Geography of Medical Knowledge: Scientific Focus, Disease Burden, and Research Response

Hongyu Zhou[1,2,*], Prashant Garg[3] and Thiemo Fetzer[4,5]

[1]University of Cambridge, Cambridge, United Kingdom
[2]University of Antwerp, Antwerp, Belgium
[3]Imperial College London, London, United Kingdom
[4]University of Warwick, Coventry, United Kingdom
[5]University of Bonn, Bonn, Germany

Extended Abstract

Medical research is the cornerstone of evidence-based healthcare, yet its global production remains strikingly uneven. Although 80% of the world's population lives in low- and middle-income countries (LMICs), the majority of medical knowledge is still generated in high-income settings. This disconnect between where diseases impose the greatest burden and where scientific attention is focused raises concerns about the external validity of research, equitable access to innovations, and the global distribution of research capacity [1].

In this paper, we offer a fine-grained global view of how medical research is distributed across space, time, and disease, and how closely that distribution tracks the geography of health needs [2]. We construct a geography-aware medical knowledge graph that links over 300,000 research articles in leading general medical journals 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 prompts: one that extracts biomedical relations and another that parses study context, mapping the resulting free-text strings to controlled vocabularies via neural embeddings. A new crosswalk aligns medical subject headings with the Global Burden of Disease taxonomy, and a companion module classifies tens of thousands of funders into public, corporate, philanthropic, and hybrid categories.

Overlaying this graph on global health metrics [3] for more than 200 countries and territories from 1990 to 2021 reveals three striking patterns. First, endemic research responsiveness, namely, the elasticity of publication output with respect to domestic disease burden, has more than doubled over the past three decades, signaling 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 a large share of study contexts for neglected tropical diseases and malaria but contribute a much smaller share of authorship. This imbalance reflects persistent authorship disparities and limited local participation in global health research. Third, funding priorities diverge sharply: philanthropic foundations disproportionately support diseases endemic to lower-income regions, whereas corporate research and development focuses on chronic conditions associated with later stages of economic development, with public agencies occupying an intermediate position.

To examine whether these patterns reflect causal responses to emerging health risks, we analyse more than three thousand WHO disease outbreak alerts as quasi-random shocks to perceived burden. Event-study designs show that alerts trigger rapid and statistically significant increases in both domestic and global research output. The surge peaks for emerging and high-profile pathogens, such as COVID-19, Ebola, and Zika, are smaller and shorter-lived for endemic or persistent conditions, such as malaria and diarrheal diseases, and have 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. Diseases such as cardiovascular conditions, neoplasms, and other profitable chronic disorders still attract outsized attention, while maternal–neonatal disorders, nutritional deficiencies, and many infectious diseases remain under-studied relative to their burden. Funding composition and research capacity, rather than disease burden alone, emerge as pivotal drivers of these gaps.

Keywords

Research Responsiveness, Geographic Disparities, Disease Burden, LLM-enabled Knowledge Graph

AI4SciSci'25: Workshop on the Artificial Intelligence and the Science of Science, December 15, 2025, DeKalb, Illinois, USA
*Corresponding author
hz488@cam.ac.uk (H. Zhou); prashant.garg@imperial.ac.uk (P. Garg); t.fetzer@warwick.ac.uk (T. Fetzer)
0000-0002-8250-5875 (H. Zhou); 0009-0009-8072-1863 (P. Garg); 0000-0002-3471-3469 (T. Fetzer)

Declaration on Generative AI

The authors used large language models (LLMs) to assist with grammar and style editing, as well as code generation and refactoring. LLMs were also employed within the research methodology for text analysis supporting the construction of the medical knowledge graph. No figures were generated using generative AI. All AI-assisted outputs were reviewed and verified by the authors, who take full responsibility for the content of this publication.

References

[1] J. A. Evans, J.-M. Shim, J. P. Ioannidis, Attention to local health burden and the global disparity of health research, PloS One 9 (2014) e90147.

[2] A. Yegros-Yegros, W. Van de Klippe, M. F. Abad-Garcia, I. Rafols, 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 (2020) 1–14.

[3] T. Vos, S. S. Lim, C. Abbafati, K. M. Abbas, M. Abbasi, M. Abbasifard, ..., Z. A. Bhutta, 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 (2020) 1204–1222.