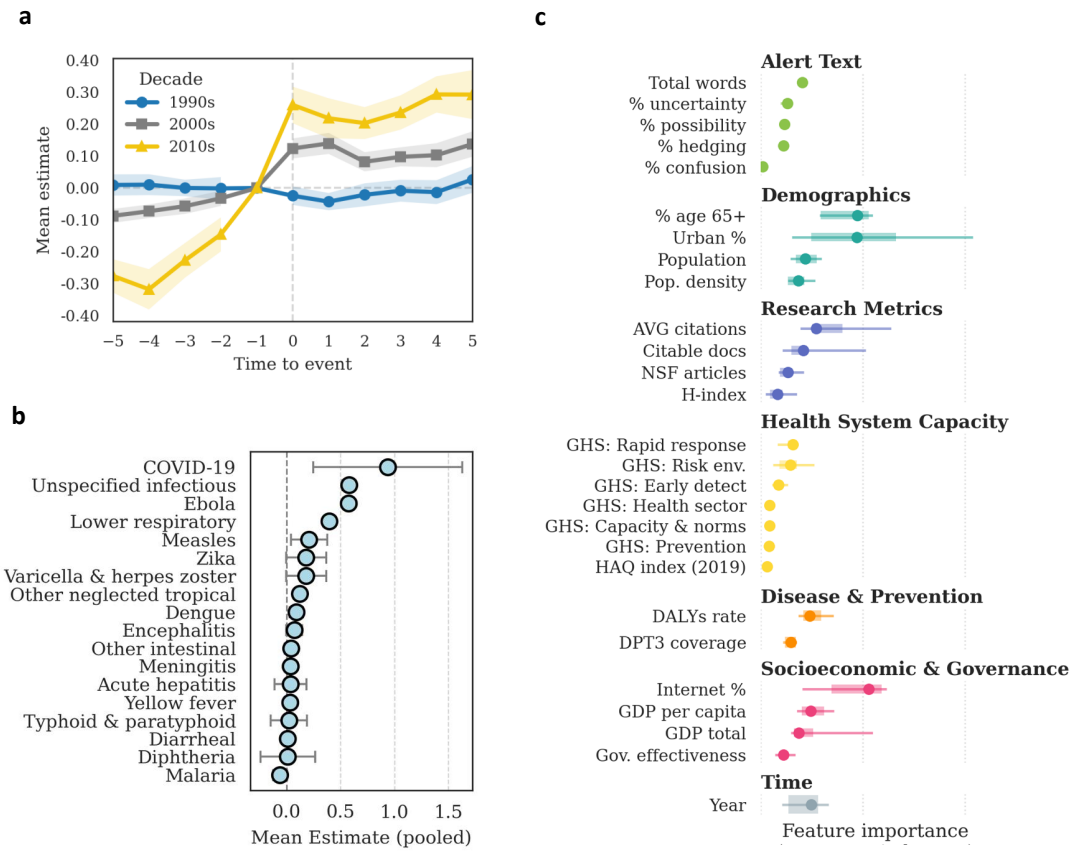
a, National endemic research responsiveness positively correlates with economic development status: elasticity rises with log GDP per capita, although wealthier countries display greater heterogeneity. Each point denotes a country or territory (coloured by continent); the dashed line is the OLS fit with 95% CI. Elasticities are estimated from regressions of log publications on log DALYs with country, disease, and year fixed effects. **b,** Responsiveness has increased over time across all income groups, with high-income countries consistently exhibiting higher elasticity. Lines show estimated elasticities from regressions with year × income-group interactions; shaded areas denote 95% confidence intervals. **c**, Relative research effort across disease groups, measured as baseline deviations from burden-expected levels derived from disease fixed effects (positive/blue = over-represented relative to burden; negative/orange = under-represented). **d,** Disease-level contributions to endemic responsiveness growth between 2014–2017 and 2018–2021, by income group. For each disease, we construct a counterfactual by holding its post-2017 publication counts constant at 2017 levels and re-estimating elasticity. The heatmap shows the percentage difference from the baseline change, indicating whether growth in that disease's research amplifies (red) or dampens (blue) the overall increase in responsiveness of the income group.

Figure 3: The preference and impact of research funders on disease research and endemic responsiveness



a, Funding tilt of five funder types across disease groups. Tilt is measured as the log-odds ratio of a disease being funded by a given type compared to the overall average. Positive values (red) indicate diseases over-represented in a funder's portfolio, while negative values (blue) indicate under-representation. Bubble size denotes the magnitude of the tilt. **b,** Contribution of funder types to changes in research responsiveness (elasticity growth) between 2014–2017 and 2018–2021, by income group. For each funder type, we construct a counterfactual by removing all papers funded by that type and re-estimating elasticity growth. Bars show the percentage difference relative to baseline growth for each income group. Positive values indicate that a funder type makes a net positive contribution to responsiveness growth in that group (removing it reduces growth), while negative values indicate that removing the funder increases growth.

Figure 4: Event–study estimates and predictors of emergency research responsiveness.



We implement an event–study difference-in-differences design for each alert–country pair to measure the growth of research on the focal disease relative to control groups. The figure shows results from the between-disease, between-country specification, which benchmarks alerted disease–country pairs against all other disease–country combinations. Alternative designs (between-disease within-country and within-disease within-country) are reported in the Appendix. The estimation yields two types of coefficients: pooled estimates, summarizing the average post-event increase in research activity, and dynamic estimates, tracing the yearly evolution of responsiveness around the outbreak. **a,** Temporal change in mean responsiveness by decade. Shaded areas denote 95% confidence intervals (same for panels b and c). **b,** Average pooled estimate across disease groups (GBD level 3). **c,** Feature importance from Random Forest regressors predicting pooled estimates of responsiveness.

Appendix

A Details on Bibliographic Dataset Construction

A.1 Journal Set Identification

To identify journals relevant to this study, we employed the 2023 edition of the *Journal Citation Reports* (JCR), provided by Clarivate (Clarivate JCR). JCR is a widely used journal evaluation tool built on the Web of Science Core Collection, reporting citation-based metrics such as the Journal Impact Factor and assigning each journal to one or more subject categories. These subject categories serve as proxies for disciplinary orientation and are commonly used in bibliometric and research evaluation studies. We linked the journals retrieved from JCR 2023 to OpenAlex (OpenAlex) using both ISSN and eISSN, to further obtain article-level bibliographic data.

- We identified **57 medicine-related categories** in JCR. Linking these journals to OpenAlex yielded **6,725 journals** with unique OpenAlex identifiers.
- We excluded journals with the word *"Review"* in the title to omit outlets primarily positioned to publish review articles rather than empirical studies. This reduced the set to **5,861 journals**.
- We focused on three broad categories that capture general medical research without explicit disease specialization:
  - *MEDICINE, GENERAL & INTERNAL*
  - *MEDICINE, RESEARCH & EXPERIMENTAL*
  - *PHARMACOLOGY & PHARMACY*

This refinement gave us **524 journals**, covering **1,065,683 papers** indexed in OpenAlex, restricted to documents of type "article" with at least one citation.

A.2 Article Corpus Filtering

Starting from 1,065,683 papers, we applied a set of filters to ensure coverage of relevant, attributable, and analyzable research outputs, while also producing a reasonable size of data for our LLM-based retrieval pipeline.

- **Domain filter:** Papers classified in OpenAlex as only belonging to the *Health Sciences* domain were retained (**741,264 papers; 69.6%**).
- **Authorship filter:** Papers with valid information on the country or territories of authors' affiliations were retained (**871,934 papers; 81.8%**).
- **Language filter:** Only papers written in English were retained (**1,050,170 papers; 98.5%**).
- **Combined filters:** Applying all three filters together produced a refined corpus of **603,447 papers (56.6%)**.

A.3 Geographic Context Filtering

To study the geographical dimension of medical research, we further filtered papers to identify those with potential geographical context in their titles or abstracts. Two complementary approaches were used:

- **GeoText method:** Applied the Python `geotext` library (https://github.com/elyase/geotext) to extract city and country mentions. This identified **202,790 papers (33.6%)**.
- **NER method:** Used a named entity recognition model (`dslim/distilbert-NER`) to extract *LOC* and *ORG* entities as potential location markers. This identified **309,950 papers (51.4%)**.
- **Combined filter:** Retaining papers identified by either method gave a total of **349,737 papers (58.0%)**.

A.4 LLM Retrieval and Linkage

We then processed the 349,737 papers with our LLM-based retrieval pipeline to enrich them with standardized biomedical and contextual information. This step successfully linked **308,135 papers (88.1%)** to at least one Medical Subject Heading (MeSH). Within this corpus, we report the following information on variables used in the study:

- **GBD categories:** Among these, **266,532 papers (86.5%)** were further linked to at least one level 2 Global Burden of Disease (GBD) category.
- **Funder information: 32,057 papers (10.4%)** had funder information disclosed and retrieved from OpenAlex.
- **Geographical context (LLM retrieval):** The LLM pipeline identified geographical context for **169,262 papers (54.9%)**.

Final dataset: The resulting curated dataset consists of **308,135 papers**, enriched with relevant disease categories, authorship information, funder information, and geographical context.

B Details on LLM Retrieval of Academic Papers

Our empirical design hinges on two ingredients for every article in our corpus:

1. *paper–data metadata*, which record whether the study uses primary or secondary data, the unit of analysis, temporal coverage, geographic scope, and ownership; and
2. *paper–disease links*, which anchor every disease mention to a controlled biomedical vocabulary so that research effort can be compared with national disease burdens.

We obtain both with a single large-language-model (LLM) call per paper. The model (`gpt-4o-mini-2024-07-18`) receives the article's title and abstract plus a structured *system prompt* that instructs it to respond with a JSON object. We use the default model hyperparameters.

Strict schema validation is enforced via OpenAI's structred output format, minimising hallucination and guaranteeing parseability.

B.1 LLM Prompt for Paper–Data Metadata Extraction

The prompt casts the model as an expert analyst of biomedical papers, tells it to ignore everything except data-related content, and spells out the required JSON keys (with explicit fall-backs such as `NA` or empty arrays) to guarantee parseable output.

System instructions We provide the following instructions to the model for each call:

```
You are an expert assistant analyzing research papers.
Task
- Given a paper's title and abstract, extract the data information mentioned in the
text.
- Return a JSON object with a single top-level field "data" containing the subfields
listed below.
Subfields
- uses_data (boolean): true if the paper uses any data, false otherwise.
- unit_of_analysis (array): e.g. individual, group, country.
- unit_of_analysis_details (string): extra detail or "NA".
- temporal_context (string): time period or "NA".
- start_year, end_year (arrays of integers): empty if absent.
- geographical_context_countries (array): ISO-3 codes or empty.
- ownership (array): choose from private company, public sector entity, academic
institution, non-profit organization, individual researcher, community-generated,
open source, other.
-ownership_details (array): extra owner details or empty.
```

```
Notes
- Extract only data-related information; ignore all other content.
- If a field is missing, use the specified default ("NA" or empty array).
-Produce only valid JSON matching the strict schema and add no extra keys.
```

User prompt For each paper, the following message is sent, where <TITLE> and <ABSTRACT> are the raw metadata:

```
Here is the title of the paper:
<TITLE>
Here is the abstract of the paper:
<ABSTRACT>
Extract the data information as specified.
```

B.2 Embedding-based Normalisation to MeSH

All disease strings appearing in the extracted metadata are converted into 1,024-dimensional embeddings using `text-embedding-3-large`. We pre-compute embeddings for the 30,836 preferred MeSH descriptor terms (including scope notes) and assign each disease the descriptor with the highest cosine similarity.

B.3 Limitations

Three caveats merit mention. First, the model works only with titles and abstracts: full-text mining could reveal additional entities but is possible only for open-access papers. Second, assigning each disease mention to a single MeSH descriptor ignores genuine polysemy. Third, information on temporal coverage, comprehensive disease attribution and data ownership is often missing; we therefore carry a dedicated missingness flag into all regressions. None of these constraints undermines the pipeline's ability to deliver a scalable, reproducible bridge from academic papers to policy-relevant, disease-coded metadata.

C Details on LLM Retrieval and Structuring of WHO Disease-Outbreak News

We download all Disease Outbreak News (DON) items published on www.who.int/emergencies/disease-outbreak-news. The snapshot used here covers alerts issued between 1996-2025 and yields 3,134 HTML pages. After stripping boiler-plate, we retain four fields per page: the alert title, publication date, canonical URL slug, and full prose.

LLM Prompt For Alert Extraction Each page is passed, one call per alert, to `gpt-4o-mini-2024-07-18` together with a system prompt that instructs the model to return an array of "alerts" containing, for every disease mention, seven slots: disease, geography, date, month, year, cases, deaths, and computed case-fatality ratio if provided. All slots default to `NA` (not available) when absent. The schema is declared as `"strict": true` and `"additionalProperties": false` so that any hallucinated field triggers an automatic retry.

System Instructions We provide the following instructions to the model for each call:

```
You are an expert assistant analyzing WHO Disease Outbreak News (DONs). Each text
describes an outbreak or suspected outbreak. Your goal is to extract every mention of
a disease along with the geography, date (day, month, year), and numerical details such
as cases, deaths, and case-fatality ratio (CFR).
Task
1. Identify every disease name explicitly mentioned.
2. Identify every geography (country, region, or area) explicitly mentioned.
3. For each mention, locate the date (if given) and any figures on cases, deaths, and
CFR. If multiple dates or data points are present, create a separate entry for each
(disease, geography, date, cases, deaths, CFR) tuple.
Output
Return an array under the key "alerts". Each item must contain: disease, geography,
date, month, year, cases, deaths, cfr. Use "NA" for any missing element. Produce only
```

```
valid JSON matching the strict schema and add no extra keys.
```

User Prompt For each alert, the following prompt is supplied, where <TITLE>, <DATE>, and <CONTENT> are the page's metadata and prose:

```
Title of the alert: <TITLE>
Published date: <DATE>
Full content:
<CONTENT>
Please extract all disease-geography-date mentions as per the instructions.
```

Normalising Diseases to MeSH Free-text disease names are embedded with `text-embedding-3-large` (1,024 dimensions) and matched to the nearest MeSH descriptor by cosine similarity, exactly as described in Section B.2. The resulting MeSH codes are subsequently linked to ICD-10 and GBD level-2 causes for empirical work.

Refining Geography Strings Alert prose often mixes coarse labels ("*Central Africa*") with granular places ("*Kasese District*"). To standardise these, we run a second LLM call on the unique set of geography strings (roughly 1,500 after de-duplication) guided by a schema that asks for: city, country (ISO-3), continent (seven-code system), region, and other. Each geography is assigned a deterministic geography identifier, enabling many-to-one joins back to the alert table.

System Instructions We provide the following instructions to the model for each call:

```
You are an expert assistant specialised in geographic data extraction and mapping. For
every input string, return a JSON object with the fields below.
- City: the city named in the string.
- Country: three-letter ISO code for the country mentioned or, if only a city is
given, inferred.
- Continent: use the codes AF, AN, AS, EU, NA, OC, SA.
```

```
- Region:  sub-national unit such as a state or province.
- Other:  any remaining context that does not fit the above.

If the input is coarse (e.g. \Africa''), fill only Continent and leave the other fields
empty.  If the input is granular (e.g. a city), fill City and, where possible, infer
Country and Continent.  Return valid JSON that matches the strict schema and contains
no extra keys.
```

User Prompt for Geography For each distinct geography string we send:

```
Geography:  <GEOGRAPHY_STRING>
Please extract City, Country (ISO3), Continent, Region and Other from the string above
and return them in the required JSON format.
```

Post-processing Steps

1. **Year back-fill.** Where the alert text lacks an explicit year ($\approx 6\%$), we impute it from the publication date.

2. **MeSH tree levels.** A unique MeSH tree cross-walk maps every descriptor to up to three hierarchical levels (mesh tree level 1–3), facilitating aggregation in event-study regressions.

Limitations Three issues deserve note. First, case and death figures follow no common template, so treating every number that precedes "cases" or "deaths" as valid can blur distinctions between suspected and confirmed counts. Second, several DON items cover more than one disease (for instance dengue and chikungunya), and our extraction rightly records each mention but inherits any ambiguity in the source text. Third, where the prose names only a continent, finer geographic slots remain empty, potentially understating sub-national variation. Even so, the pipeline yields the high-frequency, geography-resolved shocks that underpin the event-study design.

D Details on LLM classification of research funders

We classify every one of the 32,437 funders listed in the 2024 OpenAlex snapshot. Each funder's display name is sent five times to `gpt-4o-mini-2024-07-18` (temperature 0.7). The model is instructed to return a JSON object with six Boolean keys: direct government, independent public, private corporate, private philanthropic, academic institution, and hybrid. Strict schema validation (`''strict'': true`, `''additionalProperties'': false`) guarantees parseable output.

System Instructions We provide the following instructions to the model for each call:

```
You are an expert assistant in the analysis and classification of funding
organizations. Given the display name of a funder from the OpenAlex dataset, your
task is to determine its type across several categories. Specifically, based on the
provided display name, decide whether the organization is:
1. Direct Government: A government agency fully funded by public resources.
2. Independent Public: Receiving public funds but operating with independent
decision-making.
3. Private Corporate: A for-profit corporate entity, often funding research as part
of a strategic or CSR initiative.
4. Private Philanthropic: A philanthropic or non-profit organization that funds
research through endowed or donated resources.
5. Academic Institution: An academic or research institution funding research
internally or via affiliated research bodies.
6. Hybrid: An entity that combines multiple funding sources (public, private, etc.)
or does not clearly fit into the above categories.
Based solely on the funder's display name, classify it by returning boolean values for
each of the above categories. If there is any ambiguity or insufficient information,
default to "false" for categories that are not clearly applicable. Return only a
valid JSON object adhering strictly to the provided schema, with no additional keys or
commentary.
```

User Prompt For each funder, the following prompt is supplied, where "<DISPLAY_NAME>" is the name of the funder as provided by OpenAlex:

```
Funder Display Name:  <DISPLAY_NAME>
Please classify this funder into the following categories based solely on its name:
- Direct Government
- Independent Public
- Private Corporate
- Private Philanthropic
- Academic Institution
- Hybrid
Return only a JSON object following the provided schema with boolean values for each
field.
```

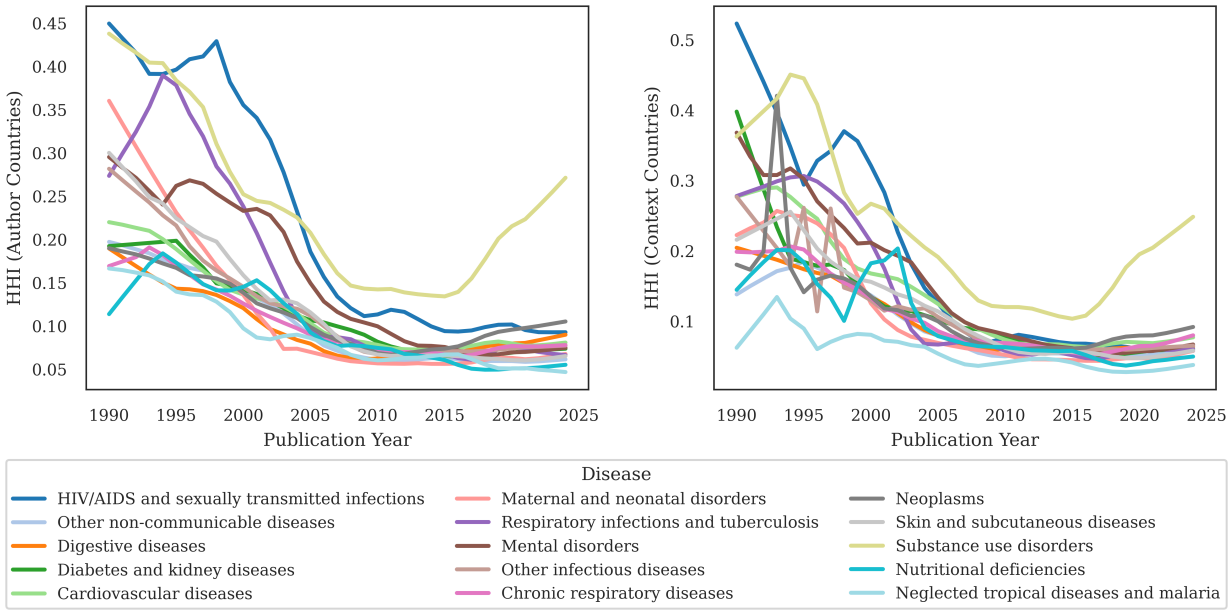
Aggregation and Consistency For each $(id, \text{category})$ pair we take the modal value across the five responses; ties are broken at random. Across the six categories, 96–99% of funders receive identical labels in all five calls, indicating high internal reproducibility despite the stochastic sampling (model temperature was set at 0.7).

Category Prevalence. In the universe of funders, 36% are labelled academic institution, 34% private philanthropic, 17% direct government, 9% private corporate, 4% independent public, and 2% hybrid. Because the categories are not mutually exclusive, a single funder can carry multiple `true` flags.

Potential Limitations Classification relies on the name alone, which can be ambiguous or language-specific; mixed-funding bodies may fit more than one bucket; and the six categories are an imperfect proxy for more granular legal forms. For our purposes (broadly contrasting public, corporate, philanthropic and hybrid sources) these coarse buckets are adequate, and the high agreement rate suggests that any residual misclassification is unlikely to change the qualitative results.

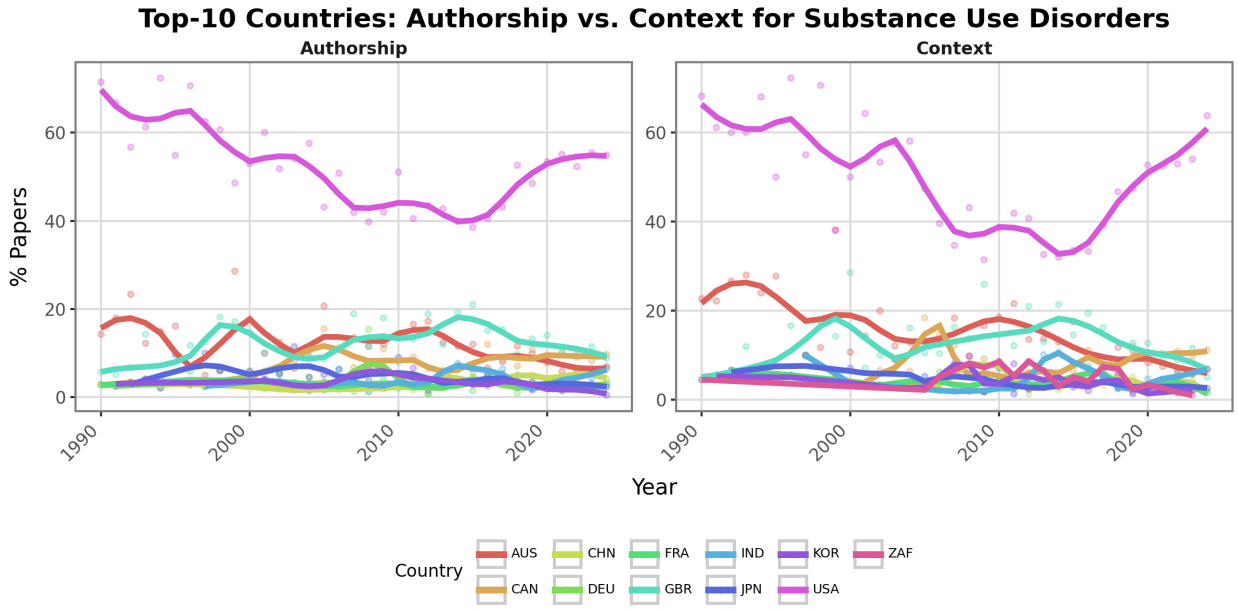
E Appendix Figures

Figure D1: The concentration of disease research output and contextual mentions of countries over time.



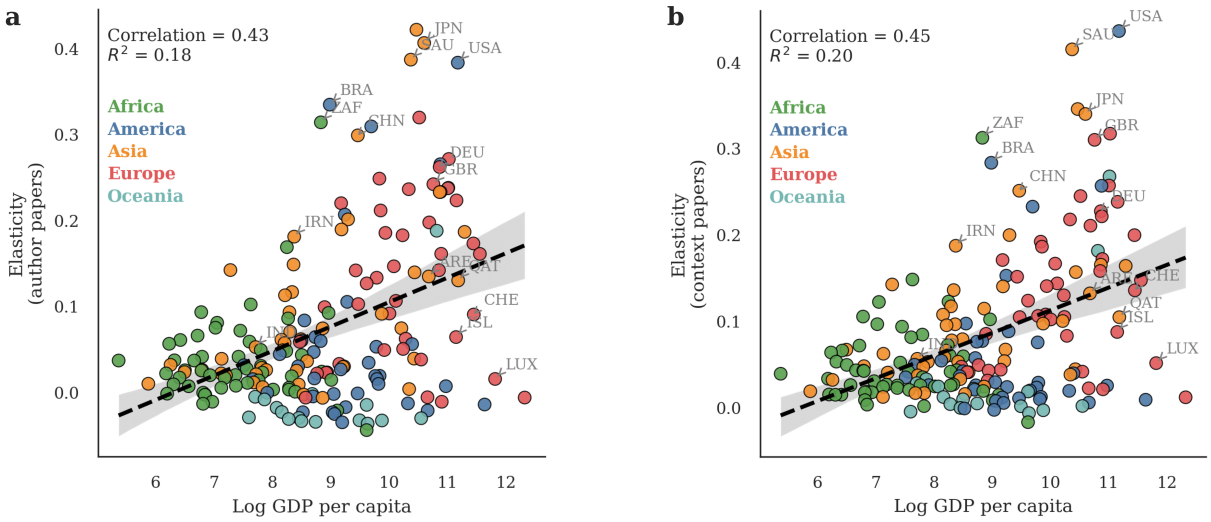
The Herfindahl-Hirschman Index (HHI) is displayed across three decades to illustrate how the concentration of country involvement evolves for various disease categories. The left panel shows HHI for countries as authors ("Author Countries"), while the right panel shows HHI for countries mentioned in titles or abstracts ("Context Countries"). Higher HHI values indicate a greater concentration among fewer countries, whereas lower values suggest broader participation or references.

Figure D2: Evolution of the top 10 most prolific countries publishing on substance use disorders.



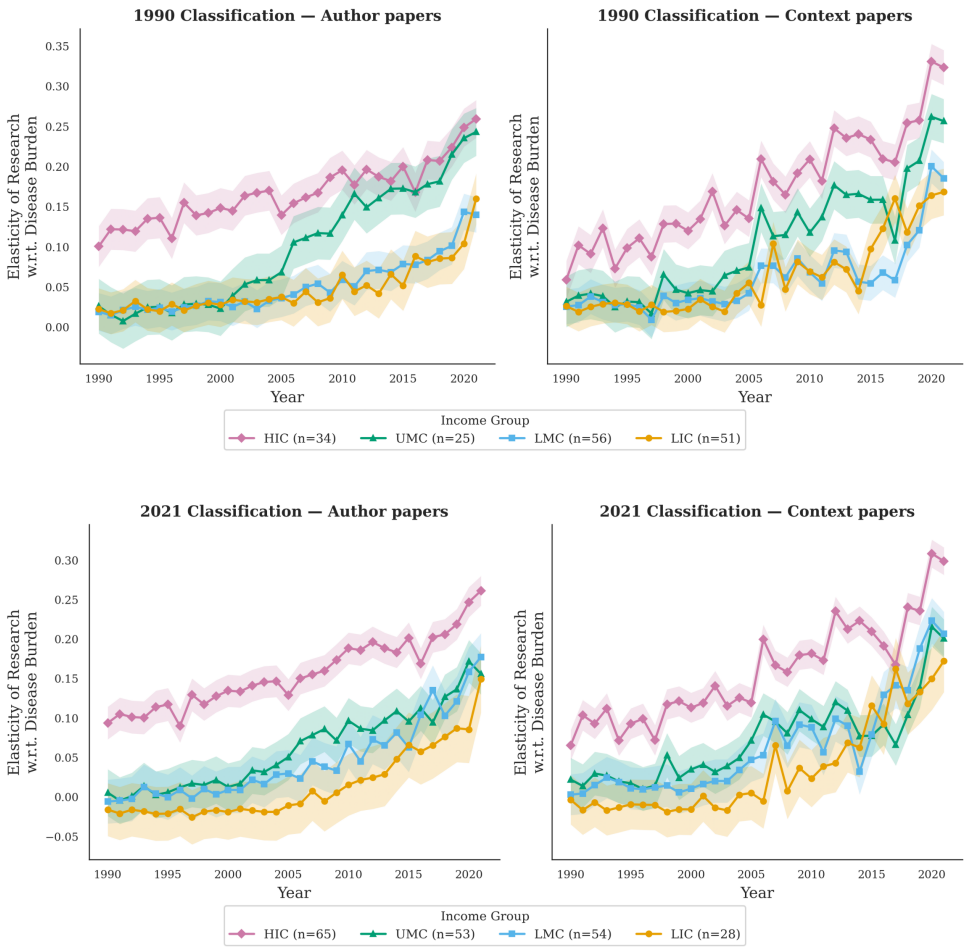
The figure depicts the percentage of articles on substance use disorders attributed to the ten most prolific countries in this area. The United States consistently accounts for the largest share of publications, with its proportion declining prior to 2015 but rising substantially in subsequent years. Since around 2015, the U.S. share has increased from approximately 40% to 57% by authorship and from 32% to 61% by country context in our dataset.

Figure D3: Country-level research responsiveness with relations to GDP per capita by authored papers and contextual mentions.



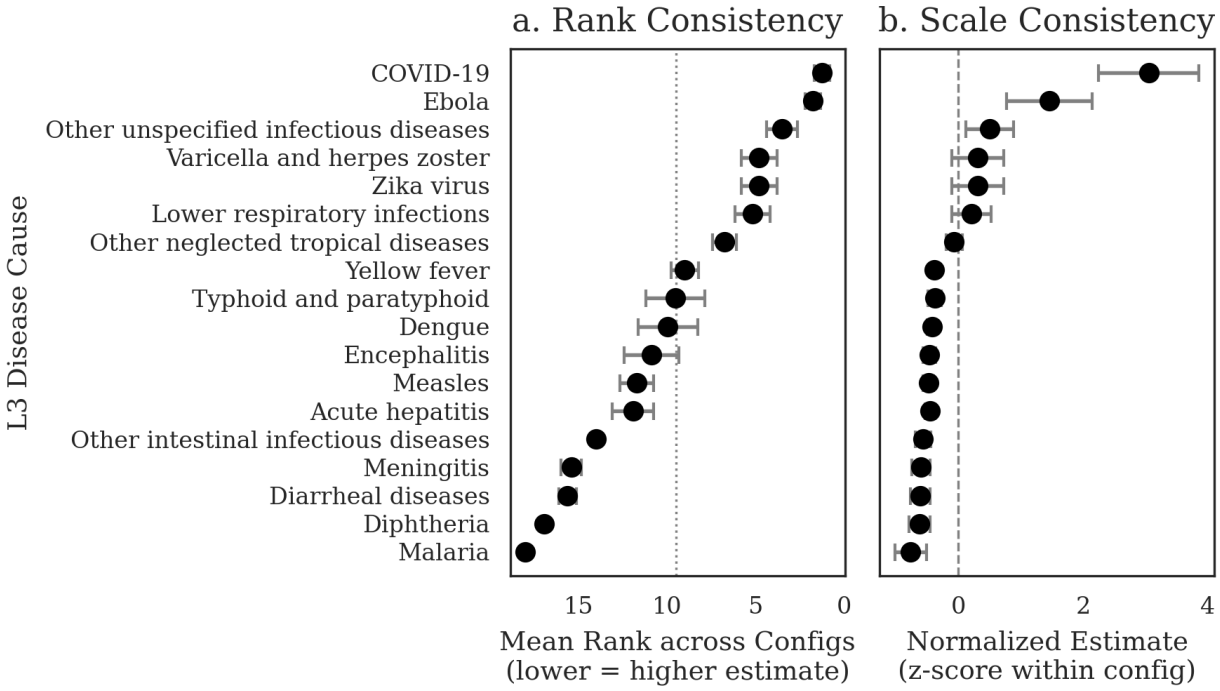
We estimate countries' responsiveness (elasticity) of research output to disease burden using a regression with an interaction between log-transformed disease burden and country, including fixed effects for year and disease category. The elasticity represents the percentage change in research output associated with a 1% change in disease burden, yielding a responsiveness coefficient for each country. We then examine its association with GDP per capita in scatter plots. Panel a reports estimates using the log of countries' authored paper counts as the dependent variable, while panel b uses the log of paper counts in which countries are mentioned in titles or abstracts. Both panels indicate a positive relationship between research responsiveness and GDP per capita, with correlations of 0.43 and 0.45, respectively.

Figure D4: Evolution of elasticity of research output to disease burden by country income group (1990 vs. 2021 classifications; author papers vs. contextual mentions).



This figures shows how the responsiveness (elasticity) of research output to disease burden has evolved across countries with different income levels. Elasticity estimates are derived from a fixed-effects regression with interaction terms for income group and year, allowing the relationship between research output and disease burden to vary by both dimensions. Specifically, the model estimates the percentage change in research output associated with a 1% change in disease burden for each income group and year, controlling for country, disease category, and year fixed effects. The lines represent the estimated elasticity for each income group over time, with shaded bands denoting 95% confidence intervals. The left-hand panels use countries' authored paper counts as the dependent variable, while the right-hand panels use counts of papers in which countries are mentioned in titles or abstracts. The top row classifies countries by their 1990 income groups (the start year of our analyses), whereas the bottom row uses 2021 classifications. The upward trend in responsiveness is robust across all income groups and dependent variable specifications.

Figure D5: Robustness of pooled event-study estimates of outbreak impacts on research effort, averaged across configurations and GBD level 3 disease categories.



Panel a presents the mean rank of disease categories across configurations (lower ranks correspond to larger estimated effects), with horizontal bars showing variability across configurations. Panel b reports the mean normalized estimates (z-scores within each configuration) for each category, again with horizontal bars indicating variability. Consistency in both rank and scale across diseases demonstrates that the observed various outbreak effects on research effort by diseases are robust to alternative model specifications.

Figure D6: Mean dynamic estimates by time to event across specifications.

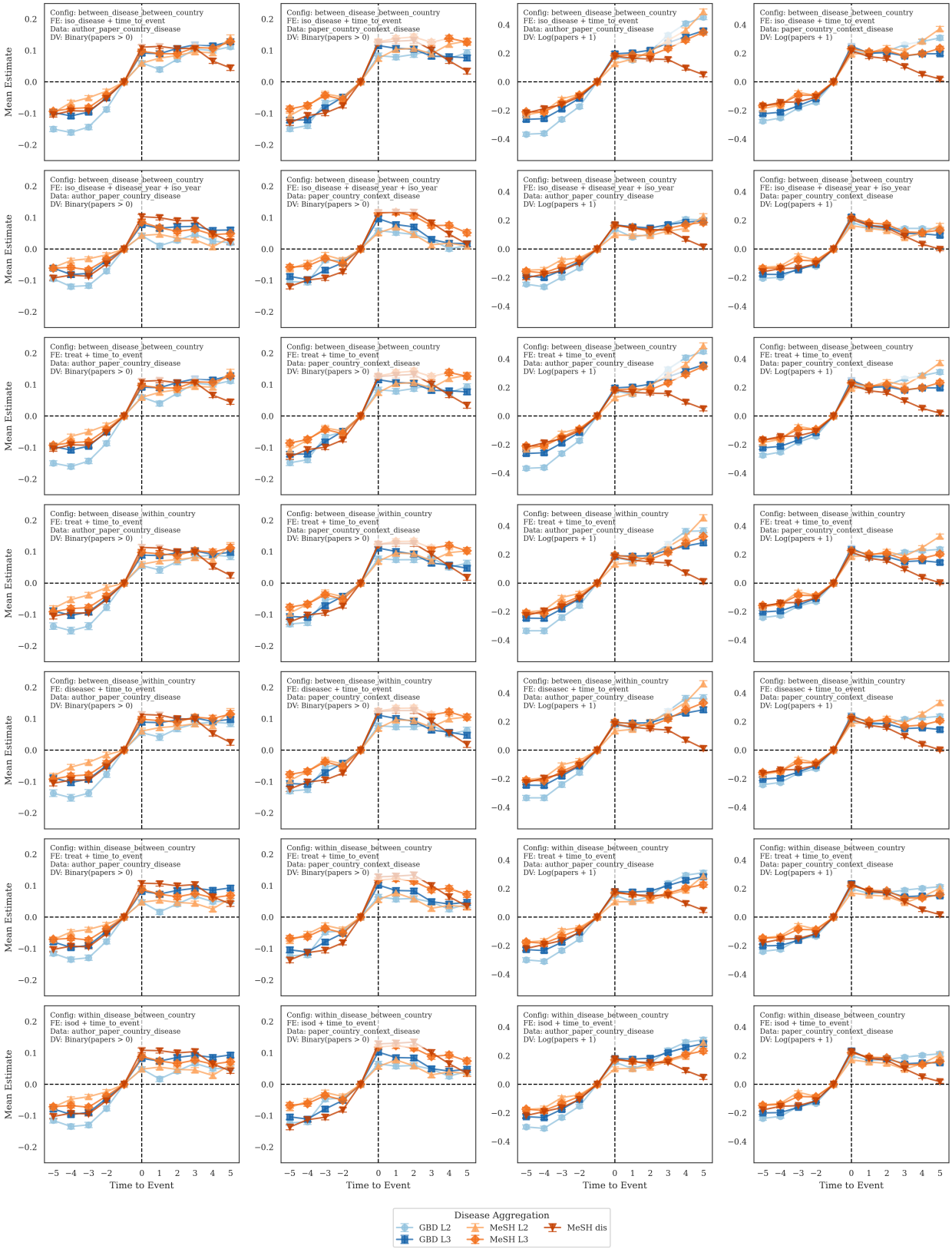


Figure D7: Mean dynamic estimates by decades across specifications.

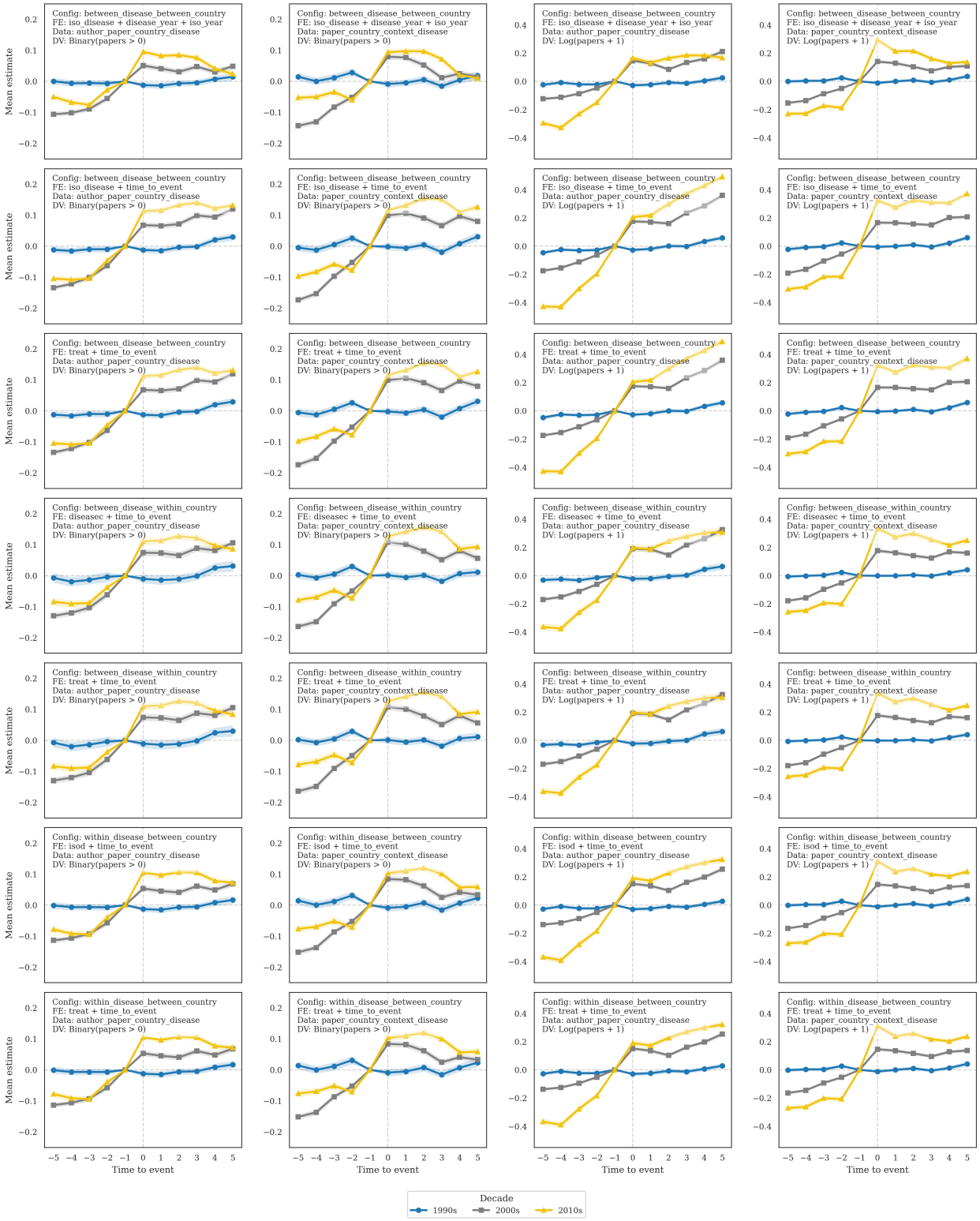
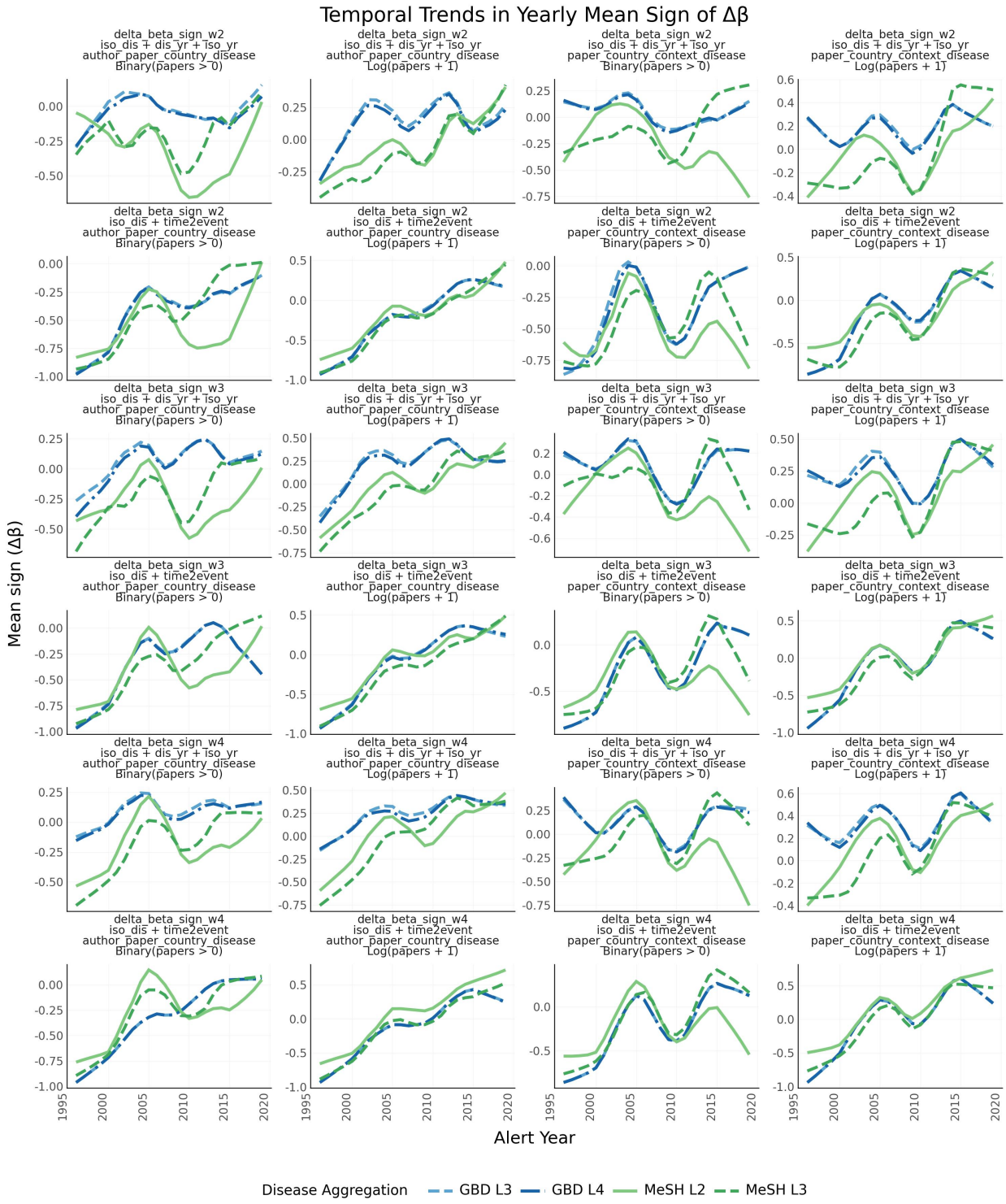


Figure D8: Robustness check on characterizing temporal shifts before and after the alert



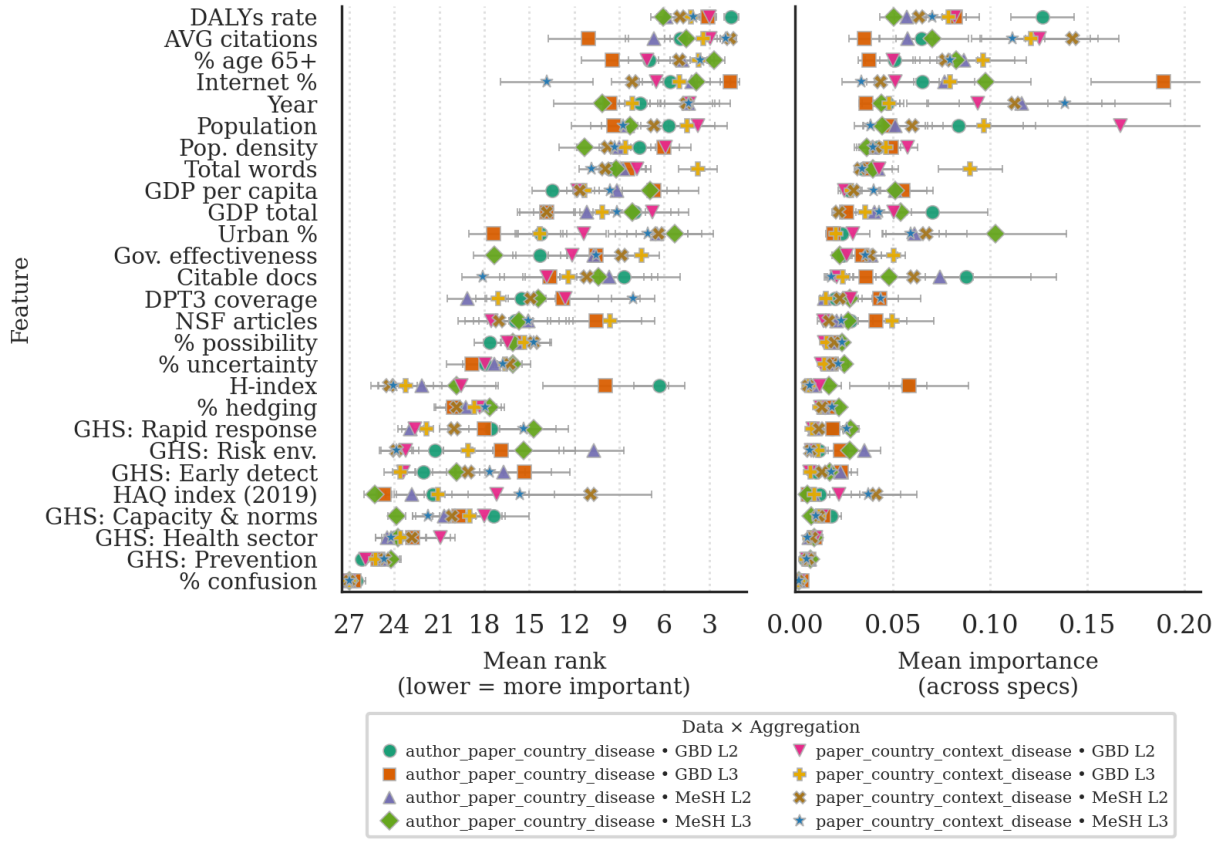
To assess how research activity shifts in response to a WHO alert, we calculate the change in the estimated event-study coefficients (β_τ) before and after the alert. For each alert event, we define the **delta beta** ($\Delta\beta$) as:

$$\Delta\beta = \frac{1}{w}\sum_{\tau=1}^{w}\beta_{\tau,\text{post}} - \frac{1}{w}\sum_{\tau=1}^{w}\beta_{-\tau,\text{pre}}$$

where w is the window size (e.g., 2, 3, or 4 years), $\beta_{\tau,\text{post}}$ are the coefficients for years after the alert, and $\beta_{-\tau,\text{pre}}$ are those for years before the alert. This statistic summarizes the average change in research output associated with the alert across different time windows.

By averaging $\Delta\beta$ across all events, we characterize the typical direction and magnitude of change in research response to alerts.

Figure D9: Robustness check on feature importance across specifications



Mean feature ranks (left, lower = more important) and mean absolute importances (right) estimated from Random Forest models. Each error bar corresponds to one dataset × disease aggregation combination and shows the mean and 95% confidence interval of pooled estimates, calculated across variations in dependent variable, fixed effects, and event–study configuration. This robustness check demonstrates which predictors remain consistently important across alternative modeling choices.

References

Abimbola, S., Asthana, S., Montenegro, C., Guinto, R. R., Jumbam, D. T., Louskieter, L., ... others (2021). Addressing power asymmetries in global health: imperatives in the wake of the covid-19 pandemic. *PLoS medicine*, *18*(4), e1003604.

Access to Medicine Foundation. (2024). *Access to medicine index 2024.* Retrieved from https://accesstomedicinefoundation.org/news/pharma-companies-still-not-maximising-their-potential-to-reach-more-patients-in-low-and-middle-income-countries

Acemoglu, D., & Linn, J. (2004). Market size in innovation: theory and evidence from the pharmaceutical industry. *The Quarterly journal of economics*, *119*(3), 1049–1090.

Adams, J. (2012). The rise of research networks. *Nature*, *490*(7420), 335–336.

Azoulay, P., Graff Zivin, J. S., Li, D., & Sampat, B. N. (2019). Public r&d investments and private-sector patenting: evidence from nih funding rules. *The Review of economic studies*, *86*(1), 117–152.

Barakati, K., Molak, A., Nelson, C., Zhang, X., Takeuchi, I., & Kalinin, S. V. (2025). Causal discovery from data assisted by large language models. *arXiv preprint arXiv:2503.13833*.

Bareinboim, E., & Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, *113*(27), 7345–7352.

Budish, E., Roin, B. N., & Williams, H. (2015). Do firms underinvest in long-term research? evidence from cancer clinical trials. *American Economic Review*, *105*(7), 2044–2085.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, *225*(2), 200–230.

Chu, J. S., & Evans, J. A. (2021). Slowed canonical progress in large fields of science. *Proceedings of the National Academy of Sciences*, *118*(41), e2021636118.

Dimitris, M. C., Gittings, M., & King, N. B. (2021). How global is global health research?

a large-scale analysis of trends in authorship. *BMJ global health*, *6*(1).

Eldridge, L., Garton, E. M., Duncan, K., & Gopal, S. (2024). Authorship of publications supported by nci-funded grants involving low-and middle-income countries. *JAMA Network Open*, *7*(3), e243215–e243215.

Evans, J. A., Shim, J.-M., & Ioannidis, J. P. (2014). Attention to local health burden and the global disparity of health research. *PloS one*, *9*(4), e90147.

Fetzer, T., Lambert, P. J., Feld, B., & Garg, P. (2024, November). Ai-generated production networks: Measurement and applications to global trade [Working Paper]. (11497 / WERPS 1528 / CAGE WP 733). Retrieved from https://ssrn.com/abstract=5050959 (Also: ECONtribute Discussion Paper No. 346, University of Bonn & University of Cologne; University of Warwick Working Paper Series (WERPS 1528).)

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., ... others (2018). Science of science. *Science*, *359*(6379), eaao0185.

Furuse, Y. (2019). Analysis of research intensity on infectious disease by disease burden reveals which infectious diseases are neglected by researchers. *Proceedings of the National Academy of Sciences*, *116*(2), 478–483.

Gao, H., Zhang, C., Wu, F., Zheng, C., Zhao, J., & Liu, H. (2025). Bootstrapping heterogeneous graph representation learning via large language models: A generalized approach. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 39, pp. 16717–16726).

Garg, P., & Fetzer, T. (2025a). Artificial intelligence health advice accuracy varies across languages and contexts. *arXiv preprint arXiv:2504.18310*.

Garg, P., & Fetzer, T. (2025b). Causal claims in economics. *arXiv preprint arXiv:2501.06873*.

Gillum, L. A., Gouveia, C., Dorsey, E. R., Pletcher, M., Mathers, C. D., McCulloch, C. E., & et al. (2011). NIH Disease Funding Levels and Burden of Disease. *PLoS ONE*, *6*(2), e16837. Retrieved from https://doi.org/10.1371/journal.pone.0016837 doi:

10.1371/journal.pone.0016837

Global Burden of Disease Collaborative Network. (2024). *Global burden of disease study 2021 (gbd 2021).* Seattle, United States: Institute for Health Metrics and Evaluation (IHME).

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of econometrics*, *225*(2), 254–277.

Hedt-Gauthier, B. L., Jeufack, H. M., Neufeld, N. H., Alem, A., Sauer, S., Odhiambo, J., … Volmink, J. (2019). Stuck in the middle: a systematic review of authorship in collaborative health research in africa, 2014–2016. *BMJ global health*, *4*(5), e001853.

Hoekman, J., Frenken, K., & Tijssen, R. J. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within europe. *Research policy*, *39*(5), 662–673.

Ioannidis, J. P., Bendavid, E., Salholz-Hillel, M., Boyack, K. W., & Baas, J. (2022). Massive covidization of research citations and the citation elite. *Proceedings of the National Academy of Sciences*, *119*(28), e2204074119.

Jiralerspong, T., Chen, X., More, Y., Shah, V., & Bengio, Y. (2024). Efficient causal graph discovery using large language models. *arXiv preprint arXiv:2402.01207*.

Kandibedala, B., Pyayt, A., Piraino, N., Caballero, C., & Gubanov, M. (2023). Covidkg. org-a web-scale covid-19 interactive, trustworthy knowledge graph, constructed and interrogated for bias using deep-learning. In *International conference on extending database technology (edbt).*

Lewis, J., Schneegans, S., Straza, T., et al. (2021). *Unesco science report: The race against time for smarter development* (Vol. 2021). Unesco Publishing.

Merton, R. K. (1968). The matthew effect in science: The reward and communication systems of science are considered. *Science*, *159*(3810), 56–63.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of web of science and scopus: a comparative analysis. *Scientometrics*, *106*(1), 213–228.

Myers, K. (2020). The elasticity of science. *American Economic Journal: Applied Economics*,

12(4), 103–134.

Naidoo, A. V., Hodkinson, P., King, L. L., & Wallis, L. A. (2021). African authorship on african papers during the covid-19 pandemic. *BMJ global health*, *6*(3).

National Library of Medicine. (2023). *Mesh tree structure.* https://www.nlm.nih.gov/mesh/intro_trees.html. (Accessed: 2025-01-28)

Nature. (2023). North–south country collaborations reveal untapped potential. *Nature*. Retrieved from https://www.nature.com/articles/d41586-023-03903-9

Priem, J., Piwowar, H., & Orr, R. (2022). Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833*.

Røttingen, J.-A., Regmi, S., Eide, M., Young, A. J., Viergever, R. F., Årdal, C., ... Terry, R. F. (2013). Mapping of available health research and development data: what's there, what's missing, and what role is there for a global observatory? *The Lancet*, *382*(9900), 1286–1307.

Sampson, M., McGowan, J., Cogo, E., Grimshaw, J., Moher, D., & Lefebvre, C. (2009). An evidence-based practice guideline for the peer review of electronic search strategies. *Journal of clinical epidemiology*, *62*(9), 944–952.

Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., ... others (2023). Large language models encode clinical knowledge. *Nature*, *620*(7972), 172–180.

Sudhahar, S., Ozer, B., Chang, J., Chadwick, W., O'Donovan, D., Campbell, A., ... Roberts, I. (2024). An experimentally validated approach to automated biological evidence generation in drug discovery using knowledge graphs. *Nature Communications*, *15*(1), 5703.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics*, *225*(2), 175–199.

Van Leeuwen, T. N., Moed, H. F., Tijssen, R. J., Visser, M. S., & Van Raan, A. F. (2001). Language biases in the coverage of the science citation index and its consequences-for international comparisons of national research performance. *scientometrics*, *51*(1), 335–346.

Vollset, S. E., Ababneh, H. S., Abate, Y. H., Abbafati, C., Abbasgholizadeh, R., Abbasian, M., ... others (2024). Burden of disease scenarios for 204 countries and territories, 2022–2050: a forecasting analysis for the global burden of disease study 2021. *The Lancet*, *403*(10440), 2204–2256.

Vos, T., Lim, S. S., Abbafati, C., Abbas, K. M., Abbasi, M., Abbasifard, M., ... others (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The lancet*, *396*(10258), 1204–1222.

Wagner, C. S., Cai, X., Zhang, Y., & Fry, C. V. (2022). One-year in: Covid-19 research at the international level in cord-19 data. *PLoS One*, *17*(5), e0261624.

Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, *46*(8), 1416–1436.

World Health Organization. (2024a). *Global report on neglected tropical diseases 2024.* Retrieved from https://www.who.int/teams/control-of-neglected-tropical-diseases/global-report-on-neglected-tropical-diseases-2024

World Health Organization. (2024b). *Number of clinical trials by year, country, WHO region and income group (1999–2024).* https://www.who.int/observatories/global-observatory-on-health-research-and-development/monitoring/number-of-clinical-trials-by-year-country-who-region-and-income-group. (Accessed 15 Jun 2025)

Wornow, M., Xu, Y., Thapa, R., Patel, B., Steinberg, E., Fleming, S., ... Shah, N. H. (2023). The shaky foundations of large language models and foundation models for electronic health records. *npj digital medicine*, *6*(1), 135.

Yegros-Yegros, A., Van de Klippe, W., Abad-Garcia, M. F., & Rafols, I. (2020). Exploring why global health needs are unmet by research efforts: the potential influences of geography, industry and publication incentives. *Health research policy and systems*, *18*, 1–14.

Zhao, W., Wang, L., & Zhang, L. (2022). How does academia respond to the burden of

infectious and parasitic disease? *Health Research Policy and Systems*, *20*(1), 89.