Supplementary Information for

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

1. List of Climate-Sensitive Diseases (CSDs)

The selection of CSDs was guided by a systematic review of existing literature and authoritative assessment reports. We primarily referenced the health-related chapters of the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) (Cissé et al., 2022), followed by a synthesis of key review articles (Agache et al., 2025; Bell et al., 2018; Ding & Yang, 2024; Haines & Ebi, 2019; Huang & Liu, 2022; Hunter et al., 2024; Mora et al., 2022; Yenew et al., 2025). To ensure standardized and interoperable classification, diseases were coded according to the International Classification of Diseases, Tenth Revision (ICD-10, version 2019). The inclusion process prioritized diseases that are frequently documented in the literature. Most entries are specified at the three-character category level (e.g., A20 for plague), while a subset of typical conditions with distinct subcategories are detailed to the fourth-character level (e.g., A69.2 for Lyme disease).

1.1 Communicable Diseases

A00-A09 Intestinal infectious diseases (especially: A00 Cholera, A01 Typhoid and paratyphoid fevers, A02 Other salmonella infections, A03 Shigellosis, A04 Other bacterial intestinal infections, and A08.1 Acute gastroenteropathy due to Norovirus);

A15-A19 Tuberculosis (especially: A15 Respiratory tuberculosis, bacteriologically and histologically confirmed, and A16 Respiratory tuberculosis, not confirmed bacteriologically or histologically);

A20-A28 Certain zoonotic bacterial diseases (especially: A20 Plague, A22 Anthrax, and A27 Leptospirosis);

A68 Relapsing fevers, and A69.2 Lyme disease;

A75-A79 Rickettsioses (especially: A75 Typhus fever, A75.3 Typhus fever due to

Rickettsia tsutsugamushi, A77 Spotted fever [tick-borne rickettsioses], and A78 Q fever);

A83 Mosquito-borne viral encephalitis (especially: A83.0 Japanese encephalitis);

A84 Tick-borne viral encephalitis;

A92-A99 Arthropod-borne viral fevers and viral haemorrhagic fevers (especially: A92.0 Chikungunya virus disease, A92.3 West Nile virus infection, A92.5 Zika virus disease, A95 Yellow fever, A97 Dengue, A98.5 Haemorrhagic fever with renal syndrome);

B15 Acute hepatitis A;

B20-B24 Human immunodeficiency virus [HIV] disease;

B30 Viral conjunctivitis, and B33.4 Hantavirus (cardio-)pulmonary syndrome [HPS] [HCPS];

B50 Plasmodium falciparum malaria, B51 Plasmodium vivax malaria, B52 Plasmodium malariae malaria, B53 Other parasitologically confirmed malaria, B54 Unspecified malaria, B55 Leishmaniasis, and B60.0 Babesiosis;

B65 Schistosomiasis [bilharziasis];

J09-J18 Influenza and pneumonia.

1.2 Non-communicable Diseases

D50 Iron deficiency anaemia;

E10-E14 Diabetes mellitus, E40-E46 Malnutrition, and E65-E68 Obesity and other hyperalimentation;

F00 Dementia in Alzheimer disease, F01 Vascular dementia, F02 Dementia in other diseases classified elsewhere, and F03 Unspecified dementia, G30 Alzheimer disease;

F30-F39 Mood [affective] disorders, F40 Phobic anxiety disorders, and F41 Other anxiety disorders;

I05-I09 Chronic rheumatic heart diseases, I10-I15 Hypertensive diseases, I20-I25 Ischaemic heart diseases, I26-I28 Pulmonary heart disease and diseases of pulmonary circulation (especially: I26 Pulmonary embolism), and I30-I52 Other forms of heart disease (especially: I48 Atrial fibrillation and flutter, and I50 Heart failure);

I60-I69 Cerebrovascular diseases (especially: I60 Subarachnoid haemorrhage, I61

Intracerebral haemorrhage, I62 Other nontraumatic intracranial haemorrhage, I63 Cerebral infarction, and I64 Stroke, not specified as haemorrhage or infarction);

J00-J06 Acute upper respiratory infections, and J20-J22 Other acute lower respiratory infections (especially: J20 Acute bronchitis);

J30 Vasomotor and allergic rhinitis, J31 Chronic rhinitis, nasopharyngitis and pharyngitis, J32 Chronic sinusitis, and J39.3 Upper respiratory tract hypersensitivity reaction, site unspecified;

J40-J47 Chronic lower respiratory diseases (especially: J40 Bronchitis, not specified as acute or chronic, J41 Simple and mucopurulent chronic bronchitis, J42 Unspecified chronic bronchitis, J43 Emphysema, J44 Other chronic obstructive pulmonary disease, J45 Asthma, and J46 Status asthmaticus);

N00-N08 Glomerular diseases, N10-N16 Renal tubulo-interstitial diseases, N30 Cystitis, N34 Urethritis and urethral syndrome, and N39.0 Urinary tract infection, site not specified;

O00-O99 Pregnancy, childbirth and the puerperium, and P00-P96 Certain conditions originating in the perinatal period (all specifically related to extreme temperatures);

L55 Sunburn;

M05-M14 Inflammatory polyarthropathies (especially: M05 Seropositive rheumatoid arthritis, and M06 Other rheumatoid arthritis);

T33-T35 Frostbite, T67 Effects of heat and light (especially: T67.0 Heatstroke and sunstroke), T68 Hypothermia, and T69.1 Chilblains;

X30-X39 Exposure to forces of nature.

The list presented here is intentionally broad but inevitably incomplete. It may include diseases with weak climate sensitivity or omit others that are highly climate-sensitive. We recognize this as a dynamic and evolving effort. We therefore encourage collaborative input from the scientific community to refine, validate, and expand it, ultimately working toward a robust, systematic, and consensus-based list of climate-sensitive diseases.

2. Composite Keywords for Searching Literature

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”. All composite keywords are listed in Table S1.

Table S1: Composite keywords for searching literature

Number	Disease type	Disease query	Connecting	Cost query
1	Enteric infectious diseases	("cholera" OR "Vibrio cholerae" OR "typhoid fever" OR "paratyphoid fever" OR "Salmonella infection" OR "shigellosis" OR "Shigella" OR "Escherichia coli infection" OR "E. coli enteritis" OR "norovirus" OR "viral gastroenteritis" OR "enteric bacterial infection")	AND	("health expenditure" OR "health spending" OR "cost of illness" OR "economic burden" OR "treatment cost" OR "medical cost" OR "disease cost" OR "price of disease" OR "direct cost" OR "indirect cost" OR "disease burden" OR "health spending by disease" OR "spending by health condition")
2	Respiratory tuberculosis	("tuberculosis" OR "pulmonary tuberculosis" OR "Mycobacterium tuberculosis" OR "TB")	AND	
3	Zootonic bacterial diseases	("plague" OR "Yersinia pestis" OR "anthrax" OR "Bacillus anthracis" OR "leptospirosis" OR "Leptospira interrogans")	AND	
4	Relapsing fever/Lyme disease	("relapsing fever" OR "Borrelia infection" OR "Lyme disease" OR "Borrelia burgdorferi")	AND	
5	Rickettsial diseases (typhus, etc.)	("typhus" OR "scrub typhus" OR "rickettsiosis" OR "spotted fever" OR "Q fever" OR "Coxiella burnetii")	AND	
6	Japanese encephalitis	("Japanese encephalitis" OR "JE virus" OR "flavivirus encephalitis")	AND	
7	Tick-borne viral encephalitis	("tick-borne encephalitis" OR "TBE virus")	AND	
8	Arthropod-borne viral fever/hemorrhagic fever	("dengue fever" OR "dengue virus" OR "DENV" OR "chikungunya" OR "Zika virus" OR "West Nile virus" OR "yellow fever" OR "HFRS" OR "hemorrhagic fever with renal syndrome" OR "Hantavirus")	AND	
9	Hepatitis A	("hepatitis A" OR "HAV infection")	AND	
10	Viral conjunctivitis and cardiopulmonary syndrome	("viral conjunctivitis" OR "adenoviral conjunctivitis" OR "Hantavirus pulmonary syndrome" OR "HPS")	AND	
11	Malaria/kala-azar/babesiosis	("malaria" OR "Plasmodium falciparum" OR "Plasmodium vivax" OR "leishmaniasis" OR "visceral leishmaniasis" OR "kala-azar" OR "babesiosis" OR "Babesia microti")	AND	
12	Schistosomiasis	("schistosomiasis" OR "Schistosoma haematobium" OR "Schistosoma	AND	

		mansoni")	
13	Influenza and pneumonia	("influenza" OR "seasonal flu" OR "H1N1" OR "pneumonia" OR "viral pneumonia" OR "bacterial pneumonia")	AND
14	HIV/AIDS	("HIV" OR "human immunodeficiency virus" OR "AIDS" OR "acquired immunodeficiency syndrome")	AND
15	Hematological system diseases	("anemia" OR "iron deficiency" OR "sickle cell")	AND
16	Diabetes/malnutrition/obesity	("diabetes mellitus" OR "type 1 diabetes" OR "type 2 diabetes" OR "malnutrition" OR "undernutrition" OR "protein energy malnutrition" OR "obesity" OR "overweight")	AND
17	Mental and behavioral disorders (dementia, depression, anxiety)	("dementia" OR "Alzheimer's disease" OR "major depressive disorder" OR "depression" OR "bipolar disorder" OR "mania" OR "anxiety disorder" OR "generalized anxiety")	AND
18	Nervous system diseases	("Alzheimer’s disease" OR "dementia of Alzheimer type" OR "neurodegenerative disease")	AND
19	Circulatory system diseases	("hypertension" OR "high blood pressure" OR "ischemic heart disease" OR "myocardial infarction" OR "heart failure" OR "arrhythmia" OR "pulmonary embolism" OR "stroke" OR "cerebrovascular disease" OR "hemorrhagic stroke" OR "ischemic stroke")	AND
20	Respiratory system diseases	("upper respiratory tract infection" OR "URTI" OR "acute bronchitis" OR "lower respiratory infection" OR "chronic bronchitis" OR "emphysema" OR "chronic obstructive pulmonary disease" OR "COPD" OR "asthma" OR "allergic rhinitis" OR "sinusitis")	AND
21	Digestive system diseases	("diarrhea" OR "gastroenteritis" OR "foodborne illness" OR "intestinal infection")	AND
22	Urogenital system	("urinary tract infection" OR "UTI" OR "kidney infection" OR "nephritis")	AND

23	Skin diseases: sunburn, etc.	("sunburn" OR "UV-related skin damage" OR "solar erythema")	AND
24	Musculoskeletal system diseases	("rheumatoid arthritis" OR "inflammatory polyarthritis" OR "chronic joint inflammation")	AND
25	Climate-related injuries and external factors	("heat stroke" OR "heat stress" OR "hyperthermia" OR "hypothermia" OR "frostbite" OR "chilblains" OR "weather-related injury" OR "natural disaster injury" OR "extreme temperature injury")	AND
26	Pregnancy complications	("pregnancy complications" OR "pregnancy outcome" OR "maternal morbidity" OR "heat exposure pregnancy")	AND
27	Perinatal diseases	("perinatal conditions" OR "neonatal complications" OR "low birth weight" OR "preterm birth")	AND
28	Symptoms/signs/unspecified diseases	("fatigue" OR "headache" OR "heat-related symptoms" OR "malaise")	AND

3. Harmonized Information Fields of COI Data

The dataset organizes extracted data into harmonized fields covering spatiotemporal context, cost metrics, population and classification characteristic, as shown in Table S2. Regional and temporal information is captured through data_country, data_region, data_year, and base_year. Monetary values and statistical denominators are recorded alongside cohesive equivalents (e.g., cost_usd_2024_unified and cost_caliber_unified). Demographic profiles encompass age, gender, income, and sample size, while insurance schemes and medical treatment settings are cataloged in cost_insurance and cost_treatment. Notably, columns appended with “_infer” contain LLM-assisted categorized values derived from the corresponding raw, descriptive text fields, preserving original source information while ensuring cross-study analytical consistency.

Data processing and harmonization prioritized cross-study comparability through systematic cleaning and inference. Disease names and ICD codes were mapped to the GBD 2021 Nonfatal ICD Code Map, targeting Level 3 granularity (~200 types) or defaulting to Level 2 when the original were overly coarse (Global Burden of Disease Collaborative Network, 2024). Statistical calibers were disentangled into temporal and population dimensions, then mathematically normalized to "per patient per year" or per "patient per incident" (e.g., dividing aggregate annual costs by sample size). Insurance schemes and treatment types were classified into predefined categories based on the dominant cost driver in each study, following mainstream health accounting frameworks (Dieleman et al., 2020). Processing and harmonization methods for other information fields are detailed in Table S2.

Table S2: Information fields and processing methods for the dataset of COI for CSDs

Number	Information cluster	Information field name	Definition	Processing methods
1		data_country	Country of the cost	Adopting the national standard name of the World Bank
2		data_region	Sub-national region of the cost	
3		data_year	Year of the cost	
4	Spatial-temporal & disease identification	disease_name_infer	Harmonized disease name	Firstly, query the ICD-10 code corresponding to disease_name (ICD-10 already reported does not need to be queried), and then correspond it to the third level disease name in GBD2021 Non fatal ICD Code Map. If it is too broad, use the second level disease name
5		disease_icd_infer	Harmonized ICD-10 codes	Prioritize referring to the ICD-10 code in the original report. If not reported, use the ICD-10 code corresponding to the code map's third level name
6	Harmonized cost outcomes	cost_caliber_unified	Harmonized statistical caliber	Unified majorly into "per patient per year", "per patient per incident" (include per visit & per treatment course), and "total one year"

7		cost_usd_2024_unified	Cost value of per patient per year and per incident converted to 2024 USD	Expenses other than per patient per year and per patient per incident are not recorded in this section
8	Key classification information	population_age	Age group of population related to the cost	
9		population_gender	Gender of population related to the cost	Gender limited to male and female only
10		cost_direct_infer	Harmonized direct/indirect classification	
11		cost_insurance_infer	Harmonized insurance reimbursement type	Unified into three categories: Social insurance, Commercial insurance, and Out of pocket
12		cost_treatment_infer	Harmonized medical treatment type	Unified into nine categories: Outpatient care, Inpatient care, Emergency department care, Nursing care, Dental care, Pharmaceuticals, Diagnostics, Devices & comsumables, General administration
13	Original extracted and process-related information	disease_name	Originally reported disease name	Try to use MeSH terminology as much as possible
14		disease_icd	Originally reported ICD-10 code	
15		base_year	Baseline year of the cost for currency	Prioritize referencing the year of the report, and if not reported, record the year of publication of the literature
16		currency	Originally reported currency unit	
17		currency_infer	Harmonized currency unit	Unified currency code for ISO 4217
18		cost_caliber	Originally reported statistical caliber	
19		cost_caliber_infer	Statistical caliber with ununified time scale	Unified format as "object (time) ", for example, per patient (2 years)

20		cost_sample	Sample size of population related to the cost	The unified format is "quantity (unit)", for example, 2000 (patients)
21		cost_value	Reported and verified cost value	
22		cost_usd_2024	Cost value converted to 2024 USD	Convert the base year to 2024 using the World Bank's Consumer Price Index (CPI) for each country; Then, using the World Bank's 2024 official market exchange rate for converting countries to US dollars, it is converted to 2024 unchanged US dollars
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	Unified into four categories: Productivity loss, Household or caregiver burden, Premature deaths, and Cross-sector burden
30	Literature information	literature_doi	DOI of the literature	
31		literature_title	Title of the literature	
32		literature_year	Publication year of the literature	

4. Eligibility Criteria and Prompts in Literature Screening

4.1 Step 1: Identify and Judge CSDs

Inclusion criterion: The study explicitly mentions at least one of the 80+ CSDs and focuses its research, analysis, and discussion on that disease;

Exclusion criterion: The study does not mention any CSDs, or only mentions them briefly without sufficient research, analysis, or discussion;

Prompt:

```
Please analyze the following abstract to accurately identify whether the literature
mentions any climate-sensitive diseases, and conducts some research, analysis, or
discussion on them. In order to determine whether the literature should be included,
excluded, or remains unclear.

The climate sensitive diseases that need to be identified are divided into two
categories, namely climate-sensitive communicable diseases and climate-sensitive non-
communicable diseases.
A total of 42 types of climate-sensitive communicable diseases need to be identified,
including: cholera; typhoid; paratyphoid; salmonella infection; dysentery;
shigellosis; bacterial intestinal infection; e. coli infection; norwalk virus; other
intestinal infection; respiratory tuberculosis; pulmonary tuberculosis; zoonotic
bacterial disease; plague; anthrax; leptospirosis; relapsing fever; lyme disease;
rickettsiosis; typhus; tsutsugamushi; spotted fever; q fever; mosquito-borne viral
encephalitis; japanese encephalitis; tick-borne viral encephalitis; arthropod-borne
viral fever and viral hemorrhagic fever; dengue; chikungunya; west nile virus; zika
virus; yellow fever; hemorrhagic fever with renal syndrome (HFRS); hantavirus (HTV);
hepatitis A; viral conjunctivitis; hantavirus pulmonary syndrome (HPS); malaria;
leishmaniasis; babesiosis; schistosomiasis; seasonal influenza and pneumonia.
A total of 47 types of climate-sensitive non-communicable diseases need to be
identified, including: diabetes; diabetes mellitus; malnutrition; nutritional
deficiency; obesity; dementia; mood disorder; mania; depression; anxiety; alzheimer's
disease; cardiovascular disease; hypertension; ischemic heart disease; myocardial
infarction; pulmonary embolism; arrhythmia; atrial fibrillation and flutter; heart
failure; other cardiovascular disease; cerebrovascular disease; stroke; hemorrhagic
stroke; cerebral infarction; other cerebrovascular disease; upper respiratory
infection; lower respiratory infection; bronchitis; rhinitis; nasopharyngitis;
pharyngitis; sinusitis; allergic; chronic lower respiratory disease; emphysema;
chronic obstructive pulmonary disease (COPD); asthma; other respiratory disease;
sunburn; rheumatoid arthritis; frostbite; effects of heat and light; heatstroke;
sunstroke; hypothermia; chilblain; extreme weather induced injury.

Inclusion criteria (assess each):
```

1. The literature mentions any of the 42 types of climate-sensitive communicable diseases listed above. Diseases not listed above should not be included.
2. Under the premise of satisfying “inclusion criterion 1”, the literature has discussed, analyzed, or researched climate-sensitive communicable diseases, including characteristics, diagnosis, pathology, epidemic patterns, treatment plans, costs, economic impacts, health policies, or health equity of these diseases.
3. The literature mentions any of the 47 types of climate-sensitive non-communicable diseases listed above. Diseases not listed above should not be included.
4. Under the premise of satisfying “inclusion criterion 3”, the literature has discussed, analyzed, or researched climate-sensitive non-communicable diseases, including characteristics, diagnosis, pathology, epidemic patterns, treatment plans, costs, economic impacts, health policies, or health equity of these diseases.

Exclusion criteria (assess each):
1. The literature does not mention any of the climate-sensitive diseases listed above.
2. The literature mentions some of climate-sensitive diseases listed above, but there has been no discussion, analysis, or research on these diseases, including their characteristics, diagnosis, pathology, epidemic patterns, treatment plans, costs, economic impacts, health policies, or health equity, etc.
3. The abstract and title of the literature are too brief or vague, lacking sufficient disease-related words or sentences to make judgements.

Output requirements:
1. For each inclusion and exclusion criterion, output one of “satisfied”, “not satisfied”, or “unclear”.
2. Finally, provide the final decision on whether to include the literature. If the literature meets someone of the “inclusion criterion 2” or “inclusion criterion 4” and does not meet any exclusion criteria, output “included”; If the literature meets any of the exclusion criteria, output “excluded”; If the literature only meets “inclusion criterion 1” or “inclusion criterion 3” and does not meet “inclusion criterion 2” or “inclusion criterion 4”, or if the information is limited and unclear, output “unclear”.

Format output of these criteria as JSON. Use this exact structure, even if output is “unclear”:
{Inclusion1 ..., Inclusion2 ..., Inclusion3 ..., Inclusion4 ..., Exclusion1 ..., Exclusion2 ..., Exclusion3 ..., Final_decision ...}
For example:
```json
{
“Inclusion1”: “satisfied”,
“Inclusion2”: “not satisfied”,
“Inclusion3”: “satisfied”,
“Inclusion4”: “satisfied”,
```

````
"Exclusion1": "not satisfied",
"Exclusion2": "not satisfied",
"Exclusion3": "not satisfied",
"Final_decision": "included"
}
```
ABSTRACT:
````

4.2 Step 2: Include Cost or Economic Studies

Inclusion criterion: The study primarily quantifies, estimates, or compares disease costs/expenditures and presents data on costs, expenditures, or economic burden;

Exclusion criterion: 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;

Prompt:

```
Analyze the following abstract to determine if the literature is likely to be included, excluded, or unclear based on economic criteria.

A typical cost/expenditure/economic burden study normally include:
1. Definition of cost components and perspective (e.g., direct, indirect, hospital, outpatient, salary loss);
2. Data collection of healthcare utilization, expenditures, or economic costs;
3. Cost/expenditure calculation, estimation, or modeling (e.g., statistical survey, accounting, top-down decomposition, bottom-up aggregation, regression prediction);
4. Statistical or economic analysis of cost/expenditure data;
5. Discussion or interpretation of cost results, economic burdens, or impacts;
6. Reporting of findings with implications for therapy/technology assessment, medical budgeting, healthcare policy, or management.

Inclusion criteria (assess each):
1. The literature presents an original, statistical, or modeling analysis of healthcare or disease-related costs/expenditures, economic burden, or financial impacts.
2. The literature focuses on quantifying, estimating, comparing, or modeling costs, expenditures, or economic consequences related to disease.

Exclusion criteria (assess each):
1. The literature does not include any cost or economic analysis or discussion.
2. The literature is a purely clinical, epidemiological, or biological study without cost or economic outcomes.
3. The literature is a review, editorial, commentary, or theoretical discussion without original cost data or quantitative results.
```
```

```
4. The literature is a book chapter or conference abstract compilation.
5. Abstract and title are too brief or vague, lacking sufficient information on cost or economic analysis to make a decision.

For each step, decide one of “satisfied”, “not satisfied”, or “unclear”;

Finally, provide a final decision:
1. If the literature meets both inclusion criteria and none of the exclusion criteria, output “included”;
2. If the literature meets one or more exclusion criteria, output “excluded”;
3. If the decision cannot be made due to limited information, output “unclear”.

Format output of all criteria as JSON. Use this exact structure, even if answer is “unclear”:
{Inclusion1 ..., Inclusion2 ..., ..., Exclusion1 ..., ..., Exclusion5 ..., Final_decision ...}
For example:
```json
{
“Inclusion1”: “satisfied”,
“Inclusion2”: “satisfied”,
“Exclusion1”: “not satisfied”,
“Exclusion2”: “not satisfied”,
“Exclusion3”: “not satisfied”,
“Exclusion4”: “not satisfied”,
“Exclusion5”: “not satisfied”,
“Final_decision”: “included”
}
```
ABSTRACT:
```

4.3 Step 3: Include Actual Incurred Costs or Expenditures

Inclusion criterion: 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) (Jo, 2014);

Exclusion criterion: The reported costs are modeled projections, expected costs, cost gaps, or other non-actual numbers, such as cost changes or benefit-cost ratios (Jönsson et al., 2023; Sontheimer et al., 2021);

Prompt:

```
Analyze the following abstract to determine if the literature is likely to be included, excluded, or unclear based on judging whether the cost/expenditure/economic impacts
```

are actually existed values.

Inclusion criteria (assess each):
1. The main costs, expenditures, or economic burdens considered in the literature are direct costs caused by diseases, such as hospital, outpatient, medication, diagnosis, or transportation costs.
2. The main costs, expenditures, or economic burdens considered in the literature are indirect costs caused by diseases, such as salary loss, labor loss, low life quality induced costs, social welfare loss.

Exclusion criteria (assess each):
1. The main costs, expenditures, or economic burdens considered in the literature are simulated, predicted, or estimated by models such as Markov models, regression analysis, scenario analysis, etc. That is, there is no actual behavior of disease diagnosis & treatment or spending/paying expenses.
2. The main costs, expenditures, or economic burdens considered in the literature are required costs, cost gap, or projected costs, meaning it is not actually incurred or expended costs/expenditures.
3. The main costs, expenditures, or economic burdens considered in the literature are change amounts (cost increases/reductions caused by adopting a certain drug, therapy, or surgery) rather than total amount.
4. The main results considered in the literature are cost-benefit ratios, cost-effectiveness results, or budget adjustments rather than existed cost/expenditure/economic burden values.

For each step, decide one of “satisfied”, “not satisfied”, or “unclear”;

Finally, provide a final decision:
1. If the literature meets one of inclusion criteria and none of the exclusion criteria, output “included”;
2. If the literature meets any one of exclusion criteria, output “excluded”;
3. If the decision cannot be made due to limited information, output “unclear”.

Format output of all criteria as JSON. Use this exact structure, even if answer is “unclear”:
{Inclusion1 ..., Inclusion2 ..., Exclusion1 ..., ..., Exclusion4 ..., Final_decision ...}
For example:
```json
{
“Inclusion1”: “satisfied”,
“Inclusion2”: “not satisfied”,
“Exclusion1”: “not satisfied”,
“Exclusion2”: “not satisfied”,
```

```
“Exclusion3”: “not satisfied”,
“Exclusion4”: “not satisfied”,
“Final_decision”: “included”
}
```
ABSTRACT:
```

# 5. Pipeline and Prompts in Cost Data Extraction

## 5.1 Extraction of Cost Tables

We apply inclusion and exclusion criteria to each table in the literature (Zheng et al., 2023). 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.

Prompt:

```
Please analyze the full text of the literature to accurately identify and record any tables or paragraphs of costs, expenses, or economic burdens associated with diseases in the literature.

Part I: identify source tables for disease costs
Analyze the full text of the literature, starting from Table 1 through the last table.
Evaluate each table to determine whether it mainly exhibits diseases costs, expenditures, or economic information.
1. Inclusion: a table should be “included” only if it satisfies both inclusion criteria:
(1) It contains keywords such as cost, expense, expenditure, spending, fee, budget, loss, financial, economic, or monetary.
(2) It includes specific currency information (e.g., RMB, Dollars, €, etc.).
2. Exclusion: if either inclusion criterion is not met, or any of the following exclusion criteria satisfies, record “excluded” for this table:
(1) It mainly exhibits clinical, epidemiological, or statistical results, rather than cost, expenditure, or economic results.
(2) It mainly exhibits information of unit prices, cost changes (due to some therapies or experiments), or cost-benefit ratios, rather than existed costs, expenditures, economic burden values.
3. Extract tables: for each table from Table 1 to the last table, if it is “included”,
```
```

record the complete table information, including table title, headers, all table content (rows), and any table tips or footnotes. Do not omit or fabricate.

Part II: identify source paragraphs for disease costs

If all tables received “excluded” in Part I, proceed to this part; otherwise, skip it.

Analyze each paragraph in the literature section of “Results”, especially in sub-sections about economic results, determine whether it mainly exhibits diseases costs, expenditures, or economic information.

1. Inclusion: a paragraph should be marked “included” only if it satisfies the criterion:

- It contains keywords such as cost, expense, expenditure, spending, fee, budget, loss, financial, and economic, and includes specific currency information (e.g., RMB, Dollars, €, etc.).

2. Exclusion: If the inclusion criterion is not met, or any of the exclusion criteria satisfies, record “excluded” for this paragraph:

(1) It mainly exhibits clinical, epidemiological, or statistical results.

(2) It mainly exhibits information of unit prices, cost changes (due to some therapies or experiments), or cost-benefit ratios.

3. Extract paragraphs: for each paragraph in the literature section of “Results”, if it is “included”, extract segments in the paragraph containing cost values. Each segment mustn’t exceed 200 English words and mustn’t include any tables.

Part III: template-based output format

1. Output table information: if any tables are extracted, output them sequentially without omission. For each table whose title includes the following content (<Table Name>): (<Column Name 1>; <Column Name 2>; …), output table name, headers, main body row by row, and any tips/footnotes. If no table extracted, output NR.

2. Output paragraph information: if no tables are recorded, but some paragraphs are extracted, output those sequentially. If no paragraph extracted, output NR.

3. Output formatting rules:

(1) Tables: compress the table representation. Column names appear only once; row data are presented as arrays. Do not delete any from the original table.

(2) Paragraphs: store extracted paragraphs in a text array. Each paragraph must include disease names and cost values.

4. Output the results of both parts in JSON format, even if the result is NR:

{disease_names …, table_information …, text_information …};

Nest table structure in table_information: {table_title …, table_header …, table_rows [] [] []…, table_tip …}, each square bracket in table_rows represents a row of the table.

For example:

```json
{
“disease_names”: “heart failure, ischemic heart disease”,
```

````
“table_information”: “{
“table_title”: “Table 2: Lifetime Mean Costs, Effectiveness, and Cost-effectiveness Ratios”,
“table_header”: “[“Strategy”, “Heart failure admissions”, “Life expectancy”, “Cost”, “QALYs”]”,
“table_rows”: “[“Enalapril indefinitely”, “2.91”, “6.93 years”, “114778 (86083 to 147136)”, “5.49 (5.47-5.51)”],
[“Enalapril for 2 mo, then sacubitril-valsartan”, “2.48”, “8.10 years”, “139456 (100800 to 168916)”, “6.45 (6.43-6.47)”]”
“table_tip”: “”
},
{
… (such as Table 3)
}”
“text_information”: “NR”
}
```
Or
```json
{
“disease_names”: “diabetes, stroke”,
“table_information”: “NR”,
“text_information”: “[“Paragraph 3: According to the findings of these studies in which patients provided their data, the average cost of an outpatient visit ranged from 5.97 million USD (in Nigeria) to 56.94 million USD 7.41 million (in Iran). Costs of annual outpatient visits were reported in these studies conducted in LMICs (24 million USD in Bangladesh and 85.05 million USD in Iran). In comparison, the annual outpatient visit cost in Nigeria was 26.88 million USD.”, “Paragraph 5: The USA had the highest inpatient costs at 9581 USD annually per patient, while Russia had the lowest at 1497 USD annually. The cost of medicine is the second highest cost component per patient (n = 12, 28%), and it exhibited the same results as the previous cost component, with the United States having the highest cost at 7884 USD yearly per patient and Finland having the lowest cost at 192 USD annually per patient.”]”
}
```
FULL TEXT:
````

5.2 Extraction of Key Information Paragraphs

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 (Jaskulski et al., 2025).

Prompt:

```
Please analyze the full text of the literature to accurately identify and record all paragraphs containing information related to the disease costs, including years, regions, population characteristics, cost types, etc.
The input consists of two parts:
1. Disease cost tables extracted from the literature, containing all numerical data related to disease costs, expenses, or economic burdens.
2. The full text of the literature.

Part I: identify paragraphs containing years and regions
1. In the literature section of “Methods”, identify and record paragraphs that contain research period information, normally with terms such as “year”, “decade”, “duration”, “period”, “from...to...”, etc.
2. In the literature section of “Methods”, identify and record paragraphs that contain research region or spot information, normally with names of country, province, state, region, city, county or hospital. Do not exceed 100 English words in total.

Part II: identify paragraphs containing inflation & base year information
In the literature section of “Methods”, especially in sub-sections about economic analysis, identify and record paragraphs indicating whether cost values were adjusted for inflation, and whether there reported a base year for multi-year costs or adjusting inflation. Such paragraphs normally contain keywords like “inflation”, “base year”, “adjust to”, “convert to”, “price index”, “present value”, etc. Do not exceed 100 English words.

Part III: identify paragraphs describing population characteristics
In the literature sub-section of “population/sample” within section of “Methods”, as well as the sub-section of “population characteristics/baseline results” within “Results”, identify and record sequentially:
1. Paragraphs containing phrases like entire population, patients with <Disease Name>, patients diagnosed with <Disease Name>, etc.
2. Paragraphs containing age groups of population like X-Y years and >X years, or age-related terms such as children, elderly, adults, etc.
3. Paragraphs containing gender-related terms such as male, female, gender-stratified, etc.
4. Paragraphs containing income-related phrases such as low/high income, socioeconomic
```

status, rich/poor, <X dollars per month, etc.
5. Any other paragraphs describing specific population subgroups.
Do not exceed 200 English words in total.

Part IV: identify paragraphs containing disease names and ICD codes
In the literature section of “Methods”, especially in sub-sections defining diseases and population selection criteria, identify and record paragraphs that mention disease names along with their ICD-10 codes. If other coding systems are used (e.g., ICD-9, ICD-9-CM, SNOMED), record those as well. Do not exceed 100 English words.

Part V: identify paragraphs describing cost types
In the literature sub-sections describing cost types and cost calculation methods within section of “Methods”, as well as adjacent paragraphs of extracted cost tables (the first input), identify and record sequentially:
1. Paragraphs containing terms like total cost, direct cost, indirect cost, etc.
2. Paragraphs containing reimbursement-related terms such as medical insurance, commercial insurance, out-of-pocket, self-paid, reimbursed by..., etc.
3. Paragraphs containing medical treatment types for direct costs, such as hospitalization, outpatient, medication, surgery, diagnosis, etc.
4. Paragraphs containing types for indirect costs, such as salary reduction, productivity loss, household burden, etc.
5. Any other paragraphs describing specific cost types.
Do not exceed 300 English words in total.

Part VII: template-based output format
Combine results from six parts above, each part’s records should be merged into a single paragraph. Output these six paragraphs together in the following JSON format: {p_year_region …, p_inflation_base …, p_population …, p_disease …, p_cost_type …, p_caliber_sample …}. For each item, if no paragraphs recorded, output NR.
For example:
```json
{
“p_year_region”: “In brief, 21374 eligible participants with acute myocardial infarction were recruited from 63 hospitals in Kerala, India, from November 2014 to November 2016. In this cross-sectional substudy, individual- and household-level cost data were collected 30 days after hospital discharge from a sample of 2114 respondents from November 2014 to July 2016.”,
“p_inflation_base”: “Costs were assessed in Indian rupees (Rs) and converted into 2015 international dollars based on purchasing power parities published by the World Bank (17.152=international$)13; thus, unless otherwise indicated, all costs reported herein as “$” represent international dollars.”,
“p_population”: “We performed age-stratified (<50 years; 50-70 years; and >70 years) and sex-stratified random sampling of a subset of participants who presented for 30-day follow-up to evaluate microeconomic (ie, individual- and household-level) costs
```

~~~~
associated with acute myocardial infarction hospitalization.”,
“p_disease”: “As a prespecified substudy of ACS QUIK, the present study collected data on individual- and household-level expenditures associated with acute myocardial infarction.”,
“p_cost_type”: “Inpatient cardiovascular hospitalization costs were calculated as the sum of self-reported spending on hospital admission, diagnostic tests, emergency department, food, treatment, ambulance, surgery, medicine, attendants, travel, and other self-reported inpatient costs. Total health care costs additionally included posthospitalization expenses on physicians, rehabilitation, home care, diagnostic tests, food, medicine, transportation, and other self-reported posthospitalization expenses. Out-of-pocket costs were calculated as the sum of all reported costs minus the total reimbursement by the insurance provider. Inpatient percentages were calculated only among patients with expenses reported.”,
“p_caliber_sample”: “In this cross-sectional substudy, individual- and household-level cost data were collected 30 days after hospital discharge from a sample of 2114 respondents from November 2014 to July 2016. The median (interquartile range) expenditure among respondents was $480.4 ($112.5-$1733.0) per acute myocardial infarction encounter, largely driven by in-hospital expenditures.”
}
```
FULL TEXT:
~~~~

5.3 Identification of 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.

Prompt:

~~~~
Task description:
You are given one table at a time, extracted from a health economics study. Table structure, headers, and layout may vary substantially. Your task is not to classify, judge, harmonize, or infer costs. Your task is to systematically extract numeric cost
~~~~

values exactly as reported, row by row, and output one JSON object per extracted cost value.
Objectives are:
1. No missing cost-bearing rows.
2. No hallucination or inference.
3. Minimal token usage.
4. One cost value → one JSON object. Each output object can be directly concatenated into a JSON array.
5. Strictly follow extraction rules. No additional contents are allowed.

Extraction rules:
1. Row-based processing:
> Process table row by row, in original order.
> Skip rows with no numeric cost.
> Never merge multiple rows.
2. One cost per object:
> Extract only one numeric cost per object.
> Prefer mean over median.
> Ignore CI, SD, IQR, ranges, brackets.
3. Currency:
> If USD and other currencies coexist, select USD.
> Otherwise, select reported currency.
> Do not convert.
4. Numeric normalization:
> Each extracted cost value MUST be output as one standalone JSON object with 2 fields: cost_value and cost_feature.
> cost_value = numeric string only.
> Do not scale, include brackets, or uncertainty.

How to construct cost_feature (detailed):
cost_feature is a single semicolon-separated string concatenating all remaining attributes.
Fixed internal order (STRICT): 11 elements inside cost_feature must appear exactly in the following order, even if some are NR: currency_raw; sample_size_raw; cost_caliber_raw; data_country_raw; data_region_raw; data_year_raw; base_year_raw; disease_name_raw; disease_code_raw; cost_category_major_raw; cost_category_minor_raw.
Extract only explicitly stated information, if not reported, fill NR.
> currency_raw: ISO-style currency if identifiable (e.g., USD, CNY, INT$), otherwise NR.
> sample_size_raw: Numeric sample size if stated.
> cost_caliber_raw: Statistical unit if stated (e.g., “mean annual cost per patient”).
> data_country_raw: Country name if explicitly stated.
> data_region_raw: Geographic region/city only (NOT care setting).

> data_year_raw: Study year(s) if stated.
> base_year_raw: Price base year if stated.
> disease_name_raw: English disease name if stated.
> disease_code_raw: ICD code or other classification if stated.

Table Format Indicators:
1. Wide format indicators:
> Multiple cost value columns.
> Column headers represent subgroups, settings, or time periods.
> Each column contains multiple related cost values.
2. Long format indicators:
> Single cost value column.
> Row label provides complete cost classification.
3. Key principle: When analyzing tables, focus on “one column, multiple values share a common attribute.” This attribute becomes the major cost category. Row labels correspond to minor categories.

Cost category hierarchy rules:
1. Core principle: Cost hierarchy is determined dynamically based on table structure, following the “one column, multiple values share a common attribute” rule.
2. Wide format (multiple cost columns): Identify columns containing numeric cost values. Column header → major cost category; Row label → minor cost category.
For example: “Columns: Inpatients, Outpatients”.
Inpatients column contains multiple costs → “Inpatients” is cost_category_major_raw, row label is cost_category_minor_raw.
3. Handling multi-level row hierarchies in wide-format tables: If a wide-format table contains two or more hierarchical levels within the row structure (e.g., section headers + indented sub-rows),
> All applicable row-level categories MUST be preserved
> Concatenate hierarchical row labels using a hyphen (-)
> The concatenated string MUST be placed entirely in cost_category_minor_raw
> Preserve the original wording and order as shown in the table
> Do NOT invent, normalize, or rephrase labels
Example logic (illustrative only):
Higher-level row: Admission facility type; Lower-level row: Township health centers → cost_category_minor_raw = “Admission facility type - Township health centers”.
4. Long format (single cost column): If row label contains hierarchical category (e.g., Direct medical costs – Medicines): cost_category_major_raw = higher-level category; cost_category_minor_raw = specific sub-category.
5. If only one category level: cost_category_major_raw = category; cost_category_minor_raw = NR.
6. General rules:
> Do not invent hierarchy.
> Do not normalize wording, and preserve original terms.

```
Output Format (STRUCTURE ONLY):
Output these 2 pieces in precise JSON format: {cost_value …, cost_feature …}. For multiple cost values in one form/paragraph, the output should be a single JSON array: [{...}, {...}, ..., {...}].
For example:
```json
[
{
“cost_value”: “150400”,
“cost_feature”: “USD; NR; Mean annual cost per patient; USA; NR; 2018; 2018; Diabetes mellitus; NR; Direct medical costs; Medicines”
}
{
… (another cost value)
}
…
]
INPUT TABLE OR PARAGRAPH:
```

5.4-1 Sub-step 1 in “Populating Specific Information”

Sub-step 1 fills in basic cost information, including year, country/region, disease name, and cost value.

Prompt:

```
Task description:
You are given two inputs for a single disease cost value:
1. A JSON object containing the cost value and its preliminary features;
2. Supplementary text paragraphs with contextual details (e.g., years, regions, populations, cost types).
Your task is to extract only the following fields for this cost value from these inputs, in strict adherence to the rules below.

Extraction rules (CRITICAL):
1. Priority order: Use Input 1 first, then Input 2.
2. Never infer unless explicitly allowed (see “Inferred fields” below).
3. If information is absent or ambiguous → fill NR.
4. Preserve original wording, do not normalize or rephrase without permission.

Raw extraction fields:
1. data_country:
> Country name the cost exists. Mandatory.
2. data_region:
> Province/state/region/city names the cost exists.
```
```

3. data_year:
> Numeric only. Consecutive years: 2015-2017. Non-consecutive: 2015, 2017.
> Partial years (e.g. Q1 2020) → full year 2020.
4. base_year:
> Price adjustment base year for multi-year data as possible. Single year → NR.
> Fill number ONLY (e.g. 2019).
5. disease_name:
> English disease name as stated in source (e.g. diabetes).
> NEVER symptoms/organs/treatments (e.g. fever or chemotherapy → NR).
6. disease_icd:
> ICD code ONLY if explicitly stated (e.g. I10).
> Codes in appendix → Fill “In appendix”.
> Non-ICD classification → brief description (e.g. ICD-9-CM: 250).
> No coding mentioned → NR.
7. cost_value:
> EXACT numeric value with unit scaling, don’t use thousand, million, and billion: one thousand → 1000; one million → 1000000; one billion → 1000000000.
> Preserve decimals (e.g. 2.73).
8. currency:
> Original currency description as stated in source.

Inferred fields (Strict enumeration ONLY):
Inference is allowed only for the following fields, using the rules below. If uncertain → fill NR.
9. disease_name_infer:
> MeSH medical term of disease name ONLY. (e.g. Diabetes mellitus).
10. currency_infer
> Use ISO 4217 code ONLY (e.g. USD, EUR).
> For international USD, label “Intl. USD”.
> Unidentifiable currency → NR.

Output format:
Combine the above information along with the literature_doi, totaling 11 items, to form specific information entry corresponding to the cost value. Format the output as JSON: {{data_country_region ..., data_year ..., base_year …, ......, currency_infer ..., literature_doi ...}}.
For example:
```json
{{
“data_country”: “France”,
“data_region”: “Haute-Seine province”,
“data_year”: “2019-2020”,
“base_year”: “2019”,
“disease_name”: “Acute stroke”,
```

```
"disease_icd": "I60-I64",
"cost_value": "1500000000",
"currency": "Euros",
"disease_name_infer": "Stroke",
"currency_infer": "EUR",
"literature_doi": "10.1097/mlr.0000000000001745"
}}

INPUT DATA (MUST USE):
Input1_cost_json:
DOI:
Cost value:
Headers:

Input2_supplementary_paragraphs:
```

5.4-2 Sub-step 2 in "Populating Specific Information"

Sub-step 2 fills in population characteristics corresponding to the cost, such as age group, gender, and income level.

Prompt:

```
Task description:
You are given two inputs for a single disease cost value:
1. A JSON object containing the cost value and its preliminary features;
2. Supplementary text paragraphs with contextual details (e.g., years, regions, populations, cost types).
Your task is to extract only the following fields for this cost value from these inputs, in strict adherence to the rules below.

Extraction rules (CRITICAL):
1. Priority order: Use Input 1 first, then Input 2.
2. Never infer unless explicitly allowed.
3. If information is absent or ambiguous → fill NR.
4. Preserve original wording, do not normalize or rephrase without permission.

Raw extraction fields:
1. population_age:
• Format rules:
  > Single age: X years (e.g., 45 years);
  > Age range: X-Y years (e.g., 18-65 years);
  > No upper limit: X years and above (e.g., 65 years and above);
  > No lower limit: 0-Y years (e.g., 0-5 years).
• Not age-specific or covers all ages → NR.
• NEVER convert non-standard descriptions to age ranges.
```

2. population_gender:
• ONLY these exact values allowed:
  > Male (male-only populations);
  > Female (female-only populations);
  > NR (mixed gender or not specified).
• NEVER use variations like M, F, men, women.

3. population_income:
• Extract EXACT income description as stated in source.
• No income stratification or covers all incomes → NR.
• Preserve original terminology (e.g., low-income, Medicaid beneficiaries).

4. population_other:
• ONLY include other population characteristics.
• If cost comes from experiment/cohort → Brief description of experiment group/specific cohort.
• EXPLICITLY PROHIBITED content:
  > Disease names;
  > Insurance types;
  > Treatment methods;
  > Cost types.
• No additional characteristics → NR.

Output format:
Combine the above information along with the literature_doi, totaling 5 items, to form specific information entry corresponding to each disease cost value. Format the output as JSON: {{population_age ..., population_gender ..., population_income …, population_other ..., literature_doi ...}}.
For example:
```json
{{
"population_age": "18-65 years",
"population_gender": "Male",
"population_income": "Low-income population",
"population_other": "NR",
"literature_doi": "10.1097/mlr.0000000000001745"
}}
```

INPUT DATA (MUST USE):
Input1_cost_json:
DOI:
Cost value:

```
Headers:

Input2_supplementary_paragraphs:
```

5.4-3 Sub-step 3 in "Populating Specific Information"

Sub-step 3 faithfully fills in cost category information, including statistical caliber, sample size, reimbursement type, and treatment type.

Prompt:

```
Task description:
You are given two inputs for a single disease cost value:
1. A JSON object containing the cost value and its preliminary features;
2. Supplementary text paragraphs with contextual details (e.g., years, regions, populations, cost types).
Your task is to extract only the following fields for this cost value from these inputs, in strict adherence to the rules below.

Extraction rules (CRITICAL):
1. Priority order: Use Input 1 first, then Input 2.
2. Never infer unless explicitly allowed.
3. If information is absent or ambiguous → fill NR.
4. Preserve original wording, do not normalize or rephrase without permission.

Raw extraction fields:
1. cost_caliber:
• Preserve statistical caliber of cost as stated in source. Containing population and time, e.g. the cost for per patient in 2 years.
• NEVER invent units or time periods not explicitly stated.
2. cost_sample:
• Preserve EXACT numeric value of sample size with unit as stated in source
• Format: "[number] ([unit])", e.g., 10000 (patients)
3. cost_insurance:
• ONLY applicable for direct costs. Indirect costs → NR.
• Preserve medical insurance description of cost as stated in source.
4. cost_treatment:
• ONLY applicable for direct costs. Indirect costs → NR.
• Preserve medical treatment type description of cost as stated in source, e.g. diagnosis and testing for outpatient care.
5. cost_indirect:
• ONLY applicable for indirect costs. Direct costs → NR.
• Preserve type description of indirect cost as stated in source, e.g. salary loss due to inability to work.
6. cost_other:
• ONLY include non-standard other cost characteristics.
```

```
• Cost comes from experiment/cohort → Brief description of experiment group/specific cohort.
• Prohibited content: population characteristics.
• No additional characteristics → NR.

Output format:
Combine the above information along with the literature_doi, totaling 7 items, to form specific information entry corresponding to each disease cost value. Format the output as JSON: {{cost_caliber ..., cost_sample ..., cost_insurance …, ......, cost_other ..., literature_doi ...}}.
For example:
```json
{{
"cost_caliber": "Cost for France in the year of 2012",
"cost_sample": "66.3 million (population)",
"cost_insurance": "The cost are all from patients covered by compulsory social medical insurance",
"cost_treatment": "Postoperative recovery in hospitalization",
"cost_indirect": "NR",
"cost_other": "NR",
"literature_doi": "10.1097/mlr.0000000000001745"
}}


INPUT DATA (MUST USE):
Input1_cost_json:
DOI:
Cost value:
Headers:

Input2_supplementary_paragraphs:
```

5.4-4 Sub-step 4 in "Populating Specific Information"

Sub-step 4 performs categorical inference on the information from Sub-step 3, involving statistical caliber, reimbursement type, and treatment type (Danhauser et al., 2025).

Prompt:

```
Task description:
You are given two inputs for a single disease cost value:
1. A JSON object containing the cost value and its preliminary features;
2. Supplementary text paragraphs with contextual details (e.g., years, regions, populations, cost types).
Your task is to extract and infer the following fields for this cost value from these
```
```

inputs, in strict adherence to the rules below.

Rules (CRITICAL):
1. Priority order: Use Input 1 first, then Input 2.
2. Adhere rules of inferred fields as shown below.
3. If information is absent or ambiguous → fill NR.
4. Preserve original wording, do not normalize or rephrase without permission.

Inferred fields (Strict enumeration ONLY):
1. cost_caliber_infer:
• Format: "[population unit] [time period]" (e.g., per patient one year).
• Population unit options:
> per patient: per patient, each patient, per case, each case, each individual
> per capita: ONLY if the cost is from total cost divided by the total population (NOT by all patients)
> country/regional total: total country/region, in total for a country/region, state-/province-wide
• Time period options:
> one year: per year, in one year, annual, yearly
> one month: per month, in one month, monthly
> X years/months: for conditions of more than one year/month
> per course/period: a course of treatment, one period of hospitalization, one episode of nursing, per inpatient admission
> per visit: per visit to clinic/hospital, one diagnosis, consult once
2. cost_direct_infer:
• ONLY these exact values allowed:
> direct: medical care, hospitalization, nursing, medicines, transport, healthcare services, etc.
> indirect: productivity loss, income loss, labor loss, household burden, etc.
• NEVER use additional descriptions.
3. cost_insurance_infer:
• ONLY these exact values allowed:
> Government & compulsory insurance: government/public/social health insurance or compulsory non-governmental insurance
> Voluntary private insurance: voluntary private/commercial insurance
> Out-of-pocket: user charges/patient-paid
• Not specified or not applicable → NR.
4. cost_treatment_infer:
• Indirect costs → NR.
• Major treatment select one from these 7 options ONLY:
> Ambulatory care: preventive, curative, and rehabilitative services provided in outpatient settings such as clinics and physician offices
> Inpatient care: all medical services and goods consumed during a hospital stay
> Emergency department care: treatment and rehabilitative services delivered in

````
hospital-based emergency departments
> Nursing facility care: nursing and residential care provided in homes or long-term care facilities
> Dental care: preventive and curative oral health services performed at dental facilities
> Retail pharmaceuticals: medications purchased from retail pharmacies, excluding drugs used in institutions
> General administration: administrative and overhead costs of managing health systems and insurance programs
5. cost_indirect_infer:
• Direct costs → NR.
• Select one from these 5 options ONLY:
  > Absenteeism: income loss due to inability to work
  > Presenteeism: productivity loss caused by working with diseases
  > Premature deaths: potential income loss due to deaths
  > Caregiver burden: income loss caused by the need to take care of patients
  > Social welfare burden: long-term social security and allowance need
• No specific type mentioned → NR.

Output format:
Combine the above information along with the literature_doi, totaling 6 items, to form specific information entry corresponding to each disease cost value. Format the output as JSON: {{cost_caliber_infer ..., cost_direct_infer ..., cost_insurance_infer …, ..., cost_indirect_infer ..., literature_doi ...}}.
For example:
```json
{{
“cost_caliber_infer”: “Country total one year”,
“cost_direct_infer”: “Direct”,
“cost_insurance_infer”: “Government and compulsory insurance”,
“cost_treatment_infer”: “Inpatient care”,
“cost_indirect_infer”: “NR”,
“literature_doi”: “10.1097/mlr.0000000000001745”
}}


INPUT DATA (MUST USE):
Input1_cost_json:
DOI:
Cost value:
Headers:

Input2_supplementary_paragraphs:
````
````

# 6. Rationale for Selecting LLMs

## 6.1 Rationale for Selecting DeepSeek-V3.2 as the Main LLM

The primary selection of DeepSeek-V3.2 for our automated screening and data extraction pipeline is driven by its superior balance of contextual capacity, structural reasoning capability, and computational efficiency. Given that a significant proportion of the retrieved literature exceeds 8k–32k tokens, the model's extensive context window was critical for mitigating the "lost in the middle" phenomenon and ensuring that cost data dispersed across tables, footnotes, and distant paragraphs are accurately captured (Liu et al., 2023). Furthermore, DeepSeek-V3.2 demonstrates robust performance in adhering to complex, multi-step prompt instructions and generating strict JSON outputs, which is essential for minimizing hallucinations and maintaining the fidelity of structured economic data during the four-step extraction process (Gupta et al., 2024; Zhang et al., 2025).

## 6.2 Rationale for Selecting Other LLMs in Quality Assessments

To address potential model-specific biases and ensure the reproducibility of our findings, we implement a cross-model validation framework using GPT-4.1 mini and Qwen3.6-Plus as benchmark models. These models are selected to represent diverse architectural families and training corpora, allowing us to assess whether screening decisions and data extractions are consistent across different LLMs or merely artifacts of a specific model's training bias (Wu et al., 2025). By triangulating results from these distinct platforms, we aim to establish a more robust consensus on eligibility criteria and data accuracy, thereby enhancing the methodological rigor of our AI-assisted systematic review (Shi et al., 2025).

This multi-model approach also serves as a practical performance benchmark against industry standards. We conduct systematic random sampling where records are independently processed by all three models, calculating inter-model agreement metrics (Cohen's Kappa) and standard performance indicators such as Precision, Recall, and F1-score against a manual audit. This comparative analysis not only quantifies the reliability of DeepSeek-V3.2 relative to established baselines but also ensures that our final dataset reflects stable, consensus-driven findings rather than model-dependent anomalies, thus validating the robustness of our automated pipeline (Danhauser et al., 2025; Gao et al., 2026).

# 7. Methods and Results of Quality Assessments

## 7.1 Comparison among LLMs of Literature Screening

To evaluate the robustness of automated literature screening, we conduct a comparison against two alternative LLMs, GPT-4.1 mini and Qwen3.6-Plus. A systematic random sample comprising 5% of the retrieved articles is selected for validation, consistent with established quality control practices for AI-assisted systematic reviews (Wu et al., 2025; Zhang et al., 2025). Each sampled abstract, combined with the identical step-specific prompts and eligibility criteria, is independently processed by both LLMs through the same three-stage screening strategy. We compute inter-model accuracy, precision, recall, false positive rate, and F1 score across all stages to quantify consistency in identifying CSDs, economic analyses, and actual incurred costs. This cross-platform benchmarking determines whether screening outcomes reflect stable decision boundaries or are model-dependent, thereby verifying the reproducibility and methodological rigor of our LLM-driven literature selection process (Cai et al., 2025).

Results in Table S3 show that, the multi-step LLM-driven screening demonstrated robust quality, characterized by consistently high accuracy (>93% in steps 2–3) and a markedly reduced false positive rate (3.5–4.0% in step 3) across LLMs. This cross-model consistency confirmed the screening’s structural stability and effective noise-filtering capacity. However, precision remained modest (23–49%) and recall gradually declined in later stages. These limitations likely stemmed from increasingly stringent exclusion criteria, which may inadvertently discard borderline-relevant studies, alongside LLMs’ inherent conservatism when processing complex contextual matching tasks. Despite this trade-off, the screening’s high specificity ensured a rigorously curated candidate pool, substantially reducing downstream manual validation burden.

The controlled recall reduction was deemed acceptable within a “high-specificity-first” framework, aligning with systematic review standards in which preventing false inclusions typically outweighs the cost of minor, retrievable omissions.

Table S3: Results of comparison among LLMs of literature screening

| LLM | Step of screening | Accuracy (%) | Precision (%) | Recall (%) | False positive rate (%) | F1 score |
|---|---|---|---|---|---|---|
| GPT-4.1 mini | Step1 | 61.43 | 32.91 | 95.85 | 46.81 | 0.49 |
| | Step2 | 93.95 | 28.22 | 86.79 | 5.86 | 0.426 |
| | Step3 | 95.48 | 23.15 | 64.10 | 3.94 | 0.34 |
| Qwen3.6-Plus | Step1 | 79.68 | 48.68 | 94.39 | 23.85 | 0.642 |
| | Step2 | 92.13 | 41.67 | 89.43 | 7.71 | 0.569 |
| | Step3 | 96.13 | 28.28 | 71.79 | 3.41 | 0.406 |

### 7.2 Manual Label Validation of Data Extraction

To validate the performance of LLM-driven cost data extraction pipeline, we implement a systematic manual label of the final extraction results (Zheng et al., 2023). A random sample comprising 2% of all extracted cost values is independently reviewed (Wu et al., 2025). 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 (Danhauser et al., 2025; Ribeiro de Faria et al., 2025). 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 (Gao et al., 2026).

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 Table S4. 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.

Table S4: Results of manual label validation of cost data extraction

| Information cluster | Information field | Accuracy (%) | Precision (%) | Recall (%) | F1 score |
|---|---|---|---|---|---|
| Reported information | data_country | 97.9 | 99.3 | 98.6 | 0.989 |
| | data_year | 94.7 | 96.9 | 97.6 | 0.973 |
| | disease_name | 96.1 | 97.3 | 98.7 | 0.980 |
| | cost_value | 67.2 | 87.7 | 74.2 | 0.804 |
| | currency | 96.3 | 97.8 | 98.4 | 0.981 |
| Inferred information | population_age | 87.0 | 93.1 | 93.0 | 0.930 |
| | cost_caliber_infer | 79.7 | 80.0 | 99.6 | 0.887 |
| | cost_insurance_infer | 89.8 | 93.2 | 96.2 | 0.946 |
| | cost_treatment_infer | 75.9 | 84.7 | 88.0 | 0.863 |

## 8. Methods and Results of Credibility and Confidence Scorings

### 8.1 Methods of Credibility Scoring for Each Publication

In order to deliver reliability information for COI evidence studies, we introduce a Credibility scoring scheme for each publication. This Credibility is mainly composed of three components: “database authority and scale”, “sample size”, and “cost calculation method” (Cotton et al., 2025). Since the Credibility of publications is relatively difficult to depict with a unified quantitative standard, we adopt a qualitative classification approach to score these three components separately.

Regarding “database authority and scale”, we categorize them into five types, considering “official national health accounts or case databases” as the most authoritative; followed by “national/multi-regional large-scale surveys and case collections”; the 3rd level is “regional-level official health accounts or case databases”; while “surveys and case collections from individual regions, hospitals, or persons” is the least authoritative. These levels are denoted as 4, 3, 2, and 1, respectively.

For “sample size”, considering the significant variation in sample sizes across different studies, it is not appropriate to directly use absolute numerical values. Therefore, we propose to categorize sample sizes into five levels based on their order of magnitude: (0, 1000), [1000, 10000), [10000, $10^5$), [$10^5$, $10^6$), [$10^6$, +∞). These levels are assigned integer scores ranging from 1 to 5 respectively.

For “cost calculation method”, it is divided into three levels, with the results obtained from top-down methods like health account decomposition having the highest credibility (Wieser et al., 2018; Wubulihasimu et al., 2015); followed by bottom-up data aggregation from surveys and case collections; and the model output data has the lowest credibility. These levels are similarly denoted as 3, 2, and 1, respectively.

To systematically extract the aforementioned information following a unified standard, we continue to employ LLM to analyze literature. Initially, we extract paragraphs containing the aforementioned three components based on the full text of the literature.

Prompt:

```
Role & Objective: You are an expert academic data extraction specialist in health economics and clinical research. Your task is to analyze the provided article on disease-related costs and accurately extract exactly four structured metadata fields. Focus exclusively on peer-reviewed economic evaluations, cost-of-illness studies, or healthcare utilization analyses.

Exaction fields and requirements:
1. doi (String | NR)
> Extract the Digital Object Identifier of the article.
> Format: Standard DOI string without URL prefixes (e.g., 10.1016/j.jval.2023.05.002).
> If only a URL is visible, extract the DOI portion. If the DOI is completely absent or unreadable, return NR.

2. data_source (String | Array of Strings | NR)
```

> Identify the primary database, survey, registry, cohort, or administrative dataset used for the cost analysis.
> Be specific and retain official acronyms/names (e.g., National Inpatient Sample (NIS), Optum Clinformatics Data Mart, Prospective multicenter cohort).
> If multiple independent sources are combined, separate them with semicolons. If the source is vague or unreported, return NR.

3. sample_size (String | NR)
> Extract the final number of participants, patients, or claims included in the cost analysis.
> Prefer a single integer and a word of "patients/participants" representing the total analytic N. If only an approximate value, range, or subgroup breakdown is provided, return it as a descriptive string (e.g., approximately 45,200 patients, 8,000–12,000 patients).
> Do not infer or calculate N from percentages. If missing, return null.

4. calculation_methodology (String | NR)
> Describe precisely how the disease cost figures were derived, aggregated, or standardized.
> Look for explicit methodological terms such as: mean/median cost per patient, bottom-up microcosting, top-down allocation, claims-based charge-to-cost conversion, full account decomposition, inflation adjustment year, PPP/currency conversion, or standardized billing code aggregation.
> Keep the description concise (1 sentence) and technically accurate. If the methodology is not explicitly detailed, return NR.

Constraints:
1. Do not hallucinate, approximate, or infer missing information. If a field is not explicitly stated in the full text, output NR.
2. Ignore funding sources, author affiliations, or secondary citations unless they are explicitly the primary data source for the cost analysis.
3. Ensure the output is parsable JSON. Escape special characters properly if necessary.

Output specification:
1. Return ONLY a valid JSON object. Do not include markdown formatting, explanations, or conversational text.
2. Use exactly the following keys: doi, data_source, sample_size, calculation_methodology.
3. Strictly enforce the data types specified above. Use NR for any field that cannot be confidently extracted from the text.

JSON Example:
{

```
  “doi”: “10.1001/jama.2022.18456”,
  “data_source”: “National Health and Nutrition Examination Survey (NHANES); Medicare Fee-for-Service Claims”,
  “sample_size”: “12,450 patients”,
  “calculation_methodology”: “Costs were derived using bottom-up microcosting based on standardized unit prices from the Medicare Physician Fee Schedule and hospital cost reports. Values represent mean annual direct medical costs per patient, adjusted to 2022 USD using the Consumer Price Index.”
}
FULL TEXT:
```

Secondly, based on the paragraphs containing key information, we categorize the above components uniformly.

Prompt:

```
Based on the extracted data_source, sample_size, and calculation_methodology from the previous step, rewrite them into the following compact format without losing factual correctness. Do not introduce new information not explicitly present in the extracted fields.

1. data_source_infer: Choose only from these four types:
> “national official health accounts or disease database”;
> “national/multi-regional surveys or case collection”;
> “one region official health accounts or disease database”;
> “one region/hospital survey or case collection”.
Other types are prohibited to fill in. If multiple sources exist, fill only the most dominant one.

2. sample_size_infer: Keep only the numeric value (integer or range). Remove words like “patients”, “participants”, “approximately”. Examples: 12450, 8000–12000, ~45000.

3. calculation_methodology_infer: Choose only from these four types:
> “top-down account decomposition” for systematically and top-down decomposing the health account to obtain costs;
> “bottom-up microcosting or sampling mean” for bottom-up summarizing or averaging the costs of multiple samples from a cohort or case database;
> “Separated single data” for the costs obtained from a single patient, a single visit, or a single trial;
> “model output data” for the cost calculated or estimated from the model, not the actual cost incurred.
Other types are prohibited to fill in. If multiple sources exist, fill only the most dominant one.

Rules:
1. Do not change doi.
```

```
2. If the original field is NR, output NR in the short version.
3. sample_size_infer must be number or string (if range like 8000–12000 keep as string).
4. Keep all outputs parsable, deterministic, and conservative (no hallucination).

Output format: JSON only, same keys as input.
Example:
json
{
  “doi": "10.1001/jama.2022.18456",
  “data_source_infer": " national/multi-regional surveys or case collection",
  “sample_size_infer": 12450,
  “calculation_methodology_infer": “bottom-up microcosting or sampling mean"
}
```

Finally, normalize the results of these three components separately:

$$Score_{i,j,norm} = \frac{Score_{i,j} - \min Score_i}{\max Score_i - \min Score_i} \quad (1)$$

Where the $Score_{i,j}$ means the score of $i$th component of $j$th publication; $\min Score_i$ and $\max Score_i$ denote minimum and maximum of $i$th component scores for all publications. Then, the normalized results are averaged to obtain the Agreement score for each publication, which is averagely divided into five levels from 1 to 5 (intervals for all levels are 0.2), with 1 representing the lowest agreement and 5 representing the highest.

### 8.2 Methods of Confidence Scoring for Each Type of Cost Set

To facilitate transparent analysis of COI, we develop a Confidence scoring scheme to evaluate the extracted cost values. For each type of cost set (median + quartiles), we apply a three-tier framework adapted from IPCC guidance to scoring the Confidence (Kause et al., 2022; Mastrandrea et al., 2010):

(1) Evidence, quantified by the counts of extracted cost values and publications contributing to this cost set;

(2) Agreement, measured as the reciprocal of relative interquartile range (RIQR), reflecting dispersion among these original cost values;

(3) Credibility, derived from publication-level Credibility scorings in Section 8.1.

This multi-dimensional disclosure enables researchers to weight and tailor dataset contents according to their specific analytical requirements and risk tolerance.

### 8.3 Results of Aggregated National/Regional COI Evidences with Decomposition Details

Based on our Credibility scoring framework for publications, we select 36 aggregated national/regional COI publications (with decomposition details) from the "cost calculation method" processing procedure. These publications were all scored with relatively high reliability, covering over 90% of all eligible ones in this top-down category. Table S5 summarizes the basic characteristics of these selected publications. The included studies span 22 countries across multiple continents, with the majority concentrated in China (including Taiwan province), Brazil, South Korea, Spain, and the US. Publication years showed a notable clustering around 2020–2024. Most studies drew on national official health accounts or disease databases, while a smaller subset used multi-regional surveys. Sample sizes varied substantially, from several hundred to over 1.4 million, reflecting the heterogeneity in database scale and disease epidemiology. The decomposition dimensions most frequently reported included care types (e.g., inpatient vs. outpatient), gender, age groups, disease subtypes, and indirect cost components, with some studies further stratifying by region, insurance type, or patient conditions. This diverse coverage ensures that the selected subset provides a comprehensive empirical foundation for researchers requiring consistently estimated, decomposable cost data (Kneale et al., 2017; Wieser et al., 2018).

Table S5: COI evidence summary for 36 aggregated publications with decomposition details

| Research region | Research year | Disease mapped to GBD hierarchies | Statistical caliber unified | Scope of cost | Main decomposition | DOI |
|---|---|---|---|---|---|---|
| Bangladesh (Dhaka City) | 2016-2018 | Typhoid and paratyphoid | per patient per incident | Median cost of illness of enteric fever from the patient and caregiver perspective | Care types | 10.1093/cid/ciaa1334 |
| Brazil | 2011 | Endocrine, metabolic, blood, and immune disorders | total per year | Sum of the direct healthcare costs attributable to obesity (BMI ≥ 30 kg/m²) and morbid obesity (BMI ≥ 40 kg/m²) | Disease subtypes | 10.1371/journal.pone.0121160 |
| Brazil | 2017-2019 | Endocrine, metabolic, blood, and immune disorders | total per year | Ranges of total direct healthcare costs attributable to the excess consumption of salt, saturated fat, and trans fat | Age groups | 10.1371/journal.pone.0278891 |
| Brazil | 1990-2019 | Endocrine, metabolic, blood, and immune disorders | per capita per year | Direct treatment cost due to non-communicable diseases attributable to high body mass index | Regions | 10.1016/j.puhe.2024.05.013 |
| Brazil (State of Santa Catarina) | 1996-2009 | Rheumatoid arthritis | per patient per incident | Total cost of high-cost outpatient procedure permits (APAC) for the pharmacological treatment of rheumatoid arthritis | Clinical procedures | 10.1016/j.rbre.2016.07.003 |
| Chile | 2005-2020 | Mental disorders | per patient per year | Total GES (Explicit Health Guarantees) plan expenditures across both public and private health insurance | Disease subtypes | 10.3389/fpubh.2022.1005033 |
| China (Gansu province) | 2018-2022 | Diseases associated with sand-dust storms | per patient per incident | Total asthma hospitalization costs attributable to sand-dust storms | Regions | 10.1016/j.envres.2024.120345 |
| China (National and Shandong province) | 2018 | Stroke | total per year | Direct and indirect economic burdens of stroke | Care types | 10.1136/bmjopen-2023-080634 |
| China (Taiwan province) | 2012-2015 | Diarrheal diseases and Other intestinal infectious diseases | total per year | Total medical cost of foodborne illnesses | Disease subtypes | 10.1016/j.jfma.2020.03.013 |
| China (Taiwan province) | 2000; 2005; 2010 | Diabetes mellitus | per patient per year | Mean total healthcare costs per type 2 diabetes patient derived from National Health Insurance (NHI) | Care types | 10.1097/MLR.0000000000000288 |
| China (Taiwan | 1998-2014 | Dengue | per patient per | Mean total costs (including both direct | Care types | 10.1016/j.jiph.2017.07.021 |

| | | | | | | |
|---|---|---|---|---|---|---|
| province) | | | incident | and indirect costs) of dengue during non-epidemic and epidemic years | | |
| France | 2015-2016 | Chronic obstructive pulmonary disease | per patient per year | Mean overall total cost of all AECOPD-related hospital stays during the six-month follow-up period | Number of hopitalized times | 10.2147/COPD.S236787 |
| France | 2015-2019 | Cardiovascular diseases; Diabetes mellitus; Mental disorders; Neurological disorders; Chronic respiratory diseases; etc. | per patient per year | Total healthcare expenditures reimbursed by the French national health insurance for major categories of health conditions | Disease types | 10.1097/MLR.0000000000001715 |
| Iran | 2019 | Diabetes mellitus | per patient per incident | 30-day treatment costs of various antidiabetic medication regimens | Treatment schedules | 10.1186/s12939-022-01791-5 |
| Italy (Emilia-Romagna region) | 2008-2011 | Other neglected tropical diseases (prevention) | per capita per year | Mean expenditure per inhabitant for performing one anti-larval treatment in public road drains | Regions | 10.3390/ijerph14040444 |
| Mexico | 2008-2013 | Diabetes mellitus | per patient per incident | Total direct cost of hospitalisations for diabetes mellitus and its complications | Disease complications | 10.1016/j.gaceta.2016.06.015 |
| Netherlands | 1994-2010 | Multiple diseases | per patient per year | Total hospital spending in the country for all diseases | Disease types | 10.1016/j.healthpol.2014.11.009 |
| New Zealand | 2007-2014 | Cardiovascular diseases; Diabetes mellitus; chronic LLK disease; Neurological disorders | per patient per year | Total health system expenditures for patients with cardiovascular disease, type 2 diabetes, chronic kidney disease, and neurological conditions | Disease types | 10.1371/journal.pmed.1002716 |
| Norway | 2011 | Diabetes mellitus | total per year | Range of the total excess cost of diabetes, including both direct and indirect costs | Care types | 10.1016/j.diabres.2016.10.012 |
| Philippines (Cebu and Manila) | 2008-2012 | Dengue | per patient per year | Average combined (public and private sector) aggregate direct medical cost of dengue | Care types | 10.4269/ajtmh.14-0139 |
| Romania (Central Region, | 2017 | Stroke | per patient per incident | Average treatment and management cost per stroke inpatient care episode | Care types | 10.3389/fpubh.2020.605919 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Southern Region, North-Western Region, and Bucharest-Ilfov Region) | | | | | | |
| South Korea | 2007; 2012 | Ischemic heart disease | per patient per year | Total economic costs comprising both direct (medical and non-medical) and indirect (productivity loss) costs of acute myocardial infarction (AMIT | Care types | 10.1371/journal.pone.0117446 |
| South Korea | 2004-2013 | Chronic obstructive pulmonary disease | per patient per year | Direct and indirect costs of COPD | Care types and indirect cost types | 10.3111/13696998.2015.1100114 |
| South Korea | 2015-2019 | Alzheimer's disease and other dementias | per patient per year | Total direct treatment costs and indirect costs of dementia | Care types and indirect cost types | 10.1186/s12877-021-02526-x |
| South Korea | 2015 | Stroke | per patient per incident | Total economic cost (direct medical, direct non-medical, and indirect costs) of stroke | Care types | 10.15171/ijhpm.2018.42 |
| Spain | 2011-2013 | Nutritional deficiencies | per patient per year | Mean hospitalization costs among older adults who were at risk of malnutrition and malnourished | Patient conditions | 10.2147/ceor.s256671 |
| Spain | 2015 | Cardiovascular diseases | per patient per incident | Range of mean direct hospital costs across the six evaluated cardiovascular complications for the matched non-diabetic cohort | Disease subtypes | 10.1136/bmjdrc-2019-001130 |
| Spain | 2011-2019 | Ischemic heart disease | per patient per incident | Mean hospital admission cost in Euros for patients with ischemic heart disease | Gender | 10.1080/14737167.2022.2108794 |
| Sri Lanka | 2002-2010 | Malaria | per patient per incident | The Anti-Malaria Campaign's expenditure of malaria during the elimination phase | Treatment budget types | 10.1186/s12936-022-04249-9 |
| Switzerland | 2011 | Infectious diseases; Cardiovascular diseases; Chronic respiratory diseases; Alzheimer's disease and other dementias; Diabetes mellitus; etc. | total per year | Healthcare expenditures nationwide across specific disease categories | Disease types | 10.1007/s10198-018-0963-5 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Switzerland | 2012-2017 | Cardiovascular diseases; Diabetes mellitus and Chronic kidney disease; Chronic respiratory diseases; etc. | total per year | Total treatment and care spending for cardiovascular diseases, diabetes and kidney diseases, and chronic respiratory diseases | Disease types | 10.1186/s12913-023-10124-3 |
| Türkiye | 2012-2016 | Chronic obstructive pulmonary disease | per patient per year | Range of the mean direct cost of patient with COPD | Care types | 10.5152/TurkThoracJ.2021.19150 |
| Türkiye | 2016-2020 | Diabetes mellitus | per patient per year | Range of average direct medical costs for different diabetes-related microvascular complications | Care types | 10.14744/AnatolJCardiol.2023.3762 |
| UK (England) | 2018 | Ischemic heart disease; Alzheimer’s disease and other dementias; Stroke | per patient per year | Total economic burden (healthcare, social care, and productivity loss) of coronary heart disease, dementia, and stroke | Care types | 10.1016/S2666-7568(24)00108-9 |
| United States | 1996-2019 | Diabetes mellitus and Endocrine, metabolic, blood, and immune diseases; Cardiovascular diseases; Infectious diseases; Neurological disorders; Chronic respiratory diseases; etc. | total per year | National health care spending for specific aggregated health categories | Disease types | 10.1001/jama.2020.0734 |
| US | 1996-2016 | Respiratory infections and tuberculosis and Chronic respiratory diseases | total per year | Total treatment healthcare spending on all respiratory conditions | Disease subtypes | 10.1164/rccm.202202-0294OC |

1282

### 8.4 Results of Case-based COI Evidences of Representative CSDs and Countries

We select five representative CSDs: COPD, stroke, diabetes mellitus, lower respiratory infections, and malaria, spanning chronic and acute conditions across major pathophysiological categories. Mainly focused countries/regions include the United States, European Union, China, India, and Ethiopia, collectively covering all World Bank income groups and major WHO regions. We analyze 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". Numerical values of specific cost types are summarized using medians and quartiles rather than arithmetic means, as these robust statistics better accommodate the right-skewed distributions of COI data while providing interpretable interval estimates (Drummond et al., 2005).

For each type of cost set (median + quartiles), the "Evidence" is composed of two indicators: the number of costs used to calculate the median and quartiles, and the number of publications from which these costs are sourced. In Figure 3 and Figure 4 of the main text, these indicators are represented by two numbers in the first row of each grid, separated by a semicolon. In Figure 3 of the main text, the Evidence data related to the costs of representative diseases in several countries by insurance categories are shown in Table S6, with (a) indicating the sample size, (b) indicating the count of publications, and (c) indicating the publication DOIs list; The Evidence data corresponding to costs by treatment categories are shown in Table S7, with (a) indicating the sample size, (b) indicating the count of publications, and (c) indicating the publication DOIs list.

Table S6: Evidence data for costs by insurance categories in Figure 3 of main text

(a) Sample size

| | United States | | EU countries | | China | | India | | Ethiopia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket |
| COPD | 24 | 13 | 189 | 16 | 45 | 6 | | | | |
| Stroke | 3 | | 22 | | 76 | 59 | 1 | | | |
| Diabetes mellitus | 14 | 21 | 64 | | 149 | 202 | | 132 | | 26 |
| Lower respiratory Infections | | | 48 | 85 | 4 | 16 | | 62 | | 27 |
| Malaria | | | | | 4 | 19 | | 104 | | 12 |

(b) Count of publications

| | United States | | EU countries | | China | | India | | Ethiopia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket |
| COPD | 3 | 2 | 11 | 2 | 4 | 3 | | | | |
| Stroke | 1 | | 8 | | 6 | 3 | 1 | | | |
| Diabetes mellitus | 5 | 3 | 7 | | 9 | 6 | | 6 | | 4 |
| Lower respiratory Infections | | | 6 | 3 | 3 | 4 | | 4 | | 2 |
| Malaria | | | | | 1 | 2 | | 2 | | 3 |

(c) Publication DOIs

| | United States | | EU countries | | China | | India | | Ethiopia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket |
| COPD | 10.2147/ceor.s80424;<br>10.18553/jmcp.2015.21.<br>7.575;<br>10.5888/pcd14.160333 | 10.1371/journal.pone.02<br>22539;<br>10.1097/jom.00000000<br>00001757 | 10.2147/copd.s123095;<br>10.2147/copd.s148051;<br>10.1016/j.rmr.2014.10.7<br>31; 10.1007/s40258-<br>018-0431-5; | 10.2147/copd.s148051;<br>10.1007/s40258-018-<br>0431-5 | 10.1186/s12890-024-<br>03368-0;<br>10.1186/s12890-017-<br>0571-7;<br>10.2147/copd.s147968; | 10.2147/copd.s118523;<br>10.1186/s12889-023-<br>15096-x;<br>10.1186/s12877-023-<br>04087-7 | | | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 10.1186/s12913-016-1371-0; 10.1080/13696998.2017.1279620; 10.1177/0269216316685029; 10.2147/copd.s236787; 10.1016/j.rmed.2020.106194; 10.2147/copd.s375190; 10.1371/journal.pone.0320949 | | 10.1186/s12877-023-04087-7 | | | | | |
| Stroke | 10.1080/0361073x.2017.1276376 | 10.1016/j.jval.2017.04.018; 10.1016/j.rehab.2023.101775; 10.3390/brainsci11060689; 10.3389/fcvm.2024.1324537; 10.3390/healthcare12141369; 10.1016/j.jval.2017.04.018; 10.3389/fpubh.2020.605919; 10.1016/j.rehab.2023.101775 | 10.34172/ijhpm.2020.111; 10.1136/bmjopen-2024-085222; 10.2337/dc19-2137; 10.3389/fphar.2020.596183; 10.3390/brainsci1102016 | 10.34172/ijhpm.2020.111; 10.3389/fphar.2020.596183; 10.1136/bmjopen-2017-020647 | 10.1080/14737167.2019.1580574 | | | | | |
| Diabetes mellitus | 10.1136/bmjdrc-2017-000471; 10.1016/j.hlpt.2020.01.005; 10.1016/j.avsg.2022.05.002; 10.2337/dc20-2833; 10.1371/journal.pone.0222539 | 10.2337/dc20-2833; 10.1371/journal.pone.0222539; 10.1161/jaha.119.015523 | 10.1155/2015/520759; 10.1007/s13300-020-00868-0; 10.1186/s12902-020-0546-1; 10.1177/23969873251319172; 10.1186/s12913-024-10840-4 | 10.1177/1010539515573831; 10.1371/journal.pone.0250200; 10.1136/bmjopen-2024-085222; 10.1186/s12889-025-21705-8; 10.1097/mlr.0000000000000288; 10.1038/s41598-020-80753-9; 10.1371/journal.pone.01 | 10.1177/1010539515573831; 10.1371/journal.pone.0250200; 10.1371/journal.pone.0159297; 10.1007/s40615-018-0488-8; 10.1136/bmjopen-2017-020647; 10.1111/jdi.12905 | 10.1186/s41256-023-00293-3; 10.1111/iwj.14552; 10.1136/bmjopen-2019-036317; 10.1016/j.vhri.2017.05.003; 10.1016/j.hlpt.2023.100807; 10.7759/cureus.70806 | 10.1371/journal.pone.0245839; 10.1136/bmjopen-2019-036317; 10.1186/s12913-022-08819-0; 10.1186/s12913-025-12952-x | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | 59297; 10.2337/dc19-2137; 10.1111/jdi.12905 | | | | | |
| Lower respiratory Infections | 10.1186/s12889-015-1885-0; 10.1186/s12889-019-7458-x; 10.1080/21645515.2017.1264829; 10.1007/s10198-015-0708-7; 10.1016/j.rmed.2018.07.001; 10.1186/s12879-024-09446-2 | 10.1186/s12889-015-1885-0; 10.1016/j.rmed.2018.07.001; 10.5114/aoms.2020.9906 | 10.3390/ijerph18126277; 10.1038/s41598-019-40206-4; 10.3390/jof7110922 | 10.3390/ijerph18126277; 10.1136/bmjpo-2023-002031; 10.3389/fpubh.2024.1364854; 10.1038/s41598-019-40206-4 | 10.1016/j.cegh.2017.12.007; 10.1093/infdis/jiab329; 10.1371/journal.pone.0278025; 10.1186/s12887-015-0510-9 | 10.4102/sajid.v34i1.109; 10.1371/journal.pmed.1004198 | |
| Malaria | | | 10.3389/fpubh.2022.994529 | 10.1186/s40249-016-0141-x; 10.3389/fpubh.2022.994529 | 10.1016/j.actatropica.2018.12.003; 10.1016/j.cegh.2021.100887 | 10.1371/journal.pone.0185315; 10.2471/blt.18.226688; 10.1186/s12936-021-03996-5 | |

Table S7: Evidence data for costs by treatment categories in Figure 4 of main text

(a) Sample size

| | | COPD | Stroke | Diabetes mellitus | Lower respiratory infections | Malaria |
|---|---|---|---|---|---|---|
| United States | Outpatient care | 28 | 1 | 18 | 30 | |
| | Inpatient care | 23 | 1 | 18 | 42 | |
| | Pharmaceuticals | 20 | | 12 | 32 | |
| EU countries | Outpatient care | 321 | 20 | 44 | 46 | |
| | Inpatient care | 144 | 180 | 31 | 131 | |
| | Pharmaceuticals | 188 | | 2 | 18 | |
| China | Outpatient care | 91 | 9 | 218 | 33 | 18 |
| | Inpatient care | 123 | 172 | 68 | 47 | 146 |
| | Pharmaceuticals | 51 | 40 | 99 | 10 | 4 |
| India | Outpatient care | 1 | | 139 | 18 | 51 |
| | Inpatient care | | 22 | 81 | 162 | 8 |
| | Pharmaceuticals | | | 3 | 1 | 12 |
| Ethiopia | Outpatient care | | | 29 | 16 | 9 |
| | Inpatient care | | | | 12 | |
| | Pharmaceuticals | | | 6 | | 3 |
| (b) Count of publications | | | | | | |
| United States | Outpatient care | 5 | 1 | 4 | 1 | |
| | Inpatient care | 5 | 1 | 5 | 3 | |
| | Pharmaceuticals | 3 | | 4 | 2 | |
| EU countries | Outpatient care | 16 | 3 | 4 | 3 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Inpatient care | 16 | 8 | 4 | 8 | |
| | Pharmaceuticals | 15 | | 1 | 6 | |
| China | Outpatient care | 8 | 3 | 9 | 3 | 1 |
| | Inpatient care | 10 | 8 | 6 | 9 | 3 |
| | Pharmaceuticals | 4 | 1 | 5 | 2 | 1 |
| India | Outpatient care | 1 | | 6 | 2 | 3 |
| | Inpatient care | | 1 | 3 | 5 | 2 |
| | Pharmaceuticals | | | 2 | 1 | 2 |
| Ethiopia | Outpatient care | | | 4 | 2 | 3 |
| | Inpatient care | | | | 2 | |
| | Pharmaceuticals | | | 2 | | 2 |
| (c) Publication DOIs | | | | | | |
| United States | Outpatient care | 10.2147/ceor.s80424; 10.1371/journal.pone.0152618; 10.1371/journal.pone.0222539; 10.2147/copd.s220009; 10.2147/copd.s513573 | 10.1080/0361073x.2017.1276376 | 10.3111/13696998.2014.969434; 10.1016/j.jacc.2017.12.064; 10.1016/j.jdiacomp.2014.10.013; 10.1136/bmjdrc-2017-000471; 10.1371/journal.pone.0222539 | 10.1097/inf.0000000000001693 | |
| | Inpatient care | 10.2147/copd.s134618; 10.1371/journal.pone.0222539; 10.2147/copd.s220009; 10.2147/copd.s513573; 10.2147/copd.s446696 | 10.1080/0361073x.2017.1276376 | 10.1016/j.jacc.2017.12.064; 10.1016/j.jdiacomp.2014.10.013; 10.2147/ceor.s117200; 10.1136/bmjdrc-2017-000471; 10.1371/journal.pone.0222539; 10.1016/j.hlpt.2020.01.005 | 10.1097/inf.0000000000001693; 10.1007/s41030-021-00169-2; 10.1007/s10900-020-00886-w | |
| | Pharmaceuticals | 10.1371/journal.pone.0222539; 10.2147/copd.s220009; 10.2147/copd.s513573 | 10.1016/j.jacc.2017.12.064; 10.1016/j.jdiacomp.2014.10.013; 10.2147/ceor.s117200; 10.2337/dc20-2833; 10.1371/journal.pone.0222539 | 10.1097/inf.0000000000001693; 10.1007/s41030-021-00169-2 | | |
| EU countries | Outpatient care | 10.2147/copd.s141852; 10.1186/s12931-017-0543-8; 10.2147/ceor.s77504; | 10.1155/2015/520759; 10.1016/j.rehab.2023.101775; 10.3390/healthcare12 | 10.1136/bmjopen-2016-012527; 10.1155/2015/520759; 10.1136/bmjopen- | 10.1186/s12889-015-1885-0; 10.1016/j.rmed.2018.07.001; | |

| | | | | |
|---|---|---|---|---|
| | 10.2147/copd.s148051; 10.1038/npjpcrm.2015.82; 10.1016/j.rmr.2014.10.731; 10.1007/s40258-018-0431-5; 10.2147/copd.s167357; 10.1186/s12913-016-1371-0; 10.1080/13696998.2017.1279620; 10.2147/copd.s222581; 10.2147/copd.s149633; 10.3390/healthcare7010035; 10.1080/15412555.2020.1868420; 10.2147/copd.s310319; 10.1186/s12913-022-07778-w 10.2147/copd.s141852; 10.1186/s12931-017-0543-8; 10.2147/ceor.s77504; 10.1038/s41533-018-0101-y; | 141369; 10.1177/23969873251319172; 10.3390/brainsci11060689 | 2017-019864; 10.1186/s12913-024-10840-4 | 10.5114/aoms.2020.99060 |
| Inpatient care | 10.1016/j.rmr.2014.10.731; 10.1007/s40258-018-0431-5; 10.2147/copd.s167357; 10.1186/s40248-016-0045-4; 10.1080/13696998.2017.1279620; 10.2147/copd.s222581; 10.1177/0269216316685029; 10.1016/j.rmed.2023.107477; 10.2147/copd.s236787; 10.2147/copd.s149633; 10.3390/healthcare7010035; 10.2147/copd.s310319 10.2147/copd.s141852; 10.1186/s12931-017-0543-8; | 10.1016/j.jval.2017.04.018; 10.3389/fpubh.2020.605919; 10.1016/j.rehab.2023.101775; 10.3390/brainsci11060689; 10.1007/s40261-024-01365-z; 10.3390/healthcare12141369; 10.1177/23969873251319172; 10.1371/journal.pone.0174861 | 10.1136/bmjopen-2016-012527; 10.1136/bmjopen-2017-019864; 10.1007/s13300-020-00868-0; 10.1186/s12913-024-10840-4 | 10.1186/s12889-019-7458-x; 10.1080/21645515.2017.1264829; 10.1016/j.anpedi.2016.10.016; 10.1111/irv.12962; 10.1186/s12879-024-09446-2; 10.5114/aoms.2020.99060; 10.1080/21645515.2024.2440206; 10.2147/copd.s92133 |
| Pharmaceuticals | 10.2147/ceor.s77504; 10.1038/s41533-018-0101-y; 10.2147/copd.s123095; 10.2147/copd.s148051; 10.1016/j.rmr.2014.10.731; 10.1007/s40258-018-0431-5; 10.2147/copd.s167357; 10.1080/13696998.2017.1279620; 10.2147/copd.s31031 | 10.1136/bmjopen-2016-012527 | 10.1186/s12889-015-1885-0; 10.1186/s12889-019-7458-x; 10.1007/s10198-015-0708-7; 10.1016/j.rmed.2018.07.001; 10.1016/j.anpedi.2016.10.016; 10.1016/j.healthpol.2024.105213 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | 9; 10.1186/s12913-022-07778-w; 10.2147/copd.s295388; 10.2147/copd.s297943; 10.1371/journal.pone.0320949 | | | | |
| China | Outpatient care | 10.2147/copd.s524647; 10.3390/jcm11071893; 10.3390/medicina59020391; 10.1186/s12890-024-03368-0; 10.1186/s12955-015-0241-5; 10.2147/copd.s118523; 10.2147/copd.s147968; 10.1186/s12889-023-15096-x | 10.12659/msm.913977; 10.1016/j.puhe.2018.07.024; 10.3390/brainsci11020161 | 10.1371/journal.pone.0250200; 10.1136/bmjopen-2024-085222; 10.1097/mlr.0000000000000288; 10.1371/journal.pone.0159297; 10.1016/j.puhe.2018.07.024; 10.1007/s40615-018-0488-8; 10.1111/jdi.12905; 10.1038/s41598-020-80753-9; 10.2337/dc19-2137 | 10.3390/ijerph18126277; 10.1016/j.vaccine.2020.12.075; 10.1080/21645515.2021.1996151 | 10.1186/s40249-016-0141-x |
| | Inpatient care | 10.2147/copd.s243595; 10.2147/copd.s524647; 10.1097/md.0000000000042274; 10.3390/healthcare10040721; 10.3390/jcm11071893; 10.3390/medicina59020391; 10.2147/copd.s118523; 10.2147/copd.s147968; 10.1186/s12889-023-15096-x; 10.1186/s12877-023-04087-7 | 10.1016/j.jstrokecerebrovasdis.2014.12.021; 10.12659/msm.913977; 10.34172/ijhpm.2020.111; 10.7189/jogh.12.11012; 10.3389/fpubh.2024.1403196; 10.3389/fpubh.2024.1442171; 10.3389/fphar.2020.596183; 10.3390/brainsci11020161 | 10.1177/1010539515573831; 10.1097/mlr.0000000000000288; 10.1371/journal.pone.0159297; 10.1007/s40615-018-0488-8; 10.1111/jdi.12905; 10.1038/s41598-020-80753-9 | 10.1111/irv.12405; 10.3390/ijerph18126277; 10.1136/bmjpo-2023-002031; 10.3389/fpubh.2024.1364854; 10.3390/jof7110922; 10.1007/s10096-018-03459-x; 10.1016/j.vaccine.2020.12.075; 10.1080/21645515.2021.1996151; 10.1186/s12879-024-10048-1 | 10.1186/s40249-016-0141-x; 10.1186/s12936-017-1934-5; 10.3389/fpubh.2022.994529 |
| | Pharmaceuticals | 10.2147/copd.s524647; 10.1186/s12890-017-0571-7; 10.2147/copd.s118523; 10.1186/s12889-023-15096-x | 10.1136/bmjopen-2024-085222 | 10.1136/bmjopen-2024-085222; 10.1186/s12889-025-21705-8; 10.1097/mlr.0000000000000288; 10.1371/journal.pone.0159297; 10.1038/s41598-020-80753-9 | 10.3390/ijerph18126277; 10.3390/jof7110922 | 10.1186/s40249-016-0141-x |
| India | Outpatient care | 10.4103/0970-2113.192878 | 10.22159/ajpcr.2017.v10i4.8510; 10.1186/s41256-023-00293-3; 10.7759/cureus.70806; 10.1111/iwj.14552; 10.1136/bmjopen-2019-036317; 10.1016/j.vhri.2017.05.003; 10.1016/j.hlpt.2023.100807 | 10.1186/s12887-015-0510-9; 10.1371/journal.pone.0278025 | 10.1016/j.actatropica.2018.12.003; 10.1016/j.cegh.2021.100887; 10.4269/ajtmh.23-0830 | |
| | Inpatient care | | 10.1080/14737167.2019.1580574 | 10.1186/s41256-023-00293-3; 10.1111/iwj.14552; 10.1016/j.vhri.2017.0 | 10.1186/s12887-015-0510-9; 10.1016/j.cegh.2017.12.007; | 10.1016/j.actatropica.2018.12.003; 10.1016/j.cegh.2021.100887 |

| | | | | |
|---|---|---|---|---|
| | | 5.003; 10.1016/j.hlpt.2023.100807 | 10.1093/infdis/jiab329; 10.3390/medsci5040033; 10.1136/bmjph-2024-001727 | |
| | Pharmaceuticals | 10.1016/j.vhri.2017.05.003; 10.1016/j.hlpt.2023.100807 | 10.3390/medsci5040033 | 10.1016/j.actatropica.2018.12.003; 10.1002/jppr.1452 |
| Ethiopia | Outpatient care | 10.1371/journal.pone.0245839; 10.1136/bmjopen-2019-036317; 10.1186/s12913-022-08819-0; 10.1186/s12913-025-12952-x | 10.4102/sajid.v34i1.109; 10.1371/journal.pmed.1004198 | 10.1371/journal.pone.0185315; 10.2471/blt.18.226688; 10.1186/s12936-021-03996-5 |
| | Inpatient care | | 10.4102/sajid.v34i1.109; 10.1371/journal.pmed.1004198 | |
| | Pharmaceuticals | 10.1371/journal.pone.0245839; 10.1186/s12913-022-08819-0 | 10.1371/journal.pone.0185315; 10.2471/blt.18.226688 | |

For each cost set (median + quartiles), "Agreement" primarily indicates the degree of dispersion within the series of costs used to calculate the median and quartiles. The smaller the degree of dispersion, the higher the Agreement. We express it using the reciprocal of RIQR, where a larger value indicates higher Agreement. The calculation for RIQR is:

$$RIQR = \frac{Q_3 - Q_1}{|Median|} \tag{2}$$

Where $Q_1$ and $Q_3$ are lower quartile and upper quartile, respectively. In Figure 3 and Figure 4 of the main text, the Agreement is expressed by the background color of each grid. The redder the background color, the higher the Agreement; the bluer, the lower the Agreement; and grey represents no data. The Agreement data related to the costs of representative diseases in several countries by insurance categories is shown in Table S8; the Agreement data corresponding to the costs by treatment categories is shown in Table S9.

Table S8: Agreement data for costs by insurance categories in Figure 3 of main text

| | United States | | EU countries | | China | | India | | Ethiopia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket |
| COPD | 1.31 | 1.88 | 0.57 | 1.18 | 0.16 | 0.64 | | | | |
| Stroke | 1.07 | | 0.07 | | 1.37 | 0.81 | 3.40 | | | |
| Diabetes | 1.38 | 0.57 | 0.17 | | 1.05 | 1.19 | | 0.42 | | 0.27 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| mellitus | | | | | | | | |
| Lower respiratory Infections | 0.02 | 0.75 | 0.08 | 0.13 | | 0.34 | | 0.33 |
| Malaria | | | 3.40 | 0.21 | | 0.18 | | 0.61 |

Table S9: Agreement data for costs by treatment categories in Figure 4 of main text

| | | COPD | Stroke | Diabetes mellitus | Lower respiratory infections | Malaria |
|---|---|---|---|---|---|---|
| United States | Outpatient care | 0.40 | 5.00 | 0.18 | 1.97 | |
| | Inpatient care | 1.10 | 5.00 | 0.29 | 0.06 | |
| | Pharmaceuticals | 1.14 | | 0.29 | 1.97 | |
| EU countries | Outpatient care | 0.17 | 0.06 | 0.03 | 0.62 | |
| | Inpatient care | 0.50 | 0.17 | 0.48 | 0.32 | |
| | Pharmaceuticals | 0.66 | | 4.97 | 0.77 | |
| China | Outpatient care | 0.67 | 1.39 | 0.52 | 0.66 | 0.73 |
| | Inpatient care | 0.83 | 0.71 | 1.04 | 0.35 | 0.26 |
| | Pharmaceuticals | 0.23 | 1.00 | 1.19 | 1.26 | 0.74 |
| India | Outpatient care | 5.00 | | 0.24 | 0.42 | 0.25 |
| | Inpatient care | | 1.14 | 0.53 | 0.26 | 0.60 |
| | Pharmaceuticals | | | 0.28 | 5.00 | 0.02 |
| Ethiopia | Outpatient care | | | 0.08 | 1.65 | 0.42 |
| | Inpatient care | | | | 1.08 | |
| | Pharmaceuticals | | | 0.44 | | 0.45 |

For each cost set (median + quartiles), the "Credibility" score is calculated through averaging Credibility scores of all publications supporting this specific type of cost, and then the average number is rounded off. Therefore, the final Credibility score for the cost set is between 1-5. For example, the average Credibility score of several publications related to the inpatient cost of COPD in EU is 3.25, and is rounded to 3 finally. The average Credibility score (unrounded) for each type of cost is shown in Table S10 and Table S11. In the main text, several plus signs ("+") are used to represent the final Credibility score.

Table S10: Average Credibility score for each insurance category of cost

| | United States | | EU countries | | China | | India | | Ethiopia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket | Compulsory insurance | Out-of-pocket |
| COPD | 4 | 3 | 3.27 | 2.5 | 3.25 | 2.67 | | | | |
| Stroke | 2 | | 3.25 | | 3.75 | 3.33 | 2 | | | |
| Diabetes mellitus | 3.2 | 3 | 2.8 | | 3.67 | 2.83 | | 2.67 | | 2.5 |
| Lower respiratory Infections | | | 3.33 | 4 | 3.33 | 3.25 | | 2.75 | | 2 |
| Malaria | | | | | 3 | 3 | | 3.5 | | 2.67 |

Table S11: Average Credibility score for each treatment category of cost

| | | COPD | Stroke | Diabetes mellitus | Lower respiratory infections | Malaria |
|---|---|---|---|---|---|---|
| United States | Outpatient care | 3.4 | 2 | 4 | 4 | |
| | Inpatient care | 3.4 | 2 | 3.5 | 3 | |
| | Pharmaceuticals | 3.33 | | 3.4 | 3.5 | |
| EU countries | Outpatient care | 2.94 | 2.8 | 2.5 | 3.33 | |
| | Inpatient care | 3.13 | 3.38 | 2.5 | 3.25 | |
| | Pharmaceuticals | 3.2 | | 3 | 3.33 | |
| China | Outpatient care | 2.38 | 2.67 | 3.44 | 3.67 | 3 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Inpatient care | 2.6 | 3 | 3.17 | 3.11 | 2.67 |
| | Pharmaceuticals | 2.75 | 4 | 4 | 3 | 3 |
| India | Outpatient care | 4 | | 2.57 | 3 | 3 |
| | Inpatient care | | 2 | 2.5 | 2.6 | 3.5 |
| | Pharmaceuticals | | | 2.5 | 2 | 2.5 |
| Ethiopia | Outpatient care | | | 2 | 2.5 | 2 |
| | Inpatient care | | | | 2 | |
| | Pharmaceuticals | | | 2 | | 2.5 |

# References


Agache, I. et al. An overview of climate change and its effects on health: from mechanisms to one health. *Journal of Allergy and Clinical Immunology: In Practice* **13**, 253–264 (2025).

Bell, J. E. et al. Changes in extreme events and the potential impacts on human health. *Journal of the Air and Waste Management Association* **68**, 265–287 (2018).

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 Medical Research Methodology* **25**, 116 (2025).

Cissé, G. et al. Health, wellbeing, and the changing structure of communities. In: *Climate change 2022: impacts, adaptation and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2022).

Cotton, C. S. et al. Canadian public health experiences during COVID-19: a new framework for assessing evidence. *Frontiers in Public Health* **13**, 1620514 (2025).

Danhauser, K. et al. Using large language models to extract information from pediatric clinical reports. *PLOS Digital Health* **4**, e0000919 (2025).

Dieleman, J. L. et al. US health care spending by payer and health condition, 1996–2016. *JAMA* **323**, 863–884 (2020).

Ding, E. & Yang, W. Focus on and response to climate-sensitive diseases (in Chinese). *Chinese Journal of Preventive Medicine* **58**, 1940–1948 (2024). 丁恩民 & 杨维中. 关注和应对气候敏感疾病. *中华预防医学杂志* **58**, 1940–1948 (2024).

Drummond, M. E. et al. *Methods for the Economic Evaluation of Health Care Programmes*, 3rd Edition (Oxford University Press, 2005).

Gao, C. et al. A large-scale database for clinical trial outcomes and features. *Nature Health* **1**, 428–

436 (2026).

Global Burden of Disease Collaborative Network. *Global burden of disease study 2021 (GBD 2021) causes of death and nonfatal causes mapped to ICD codes* (Institute for Health Metrics and Evaluation, 2024) https://ghdx.healthdata.org/record/ihme-data/gbd-2021-cause-icd-code-mappings.

Gupta, S. et al. Data extraction from polymer literature using large language models. *Communications Materials* **5**, 269 (2024).

Haines, A. & Ebi, K. The imperative for climate action to protect health. *New England Journal of Medicine* **380**, 263–273 (2019).

Huang, C. & Liu, Q. Interpretation of IPCC AR6 on climate change and human health (in Chinese). *Climate Change Research* **18**, 442–451 (2022). 黄存瑞 & 刘起勇. IPCC AR6 报告解读：气候变化与人类健康. *气候变化研究进展* **18**, 442–451 (2022).

Hunter, R. F. et al. The emerging syndemic of climate change and non-communicable diseases. *Lancet Planetary Health* **8**, e430–e431 (2024).

Jaskulski, P. et al. Reliability of large language models as a tool for knowledge extraction from biographical dictionaries: the case of the Polish Biographical Dictionary. *Digital Scholarship in the Humanities* **40**, 538–548 (2025).

Jo, C. Cost-of-illness studies: concepts, scopes, and methods. *Clinical and Molecular Hepatology* **20**, 327–337 (2014).

Jönsson, L. et al. The costs of dementia in Europe: an updated review and meta-analysis. *PharmacoEconomics* **41**, 59–75 (2023).

Kause, A. et al. Confidence levels and likelihood terms in IPCC reports: a survey of experts from different scientific disciplines. *Climatic Change* **173**, 2 (2022).

Kneale, D. et al. The use of evidence in English local public health decision-making: a systematic scoping review. *Implementation Science* **12**, 53 (2017).

Liu, N. F. et al. Lost in the middle: how language models use long contexts. Preprint at: *arXiv* https://arxiv.org/abs/2307.03172 (2023).

Mastrandrea, M. D. et al. *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties.* Intergovernmental Panel on Climate Change https://www.ipcc.ch/site/assets/uploads/2017/08/AR5_Uncertainty_Guidance_Note.pdf (2010).

Mora, C. et al. Over half of known human pathogenic diseases can be aggravated by climate change. *Nature Climate Change* **12**, 869–875 (2022).

Ribeiro de Faria, J., Xie, H. & Steffek, F. Information extraction from employment tribunal judgments using a large language model. *Artificial Intelligence and Law*, (2025).

Shi, X. et al. Evidence triangulator: using large language models to extract and synthesize causal evidence across study designs. *Nature Communications* **16**, 7355 (2025).

Sontheimer, N., Konnopka, A. & König, H. H. The excess costs of dementia: a systematic review and meta-analysis. *Journal of Alzheimer's Disease* **83**, 333–354 (2021).

Wieser, S. et al. How much does the treatment of each major disease cost? a decomposition of Swiss National Health Accounts. *European Journal of Health Economics* **19**, 1149–1161 (2018).

Wu, S. et al. Automated literature research and review-generation method based on large language models. *National Science Review* **12**, nwaf169 (2025).

Wubulihasimu, P. et al. Trends in Dutch hospital spending by age and disease 1994–2010. *Health Policy* **119**, 316–323 (2015).

Yenew, C. et al. Scoping review on assessing climate-sensitive health risks. *BMC Public Health* **25**, 914 (2025).

Zhang, J. et al. Extracting multifaceted characteristics of patients with chronic disease comorbidity: framework development using large language models. *JMIR Medical Informatics* **13**, e70096 (2025).

Zheng, Z. et al. ChatGPT chemistry assistant for text mining and the prediction of MOF synthesis. *Journal of the American Chemical Society* **145**, 18048–18062 (2023).