SWAR 69: AI-assisted data extraction with large language models in climate change and infectious diseases

Objective of this SWAR
(1) To develop a structured data extraction prompting framework applicable to three large language models for climate change and infectious diseases, including ChatGPT, Claude, and DeepSeek
(2) To evaluate the efficiency and accuracy of ChatGPT, Claude, and DeepSeek, and analyze potential error classification for LLM-assisted workflows
(3) To dynamically update the prompting framework based on performance feedback.

Study area: Statistical Analysis, Data extraction, Climate Change
Sample type: Publications
Estimated funding level needed: Low

Background
Climate change has emerged as an urgent challenge to human health in the 21st century.[1] The burden of disease associated with extreme weather exacts a massive toll on human health when climate drivers exceed critical thresholds. The World Health Organization global air quality guidelines provide evidence-informed threshold recommendations, including Air Quality Guideline (AQG) levels for pollutants such as PM2.5 and PM10 together with the corresponding interim targets. Evidence synthesis is essential in this development of the guideline. In 2014, Elliott et al. proposed the methodology of living evidence (LE) synthesis, which constantly updating a systematic synthesis of evidence to incorporate newly available.[2] LE synthesis involves four key phases: database searching and eligibility assessment, data extraction and risk of bias assessment, evidence synthesis and analysis, and publication update.[3-5] Data extraction is a critical and highly labor-intensive step. Manual data extraction has been found to account for approximately 20% of the total timeline, with an error rate exceeding 20% for single extraction and although double extraction yields a noticeable reduction in error rate, it is at the cost of doubled labour input.[6-9]

Large Language Models (LLMs) such as ChatGPT and Claude have been widely applied through prompt engineering to extract data in clinical medicine, and studies confirm their strong performance range from 90% to 96% accuracy in core data extraction, high test-retest reliability, and approximately 41 minutes saved per study compared with manual work.[10-12] However, there is a lack of research into the application of LLMs for data extraction in climate change and infectious diseases with no targeted studies conducted to date (February 2026) in this specific domain. Therefore, building a dedicated framework for climate change and infectious diseases research, including domain-adapted prompting strategies, specialized terminology matching rules, and standardized LLM-assisted workflows, is essential. This Study Within a Review (SWAR) [13] will help to do this.

The first step will be to compile a dataset which will comprise five systematic reviews focusing on the associations between climate change and infectious diseases conducted by Lanzhou University. The datasets were selected for independent data extraction by two domain experts and a consensus extraction is available, with a domain-adapted structured data extraction framework for dataset construction. This is followed by prompt development using the following three steps: (a) Parallel data extraction of the same sample studies by multiple LLMs using an initial draft of domain-adapted prompts; (b) Comparison of LLM-extracted outputs with the human double-extracted gold standard to identify discrepancies and error patterns; and (c) Iterative refinement and optimization of prompts by domain experts based on identified differences until predefined performance criteria were met. The SWAR will then compare data extraction by three LLMs with the data extraction by two humans.

Interventions and Comparators
Intervention 1: Data extraction using ChatGPT.
Intervention 2: Data extraction using Claude.
Intervention 3: Data extraction using DeepSeek.
Intervention 4: Human double-extracted (gold standard).

Index Type: Data extraction

Method for Allocating to Intervention or Comparator:
Non-Random

Outcome Measures
Primary: Accuracy (Recall & Precision & F1-score); Efficiency (Extraction time & workload ratio); Error classification (Frequency of error types).
Secondary:

Analysis Plans
We will compare the results of LLM-assisted data extraction with the human double-extracted gold standard and calculate consistency, accuracy, efficiency, and frequencies of different error types. All proportion estimates will be accompanied by 95% confidence intervals. The Kruskal-Wallis test will be used to compare performance metrics across the three LLMs, while the Chi-square test will be applied to examine differences in error type distributions. Cohen's Kappa and ICC will be used to assess the agreement between LLM outputs and the human gold standard.

Possible Problems in Implementing This SWAR
None anticipated.

References
1. Romanello M, Walawender M, Hsu SC, et al. The 2024 report of the Lancet Countdown on health and climate change: facing record-breaking threats from delayed action. Lancet 2024;404(10465):1847-96.doi:10.1016/S0140-6736(24)01822-1.
2. Elliott JH, Turner T, Clavisi O, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. PLoS Medicine 2014;11(2):e1001603. doi: 10.1371/journal.pmed.1001603.
3. Shrikhande S, Wolf J, Vert C, et al. World Health Organization repository of systematic reviews on interventions in environment, climate change and health: a new resource for decision makers, intervention implementers, and researchers. Environmental health: a global access science source 2024;23(1):88. doi: 10.1186/s12940-024-01105-y.
4. World Health Organization. (2024, February 29). WHO launches a Repository of systematic reviews on interventions in environment, climate change and health. WHO News Release. https://www.who.int/news/item/29-02-2024-who-launches-a-repository-of-systematic-reviews-on-interventions-in-environment--climate-change-and-health
5. Song X, Lian Z, Wang R, et al. The Phases of Living Evidence Synthesis Using AI: Living Evidence Synthesis (Version 1). Journal of Medical Internet Research 2026;28:e76130.doi:10.2196/76130.
6. Holub K, Hardy N, Kallmes K. Toward Automated Data Extraction According to Tabular Data Structure: Cross-sectional Pilot Survey of the Comparative Clinical Literature. JMIR formative research 2021:5(11):e33124. doi: 10.2196/33124.
7. O'Mara-Eves A, Thomas J, McNaught J, et al. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Systematic reviews 2015;4(1):5. doi: 10.1186/2046-4053-4-5.
8. Schmidt L, Finnerty Mutlu AN, Elmore R, et al. Data extraction methods for systematic review (semi)automation: Update of a living systematic review. F1000Research 2021;10:401. doi: 10.12688/f1000research.51117.3.
9. Mathes T, Klaßen P, Pieper D. Frequency of data extraction errors and methods to increase data extraction quality: a methodological review. BMC medical research methodology 2017;17(1):152. doi: 10.1186/s12874-017-0431-4.

10. Andersen TH, Marcussen TM, Termannsen AD, et al. Using artificial intelligence tools as second reviewers for data extraction in systematic reviews: A performance comparison of two AI tools against human reviewers. Cochrane Evidence Synthesis Methods 2025;3(4):e70036. doi: 10.1002/cesm.70036.
11. Gartlehner G, Kahwati L, Hilscher R, et al. Data extraction for evidence synthesis using a large language model: A proof of concept study. Research Synthesis Methods 2024;15(4):576589. doi:10.1002/jrsm.1710.
12. Gartlehner G, Kugley S, Crotty K, et al. Artificial Intelligence-Assisted Data Extraction With a Large Language Model: A Study Within Reviews. Ann Intern Med. 2025;178(12):1763-1771. doi:10.7326/ANNALS-25-00739.
13. Devane D, Burke NN, Treweek S, et al. Study within a review (SWAR). Journal of Evidence-Based Medicine 2022;15(4):328-32.doi:10.1111/jebm.12505.

Publications or presentations of this SWAR design

Examples of the implementation of this SWAR

People to show as the source of this idea: Wenlu Cai
Contact email address: cwenlu2025@lzu.edu.cn
Date of idea: 01/02/2026
Revisions made by: Wenlu Cai
Date of revisions: 15/02/2026