Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Global Dataset of Cost-of-illness for Climate-sensitive Diseases through Automated Literature Extraction

Congkai Hong
Tsinghua University https://orcid.org/0009-0008-3287-8439

Han Zhao
Tsinghua University

Zhenping Zhao
Chinese Center For Disease Control and Prevention https://orcid.org/0000-0003-4227-3034

Wenxi Tang
China Pharmaceutical University https://orcid.org/0000-0002-3564-3626

Jianxiang Shen
Tsinghua University https://orcid.org/0000-0003-2685-3813

Rentao Zhang
Lanzhou University https://orcid.org/0009-0000-8498-2117

Wanxin Zhou
The Ohio State University

Mingxin Guo
University of Sydney https://orcid.org/0009-0007-2462-7456

Hanyue Hua
Shanghai International Studies University https://orcid.org/0009-0001-6344-9820

Hanyi Wu
Tsinghua University https://orcid.org/0000-0001-6220-7179

Yuhan Zhu
Tsinghua University

Yihui Liu
Tsinghua University

Yusheng Guan
Tsinghua University https://orcid.org/0000-0003-4676-6469

Siqi Jin
Peking University

Shihui Zhang
Renmin University of China

Mengzhen Zhao

Beijing Institute of Technology https://orcid.org/0009-0008-9043-7279

Wenjia Cai

wcai@tsinghua.edu.cn

Tsinghua University https://orcid.org/0000-0002-4436-512X

Data Note

Keywords:

Posted Date: September 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10939050/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

A Global Dataset of Cost-of-illness for Climate-sensitive Diseases through Automated Literature Extraction

Congkai Hong[1,2], Han Zhao[1,2], Zhenping Zhao[3], Wenxi Tang[4], Jianxiang Shen[1,2], Rentao Zhang[5,6], Wanxin Zhou[7], Mingxin Guo[8], Hanyue Hua[9], Hanyi Wu[1,2], Yuhan Zhu[1,2], Yihui Liu[1,2], Yusheng Guan[10], Siqi Jin[11], Shihui Zhang[12], Mengzhen Zhao[13,14], Wenjia Cai[1,2,*]

[1] Department of Earth System Science, Institute for Global Change Studies, Ministry of Education Key Laboratory for Earth System Modeling, Tsinghua University, Beijing, China.
[2] Institute for Carbon Neutrality, Tsinghua University, Beijing, China.
[3] National Center for Chronic and Noncommunicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine, Beijing, China.
[4] Department of Public Management, School of International Pharmaceutical Business, China Pharmaceutical University, Nanjing, Jiangsu, China.
[5] College of Atmospheric Sciences, Lanzhou University, Lanzhou, Gansu, China.
[6] Key Laboratory of Semi-Arid Climate Change, Ministry of Education, Lanzhou University, Lanzhou, Gansu, China.
[7] Data Analytics Major, College of Arts and Sciences, The Ohio State University, Columbus, OH, USA.
[8] School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia.
[9] School of International Relations and Public Affairs, Shanghai International Studies University, Shanghai, China.
[10] School of Environment, Tsinghua University, Beijing, China.
[11] MEEKL-AERM, College of Environmental Sciences and Engineering & Center for Environment and Health, Peking University, Beijing, China.
[12] School of Ecology and Environment, Renmin University of China, Beijing, China.
[13] School of Management, Beijing Institute of Technology, Beijing, China.
[14] School of Global Governance, Beijing Institute of Technology, Beijing, China.
* Corresponding author. E-mail: wcai@tsinghua.edu.cn

Acknowledgements

This project is funded by the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (No. JYB2025XDXM905), the Inter-Organizational International Cooperation Program of the National Natural Science Foundation of China (in partnership with the Bill & Melinda Gates Foundation, No. 7231101367/2023YFVA1002), the National Natural Science Foundation of China (No. 72204137 & No. 72403022), and the Beijing Association for Science and Technology Youth Talent Support Program (No. 20250829). We also

appreciate Yuze Sun & Qihan Ma (Department of Earth System Science, Tsinghua University) for providing advice on the methodology, and Yang Yang (School of Nursing, Peking University) for curating the research data.

Author contributions

H.Z., and W.C. conceptualized this project; C.H., H.Z., Z.Z., W.T., J.S., Y.G., and W.C. designed the methodology of this project; C.H., R.Z., and Y.G. developed the software for extraction; W.T., R.Z., M.G., and H.H. provided resources of data; C.H., R.Z., W.Z., M.G., H.H., H.W., Y.Z., Y.L., and S.J. curated the research data; C.H., H.Z., R.Z., W.Z., M.G., H.H., and H.W. conducted the formal analysis; H.Z., Z.Z., W.T., J.S., Y.Z., Y.L., and W.C. validated the analysis results; C.H., J.S., W.Z., and H.W. designed the visualization scheme; C.H., and H.Z. drafted the original manuscript; Z.Z., J.S., S.Z., M.Z., and W.C. reviewed & edited the manuscript; Z.Z., and W.C. supervised this project; W.T., S.Z., M.Z., and W.C. were responsible for the funding acquisition.

A Global Dataset of Cost-of-illness for Climate-sensitive Diseases through Automated Literature Extraction

Abstract

As climate change intensifies, extreme weather induced health burden has never been felt so strongly in every corner of the world. However, most of health economic burdens currently relies on value-of-a-statistical-life (VSL) methods, which have long been criticized for having no counterpart in real economic accounts. Cost-of-illness (COI), by contrast, captures directly observable economic costs. Yet a globally comparable COI dataset has remained obscure, owing to fragmented evidences across countries and diseases, diverse classification standards, and the substantial effort required for harmonization. Therefore, we apply multi-step large language model extraction pipelines to present a global COI dataset for over 80 climate-sensitive diseases (CSDs), and design a three-tier confidence scoring framework to reveal evidence transparency and quality. By screening peer-reviewed literature, we extract and harmonize 27,489 cost observations from over 890 publications from January 2015 to July 2025. All monetary values are standardized to 2024 USD, mapped to Global Burden of Disease (GBD) hierarchies, and normalized to per patient per year or per incident. The dataset also includes spatial-temporal, demographic, insurance, and treatment categories information. Transforming fragmented evidences into a harmonized and interoperable resource, this dataset provides the broadest geographic and CSD coverage and the most comparable cost categories among current databases, enabling robust cross-country cost comparisons, climate health economic modeling, and evidence-based climate health policy design.

Introduction

Climate change exacerbates burden of CSDs such as dengue, asthma, and stroke[1,2]. However, current economic assessments remained inadequate. They have largely relied

on abstract proxies like VSL with limited real-world policy relevance[3]. An alternative to resolve this limitation is the COI framework, which reveals health burdens in tangible fiscal realities[4]. As climate vulnerabilities escalate, sensible economic assessments become indispensable for quantifying health co-benefits of climate actions, grounding policymaking in concrete evidence rather than abstract projections.

However, existing COI studies remained misaligned with the requirements for a global and comparable dataset. On the spatial coverage, studies were fragmented and confined to specific diseases, regions, or limited cost types[5–7]. This scope was often designed for localized policy, difficult to obtain a panorama of global disease economic burdens[8]. It cannot meet research demands of matching disease burdens with large-scale external databases. For instance, analyzing the economic impacts of extreme weather across multiple regions and disease types necessitates a unified dataset[9–11]. On the data comparability, the profound methodological heterogeneity across studies became a critical barrier to synthesize COI evidences[12], including variations in cost categories, inclusion criteria, and currencies[13,14]. Even the Global Health Expenditure Database maintained by the World Health Organization (WHO) could only relatively coarse granularity in disease and treatment categories, such as infectious & parasitic diseases and non-communicable diseases. Historically, the lack of advanced tools, particularly before the maturation of large language models (LLM), made large-scale harmonization and classification labor-intensive, preventing cross-study comparisons. Consequently, a major gap persisted in comparable data structures capable of supporting large-scale epidemiological-economic modeling[12,14]. Therefore, researchers face significant obstacles in assessing how climate variability causes economic burdens across health outcomes[15], hindering large-scale comparative studies.

To address these limitations, we aim to establish a global and comparable COI dataset for CSDs, representing a critical advancement over various national/regional evidences. By applying harmonized health accounting frameworks across diverse health systems, it could render health impacts tangible in fiscal terms while ensuring cross-national comparability[16,17], a feature previously absent in COI publications[6]. This harmonized approach would bridge communication gaps among finance, health, and

environmental sectors with a shared language, allowing grasping the magnitude of climate change economic risks in a more sensible way[18,19]. Furthermore, it would enable cross-country cost-benefit analyses of climate policies, facilitating direct and meaningful comparisons between costs and health co-benefits[20–22]. Moving from fragmented reports to a unified global dataset grounds ambitious, cross-sectoral climate health actions in concrete, actionable evidence.

To build the comparable global dataset of COIs for major CSDs, in Step 1, we search several literature databases for peer-reviewed literature published from January 2015 to July 2025, covering more than 80 CSDs across 94 countries. Step 2, we employ three-step LLM-powered automated screening with tailored prompts and repeated independent assessments to efficiently identify eligible publications[23–25]. We collect actual incurred cost evidence while excluding publications without costs or relying solely on modeled estimates. Step 3, we develop a four-step LLM-assisted extraction pipeline to systematically collect cost values and related classification information from full texts. Step 4, to alleviate methodological heterogeneity, we harmonize extracted data through GBD aligned coding, statistical caliber, insurance, and treatment categories, and convert to constant 2024 USD. It avoids labor-intensive and standard-inconsistent manual harmonization and enables fundamental comparisons across countries[26]. Step 5, we design a Confidence scoring system that rates each type of cost for certain disease in certain countries (such as the inpatient cost of stroke in United States, or the out-of-pocket cost of malaria in India); we also design a Credibility scoring system that reveals the sample size, data source reliability, and the calculation method for each publication. These two scoring systems help researchers choose detailed evidences with specific confidence and credibility levels suitable to their specific requirements[27]. The dataset is also designed for open-access with dynamic updates, allowing researchers to efficiently retrieve specific COI metrics through structured queries[16,28]. It transforms fragmented evidence into a comparable, interoperable tool for fundamental international comparisons, climate health economic analyses, and evidence-informed resource allocation.

Results

Structures of COI dataset

The automated extraction pipeline identifies 891 peer-reviewed studies, yielding 27,489 unique COI observations spanning over 80 CSDs. A curated sub-dataset further integrates officially reported national disease costs from several countries. Quality assessment confirms the robustness of the multi-step LLM-driven process. Accuracy exceeds 93% in literature screening steps, with false positive rate of 3.5–4.0%. For basic metadata extraction (countries, years, currencies), accuracy surpasses 90%, while the accuracy of inferred harmonized fields only range 67–87% caused by obvious hallucination. Recall exceeds or approaches 92% across most fields, indicating near-complete capture of LLM compared to manually identifying.

The COI dataset follows established health economic frameworks[29,30], encompassing direct medical costs (hospitalization, medications, outpatient care) and direct non-medical expenses (transportation, informal caregiving). Indirect costs such as productivity losses are captured but excluded from analyses. All monetary values are harmonized to "per patient per year" to focus on the average cost for each patient with specific disease, or "per patient per incident" to measure the cost of each hospital visit or treatment episode, but alternative denominators (total per year, per household per year) are still preserved and annotated. Structurally, the dataset organizes extracted information into harmonized fields covering spatiotemporal context (country, region, year, baseline year), unified cost metrics (standardized statistical calibers and 2024 USD), population (age and gender), and classification characteristics (insurance and treatment categories), as shown in Disease classifications are mapped to the GBD 2021 Nonfatal ICD Code Map at Level 3 granularity (~200 types), or defaulting to Level 2 when the original is overly coarse31. Detailed data processing and harmonization procedures are described in Methods.

Table 1. Notably, columns appended with "_infer" contain LLM-assisted categorized values, preserving original source information while ensuring cross-study analytical consistency. Disease classifications are mapped to the GBD 2021 Nonfatal ICD Code

Map at Level 3 granularity (~200 types), or defaulting to Level 2 when the original is overly coarse[31]. Detailed data processing and harmonization procedures are described in Methods.

Table 1 | Information framework for the dataset of COI for CSDs

Number	Information cluster	Information field name	Definition
1	Spatial-temporal & disease identification	data_country	Country of the cost
2		data_region	Sub-national region of the cost
3		data_year	Year of the cost
4		disease_name_infer	Harmonized disease name
5		disease_icd_infer	Harmonized ICD-10 codes
6	Harmonized cost outcomes	cost_caliber_unified	Harmonized statistical caliber
7		cost_usd_2024_unified	Cost value of per patient per year and per incident converted to 2024 USD
8	Key classification information	population_age	Age group of population related to the cost
9		population_gender	Gender of population related to the cost
10		cost_direct_infer	Harmonized direct/indirect classification
11		cost_insurance_infer	Harmonized insurance reimbursement type
12		cost_treatment_infer	Harmonized medical treatment type
13	Original extracted and process-related information	disease_name	Originally reported disease name
14		disease_icd	Originally reported ICD-10 code
15		base_year	Baseline year of the cost for currency
16		currency	Originally reported currency unit
17		currency_infer	Harmonized currency unit
18		cost_caliber	Originally reported statistical caliber
19		cost_caliber_infer	Harmonized caliber with ununified time scale
20		cost_sample	Sample size of population related to the cost
21		cost_value	Reported and verified cost value
22		cost_usd_2024	Cost value converted to 2024 USD
23		population_income	Income level of population related to the cost
24		population_other	Other population characteristics related to the cost

25		cost_insurance	Originally reported insurance reimbursement type
26		cost_treatment	Originally reported medical treatment type
27		cost_indirect	Originally reported type related to indirect cost
28		cost_other	Other cost categorical types
29		cost_indirect_infer	Harmonized type related to indirect cost
30	Literature information	literature_doi	DOI of the literature
31		literature_title	Title of the literature
32		literature_year	Publication year of the literature

Spatial-temporal, disease, and categorical distributions of COI evidences

Figure 1 presents the spatial-temporal distributions of publications and extracted COI evidence, stratified by WHO regions (Africa, Americas, Eastern Mediterranean, Europe, South-East Asia, and Western Pacific) and World Bank income groups (High income, Upper-middle income, Lower-middle income, and Low income). Geographically, COI evidence in exhibited pronounced concentration, with cost values heavily dominated by China, India, United States, and countries in the region of Western Pacific, whereas Europe demonstrated dense publication activity but lower per study cost granularity. Pronounced evidence voids persisted across Africa, South America, and the Eastern Mediterranean. Temporally, both publications and cost observations showed sustained growth from 2015 to 2025, with most regions peaking during 2019–2020, followed by a notable decline in individual cost evidence thereafter. It suggested a time lag between health economic data accessibility and COI research outputs.

Income-stratified analyses revealed that high- and upper-middle-income countries drove the majority of data production, and lower-middle-income ones began a gradual catch-up only after, while low-income ones remained substantially underrepresented across years. These geographic disparities likely stemmed from systemic inequities in health research funding, institutional capacity for economic evaluation, and routine health information systems[32–34]. The variations in health financing structures and data reporting practices may also explain the high publication volumes but limited per-study

evidence in Europe.

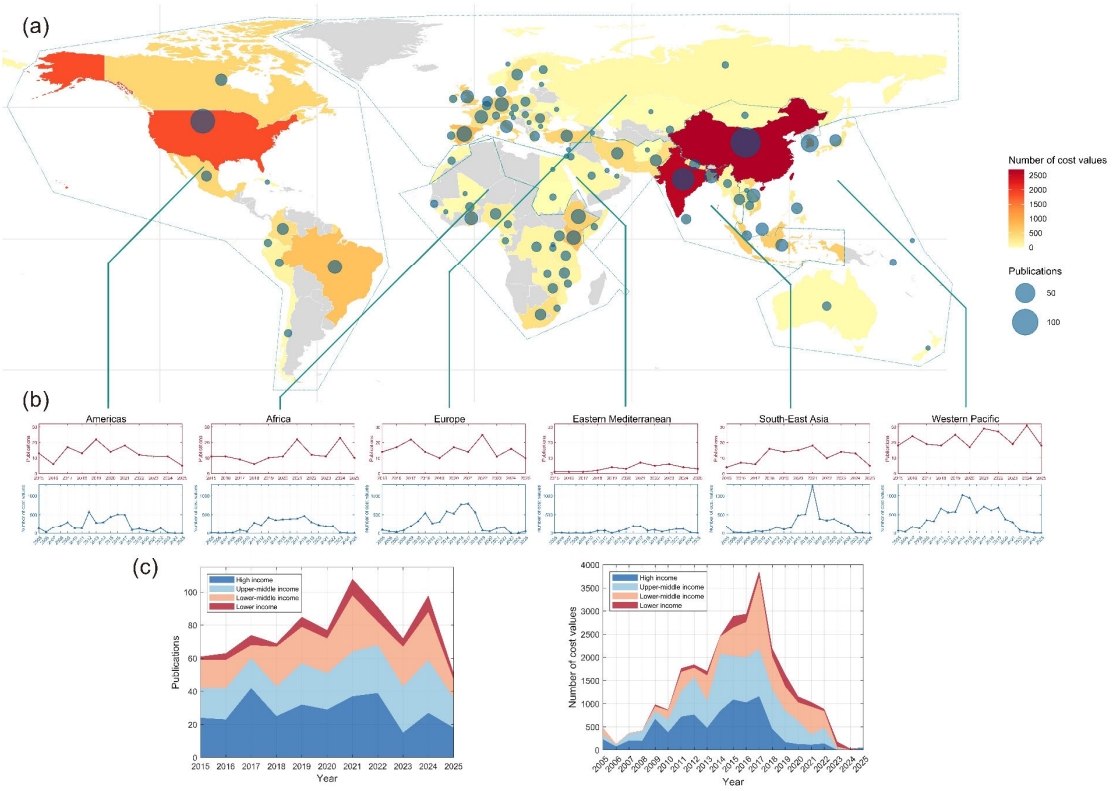


Figure 1 | Spatial-temporal distribution of CSD-COI publications and cost values. *(a) Country-level distribution of publications and cost values. (b) Spatial trends of publications and cost value counts in WHO's six regions. (c) Spatial trends of publications and cost value counts in World Bank country income groups.*

Disease categories are clustered following the GBD Nonfatal ICD Code Map Level 2 classification in Figure 2 (a), consolidated into ten primary CSD categories: respiratory infections and tuberculosis, chronic respiratory diseases, cardiovascular diseases, neglected tropical diseases and malaria, diabetes and kidney diseases, enteric infections, neurological disorders, mental disorders, other non-communicable diseases, and other infectious diseases[1,2]. COI evidence distribution revealed marked heterogeneity, wherein respiratory, cardiovascular, and metabolic diseases dominate the dataset, while other CSDs such as neurological and mental disorders remained sparsely represented. Cross-categorical stratification by age group, insurance, and treatment in Figure 2 (b) highlighted that acute infectious diseases concentrated in pediatric populations whereas chronic NCDs predominated in middle-aged and older adults. Out-of-pocket costs constituted the most frequently reported insurance type across all disease groups, with compulsory insurance showing relatively greater coverage for

chronic conditions. Inpatient and outpatient care costs formed the majority of cost reporting, with pharmaceutical costs constituting a notable secondary component particularly in diabetes and other chronic disease management. Besides, other treatment categories including diagnostics, nursing care, and general administration remained rarely disaggregated, and the gaps in reporting population age and insurance type persisted, collectively reflecting both research prioritization biases and systematic reporting limitations inherent to global health economic studies[13,26].

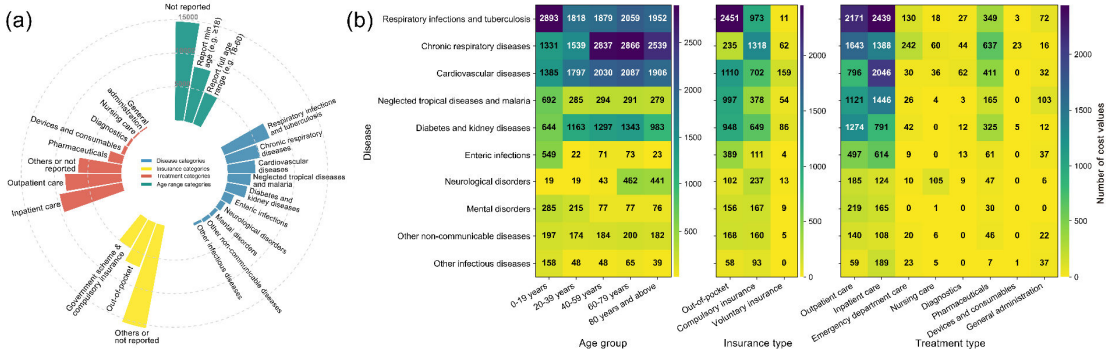


Figure 2 | Count of cost values for CSDs across categories. *(a) Count for each category. (b) Count for each major disease-age/insurance/treatment category intersection.*

Aggregated COI evidences with decomposition details

A distinct and highly valuable category of COI literature focuses on aggregated national or regional economic burdens. Unlike studies that estimate costs for specific patient cohorts, these aggregated studies adopt a top-down approach: they begin with a macro-level total cost, such as national or regional total health expenditure, and subsequently allocate it across specific subcategories (e.g., disease types, care types, or age groups), ensuring that the sum of the components exactly equals the reported total. Such decomposition details are crucial for researchers aiming to obtain a holistic view of total economic burdens. To systematically identify these studies from the massive literature pool, we apply LLM to extract and classify the "cost calculation method" from each included publication (detailed prompts are provided in the Supplementary information). By analyzing these extracted methodological details, we successfully identify 36 publications that report aggregated COI evidence with explicit decomposition details. Evidences are summarized in Table S5 in Supplementary information.

The 36 aggregated publications span 24 countries/regions across multiple WHO regions, with more emphasis on relatively developed ones like China, France, and Spain. These publications cover both communicable and non-communicable diseases, especially cardiovascular diseases, chronic obstructive pulmonary disease (COPD), diabetes, and malaria, with research periods ranging from 1990 to 2022. Notably, even for the same disease, statistical caliber, and treatment type, cost values varied dramatically across regions. For instance, per patient per year costs for COPD ranged from 350 USD in South Korea to over 24,000 USD in France. These disparities are largely attributable to differences in the scope of cost: the South Korean COPD study included both direct and indirect costs for general COPD, whereas the French study captured only COPD-related hospitalization costs. Consequently, even for the same disease, cost category, and statistical caliber, values cannot be directly substituted across regions due to differing statistical definitions, healthcare systems, and socioeconomic development. When using these costs, it is important to carefully delineate the specific indications and scope of each cost measure. Additionally, researchers must be very careful if necessary to transfer the cost of a certain disease from one country to another (similar to some studies transferring VSL from one country to another). The current data quality and quantity are not sufficient to verify whether this transfer is reliable or reasonable.

Having said that, aggregated evidences with decomposition details are particularly suitable in practice when researchers aim to obtain a holistic view of total economic burden for a given disease or region. They can also perform arithmetic operations across cost categories, such as summing inpatient and outpatient costs[8,35]. When retrieving such data from original publications, we recommend the following search strategy: use keywords such as “total cost”, “aggregate cost”, or “economic burden” combined with the target disease and region; examine methods section for phrases like “aggregated”, “top-down”, or “decomposition”; and verify that the cost year, currency, and discounting details are clearly reported to enable proper harmonization. It helps ensure both data consistency and methodological transparency.

Case-based COI evidences of representative CSDs and countries

For each type of cost, we adopt a three-tier confidence framework adapted from Intergovernmental Panel on Climate Change (IPCC) guidance[36,37]: (1) Evidence, quantified by the count of values and publications composing each type of cost; (2) Agreement, measured as relative interquartile range (RIQR=IQR/median), reflecting the statistical dispersion among extracted cost values; (3) Credibility, which evaluates the methodological reliability of the underlying publications through three sub-components: database authority and scale (assessing the official credibility and spatial coverage of data sources), sample size (rating the magnitude of the study population), and cost calculation method (classifying the completeness of the cost estimation approach, in Section 2.3), each rated on a 1–5 integer scale and averaged into an overall publication credibility score[38]. This multi-dimensional scoring provides each type of cost with confidence descriptors:(1) The two numbers for Evidence (the first number denoting count of values and the second denoting count of publications), (2) Background colors for Agreement, and (3) Multiple "+" signs for Credibility. Researchers could weigh the reliability of evidences according to their requirements and risk tolerance[27,39].

For researchers with focusing on cross-country cost comparisons or intra-country comparisons across diseases rather than requiring high completeness of cost subtypes, we calculate the median and interquartile range for extracted cost values sharing similar definitions, as these robust statistics better accommodate the right-skewed distributions of COI data[40]. To provide representative cases, we select five CSDs: COPD, stroke, diabetes mellitus, lower respiratory infections, and malaria, spanning chronic and acute conditions across major diseases. Focused countries/regions include the United States, European Union, China, India, and Ethiopia, collectively covering all World Bank income groups and major WHO regions. For each disease-country pair, we identify all values extracted from case-based bottom-up publications and got two or more "+" signs. They are aggregated costs from individual patient- or treatment-level sources rather than decomposed costs from total amount. Stratifying cost values by insurance and treatment categories, we exhibit chronic diseases (COPD, stroke, and diabetes mellitus)

using “per patient per year” calibers and acute infections (lower respiratory infections and malaria) using “per patient per incident”.

Representative results of government or compulsory insurance and out-of-pocket costs are shown in Figure 3. Global COI distributions revealed pronounced geographic disparities: high-income countries exhibit substantially higher median costs than low-income counterparts. For instance, the median out-of-pocket cost of diabetes in US was 860 USD, versus 389 and 23 USD in India and Ethiopia. Evidence of COPD and stroke dominated US and EU data, whereas malaria and lower respiratory infections prevailed in India and Ethiopia. Notably, chronic diseases drove upper-quartile cost extremes (35,371 for US COPD and 22,625 USD for diabetes), exposing both insurers and households to substantial financial risks.

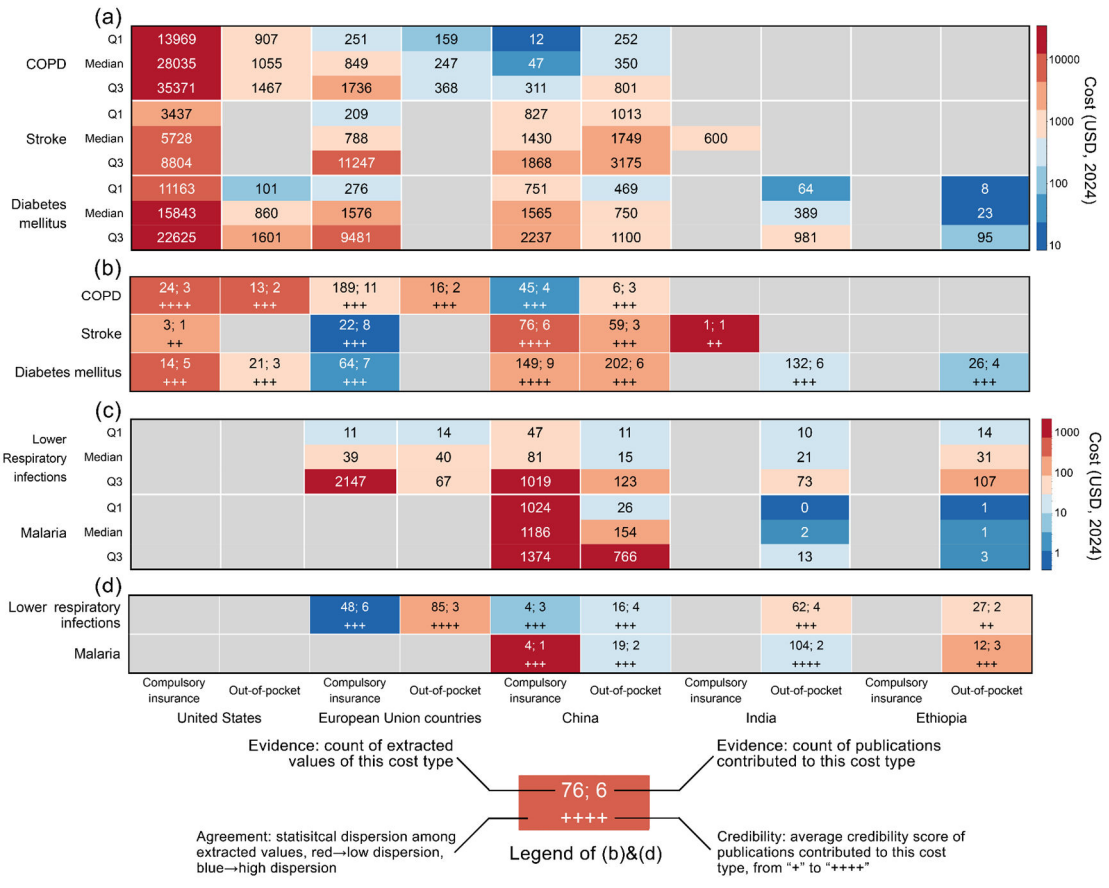


Figure 3 | Medians and quartiles of major insurance costs on representative diseases in several countries. *(a)(b) Costs and related confidence information on chronic diseases. (c)(d) Costs and related confidence information on acute infections. Costs in (a) are per patient per year, in (c) are per patient per incident. Q1 means lower quartile, Q3 means upper quartile.*

Representative results of outpatient care, inpatient care, and pharmaceuticals costs

are shown in Figure 4. Inpatient costs consistently exhibited the highest median values across countries, for example, inpatient median costs for COPD, stroke, and diabetes were 5,484, 5,728, and 12,324 USD in US, versus outpatient ones of 4,984, 1,146, and 3,643 USD. Cross-country comparisons revealed that the US had substantial cost gaps between inpatient and others, while India and Ethiopia displayed compressed ranges with extreme of 436 USD. Notably, outpatient and pharmaceutical cost medians approached or exceeded inpatient values in certain chronic diseases such as the COPD in US and the stroke in China, aligning with evolving care models emphasizing decentralized, continuous management[41–43].

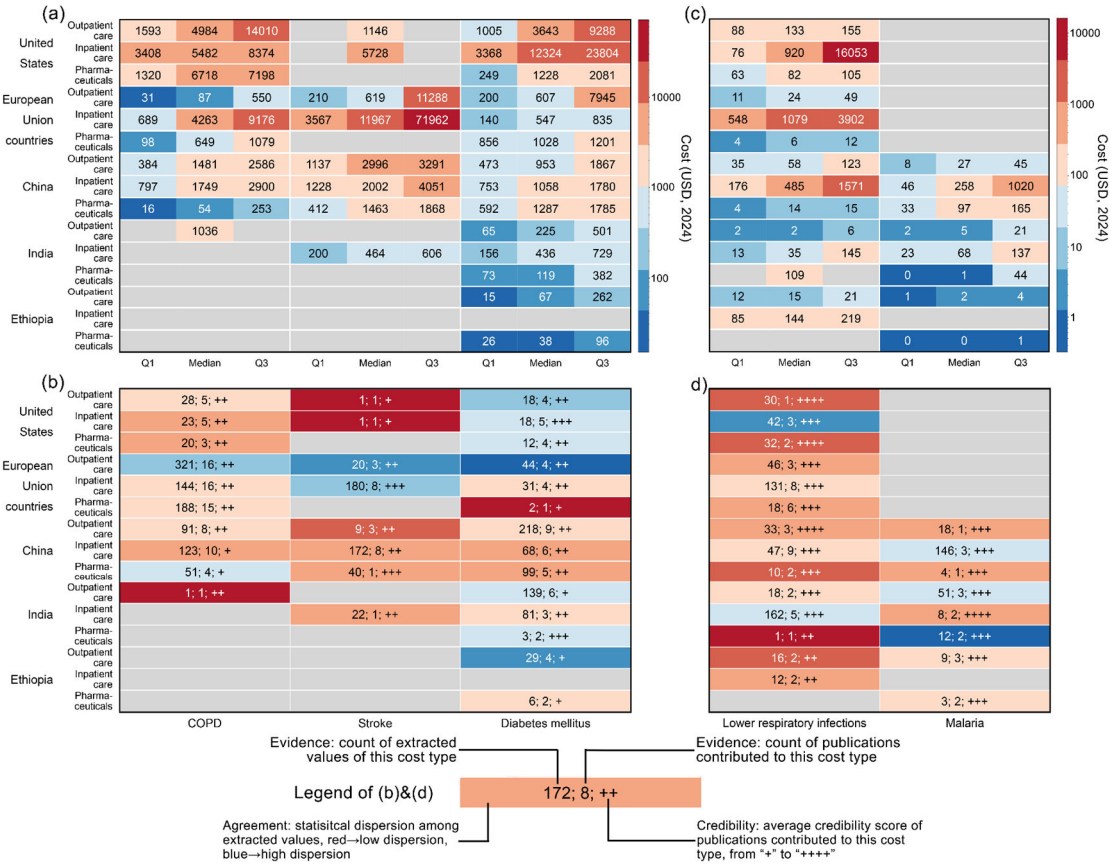


Figure 4 | Medians and quartiles of major treatment costs on representative diseases in several countries. *(a)(b) Costs and related confidence information on chronic diseases. (c)(d) Costs and related confidence information on acute infections. Costs in (a) are per patient per year, in (c) are per patient per incident. Q1 means lower quartile, Q3 means upper quartile.*

For other researchers requiring detailed data without the need for large-scale comparisons, the full COI dataset is available for targeted searches and customized matching. In addition, we have compiled a separate subset of officially reported COI data from several countries. Researchers with particularly strict requirements for

authoritative sources may consult this dedicated mini-dataset for retrieval.

Case-based data are particularly suitable for researchers who require detailed, subcategory cost structures, such as the breakdown of total cost into inpatient, outpatient, and pharmaceutical costs, rather than merely a single aggregate value. This harmonized global COI dataset is well-suited for researchers aiming to estimate the economic burden of a specific disease in a specific country or to compare costs across different subcategories (e.g., age groups, genders, and treatment types). For practical application, researchers can multiply local incidence or prevalence data by the median per patient/incident cost to approximate total national/regional expenditures, with interquartile ranges providing plausible uncertainty bounds. The dataset also remains accessible for precise matching of costs to specific insurance, treatment categories and patient demographics. However, due to highly heterogeneous sample sizes, diverse data sources, and uneven demographic distributions across included publications, arithmetic operations between values are strongly discouraged. Instead, analyses should prioritize order-of-magnitude comparisons, directional trend assessments, or scenario-based sensitivity analyses that respect the underlying variability of COI data[44].

Discussion

We provide the first global, comprehensive, and relatively comparable dataset dedicated to the economic burden of climate-sensitive health conditions, built through an automated literature screening and extraction pipeline powered by LLMs, with harmonized cost categorization, inflation adjustment, and uniform disease coding. Although the current dataset is modest in scale and lacks the granular clinical or microeconomic detail typical of targeted primary studies, it represents a substantial empirical and methodological advancement over prior research, which has largely been confined to single diseases, specific geographic regions, or narrow literature pools, limiting cross-study comparison and policy generalization[5–7]. We acknowledging the need for continuous expansion and more refined processing as primary COI studies increase, and we have established a foundational infrastructure to inform climate-

adaptive health financing, resource allocation, and future burden modeling[3,16].

To enhance the interpretability and utility of this resource, we introduce a transparent Credibility scoring framework that evaluates each publication across three dimensions: database authority and scale, sample size, and cost calculation method, rated on a 1–5 integer scale[38]. These components are then averaged into an overall credibility value, and act as an input for the three-tier Confidence framework of Evidence, Agreement, and Credibility applied to each type of cost. This scoring mechanism allows researchers to explicitly weigh data quality when selecting evidence for their specific analytical needs[45]. It provides a direct, transparent gauge of evidence quality when interpreting results, supporting systematic comparisons of costs across diverse epidemiological and socioeconomic conditions[38,46].

The harmonized cost categories and comparable metrics of this dataset substantially enhance its utility for downstream research across climate change and health economics. By aligning costs with GBD disease hierarchies, uniform currency, and WHO System of Health Accounts (SHA)-related insurance and treatment categories, the dataset enables robust international and multi-disease comparisons[8,12]. The comparable cost classification further offers actionable evidence for evaluating adaptation interventions, prioritizing resource allocation, and designing resilient financing mechanisms for climate vulnerable populations[47]. Ultimately, this infrastructure bridges disparate analytical traditions and translates climate health economic evidence into equitable, data-driven adaptation strategies, while also enabling empirical quantification of climate-attributable health costs and projecting future trajectories under warming and socioeconomic scenarios.

Despite the advantages of our automated pipeline, several method limitations warrant acknowledgment. First, because most cost values were extracted from complex source tables, omission errors occasionally occurred when tables contained highly granular or numerous cost categories, leading the LLM to overlook specific types or aggregated totals. Second, misclassification occasionally arose when details appeared in “Methods” sections rather than explicitly repeated within cost tables, thereby increasing the likelihood of LLM hallucination during cross-referencing[48,49]. Third,

inference and categorization proved challenging for studies employing highly detailed cost breakdowns (e.g., specific drug or procedure names) and unconventional frameworks (e.g., symptom-based categories), which often forced coarse or imprecise mapping onto our harmonized insurance and treatment categories. To address these inconsistencies, we implement a series of pragmatic assumptions and data processing compromises, including conservative fallback rules, hierarchical categorization, and structured validation protocols. While these constraints inevitably introduce some measurement uncertainty, they do not fundamentally compromise the dataset's overall completeness or robustness. The automated framework successfully harmonizes fragmented economic evidence across diverse epidemiological and geographic conditions, maintaining a high degree of scalability. Importantly, these limitations reflect inherent challenges in large-scale literature mining and comparable COI synthesis rather than flaws in our methodology design, and they do not diminish the novelty of our work. As AI-driven extraction techniques continue to mature, iterative refinement of prompt architectures and hybrid human-AI validation will further mitigate these issues, ensuring that the current dataset remains a reliable and foundational structure for climate health economic assessments[50,51].

We envision the next phase of climate health COI or economic burden dataset as an empirical, collaborative, and continuously evolving infrastructure. To operationalize this vision, we will deploy the dataset on a dedicated open-access platform featuring version-controlled updates and a structured community contribution mechanism. Researchers will be able to submit newly extracted COI estimates, annotate extraction ambiguities, and co-develop classification protocols through peer-moderated workflows. Such a living-database model aligns with emerging open-science paradigms and will continually improve data currency, transparency, and geographic equity[16] (https://pandemichub.who.int/news-room/news/13-04-2026-who-hub-in-berlin-announces-the-launch-of-first-global-repository-of-epidemiological-parameters).

Since we have found that there are still large missing in less-developed countries and in detailed cost categories, targeted expansion is urgently needed to close persistent coverage gaps, particularly for low- and middle-income countries, underrepresented

non-communicable conditions, and emerging CSDs. Improving data acquisition and classification accuracy will require iterative refinement of hybrid human-AI extraction pipelines, standardized validation benchmarks, and systematic integration of national health accounts and institutional reports that remain underutilized in current syntheses[51]. To thoroughly improve harmonization and comparison, we strongly advocate for the establishment of internationally unified reporting standards for COI research. The pervasive heterogeneity in statistical calibers, insurance categories, and indirect cost accounting severely constrains multi-level comparisons and meta-analytic robustness. By adopting consensus guidelines for cost classification, inflation adjustment, and cost processing, the global research community can construct scalable, interoperable COI databases that seamlessly link with major epidemiological and health economic infrastructures, including the GBD, the UK Biobank (UKB), and national health expenditure databases[52,53]. Comparable COI data would not only enable precise attribution of climate attributable economic burdens across demographic strata but also empower health economists and policymakers to model the cost-benefits of climate adaptation interventions with empirical evidence. Therefore, we call upon the academia, funding agencies, and international health organizations to incentivize open-data contributing & sharing, mitigate data gaps & imbalance, develop more precise AI extraction methods, mandate standardized economic values reporting, and invest in coordinated data infrastructure that bridges climate science, health economics, and equitable policy planning.

Methods

Study design

This study is designed and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The overall research flow is plotted in Figure 5. We first search literature using predefined climate-sensitive disease and cost-related keywords, then apply a two-step extraction: parsing cost tables and capturing key information from paragraphs. All extracted values are

harmonized via inflation adjustment, GBD-aligned disease coding, and statistical caliber classification. Finally, we introduce a credibility scoring system across database authority, sample size, and cost calculation method to enable transparent evidence quality assessment. This reproducible flow supports transparent evidence quality assessment and flexible data use for various research and policy applications.

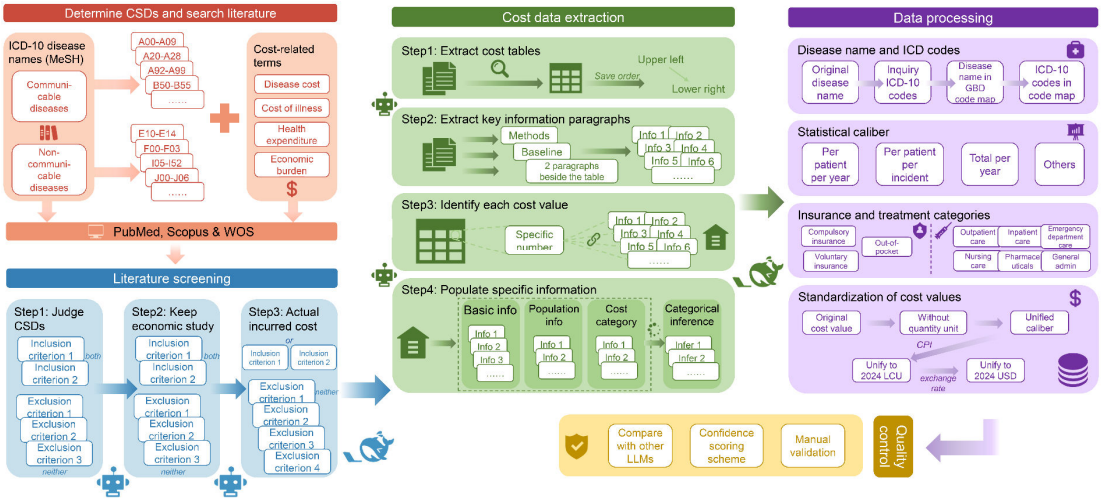


Figure 5 | Research flow of this research containing searching literature, literature screening, cost data extraction, data processing, and quality control.

Search strategy

We conduct systematic searches across PubMed, Scopus, and Web of Science (WOS) to synthesize peer-reviewed literature on the costs, expenditures, and economic burden of CSDs. First, we establish a list of CSDs based on the WHO's ICD-10 classification and existing literature[2,54,55]. These diseases are categorized into communicable and non-communicable diseases, comprising more than 80 conditions such as plague, dengue fever, malaria, diabetes, cardiovascular diseases, and COPD. Specific names and codes are provided in Supplementary information.

Second, we group these 80+ diseases into 28 types and combined disease names with cost/expenditure-related terms to create 28 composite keyword sets for database queries. Within each set, disease names are linked using "OR", and the same as cost-related terms, while the two groups are connected using "AND". Disease names primarily utilized Medical Subject Headings (MeSH), while cost-related terms included "health expenditure", "health spending", "cost of illness", "economic burden", and "disease cost". The specific composite keywords are listed in Supplementary

information.

We perform subject searches in each database (though some are not called "subject search"), limiting the publication period from January 2015 to July 2025 without language restrictions. The majority of records were in English, while a small number of French and Spanish documents indexed in the databases were also included. We export metadata including authors, titles, sources, DOIs, abstracts, and keywords for all retrieved records, and remove duplicates to establish the baseline for the screening phase.

Eligibility criteria and literature screening

Given the large number of CSDs and the substantial volume of retrieved literature, we design a three-step screening process to precisely identify CSDs, exclude non-cost/economic studies, and filter for actual incurred costs/expenditures[18,56,57]. The eligibility criteria for these steps are as follows:

Inclusion Criteria (meet all):

(1) The study explicitly mentions at least one of the 89 CSDs and focuses its research, analysis, and discussion on that disease;

(2) The study primarily quantifies, estimates, or compares disease costs/expenditures and presents data on costs, expenditures, or economic burden;

(3) The reported disease costs represent actual incurred costs, including direct costs (e.g., hospitalization, medication, transportation) and indirect costs (e.g., reduced wages and productivity losses)[30].

Exclusion Criteria (meet any one):

(1) The study does not mention any CSDs, or only mentions them briefly without sufficient research, analysis, or discussion;

(2) The study is focused on biology, clinical medicine, or epidemiology without any cost or economic analysis, or is a review, commentary, theoretical discussion, or book chapter;

(3) The reported costs are modeled projections, expected costs, cost gaps, or other non-actual numbers, such as cost changes or benefit-cost ratios[18,57].

Given the large volume of literature, we employ LLM to batch-screen abstracts

based on the aforementioned inclusion and exclusion criteria[23–25]. The critical factor in using LLM for batch analysis is crafting appropriate prompts that fully articulate task requirements while effectively minimizing hallucinations, a practice known as "prompt engineering"[58]. We design specific prompts for each of the three screening steps, adhering to the following principles: first, clearly defining or explaining key concepts; second, breaking down inclusion and exclusion criteria into distinct checkpoints for step-by-step verification to clarify screening rationale; and third, using formatted outputs with examples (in-context learning) to reduce the likelihood of hallucinations[59–61]. Specifically, our prompts list CSDs, define general characteristics of cost studies, decompose each step's criteria into 2–5 checkpoints for logical evaluation, and specify JSON format for output with a one-shot example[62]. The detailed prompts are provided in Supplementary information.

We use the official Application Programming Interface (API) of DeepSeek to call the DeepSeek-V3.2 model for batch screening, setting the temperature parameter to 0.5 to slightly reduce model creativity. The rationale for selecting this LLM is stated in Supplementary information. The process is implemented using Python 3.14. To ensure the robustness of the screening results, each article undergoes three independent screening rounds at each step. Only articles that pass at least two of the three rounds are retained for subsequent steps; others are excluded[60]. Detailed information about literature screening is shown in Figure 6.

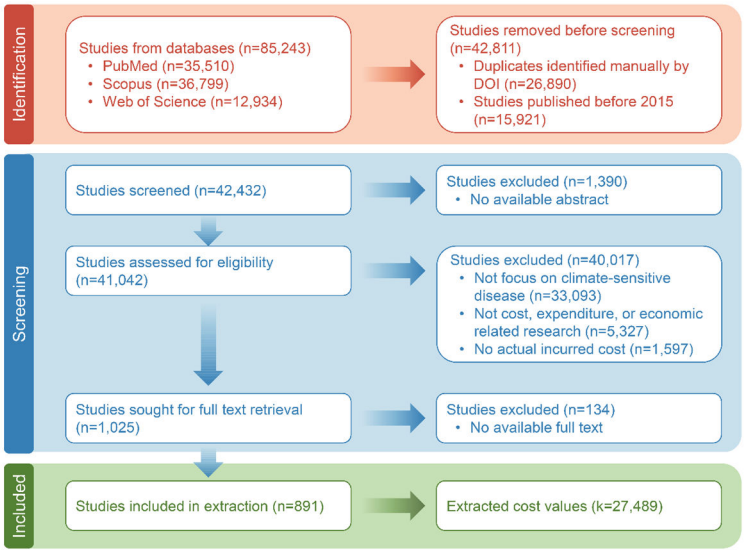


Figure 6 | PRISMA flowchart of searching and screening literature

Cost data extraction pipeline

Based on the screening results, we download the full texts of eligible articles when possible. We apply publisher-specific channels, including the Elsevier Article (Full Text) Retrieval API, NCBI Entrez Programming Utilities (E-utilities, for open-access articles only), Springer Nature Open Access API, and Wiley Text and Data Mining (TDM) API, to batch-download full texts and save them as XML or TXT files. A small number of articles are excluded due to constraints such as limited publisher coverage, availability only in PDF format, or lack of subscription access. The batch downloading process is implemented using Python 3.14 and R for Windows 4.5.1.

We design a specialized information extraction template to extract data on the costs of CSDs and related information from each article. In addition to cost values and currency units, the extracted information include: (1) population characteristics: country/region, year, age group, gender, income level, etc.; (2) cost categorization: disease name, ICD code, statistical caliber & sample size, insurance reimbursement type, treatment type, and indirect cost type; and (3) bibliographic details: title, publication year, authors, and article type. The complete information extraction template is provided in Supplementary information.

It is a significant challenge to extracting comprehensive cost/expenditure data for CSDs and to accurately identify various associated information from a vast volume of literature with diverse disciplines, structures, and writing styles[61]. Literature indicates

that excessively long prompts and complex tasks significantly increase the likelihood of LLM hallucinations[48]. Therefore, we design a four-step LLM extraction pipeline, with specialized prompts for each step, to perform functions such as extracting cost tables and key paragraphs, identifying individual numerical values, and populating specific items:

(1) Extraction of Cost Tables: Since LLMs struggle to directly identify cost values from full texts containing thousands or tens of thousands of words, we first extract tables or paragraphs primarily presenting disease costs/expenditures[25]. We apply inclusion and exclusion criteria to each table in the literature. Inclusion criteria require the presence of cost-related keywords (e.g., cost, expenditure, spending, economic) and currency units. Exclusion criteria are tables lacking economic analysis or primarily containing unit prices, cost changes, or benefit-cost ratios. If no tables meet the criteria, we sequentially evaluate each paragraph in the "Results" section using the same criteria. Finally, all included tables or paragraphs are output in JSON format.

(2) Extraction of Key Information Paragraphs: Given that information related to disease costs is complex but concentrated in the "Methods", "Baseline results", and sections adjacent to cost tables, we decide to extract these paragraphs for subsequent centralized information extraction[24]. First, we categorize cost-related information from the extraction template into six groups: "year and region", "inflation and base year", "population characteristics", "disease name and code", "cost categories", and "statistical caliber and sample size". Second, we design 2–5 key information points for each group and instruct the LLM to sequentially locate the text paragraphs containing each point, focusing on the "Methods", "Baseline results" and the two paragraphs preceding and following cost tables. Finally, we output the paragraphs corresponding to the six groups in JSON format. To prevent LLM hallucinations and context overflow due to excessive length, we control the total length of output for each group[63].

(3) Identification of Cost Values: Building on the relatively structured cost tables and paragraphs, we establish a precise and comprehensive process for identifying individual cost values. First, we restrict the identification path to traverse every number

from top-left to bottom-right. Second, we design inclusion and exclusion criteria: inclusion requires the presence of cost-related terms (e.g., cost, expenditure, economic) and currency symbols (including in table titles or footnotes); exclusion applies to statistics (confidence intervals, p-values, etc.), clinical or epidemiological results, unit prices, and change amounts. Additionally, we extract any cost category information appearing in the tables (mainly headers and footnotes), such as disease name, statistical scope, country/region, year, reimbursement type, and treatment type, recording them in a standardized hierarchical format that prioritized fidelity to the original text over speculation. Finally, each identified value and its associated cost category information are output in JSON format.

(4) Populating Specific Information: For each cost value identified in Step 3, we use the original table/paragraph, the key information paragraphs extracted in Step 2, and the cost category information recorded in Step 3 as inputs to populate the complete information extraction template. To correspond to types of information requiring completion, this step is subdivided into four sub-steps[60]: Step 1 fills in basic cost information, including year, country/region, disease name, and cost value; Step 2 fills in population characteristics corresponding to the cost, such as age group, gender, and income level; Step 3 faithfully fills in cost category information, including statistical caliber, sample size, reimbursement type, and treatment type; and the Final Step performs categorical inference on the information from Step 3, involving statistical caliber, reimbursement type, and treatment type[59].

Literature indicates that pre-specifying filling formats or providing limited options helps ensure accurate completion and structured output[62,64]. Therefore, in the prompts for each sub-step, we define the content, format, and allowable options for every information item. Regarding content, for example, disease_name is restricted to official English disease names, excluding symptoms, organs, or risk factors. For format, population_age is required to follow patterns such as “X-Y years” or “X years and above”. For options, cost_insurance is limited to three types: “government and compulsory insurance”, “voluntary private insurance”, and “out-of-pocket”. Finally, each cost value and its associated information are output in JSON format to facilitate

554 the construction of a structured data table. The detailed prompts are provided in
555 Supplementary information.

556 **Quality assessment**

557 To evaluate the robustness of automated literature screening, we conduct a
558 comparative assessment against two alternative LLMs, GPT-4.1 mini and Qwen3.6-
559 Plus. A systematic random sample comprising 5% of the retrieved articles is selected
560 for validation, consistent with established quality control practices for AI-assisted
561 systematic reviews[60,61]. Each sampled abstract, combined with the identical step-
562 specific prompts and eligibility criteria, is independently processed by both LLMs
563 through the same three-stage screening strategy. We compute inter-model accuracy,
564 precision, recall, false positive rate, and F1 score across all stages to quantify
565 consistency in identifying CSDs, economic analyses, and actual incurred costs. This
566 cross-platform benchmarking determines whether screening outcomes reflect stable
567 decision boundaries or are model-dependent, thereby verifying the reproducibility and
568 methodological rigor of our LLM-driven literature selection process[23].

569 Results in Supplementary information show that, the multi-step LLM-driven
570 screening demonstrated robust quality, characterized by consistently high accuracy
571 (>93% in steps 2–3) and a markedly reduced false positive rate (3.5–4.0% in step 3)
572 across LLMs. This cross-model consistency confirmed the screening's structural
573 stability and effective noise-filtering capacity. However, precision remained modest
574 (23–49%) and recall gradually declined in later stages. These limitations likely
575 stemmed from increasingly stringent exclusion criteria, which may inadvertently
576 discard borderline-relevant studies, alongside LLMs' inherent conservatism when
577 processing complex contextual matching tasks. Despite this trade-off, the screening's
578 high specificity ensured a rigorously curated candidate pool, substantially reducing
579 downstream manual validation burden. The controlled recall reduction was deemed
580 acceptable within a "high-specificity-first" framework, aligning with systematic review
581 standards in which preventing false inclusions typically outweighs the cost of minor,
582 retrievable omissions.

583 To validate the performance of LLM-driven cost data extraction pipeline, we

implement a systematic manual audit of the final extraction results[25]. A random sample comprising 2% of all extracted cost values is independently reviewed[60]. For each sampled value, trained reviewers cross-reference the extracted metadata—including research country, year, disease name, original cost value, currency, age group, statistical caliber, insurance, and treatment type—against the literature full texts. Each information item is adjudicated into one of four categories: correct, incorrect, omission, or extraneous. The "correct" is denoted as true positive (TP); "incorrect" is false positive (FP); "omission" is false negative (FN); "extraneous" is also FP; and there is no situation of true negative (TN). Rigorous matching criteria are enforced: deterministic items including values and currency require exact correspondence to be deemed correct, whereas conceptual items (e.g., insurance and treatment types) are considered correct if their clinical or economic meaning is semantically equivalent to the original text[59,65]. Item-specific accuracy, precision, recall, and F1 score are subsequently calculated across all audited fields. This human-in-the-loop verification ensures that automated extraction maintains high fidelity and mitigates algorithmic bias prior to downstream economic synthesis[66].

The results of the samples (n=561) are independently examined by three researchers (Y. Z., Y. L., & S. J.) in 9 information fields: data_country, data_year, disease_name, cost_value_correct, currency, population_age, cost_caliber_infer, cost_insurance_infer, cost_treatment_infer. We categorize results into four types, and calculate indicators including accuracy, precision, recall, and F1 score. Results are shown in Supplementary information. Overall, the LLM-driven extraction framework showed high agreement with full manual extraction. For extraction of basic information like countries, years, and currencies, accuracy surpassed 90%. But for information fields processed by categorical inference and harmonization, the accuracy declined, ranged only 67%–87% caused by obvious hallucination. Precision stayed consistent with accuracy, and recall exceeded or approached 92% in most of fields, indicating the LLM captured nearly all manually identified information.

Data processing

After extracting information related to the costs of CSDs, we harmonize and

categorize specific items to ensure consistency and comparability in subsequent analyses. We design specific processing rules for most information items, with key items including disease names & codes, statistical calibers, reimbursement and treatment types, and currency conversion.

For disease names & codes, we employ the GBD Code Map for standard processing and categorization[31]. First, we normalize original disease names using MeSH and identify their corresponding ICD-10 codes by referencing databases like ICDcodes.ai (https://icdcodes.ai/) and ICDLIST.com (https://icdlist.com/); this step is skipped if ICD-10 codes have already been extracted from the literature. Second, we reference the GBD 2021 Nonfatal ICD Code Map to identify the Cause Hierarchy Level 3 disease name containing the specific ICD-10 code, adopting this as the standard disease name; if multiple Level 3 diseases are involved, we list no more than three. Finally, if no ICD-10 code is extracted from the literature, we adopt the complete ICD-10 codes corresponding to this Level 3 disease as the standard disease code; if ICD-10 codes are explicitly extracted, we use that extracted codes as the standard disease code.

Next, we categorize the statistical caliber of the costs. Extracted cost values are classified into calibers such as "per patient per year", "per patient per incidence/course", "total per year", and other minor calibers (e.g., per household per year, per patient lifetime). Where possible, we convert values based on time periods and sample sizes to align the majority of cost data with either "per patient per year" or "per patient per incidence/course". Values that could not be classified into these two primary calibers are retained under their original calibers.

We also categorize reimbursement and treatment types. Reimbursement types are restricted to three types from SHA: "government and compulsory insurance", "voluntary private insurance", and "out-of-pocket"[67]; unreported or mixed reimbursement types are marked as NR. Treatment types are classified according to the "Functions" classification of healthcare services in the US National Health Accounts (NHA) framework, encompassing six types: "outpatient care", "inpatient care", "emergency department care", "nursing facility care", "pharmaceuticals", and "general administration"[68]. We add two additional types: "diagnostics", and "devices and

consumables". Treatment items that are unreported or cannot be classified into the above types are marked as NR.

To eliminate heterogeneity in quantity units, currency units, and price levels across the raw cost data, we convert all extracted cost values. First, we remove quantity unit descriptors (e.g., thousand, million) from the original data and convert them into absolute numerical values. Second, for currency and price level adjustments, we set 2024 as the target base year. Based on the cost base year and country identified in each study, we deflate cost values to 2024 levels using the Consumer Price Index (CPI) for each country released by the World Bank (https://data.worldbank.org/indicator/FP.CPI.TOTL); for countries where CPI was not reported, we use the GDP deflator as a substitute. For Eurozone countries, we use individual national CPIs rather than the Harmonized Index of Consumer Prices (HICP). For costs reported in international dollars, we apply CPIs of USA reported by the World Bank. Finally, we convert the costs to constant 2024 USD using the market exchange rates between the USD and local currencies (LCU) for 2024, as also released by the World Bank (https://data.worldbank.org/indicator/PA.NUS.FCRF). If the original literature did not explicitly report the cost base year, we use the publication year of the article as the proxy.

Confidence scoring scheme

To facilitate transparent analysis of COI, we develop a confidence scoring scheme to evaluate both the literature and the extracted cost values. For each publication, we apply the LLM to rate three components: database authority and scale, sample size, and cost calculation method. An overall literature Credibility score is then calculated as the normalized average of these three component scores, which is further grouped into five credibility levels. For each type of cost, the Confidence score is composed of Evidence, Agreement, and Credibility. The Evidence and Agreement denote the amount of evidences and the statistical consistency within extracted values of this cost type. The Credibility is calculated by averaging credibility scores of all publications contributed to this cost. This hierarchical scoring system allows researchers to evaluate evidence quality at both levels, facilitating transparent data selection from broad comparative

analyses to high-certainty policy applications[36,46].

To provide another option of disease cost data, we identify and compile official COI datasets from several countries, alongside publications that explicitly utilized national health accounts. The selected countries include China, US, UK, Australia, and several less developed countries. These official figures retain their original monetary values, base years, and native classification schemes without algorithmic harmonization or imputation. They are stored as a separate supplementary sub-dataset and explicitly excluded from all core analyses, including median estimations and cross-stratified comparisons, to maintain the methodological integrity of the primary database. This curated subset provides a policy-grade reference for researchers seeking authoritative national benchmarks, while ensuring that the main dataset remains strictly based on peer-reviewed, transparently estimated evidence.

Data availability

The main COI dataset "Standardized Global Dataset of Cost-of-illness (COI) for Climate-sensitive Diseases" is publicly available via GitHub at https://github.com/congkaihong-thudess/Global_Cost-of-Illness_Dataset_for_Climate-Sensitive_Diseases_COI-CSD_v0.git. We plan to update it regularly, depending on the availability of APIs for the LLM and publishers. We also welcome contributions from the academic communities, including recent COI publications/data resources, rigorous classification & harmonization frameworks, and improved LLM extraction pipelines.

Code availability

The codes of Python (and part of R) for literature screening, cost data extraction, and confidence scoring are available via GitHub at https://github.com/congkaihong-thudess/Global_Cost-of-Illness_Dataset_for_Climate-Sensitive_Diseases_COI-CSD_v0.git. The prompts of LLMs for literature screening, cost data extraction, and confidence scoring are provided in Supplementary information.

References

1. Haines, A. & Ebi, K. The Imperative for Climate Action to Protect Health. *N. Engl. J. Med.* **380**, 263–273 (2019).
2. Intergovernmental Panel On Climate Change (Ipcc). *Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. (Cambridge University Press, 2023). doi:10.1017/9781009325844.
3. Shen, J. *et al.* Improving cost–benefit analyses for health-considered climate mitigation policymaking. *Nat. Clim. Change* **15**, 709–718 (2025).
4. Tarricone, R. Cost-of-illness analysis: What room in health economics? *Health Policy* **77**, 51–63 (2006).
5. Hex, N. *et al.* Estimation of the direct health and indirect societal costs of diabetes in the UK using a cost of illness model. *Diabet. Med.* **41**, e15326 (2024).
6. Kinge, J. M. *et al.* Disease-specific health spending by age, sex, and type of care in Norway: a national health registry study. *BMC Med.* **21**, 201 (2023).
7. Shakya, S. *et al.* Global comparison of the economic costs of coronary heart disease: a systematic review and meta-analysis. *BMJ Open* **15**, e084917 (2025).
8. Wieser, S. *et al.* How much does the treatment of each major disease cost? A decomposition of Swiss National Health Accounts. *Eur. J. Health Econ.* **19**, 1149–1161 (2018).
9. Lewis, B. United Kingdom Standards for Official Statistics on Climate-Health Interactions (SOSCHI): update. (2025).
10. Dahil, A. *et al.* Examining the health impacts of climate change through electronic health records: A rapid review. *J. Clim. Change Health* **28**, 100667 (2026).
11. Liao, S. *et al.* Temperature-related hospitalization burden under climate change. *Nature* **644**, 960–968 (2025).
12. Wondmagegn, B. Y., Xiang, J., Williams, S., Pisaniello, D. & Bi, P. What do we know about the healthcare costs of extreme heat exposure? A comprehensive literature review. *Sci. Total Environ.* **657**, 608–618 (2019).
13. Onukwugha, E. *et al.* Cost-of-Illness Studies: An Updated Review of Current Methods. *PharmacoEconomics* **34**, 43–58 (2016).
14. Rosen, A. B. & Cutler, D. M. Challenges in Building Disease-Based National Health Accounts. *Med. Care* **47**, S7–S13 (2009).
15. Limaye, V. S., Max, W., Constible, J. & Knowlton, K. Estimating the Health-Related Costs of 10 Climate-Sensitive U.S. Events During 2012. *GeoHealth* **3**, 245–265 (2019).
16. Bourke, E., Grimshaw, S. & Blakely, T. What is the value of disease expenditure studies? An argument for an international database of spending estimates. *BMC Glob. Public Health* **1**, 23 (2023).
17. Simon, N. B. *et al.* Policy Brief—What's in a Name? A Search for Alternatives to "VSL". *Rev. Environ. Econ. Policy* **13**, 155–161 (2019).
18. Jönsson, L., Tate, A., Frisell, O. & Wimo, A. The Costs of Dementia in Europe: An Updated Review and Meta-analysis. *PharmacoEconomics* **41**, 59–75 (2023).
19. Ebi, K. L., Ogden, N. H., Semenza, J. C. & Woodward, A. Detecting and Attributing Health

Burdens to Climate Change. *Environ. Health Perspect.* **125**, 085004 (2017).
20. Rachas, A. *et al.* The Economic Burden of Disease in France From the National Health Insurance Perspective: The Healthcare Expenditures and Conditions Mapping Used to Prepare the French Social Security Funding Act and the Public Health Act. *Med. Care* **60**, 655–664 (2022).
21. Stingone, J. A. *et al.* Unlocking the Power of Data Harmonization in Environmental Health Sciences: A Comprehensive Exploration of Significance, Use Cases, and Recommendations for Standardization Efforts. *Environ. Health Perspect.* **134**, 240–251 (2026).
22. Sharifi, A., Pathak, M., Joshi, C. & He, B.-J. A systematic review of the health co-benefits of urban climate change adaptation. *Sustain. Cities Soc.* **74**, 103190 (2021).
23. Cai, X. *et al.* Utilizing Large language models to select literature for meta-analysis shows workload reduction while maintaining a similar recall level as manual curation. *BMC Med. Res. Methodol.* **25**, 116 (2025).
24. Polak, M. P. & Morgan, D. Extracting accurate materials data from research papers with conversational language models and prompt engineering. *Nat. Commun.* **15**, 1569 (2024).
25. Zheng, Z., Zhang, O., Borgs, C., Chayes, J. T. & Yaghi, O. M. ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis. *J. Am. Chem. Soc.* **145**, 18048–18062 (2023).
26. Mitchell, A. J. *et al.* Estimating global direct health-care spending on neurological and mental health between 2000 and 2019: a modelling study. *Lancet Public Health* **10**, e401–e411 (2025).
27. Earnshaw, J. & Lewis, G. NICE Guide to the Methods of Technology Appraisal: Pharmaceutical Industry Perspective. *PharmacoEconomics* **26**, 725–727 (2008).
28. Drew, J., Christie, S. D., Rainham, D. & Rizan, C. HealthcareLCA: an open-access living database of health-care environmental impact assessments. *Lancet Planet. Health* **6**, e1000–e1012 (2022).
29. Turner, H. C. *et al.* An Introduction to Costing and the Types of Costs Used within Health Economic Studies. *PharmacoEconomics - Open* **9**, 849–868 (2025).
30. Jo, C. Cost-of-illness studies: concepts, scopes, and methods. *Clin. Mol. Hepatol.* **20**, 327 (2014).
31. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2021 (GBD 2021) Causes of Death and Nonfatal Causes Mapped to ICD Codes. https://doi.org/10.6069/vd7t-zt25 (2026).
32. Charani, E. *et al.* Funders: The missing link in equitable global health research? *PLOS Glob. Public Health* **2**, e0000583 (2022).
33. Mbau, R., Vassall, A., Gilson, L. & Barasa, E. Factors Influencing the Institutionalization of Health Technology Assessment: A Scoping Literature Review. *Health Syst. Reform* **9**, 2360315 (2023).
34. Nyatuka, D., Jabin, M. S. R. & Dionne-Morris, L. Contextual Barriers to Health Information Systems Optimization in Underserved Communities in Kenya: Qualitative Study Informed by Frugal Innovation and Information and Communication Technologies for Development. *JMIR Form. Res.* **10**, e78950–e78950 (2026).
35. Wubulihasimu, P., Gheorghe, M., Slobbe, L., Polder, J. & Van Baal, P. Trends in Dutch hospital spending by age and disease 1994–2010. *Health Policy* **119**, 316–323 (2015).
36. Mastrandrea, M. D. *et al.* Guidance Note for Lead Authors of the IPCC Fifth Assessment

Report on Consistent Treatment of Uncertainties.
37. Kause, A. *et al.* Confidence levels and likelihood terms in IPCC reports: a survey of experts from different scientific disciplines. *Clim. Change* **173**, 2 (2022).
38. Cotton, C. S., LaBarge, M. C. & Nordstrom, A. Canadian public health experiences during COVID-19: a new framework for assessing evidence. *Front. Public Health* **13**, 1620514 (2025).
39. Eddy, D. M. *et al.* Model Transparency and Validation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Value Health* **15**, 843–850 (2012).
40. Drummond, M. E., Sculpher, M. J., Torrance, G. W., O'Brien, B. J. & Stoddart, G. L. *Methods for the Economic Evaluation of Health Care Programmes*. (Oxford Academic, Oxford, 2023). doi:10.1093/oso/9780198529446.001.0001.
41. Ramamoorthy, V. *et al.* Healthcare expenditure trends among adult stroke patients in the United States, 2011-2020. *J. Stroke Cerebrovasc. Dis.* **32**, 107333 (2023).
42. Li, X. Impact of integrated care models on inpatient costs and health services efficiency: evidence from a difference-in-differences analysis in China.
43. Kaur, M. *et al.* Workforce Innovation Reduces Medicaid Costs in Chronic Care.
44. Lin, Y. *et al.* Estimating the economic burden of lower respiratory infections in China from 2020 to 2040: a health-augmented macroeconomic modelling study. *Lancet Reg. Health - West. Pac.* **69**, 101855 (2026).
45. Aromataris, E. *et al.* Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. *Int. J. Evid. Based Healthc.* **13**, 132–140 (2015).
46. Kneale, D., Rojas-García, A., Raine, R. & Thomas, J. The use of evidence in English local public health decision-making: a systematic scoping review. *Implement. Sci.* **12**, 53 (2017).
47. Romanello, M. *et al.* The 2025 report of the Lancet Countdown on health and climate change: climate change action offers a lifeline. *The Lancet* **406**, 2804–2857 (2025).
48. Liu, N. F. *et al.* Lost in the Middle: How Language Models Use Long Contexts. Preprint at https://doi.org/10.48550/arXiv.2307.03172 (2023).
49. Guo, T. *et al.* Leveraging inter-chunk interactions for enhanced retrieval in large language model-based question answering. *Neurocomputing* **650**, 130931 (2025).
50. Kataoka, Y. *et al.* Automating the data extraction process for systematic reviews using GPT-4o and o3. *Res. Synth. Methods* **17**, 42–62 (2026).
51. Wang, G. *et al.* Human–large language model collaboration in clinical medicine: a systematic review and meta-analysis. *Npj Digit. Med.* **9**, 195 (2026).
52. McDuffie, E. *et al.* GBD-MAPS-Global: Gridded Fractional Source Contribution Results. https://doi.org/10.5281/zenodo.4739100 (2021).
53. Armstrong, J. *et al.* Dynamic linkage of COVID-19 test results between Public Health England's Second Generation Surveillance System and UK Biobank. *Microb. Genomics* **6**, (2020).
54. Alcayna, T. *et al.* Climate-sensitive disease outbreaks in the aftermath of extreme climatic events: A scoping review. *One Earth* **5**, 336–350 (2022).
55. Friel, S. *et al.* Climate Change, Noncommunicable Diseases, and Development: The Relationships and Common Policy Opportunities. *Annu. Rev. Public Health* **32**, 133–147 (2011).
56. Brodszky, V. *et al.* Cost-of-illness studies in nine Central and Eastern European countries. *Eur.*

J. Health Econ. **20**, 155–172 (2019).
57. Sontheimer, N., Konnopka, A. & König, H.-H. The Excess Costs of Dementia: A Systematic Review and Meta-Analysis. *J. Alzheimer's Dis.* **83**, 333–354 (2021).
58. Zaghir, J. *et al.* Prompt Engineering Paradigms for Medical Applications: Scoping Review. *J. Med. Internet Res.* **26**, e60501 (2024).
59. Danhauser, K. *et al.* Using large language models to extract information from pediatric clinical reports. *PLOS Digit. Health* **4**, e0000919 (2025).
60. Wu, S. *et al.* Automated literature research and review-generation method based on large language models. *Natl. Sci. Rev.* **12**, nwaf169 (2025).
61. Zhang, J., Zhou, J., Zhou, L. & Ba, Z. Extracting Multifaceted Characteristics of Patients With Chronic Disease Comorbidity: Framework Development Using Large Language Models. *JMIR Med. Inform.* **13**, e70096 (2025).
62. Gupta, S., Mahmood, A., Shetty, P., Adeboye, A. & Ramprasad, R. Data extraction from polymer literature using large language models. *Commun. Mater.* **5**, 269 (2024).
63. Jaskulski, P., Latos, T., Ryńca, M. & Zapała, A. Reliability of large language models as a tool for knowledge extraction from biographical dictionaries: the case of the Polish Biographical Dictionary. *Digit. Scholarsh. Humanit.* **40**, 538–548 (2025).
64. Yamagishi, Y., Nakamura, Y., Hanaoka, S. & Abe, O. Large Language Model Approach for Zero-Shot Information Extraction and Clustering of Japanese Radiology Reports: Algorithm Development and Validation. *JMIR Cancer* **11**, e57275–e57275 (2025).
65. Ribeiro De Faria, J., Xie, H. & Steffek, F. Information extraction from employment tribunal judgments using a large language model. *Artif. Intell. Law* https://doi.org/10.1007/s10506-025-09443-z (2025) doi:10.1007/s10506-025-09443-z.
66. Gao, C., Pradeepkumar, J., Das, T., Thati, S. & Sun, J. A large-scale database for clinical trial outcomes and features. *Nat. Health* **1**, 428–436 (2026).
67. OECD, World Health Organization, & Eurostat. *A System of Health Accounts: 2011 Edition*. (OECD, 2011). doi:10.1787/9789264116016-en.
68. Dieleman, J. L. *et al.* US Health Care Spending by Payer and Health Condition, 1996-2016. *JAMA* **323**, 863 (2020).

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence this work.

Additional information

Supplementary information

The online version of supplementary information is available.

Supplementary Files

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

- Supplementaryinformation.pdf