

Case Study Summary

Title: Scientific sufficiency of causal claims: A practical approach

Version: Draft for ARA Workshop XVI, September 2026

Presented by: Louis Anthony (Tony) Cox Jr., Cox Associates

Panel Advisor: To be assigned

1. Provide a few sentences summarizing the method illustrated by the case study.

The case study illustrates a structured method for evaluating whether scientific evidence is sufficient to support stated causal conclusions about health risks. The method begins with five broad domains of necessary conditions for valid causal inference from observational evidence: (A) conceptual and definitional clarity; (B) study-design prerequisites; (C) analytic conditions for internal validity; (D) sound interpretation; and (E) robustness, generalizability, and applicability to the target population and conditions.

The practical implementation is an AI-assisted Scientific Paper review system (ASP). ASP identifies explicit and implied causal claims in a document, applies modular methodological review questions to the supporting evidence, and produces linked evidence bundles that preserve the path from source text to methodological judgments and higher-level conclusions. Its intended role is to assist, not replace, expert review by making the inferential chain more systematic, transparent, inspectable, and auditable.

2. Describe the problem formulation(s) the case study is designed to address. How is the method described in the case useful for addressing the problem formulation?

Risk assessments and evidence syntheses often combine observational epidemiology, systematic reviews, mechanistic or toxicological evidence, and policy-facing interpretations. In practice, the steps from observed association to causal interpretation and then to predictions about the effects of interventions may be incompletely specified or difficult to audit.

The case study asks a practical question: What evidence and assumptions are needed before a causal claim is scientifically well supported? The five-domain framework provides a structured way to formulate and evaluate that question. ASP operationalizes the framework by extracting causal claims, asking pre-specified methodological questions, identifying assumptions and limitations, and linking conclusions back to supporting passages. This can help expert reviewers distinguish observed associations, causal interpretations, mechanistic plausibility, and intervention or policy claims.

3. Comment on whether the method is general enough to be used directly, or if it can be extrapolated, for application to other chemicals and/or problem formulations. Please explain why or why not.

The underlying method is intended to be broadly applicable because it evaluates the logic and methodological support for causal claims rather than depending on a chemical-specific model. The same five domains can be applied to different chemicals, exposures, endpoints, study designs, and

evidence streams.

ASP is modular: topic-specific or method-specific question libraries can be added or revised while retaining the same basic architecture. Applications described in the accompanying papers include vinyl chloride, glyphosate, alcohol and coronary heart disease, ultra-processed foods, and air-pollution-related evidence. Chemical-specific biological and toxicological knowledge remains essential, and human expert review is needed to judge context, relevance, and scientific plausibility.

4. Discuss the overall strengths and weaknesses of the method.

Strengths include explicit decomposition of causal reasoning; consistent application of methodological questions; traceability from conclusions to source text; the ability to review heterogeneous evidence types; rapid generation of reproducible first-pass reviews; and a structure that makes disagreements and untested assumptions easier to locate and discuss.

Limitations include dependence on the quality and completeness of the source documents; possible errors or inconsistent judgments by large language models; sensitivity to question wording and implementation details; and the need for subject-matter expertise to evaluate biological plausibility, context, and difficult methodological issues. ASP should therefore be treated as a methodological assistant or screening and audit tool, not as an autonomous authority or substitute for expert judgment.

5. Outline the minimum data requirements and describe the types of data sets that are needed.

For methodological review, the minimum input is the full text of a scientific document containing the evidence, analyses, and conclusions to be evaluated. Better review is possible when methods, assumptions, tables, appendices, and supporting materials are available. Collections of papers or evidence reviews can also be evaluated document by document and then compared or synthesized.

ASP does not require access to the underlying individual-level study data to perform a document-based methodological audit. However, some methodological questions cannot be resolved from a publication alone. Reanalysis, quantitative bias analysis, model checking, or independent causal-effect estimation may require the original study data, code, validation data, or external information.

A. Describe the dose-response relationship in the dose range relevant to human exposure?

Not directly. The method is not itself a dose-response model. It can evaluate whether a source document has adequately defined, estimated, tested, and interpreted a dose-response relationship, including whether model form, exposure measurement, relevant dose range, uncertainty, and causal interpretation are adequately supported.

B. Address human variability and sensitive populations?

Yes, as an evaluation question. The framework asks whether the target population is clearly defined, whether heterogeneity and effect modification are addressed where relevant, and whether estimated effects can be generalized or transported to the population of interest. It does not itself generate missing subgroup data.

C. Address background exposures or responses?

Yes, where relevant to the causal question. ASP can prompt review of co-exposures, baseline risks, confounding, measurement issues, and alternative explanations. Whether these issues can be resolved depends on the evidence and data available in the underlying studies.

D. Address incorporation of existing biological understanding of the likely mode of action?

Yes. One ASP module specifically reviews toxicological and mechanistic evidence, including biological pathways, relevance of experimental concentrations or doses, coherence, alternative explanations, and translation across systems, doses, time scales, or species. Biological understanding is used as evidence to be evaluated, not as a substitute for causal identification in human data.

E. Address other extrapolations, if relevant – insufficient data, including duration extrapolations, interspecies extrapolation?

Yes, as part of external-validity and applicability review. The framework asks whether conclusions require extrapolation across populations, settings, doses, durations, species, or other conditions, and whether assumptions supporting those extrapolations are stated and justified. ASP does not itself supply a universally valid extrapolation model.

F. Address uncertainty?

Yes. The framework explicitly asks about uncertainty due to confounding, measurement error, missing data, model specification, unmeasured or latent variables, robustness to alternative analyses, and uncertainty in generalization or transportability. It can also identify key assumptions that have not been tested. Quantitative characterization of uncertainty depends on what information is available in the underlying evidence.

G. Allow the calculation of risk (probability of response for the endpoint of interest) in the exposed human population?

Not by itself. ASP is an evidence-evaluation and causal-methodology review system, not a risk-calculation engine. It can assess whether a quantitative risk estimate is supported by an appropriate causal estimand, study design, analysis, dose-response model, and transportability assumptions, and it can identify gaps that would limit the credibility of such a calculation.

H. Work practically? If the method still requires development, how close is it to practical implementation?

Yes, in proof-of-concept form. Current ASP implementations can review individual papers and collections of documents and generate linked methodological reports, claim inventories, evidence summaries, and audience-specific briefs in minutes. The accompanying papers provide examples and online demonstrations.

Further development is needed for systematic validation across topics, models, and expert reviewers; improved retrieval and checking of cited references; refinement of question libraries; and clearer procedures for human verification and correction. The most appropriate current use is as a rigorous and efficient first step that structures and accelerates expert review.

References

Cox LA Jr. Toward AI-Assisted Methodological Evaluations of Health Risk Analyses Using Diverse Types of Evidence. *Risk Analysis*. 2026;46:e70356. doi:10.1111/risa.70356.

Cox LA Jr. Necessary conditions for valid causal inference from observational data. *Critical Reviews in Toxicology*. 2025;55(10):995-1110. doi:10.1080/10408444.2025.2611827.

Cox LA Jr. AI-Assisted Methodological Review of Documents Making Causal Claims About Health Risks. *Perspective manuscript*, 2026.

Cox LA Jr. An AI assistant to help review and improve causal reasoning in epidemiological documents. *Global Epidemiology*. 2023;7:100130. doi:10.1016/j.gloepi.2023.100130.