

AI-assisted evaluation of causal evidence

From the causal revolution to reusable methodological review and cross-paper synthesis

Tony Cox | tcoxdenver@aol.com

Key points

- Current causal inferences and claims in scientific publication, regulation, litigation, and news are a mess.
 - Different concepts and definitions of causation
 - Evidence of association mixed with evidence of causation
- It is possible to do better now.
 - The “Causal Revolution” has given us conceptual and methodological clarity and effective tools
- AI can help
 - Reduce the learning curve
 - Provide quick, fairly competent, assistance
 - This talk: “Five-domain” framework & software
 - Work in progress: AI as intern, not oracle

Current situation: Room for improvement

Alcohol: biomarker changes → lower CHD risk?

Causal Claim

“Alcohol intake is **causally related** to lower risk of coronary heart disease through changes in lipids and haemostatic factors.”

Supporting Evidence and Rationale

“**On the basis of published associations** between these biomarkers and risk of coronary heart disease **30 g of alcohol a day would cause an estimated reduction of 24.7%** in risk of coronary heart disease.”

Later Observations & Claims

“immediately higher cardiovascular risk following any alcohol consumption”
— Mostofsky et al. (2016)

“protective effects of moderate alcohol intake against stroke are **largely non-causal**”
— Millwood et al. (2019)

Question: Do biomarker changes identify the effect on CHD events?

Gas stoves: observational Odds Ratio → preventable asthma?

Causal Claim

“We found that **12.7% (95% CI = 6.3–19.3%)** of current childhood asthma in the US is **attributable to** gas stove use.”

“The proportion of childhood asthma that could be **theoretically prevented** if gas stove use was not present...”

Supporting Evidence and Rationale

“Effect sizes previously reported by meta-analyses for current asthma (**Odds Ratio = 1.34**, 95% Confidence Interval (CI) = 1.12–1.57) were utilized in the **PAF estimations**.”

Later Observations & Claims

“In the case of gas cooking exposure and childhood asthma, a **causal relationship is not supported** by the epidemiology literature.”
— Li et al. (2024)

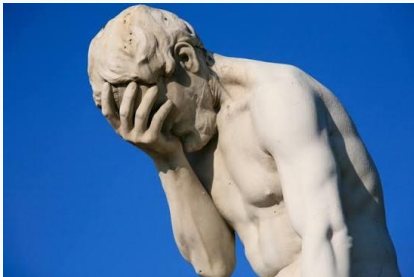
Question: Does this OR identify cases preventable by stove removal?

Screen time: association → preventing ADHD?

Causal Claim

“Our study found a **positive correlation** between screen time and the risk of ADHD. ...

Therefore, it is necessary to reduce screen time per day in children to **prevent** the occurrence of ADHD.”



Supporting Evidence and Rationale

“There were **81,234 children in the nine studies** that made up this meta-analysis...”

“...the OR (95 % CI) value of screen time and ADHD in the screen time ≥ 2 h/d group was **1.51 (1.20–1.90)**.”

Later Observations & Claims

A 2026 meta-analysis reported a pooled cohort association (**OR 1.48**), with **very high heterogeneity ($I^2 = 97.4\%$)**.

“**Both directions** likely operate simultaneously...”
— Zhao et al. (2026)

Question: Does reducing screen time prevent or reduce ADHD?

Screen time and childhood attention deficit hyperactivity disorder: a meta-analysis

Hezuo Liu¹, Xiaorong Chen¹, Minhui Huang¹, Xiaomin Yu², Ye Gan¹, Ji Wang¹, Qiujin Chen³, Zhipeng Nie⁴, Haijing Ge¹

 Summarize this page using Acrobat AI Assistant. X

Affiliations + expand

PMID: 37163581 DOI: [10.1515/reveh-2022-0262](https://doi.org/10.1515/reveh-2022-0262)

Abstract

Objectives: The association between screen time and attention deficit hyperactivity disorder (ADHD) has been controversial. This study sheds light on the contentious correlation between screen time and ADHD.

Content: Until August 2022, electronic searches of the PubMed, Embase, and Web of Science databases were carried out. The combined effect value odds ratios (OR) and 95 % confidence interval (95 % CI) were calculated for the meta-analysis using Stata 12.0. There were 81,234 children in the nine studies that made up this meta-analysis which included 28,997 children with ADHD and 52,237 healthy controls. When compared with the screen time <2 h/d, the OR (95 % CI) value of screen time and ADHD in the screen time ≥2 h/d group was 1.51 (1.20-1.90).

Summary and outlook: Based on the current meta-analysis results, our study found a positive correlation between screen time and the risk of ADHD. Excessive screen exposure may significantly contribute to the development of ADHD in children. Therefore, it is necessary to reduce screen time per day in children to prevent the occurrence of ADHD.

Keywords: attention deficit hyperactivity disorder; children; meta-analysis; screen time.

Conclusions may be true or false, but the reasoning supporting them should be sound, traceable, and independently verifiable.

Large Cohort Studies Find Little-to-No Evidence of Association Between ADHD and Digital Media Screen Time



- Rival causal models drive rival causal claims
- The most valuable evidence is often evidence that discriminates among rival causal models.

What additional inference would each claim require?

Observed findings do not automatically identify the effects of changing exposure.

Alcohol → CHD

EVIDENCE

Short-term changes in lipid and haemostatic biomarkers.

CAUSAL BRIDGE

Do the biomarker changes translate into fewer CHD events?

Gas stoves → asthma

EVIDENCE

Pooled observational OR 1.34; modeled PAF 12.7%.

CAUSAL BRIDGE

Would removing or ventilating stoves reduce asthma in U.S. children?

Screen time → ADHD

EVIDENCE

Positive correlations reported by a meta-analysis.

CAUSAL BRIDGE

Would reducing screen time prevent ADHD—or might ADHD increase screen time?

Challenge: There is a methodological implementation gap

Questionable scientific causal claims abound

Major journals often publish influential exposure–health associations and dubious causal interpretations that may be changed and/or contested later.

Useful methodology exists

Causal estimands, DAGs, study designs, estimation methods, robustness, transportability, and falsification are available.

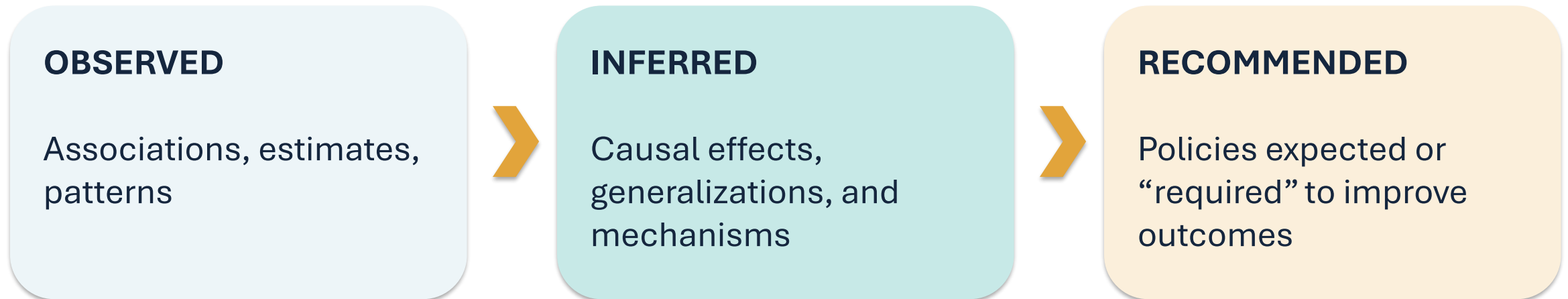
Limited impact on routine causal determinations compared with judgment-based approaches (WoE, Bradford Hill).

Practical opportunity

AI can systematically ask prespecified questions, expose and document support chains, and help humans locate consequential gaps and assumptions.

Research: Can structured AI assistance improve the completeness, accuracy, transparency, and practical usefulness of scientific evaluations of causal claims?

Problem: weak evidence often becomes stronger claims



Each transition calls for additional assumptions and evidence. Reviews often blend the three, obscuring the inferential chain.

Challenge: abundant evidence, weak causal clarity

What do conflicting study findings and policy claims validly establish?

Evidence explosion

RCTs, cohort and case-control studies, mechanistic studies, reanalyses, systematic reviews, and regulatory reports often address different causal hypotheses.

Causal ambiguity

Associations, bias, dose-response, human relevance, and decision implications are often collapsed into one causal conclusion.
Hazard versus Risk?

Expert bottleneck

High-quality synthesis is slow, costly, difficult to audit, and quickly becomes stale as evidence changes.

Proliferation of self-labeled “systematic reviews.”

High-stakes decisions fail when evidence is abundant but causal understanding is weak.

Why conventional review methods may leave gaps in causal understanding

Reporting standards

STROBE improves reporting transparency, but does not adjudicate each substantive claim.

Risk-of-bias / evidence grading

Often centered on whole studies or evidence bodies rather than individual claims.

Heterogeneous evidence

Epidemiology, toxicology, mechanisms, reviews, and policy arguments require different checks.

Narrative synthesis

Evidence, interpretation, and judgment can be hard to disentangle or independently reproduce.

Needed: a claim-centered, source-linked, modular review process.

Sound evidence must be assessed, integrated and tested

What do high-stakes decision-makers need?

What is supported?

Which causal proposition, if any, is actually supported by the evidence?

Under what conditions?

In what populations, for what outcomes, and at what human-relevant exposures does a causal conclusion apply?

How strong is it?

What critical gaps or uncertainties, alternative explanations, and evidence dependencies remain?

What would change the answer?

Which reanalysis, new evidence, or targeted study would improve the next decision? Which could reverse it?

Clear answers to decision-relevant causal questions raise validity and confidence.

Vision: The AI-assisted Scientific Paper-reviewer (ASP)

From literature to decision-ready causal science

1. Ingest

Papers, reviews, mechanisms, exposure-response data and models.

2. Structure

Extract causal claims, designs, assumptions, evidence links.

3. Evaluate

Assess what each study does—and does not—establish about each claim.

4. Synthesize

Combine evidence; expose dependencies and uncertainty.

5. Guide action

Identify robust conclusions and high-value next steps.

ASP keeps propositions, evidence, assumptions, and conclusions distinct.

What makes ASP scientifically different?

A causal-evidence audit rather than another literature summary.

1. Intervention question

What would changing exposure actually change—and under what conditions?

2. Rival models

Make competing causal explanations explicit and derive testable predictions.

3. Five-domain audit

- Conceptual clarity
- study design
- causal analysis
- interpretation
- applicability.

4. Human + AI

AI supplies breadth and traceability; scientists interrogate, correct, and refine.

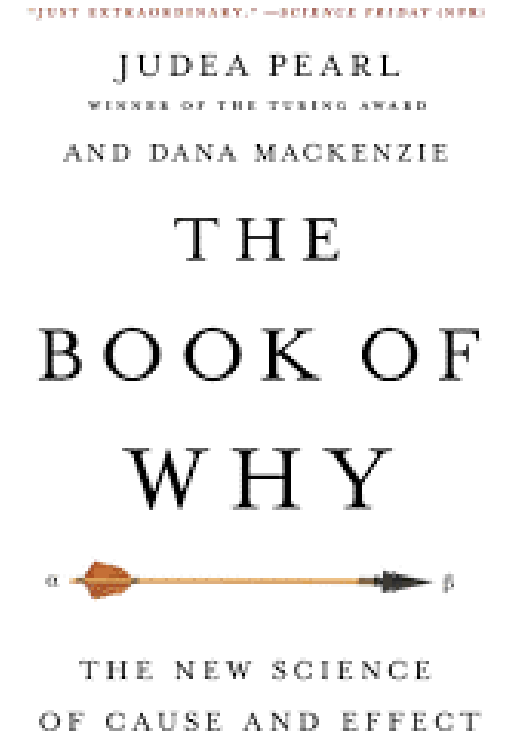
Objective: the strongest defensible causal conclusion the current evidence supports—and what would most change the answer.

Causality 101: Quick refresher

Key ideas from the causal revolution

Causality is about *change*

- How would changing inputs change outcomes (or their probabilities)?
 - Causal model: $P(\text{outputs} \mid \text{inputs})$
 - Sometimes written $P(\text{outputs} \mid \text{do}(\text{inputs}))$
 - Causal models: Simulation, Causal Bayesian Networks, Structural Equation Models (SEMs), path diagrams, etc.
 - Causal network: Children depend on parents
 - Inputs = conditions + events + **actions**
 - Decision analysis and optimization: How *should* we change (or set or “do”) inputs to cause desired changes in outputs?
- Descriptive: What *is* true?
- Causal: What can be *made* true (probably) by changing decision inputs? How?



Simpson's Paradox

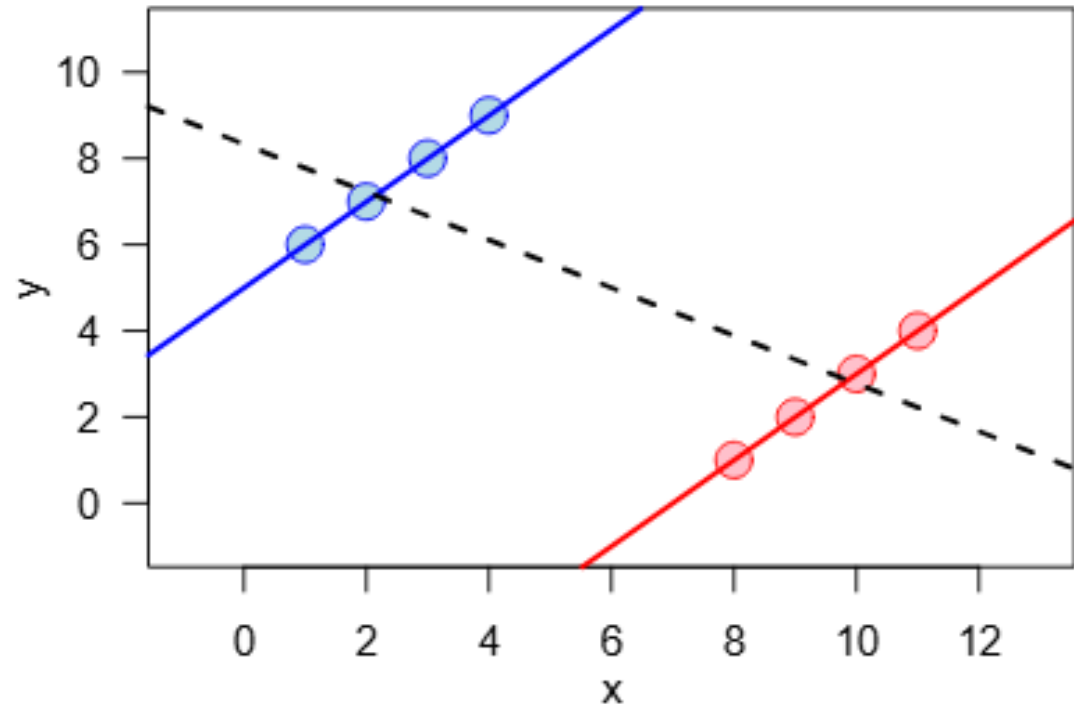
- How does Y depend on X?
- Does Y decrease with X?

Yes, for Y vs. X overall (black dashed line)

No, for Y vs. X within each group (Blue or Red)

Answer depends on what we condition on.

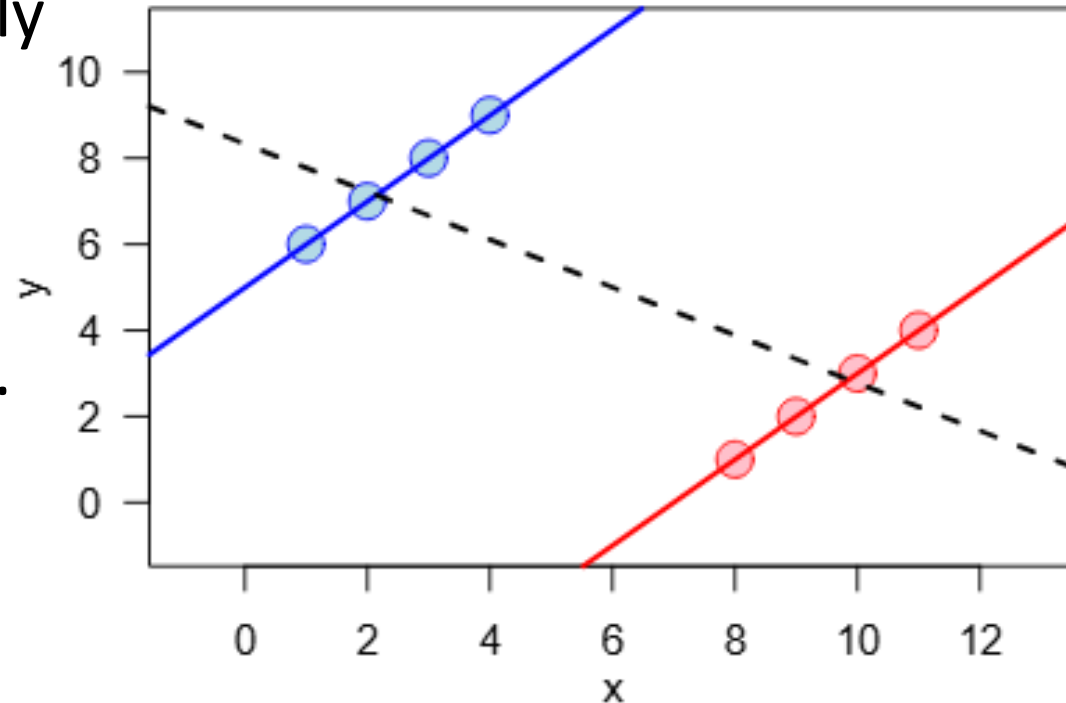
Causation $\neq$ association



Example: How would increasing x from 1 to 5 change y for case 1?

Identifiability: Do the **data** and our causal **assumptions** uniquely determine the **effect** of changing X on Y ?

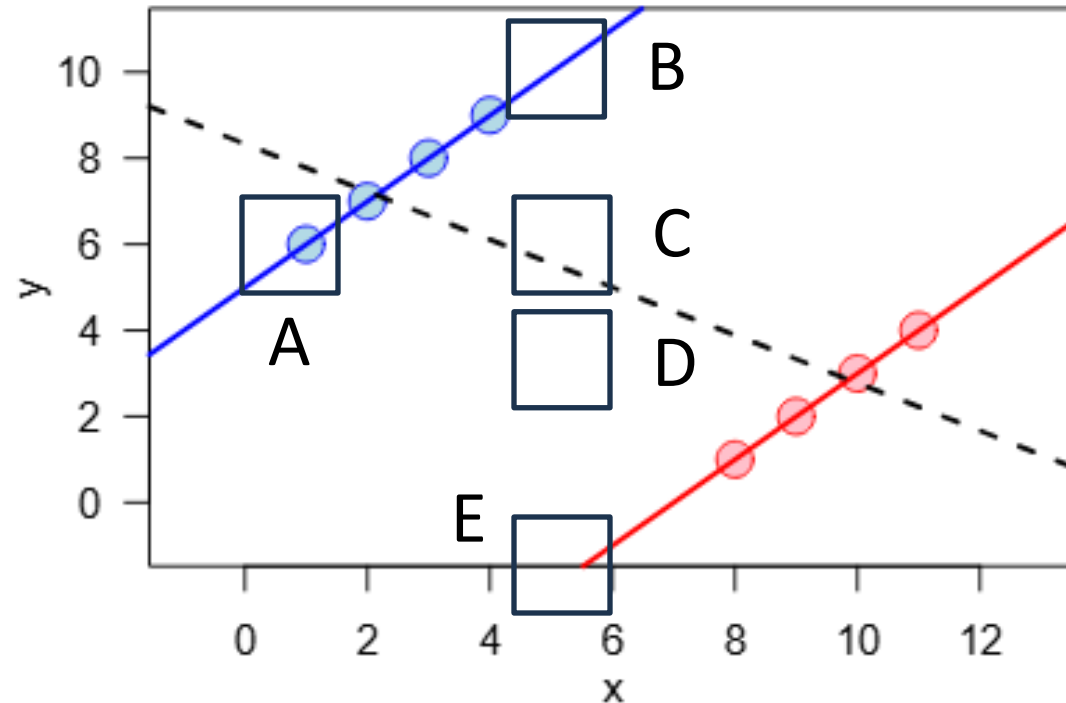
No—not from these data alone.



How would increasing x from 1 to 5 change y for case 1?

Defining “effect”: Compare case 1's outcome at $X=1$ with what its outcome would be if we changed X to 5. Where would it go: B, C, D, or E?

Might different units (cases) respond differently?
 (“Heterogeneous Treatment Effects”, HTEs)



From causal question to causal estimate

- **Define** the causal effect: Specify the **causal contrast** we want to know.
 - Example: *How would average sales differ if we lowered the price by 10% versus if we did not?*
 - A causal effect compares outcomes under two alternative interventions or conditions.
 - **The fundamental problem:** For the same person, customer, firm, etc., we usually observe the outcome under **only one** of those alternatives.
 - The other outcome is **counterfactual**: what *would have happened* instead.
- **Identify** the effect: Can the desired causal contrast be learned from the data we have, given appropriate **causal assumptions**?
 - What must we control for? Is there confounding or selection bias?
 - Identification tells us **what statistical quantity would equal the causal effect** *if* our assumptions are correct.
- **Estimate** the effect: Use the observed data and a statistical method to estimate the identified quantity.
 - Regression, matching, weighting, experiments, instrumental variables, etc.
 - Estimation introduces **sampling/statistical uncertainty**.
- **Validate** the effect estimate across settings; establish robustness
- **Key distinction: Good estimation cannot rescue bad identification.**
 - A precise estimate of the wrong quantity is not a valid causal-effect estimate.

Causal frameworks and terms

- **Interventional causality** (our focus): Changes in causes *change* (the probability distributions of) their effects
 - What can be made true (or more probable) by taking actions?
- **Predictive causality**: Causes help to *predict* their effects
 - Information flows from causes to their effects over time
- **Structural causality**: Effects depend (only) on their direct causes
 - Causes determine the *probabilities* of their effects.
 - Effect probabilities are determined by their **direct causes**.

Other causal frameworks

- **Potential outcomes:** Differences in causes make effect probabilities *different* from what they otherwise would have been
- **Mechanistic causality:** Explains *how changes propagate* through a network of variables
 - Dominos
- (Weight of evidence causality: Strong, consistent, temporal, etc. associations might be causal)
- (Attributive causality: Attribute effects to causes)
- Example: Do nicotine-stained fingers “cause” increased lung cancer risk?

Example 1:

Dublin air pollution quasi-experiment

Example: Dublin intervention study

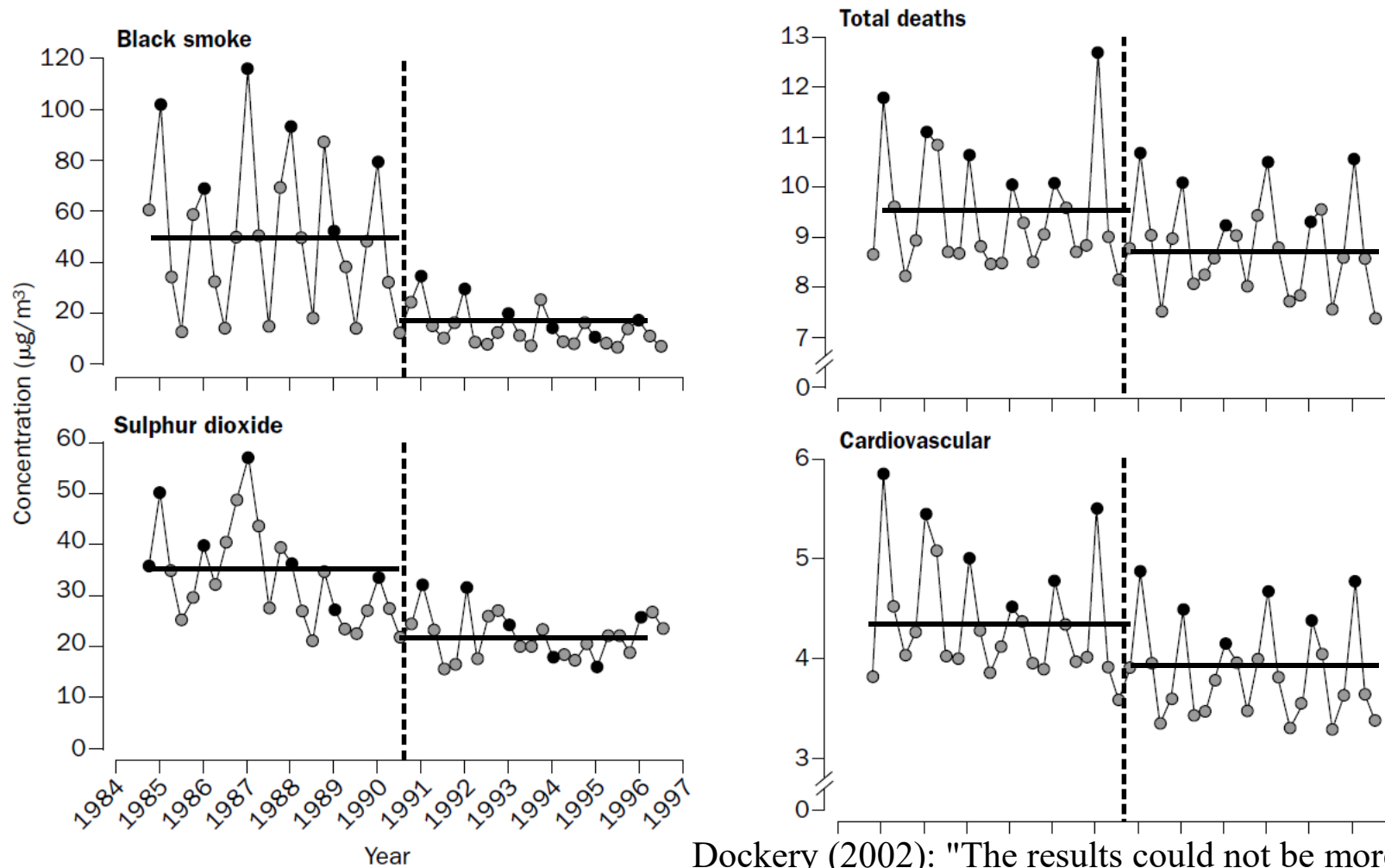


Figure 1: **Seasonal mean black smoke (upper) and sulphur dioxide (lower) concentrations, September 1984–96**

Vertical line shows date sale of coal was banned in Dublin County Borough. Black circles represent winter data.

Dockery (2002): "The results could not be more clear: Reducing particulate air pollution reduces the number of respiratory and cardiovascular-related deaths immediately."

www.news.harvard.edu/gazette/2002/10.24/14-coal.html

Example: Dublin intervention study

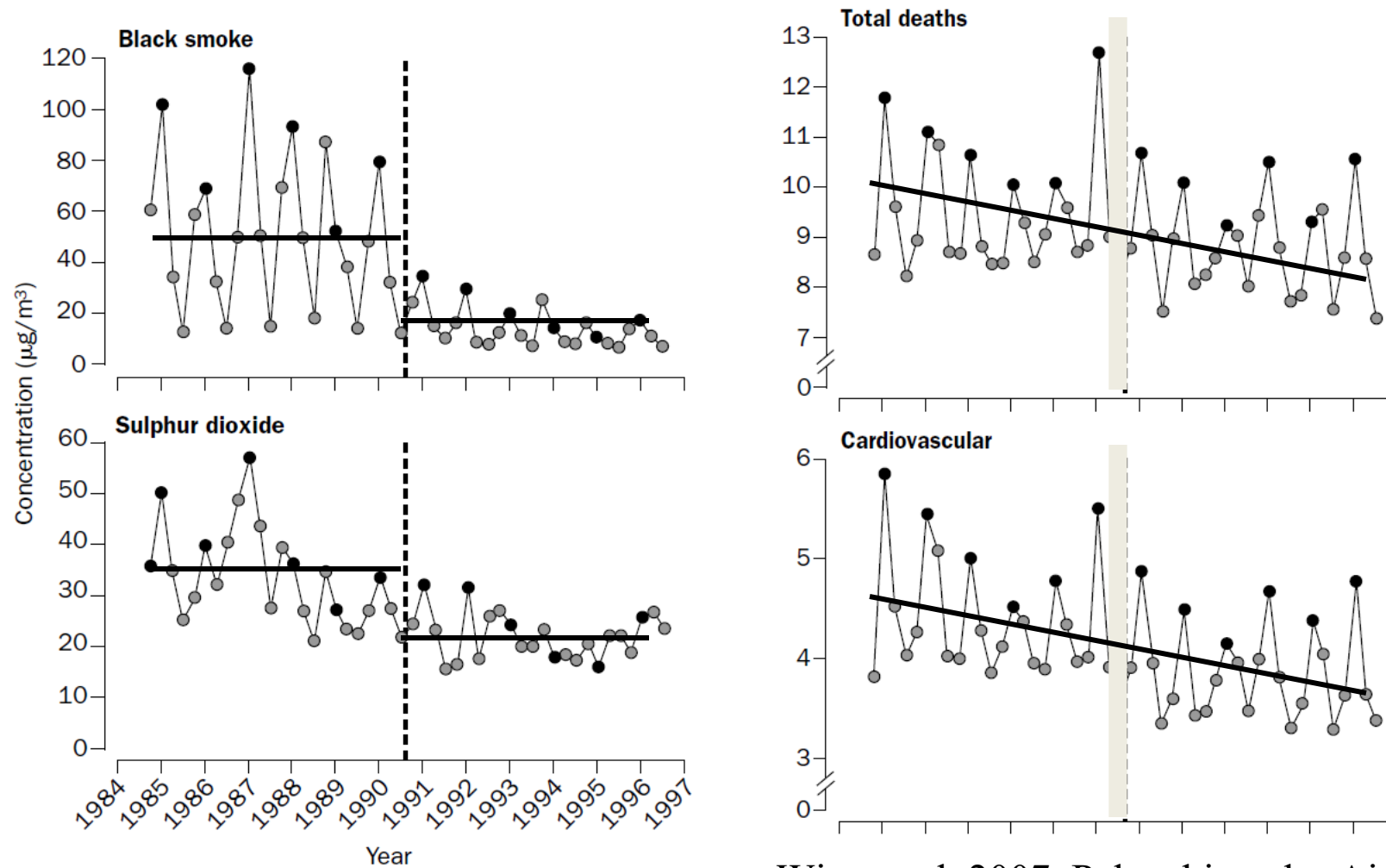
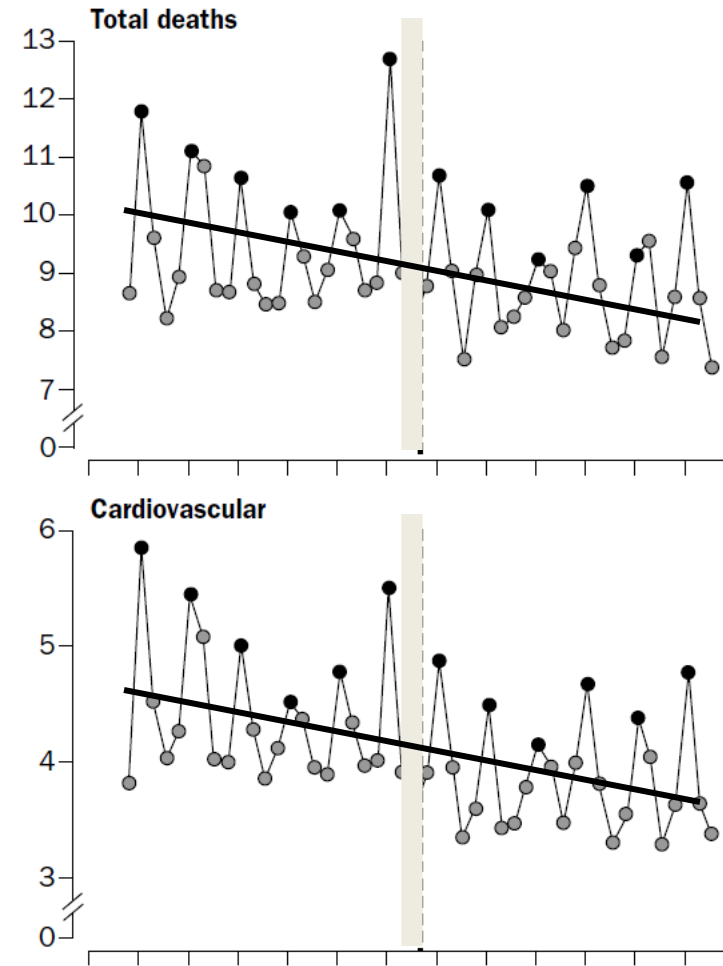


Figure 1: **Seasonal mean black smoke (upper) and sulphur dioxide (lower) concentrations, September 1984–96**
Vertical line shows date sale of coal was banned in Dublin County Borough. Black circles represent winter data.

Wittmaack 2007, Pelucchi et al.: Air pollution reduction has no detectable effect on mortality rate trend www.ncbi.nlm.nih.gov/pubmed/17365039

Right causal evidence questions

- Q1: Did risk go down faster *where* particulate pollution was greatly reduced than where it wasn't?
- Answer: No (Dockery et al., 2013)
- Q2: Did risk go down faster *after* particulate pollution was greatly reduced than before?
- Answer: No (Zigler & Dominici, 2014)
- Interventional view: Evidence about *associations* between PM2.5 levels and mortality is not evidence about whether *changes* in one cause changes in the other.
 - Evidence about changes caused by interventions is needed to inform policy



Dockery et al., 2013.: Air pollution reduction has no detectable effect on mortality rate trend

2026 Perspective

[Atmiq](#) / [Guides](#) / Ireland's coal bans: what the HEI reassessment changed

Ireland's coal bans: what the HEI reassessment changed

A wider evaluation separated changes in black smoke and respiratory mortality from less certain changes in other outcomes.

Atmiq editorial · **October 20, 2026** · 2 min

Sources

[Revisiting Dublin and extending the comparison ↓](#)

[Cleaner air, with outcome-specific results ↓](#)

[The difficulty of isolating an intervention ↓](#)

[Keep early and later evidence together ↓](#)

- [Dockery DW, Rich DQ, Goodman PG, et al.. Effect of Air Pollution Control on Mortality and Hospital Admissions in Ireland. Health Effects Institute, Research Report 176 \(2013\) ↗](#)
- [Clancy L, Goodman P, Sinclair H, Dockery DW. Effect of air-pollution control on death rates in Dublin, Ireland: an intervention study. The Lancet \(2002\). DOI: 10.1016/S0140-6736\(02\)11281-5 ↗](#)

**Put the
information to use.**

Check your place, sources and station readings on the Atmiq map.

<https://atmiq.xyz/en/guides/ireland-coal-bans-reassessment-2013>

2026 Perspective

“Revisiting Dublin and extending the comparison

- HEI Research Report 176, led by Dockery and colleagues in 2013, revisited Dublin’s 1990 ban and examined later coal bans in 11 additional Irish cities. It compared pollution and health before and after the bans, using comparison populations and more extensive treatment of background trends.

Cleaner air, with outcome-specific results

- Black smoke fell clearly after the bans, particularly during the heating season. Respiratory mortality decreased by 17% after the 1990 Dublin ban. **Unlike the early study, the reassessment did not find reductions in total or cardiovascular mortality after the Dublin or later bans.** The distinction between outcomes is central to interpreting the report.

The difficulty of isolating an intervention

- Long-term mortality can change for many reasons at the same time as a policy. Comparison areas and trend adjustment help but cannot recreate a randomized trial. A result that is not statistically supported is not proof that the true effect is exactly zero, and it does not erase the measured improvement in air quality.

Keep early and later evidence together

- Use this report alongside Clancy 2002 whenever discussing the Dublin example. Avoid turning a complex reassessment into either a guaranteed mortality reduction or a claim that cleaner air is useless. The report concerns population policies and cannot estimate a personal benefit from closing a window today.”

<https://atmiq.xyz/en/guides/ireland-coal-bans-reassessment-2013>

Can AI help us get there in minutes instead of decades?

Critical Synthesis: Causal and Policy Inference from the Dublin Coal Ban Natural Experiment

Evidence-calibrated, source-traceable cross-document critical synthesis based on 5 normalized propositions from 1 papers.

Traceability rule: Representative quotations are inserted verbatim from the standardized ASP source-evidence registry. Author statements are distinguished from the evidence needed to establish them.

Overall Assessment

The body of evidence from the Dockery & Clancy Dublin study provides strong descriptive evidence that the 1990 citywide coal sales ban was immediately followed by large reductions in black smoke particulate pollution and statistically significant decreases in respiratory and cardiovascular mortality. Adjustment for measured confounders (age, weather, influenza seasons, national mortality trends) increases confidence that these reductions were causally linked to the intervention. However, the non-randomized, aggregate-level interrupted time series design, lack of sensitivity or bias analyses, possible exposure measurement error, incomplete assessment of missing data, and the persistent possibility of unmeasured or inadequately controlled confounding (e.g., Dublin-specific healthcare, migration, or risk factor changes) preclude strong causal certainty. Attribution of exact numbers of deaths averted is model-dependent and assumes without direct proof that all the observed change was due to the intervention alone. Extension of these findings to general policy claims about urban air pollution control lacks direct empirical support for external validity or transportability beyond Dublin's specific context. The evidence is most appropriately described as highly suggestive (probable), but not robustly causal, and policy extrapolation should be made with caution.

Conclusions Not Supported by the Current Evidence

- Full causal certainty that all observed mortality reductions were solely due to the ban.
 - Quantitative attribution of exact numbers of deaths averted as being solely due to the intervention.
 - Generalization that all other urban areas will realize similar benefits from comparable interventions.
-

Priority Future Evidence

- Prospective or retrospective replication of similar interventions in comparably polluted cities with robust causal designs.
 - Individual-level linkage of exposure to outcome pre/post intervention.
 - Formal sensitivity analysis for confounding, missing data, and model specification.
 - Negative control outcomes or placebo time/location analyses.
 - Fine-grained spatial monitoring for exposure heterogeneity.
-

Bottom Line

The 1990 Dublin coal ban provides a paradigmatic case of a large, policy-driven reduction in air pollution, temporally aligned with major public health benefits as shown by ecological data and covariate-adjusted trends. While the evidence for a probable causal link is strong, it is not robust in the face of plausible alternative explanations due to aggregate design, unaddressed confounding, and lack of analytic sensitivity checks. Policy recommendations to reduce urban particulate pollution are reasoned but cannot claim certainty or full generalizability on the basis of this study alone.

Example 2:

Ultra processed foods and dementia

We know junk food isn't healthy, but scientists are trying to determine if it's also bad for our brains

New research suggests brain health could be **affected** by consuming too much low-nutrient processed food

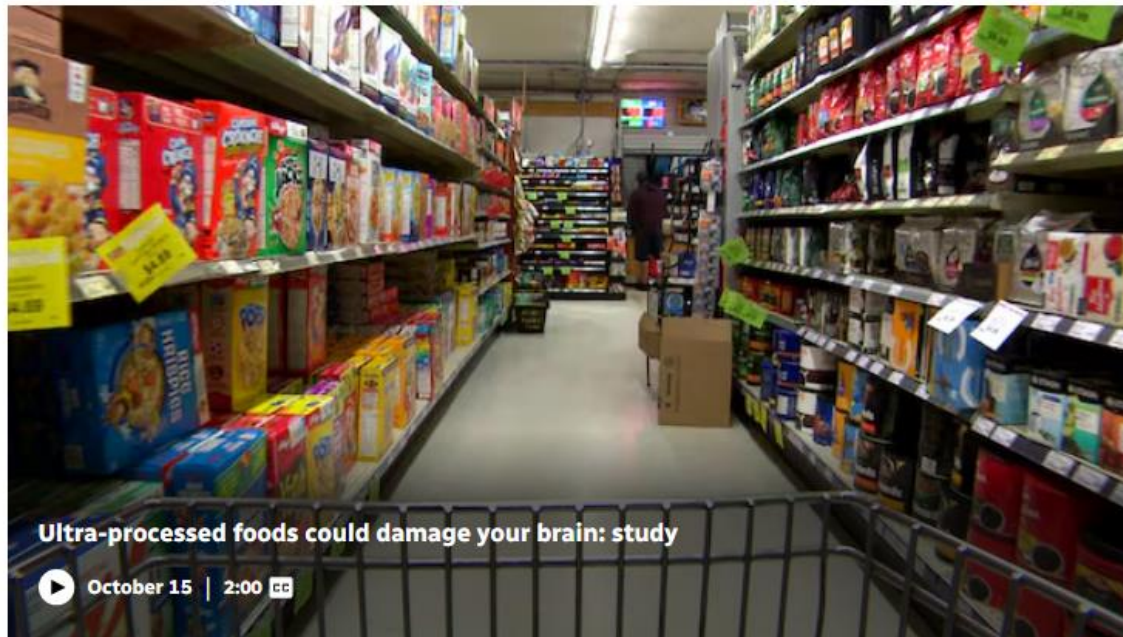


Nick Logan · CBC News · Posted: Oct 15, 2026 9:00 AM MDT | Last Updated: October 16



Listen to this article ⓘ

Estimated 6 minutes



New Canadian **research** suggests that eating too many ultra-processed foods, such as soda, chips and processed meats, could **damage** your brain. Researchers urge Canadians to cut down on such products.

Really?

MORE IN STROKE Symptoms Types Diagnosis Treatment Causes & Risk Factors Complications Related Conditions Prevention

HEALTH NEWS Fact Checked

Ultra-Processed Food Intake May Increase Dementia, Stroke Risk

Add Healthline on Google



Written by [Mandy French](#) on October 18, 2026 —
Fact checked by [Jill Seladi-Schulman, Ph.D.](#)



Really?

Recent research has shown that high intake of ultra-processed foods may be associated with an increased risk of stroke.

Original Investigation | Cardiology

Ultraprocessed Food Intake and Adiposity, Vascular Brain Injury, and Cognition in Middle-Aged Adults

Russell J. de Souza, RD, ScD^{1,2,3}; Karleen M. Schulze, MMath^{3,18}; Dipika Desai, MSc^{2,3}; [et al](#)

» [Author Affiliations](#) | [Article Information](#)

[RELATED ARTICLES](#) [FIGURES](#) [SUPPLEMENTAL CONTENT](#)

Key Points

Question Is higher consumption of ultraprocessed foods associated with adiposity, subclinical vascular brain injury, and cognitive performance among middle-aged adults in Canada?

Findings In this **cross-sectional study** of 8783 adults, higher ultraprocessed food intake was associated with greater percentage of body fat and visceral adipose tissue and with higher odds of magnetic resonance imaging-defined vascular brain injury. Ultraprocessed food intake was not associated with cognitive performance.

Meaning In this study, higher ultraprocessed food intake was associated with markers of adiposity and subclinical cerebrovascular injury that may precede detectable cognitive deficits, but longitudinal studies are needed to establish temporality.

Main Outcomes and Measures The main outcomes were body fat (BF) percentage (bioelectrical impedance), visceral adipose tissue (VAT) volume, hepatic fat fraction (HFF), VBI defined as covert brain infarcts or high white matter hyperintensities (HWMHs) (measured by MRI), and cognitive performance measured by the Montreal Cognitive Assessment (MOCA) and the Digital Symbol Substitution Test (DSST).

Results In this cross-sectional study of 8783 participants (mean [SD] age, 57.7 [8.7] years; 4917 females [56.6%]), those in the highest quartile of UPF intake had a BF percentage that was 1.7 (95% CI, 1.2-2.2) percentage points higher and a 5.7 (95% CI, 3.2-8.3) mL greater VAT volume compared with those in the lowest quartile, but no significant difference in HFF. Compared with the lowest quartile, the highest UPF quartile had higher odds of VBI (adjusted OR, 1.40; 95% CI, 1.10-1.77; *P* for trend = .009). No significant associations were observed between UPF intake and MOCA or DSST performance.

Conclusions and Relevance In this cross-sectional study of middle-aged adults in Canada, higher UPF intake was associated with greater adiposity and higher odds of MRI-defined VBI but not cognitive performance. Associations with VBI were sensitive to UPF exposure definition. Longitudinal studies are needed to establish temporality.



Back to Top



Sections



PDF



Share



Example 2:

Ultra processed foods and dementia

https://moirai.shinyapps.io/asp_monograph_de_souza_2026/

ASP Monograph Evidence Portal

Review topic: de Souza 2026

Automated paper-level methodological evidence review.

1
Papers staged

1
Papers completed

1
Complete 9/9 core reviews

0
Failed papers

Individual Paper Evidence Bundles

Paper	Status	Core Reports	Outputs
de Souza 2026	completed	9/9	Open Evidence Bundle Open Briefs

de Souza et al. (2026): Causal Evidence Assessment

[Open Monograph](#)[Open Critical Synthesis](#)

Paper: de_Souza_2026

Available items: 16

Reusable single-paper output folder

Evidence Bundle and Core Reports

Source PDF

[Open](#)

Open ASP Bundle

[Open](#)

Study Summary

[Open](#)

Extracted Causal Claims

[Open](#)

Grouped STROBE Review

[Open](#)

Causal Critique

[Open](#)

Statistical Methods Review

[Open](#)

SR/MA Review

[Open](#)

Risk-of-Bias Review

[Open](#)

Mechanistic Review

[Open](#)

Cross-Module Synthesis

[Open](#)

Key Assumptions

[Open](#)

Key References

[Open](#)

Reference Support Audit

[Open](#)

Integrated Assessment

[Open](#)

Audience Briefs

[Open](#)

This folder is self-contained. Copy or rename the whole folder to save it permanently.

Critical Synthesis: de Souza 2026—Association of Ultraprocessed Food Intake with Adiposity, Subclinical Brain Injury, and Cognitive Outcomes in Middle-Aged Canadian Adults

Evidence-calibrated, source-traceable cross-document critical synthesis based on 5 normalized propositions from 1 papers.

Traceability rule: Representative quotations are inserted verbatim from the standardized ASP source-evidence registry. Author statements are distinguished from the evidence needed to establish them.

Overall Assessment

de Souza 2026 provides cross-sectional evidence of associations between ultraprocessed food (UPF) intake and higher adiposity and subclinical vascular brain injury (VBI), but no association with cognitive performance in middle-aged Canadian adults. However, the evidence is undermined by the study's cross-sectional design, exposure measurement challenges, analytic sensitivity (especially for VBI), and risks of confounding and reverse causation. **No causal or intervention effects can be established from this evidence, and public health or clinical recommendations about UPF intake are not supported.** The primary value of this study is as hypothesis-generating, underscoring the need for longitudinal and mechanistic research.

The AI's conclusion is strong.

How trustworthy is it? Why or why not?

Priority Future Evidence

- Longitudinal studies examining UPF intake, adiposity, VBI, and cognitive trajectories.
 - Randomized or natural experiments reducing UPF intake to test intervention effects.
 - Use of validated biomarkers for UPF intake and gold-standard outcome ascertainment.
 - Advanced bias analysis or negative control approaches to address confounding and measurement error.
-

Bottom Line

de Souza 2026 yields cross-sectional evidence for associations of UPF intake with adiposity and subclinical VBI, but not cognition, in middle-aged Canadian adults. Given substantial methodological limitations—especially the study's inability to establish temporality, vulnerability to confounding and measurement error, and analytic instability—no causal claims or practice/policy recommendations are supported. The study should be regarded as hypothesis-generating, not directive: its main contribution is in motivating more rigorous longitudinal, interventional, or mechanistic research to clarify whether and how ultraprocessed foods affect health.

Ultra-Processed Food Intake May Increase Dementia, Stroke Risk

Add Healthline on Google



Recent research has shown that high intake of ultra-processed foods may be associated with an increased risk of stroke.

Really?

AI: No valid supporting evidence presented

How to **communicate** important technical and methodological caveats to **specialists**?

Detailed summary + easy drill-down may help

Paper: de_Souza_2026

Available items: 16

Reusable single-paper output folder

Evidence Bundle and Core Reports

Source PDF

Open

Open ASP Bundle

Open

Study Summary

Open

Extracted Causal Claims

Open

Grouped STROBE Review

Open

Causal Critique

Open

Statistical Methods Review

Open

SR/MA Review

Open

Risk-of-Bias Review

Open

Mechanistic Review

Open

Cross-Module Synthesis

Open

Key Assumptions

Open

Key References

Open

Reference Support Audit

Open

Integrated Assessment

Open

Audience Briefs

Open

This folder is self-contained. Copy or rename the whole folder to save it permanently.

Document: de_Souza_2026

Generated: 2026-10-21 13:18:51 MDT Model: gpt-4.1

Inputs: synthesis, causal critique, key assumptions, references, and reference support audit where available.

Overall Assessment

The document **de_Souza_2026** provides observational cross-sectional evidence from a large, well-characterized Canadian cohort, examining the association between ultraprocessed food (UPF) intake and outcomes including adiposity, subclinical vascular brain injury (VBI) (defined by MRI), and cognitive function. The findings are robust for associations between UPF and adiposity, with partial and fragile support for VBI, and null results for cognition. Major conclusions regarding association are well-supported by the study design and analytic approach, **but claims of causality, mechanism, or policy relevance go beyond the strongest evidence.** Cited references generally support context, plausibility, or partial reasoning; none provide full triangulation or direct replication for all claimed findings. The strongest evidence applies to associations, not causal or interventional implications.

Key Conclusions and Support

Conclusion or Recommendation	Support Level	Best Supporting Evidence	Main Caveats	Assessment
Higher UPF intake is associated with higher body fat percentage and greater visceral adipose tissue (VAT) volume among middle-aged Canadian adults.	Strong	Consistently positive, statistically robust associations in primary study; directly supported by similar findings in international cohorts/trials (e.g., Konieczna 2021, Liu 2023, Hall 2019).	Association only—cross-sectional design cannot establish causality or temporality. Possible residual confounding and exposure misclassification risk.	Study data support clear association, echoed in external observational studies and RCTs for related endpoints; causality not established.
Higher UPF intake is associated with increased odds of subclinical vascular brain injury (VBI) as measured by MRI.	Partial	Association present in main UPF metrics in de_Souza_2026; mechanistic plausibility (Morys 2025). Prior cohort (Anand 2022) validates imaging/methods, not UPF effect.	Findings are sensitive to how UPF exposure is defined; no direct external replication for this outcome. Potential for confounding, misclassification, and lack of temporality.	Evidence for association exists but is fragile and not consistently supported externally or across analytic definitions. Causality and mechanism unproven.
Higher UPF intake is not associated with cognitive performance as assessed by MOCA and DSST in the main analyses.	Strong	Null results in this large, well-controlled cross-sectional sample; comparable to findings in Cardoso 2022. External studies (Gomes Gonçalves 2023) vary (mostly in older samples).	Possible test insensitivity, ceiling effects, high-education cohort. Cognitive impairment may not yet be manifest in midlife; measurement artifacts possible.	No evidence for association in this context; conclusions valid for similar populations and measures, less so for long-term cognitive risks.
The associations between UPF intake and VBI/adiposity may be stronger among females.	Partial/Contextual	Exploratory sex-stratified analyses in de_Souza_2026 show stronger associations; effect modification not primary outcome, not replicated externally.	These findings are not uniformly robust and not prespecified; not replicated in external studies.	Possible effect modification observed but should be interpreted cautiously.
UPF intake is not universally associated with VBI or adiposity across all UPF metrics/definitions.	Strong	de_Souza_2026 sensitivity analyses; alternate UPF definitions yield non-significant results.	Highlights risk of overgeneralization; underscores measurement-dependence of findings.	Findings robustly demonstrate that associations are contingent on exposure metric.
The study demonstrates that cross-sectional analyses can identify early associations between dietary patterns and subclinical organ damage before cognitive decline.	Strong	Study design and analytic transparency; consistent null cognitive results, associations for VBI/adiposity.	Cannot establish temporality or pathway from cross-sectional data.	As a descriptive/epidemiologic insight, well supported.
Findings highlight a need for longitudinal studies to clarify causality and mechanisms linking UPF intake with cardiometabolic and brain health.	Strong	Omnipresent and explicit in all discussion, summary, and limitations sections; all reviews confirm appropriateness.	—	This is the main actionable implication given current study's limits.

Key Assumptions Affecting Interpretation

- **Cross-sectional associations reflect underlying relationships:** If reverse causation is present (e.g., those with higher adiposity/VBI change diet), associations may not reflect the hypothesized direction.
- **Covariate adjustment is sufficient:** Unmeasured or residual confounders (especially behavior/social factors) could explain observed associations.
- **Exposure measurement accuracy:** Classification and recall errors are possible (FFQ–NOVA misalignments); no biomarker calibration was performed.
- **Single time-point exposure and outcome measurement:** May not accurately capture habitual or cumulative processes; misalignment with etiologic windows possible.
- **Appropriateness of analytic model:** Some associations (notably VBI) are sensitive to exposure operationalization, raising questions about robustness.
- **Selection bias:** Complete-case approach may limit generalizability if those excluded differ by both exposure/outcome.
- **Effect modification unrecognized beyond sex:** If present, overall results may not apply to all population subgroups.

Reference-Support Issues

- **Cited references primarily support** background, measurement methodology, or partial context for individual association (e.g., UPF–adiposity, UPF–brain structure in related populations). No key reference, other than *de_Souza_2026* itself, directly supports the integrated conclusions about UPF, MRI-defined VBI, and cognitive function in a Canadian cohort.
- For **UPF and adiposity**, associations are externally validated in similar population-based and prospective studies (Konieczna 2021, Liu 2023) and, for causality, in short-term RCTs (Hall 2019).
- For **UPF and VBI**, references are either for related (not identical) outcomes, support mechanistic plausibility, or provide context—not direct demonstration.
- For **cognitive outcomes**, cited studies are divergent; the current study's null findings are supported for mid-life, while some external studies in older adults find associations.
- **No direct external replication** for the complete set of findings in the same setting, exposure, and outcomes. Partial support only for segments of the reasoning.
- **Policy and causal language** is not matched by reference or primary data support; recommendations for prevention/intervention are not justified by interventional evidence.

Bottom-Line Appraisal

de_Souza_2026 provides careful, methodologically strong descriptive epidemiologic evidence for an association between higher ultraprocessed food intake and greater adiposity, with possible but less certain association with subclinical vascular brain injury, and no observed effect on cognitive function in a well-defined Canadian cohort. The document is commended for its transparent reporting, explicit discussion of limitations, and caution in interpretation in most sections. All modules and reviews converge: **the observed associations are not sufficient to infer causality or justify policy interventions**. Cited references lend useful context and partial external support, but do not directly support the strongest claims for VBI or cognition, nor do they provide causal proof. Where evidence is strongest (UPF–adiposity association), claims are appropriately reported as associations. Where evidence is weaker (UPF–VBI, mechanistic/causal/policy claims), the document's conclusions or recommendations may overstep what the evidence and references directly support. The main value of this study is in generating hypotheses and identifying the need for rigorous longitudinal and experimental follow-up on UPF and neurologic outcomes.



How to communicate important technical and methodological caveats to **non-specialists**?

Answering audience-specific questions may help

de Souza et al. (2026): Causal Evidence Assessment

Open Monograph

Open Critical Synthesis

de_Souza_2026 ASP Evidence Bundle

Self-contained paper-level ASP evidence review bundle.

Generated from a single selected paper.

Paper: de_Souza_2026

Available items: 16

Reusable single-paper output folder

Evidence Bundle and Core Reports

Source PDF

Open

Open ASP Bundle

Open

Study Summary

Open

Extracted Causal Claims

Open

Grouped STROBE Review

Open

Causal Critique

Open

Statistical Methods Review

Open

SR/MA Review

Open

Risk-of-Bias Review

Open

Mechanistic Review

Open

Cross-Module Synthesis

Open

Key Assumptions

Open

Key References

Open

Reference Support Audit

Open

Integrated Assessment

Open

Audience Briefs

Open

This folder is self-contained. Copy or rename the whole folder to save it permanently.

49

75°F Partly sunny

10/21/2026

de_Souza_2026 - Audience Evidence Briefs

Experimental AI-generated audience briefs from ASP reports. Inspect before use.

Briefs

- [Scientist Brief](#)
- [Legal Brief](#)
- [Judge Brief](#)
- [Reporter Brief](#)
- [Policy Brief](#)

Prompt bundles

- [Scientist Prompt](#)
- [Legal Prompt](#)
- [Judge Prompt](#)
- [Reporter Prompt](#)
- [Policy Prompt](#)

https://moirai.shinyapps.io/asp_monograph_de_souza_2026/

de_Souza_2026 - Audience Evidence Briefs

Experimental AI-generated audience briefs from ASP reports. Inspect before use.

Briefs

- [Scientist Brief](#)
- [Legal Brief](#)
- [Judge Brief](#)
- [Reporter Brief](#)
- [Policy Brief](#)

Prompt bundles

- [Scientist Prompt](#)
- [Legal Prompt](#)
- [Judge Prompt](#)
- [Reporter Prompt](#)
- [Policy Prompt](#)

https://moirai.shinyapps.io/asp_monograph_de_souza_2026/

Evidence Brief: Ultraprocessed Food Intake and Health in Middle-Aged Canadians

Audience: Reporter

- **What did the study actually find?**

In a large cross-sectional study of Canadian adults aged 35–69, higher intake of ultraprocessed foods (UPFs) was statistically associated with higher body fat and more visceral (abdominal) fat. There was also a statistical association between higher UPF intake and greater odds of having silent vascular brain injuries (as seen on MRI scans), but this finding was sensitive to how UPF intake was measured. No association was found between UPF intake and cognitive test performance in this age group.

- **What should a headline say without overstating the evidence?**

Study reports correlation between ultraprocessed food intake and body fat, possible link to silent brain injury in Canadian adults

- **What should a headline not say?**

Headlines should not claim that UPFs cause obesity, brain injury, or cognitive decline, or that reducing UPF intake will prevent these outcomes, based on this study alone.

- **What caveats must appear high in the story?**

- This was an observational, cross-sectional study—meaning it looked at data from one point in time and cannot show cause and effect.
- It is possible that people with higher body fat or brain changes already eat more UPFs, rather than UPFs causing these changes.
- The association with brain injury was not consistent across all ways of measuring UPF intake.
- Diet was self-reported and classified after the fact, which could introduce errors.
- The study population was mostly White and urban, which may limit generalizability.

- **What is the difference between association and causation here?**

Association means that two things are observed together (e.g., higher UPF intake and higher body fat), but it does not prove that one causes the other. Causation would require evidence that eating more UPFs directly leads to these health outcomes, which this study cannot provide.

- **What context would help readers avoid misunderstanding the result?**

- Other studies have also found associations between UPF intake and obesity, but evidence for direct effects on brain health is less clear.
- This study did not find a link between UPF intake and cognitive performance in middle-aged adults, possibly because cognitive decline may take longer to appear.
- Calls for policy or clinical action should be based on stronger evidence, such as long-term or experimental studies.

- **What is a concise, accurate one-paragraph summary?**

A new Canadian study of nearly 9,000 middle-aged adults found that those who ate more ultraprocessed foods tended to have higher body fat and more abdominal fat, and were more likely to show silent signs of vascular brain injury on MRI scans. However, the study did not find a link between ultraprocessed food intake and cognitive test scores. Because the research only looked at data from a single point in time, it cannot show that ultraprocessed foods cause these health problems. The findings highlight the need for longer-term studies to clarify whether eating more ultraprocessed foods leads to obesity or brain changes, or if other factors are involved.

de_Souza_2026 - Audience Evidence Briefs

Experimental AI-generated audience briefs from ASP reports. Inspect before use.

Briefs

- [Scientist Brief](#)
- [Legal Brief](#)
- [Judge Brief](#)
- [Reporter Brief](#)
- [Policy Brief](#)

Prompt bundles

- [Scientist Prompt](#)
- [Legal Prompt](#)
- [Judge Prompt](#)
- [Reporter Prompt](#)
- [Policy Prompt](#)

https://moirai.shinyapps.io/asp_monograph_de_souza_2026/

Status: Experimental AI-generated evidence brief. Review before relying on conclusions.

Evidence Brief: Ultraprocessed Food Intake, Adiposity, Vascular Brain Injury, and Cognition in Middle-Aged Adults (de_Souza_2026)

Audience: Policy Decision-Makers

- **What decision-relevant question does this paper inform?**
Whether higher intake of ultraprocessed foods (UPFs) is associated with increased risk of obesity, subclinical vascular brain injury (VBI), or cognitive impairment in middle-aged adults, and whether this evidence supports policy or public health interventions targeting UPF consumption.
- **What action, if any, could this evidence support now?**
The evidence supports further research and monitoring, not direct policy or clinical interventions to reduce UPF intake for prevention of VBI, adiposity, or cognitive decline. The study is best used to generate hypotheses and justify longitudinal studies.
- **How strong is the evidence for causal intervention effects?**
The evidence is observational and cross-sectional. It shows statistical associations, not causality. No causal or interventional effects can be inferred for UPF reduction on VBI, adiposity, or cognition from this study alone.
- **What are the most important uncertainties for policy or organizational decision-making?**
Key uncertainties include: inability to determine causality or temporality (reverse causation possible), potential for residual confounding and measurement error, sensitivity of findings to how UPF intake is defined, and limited generalizability beyond the studied Canadian population.
- **What would be the risks of overreacting or underreacting to this evidence?**
Overreacting (e.g., implementing population-level UPF restrictions based on this study alone) risks acting on non-causal associations, possible unintended consequences, and misallocation of resources. Underreacting could delay needed research to clarify potential health risks of UPFs.
- **What monitoring, research, communication, or precautionary steps are reasonable?**
Reasonable steps include supporting well-designed longitudinal and mechanistic studies, monitoring UPF consumption trends, and communicating the current uncertainty and need for stronger evidence before policy action. Precautionary advice should be clearly labeled as such.
- **What additional evidence would most improve the decision?**
High-quality longitudinal studies or randomized controlled trials examining UPF intake and incident VBI, adiposity, and cognitive outcomes, with robust exposure measurement and bias control, would be most informative for policy decisions.

Bottom Line

This large Canadian cross-sectional study finds that higher ultraprocessed food intake is statistically associated with greater body fat and higher odds of subclinical vascular brain injury, but not with cognitive impairment in middle-aged adults. However, these findings do not establish causality, are sensitive to how UPF intake is measured, and may be affected by confounding and measurement error. The evidence does not support direct policy or clinical interventions to reduce UPF intake for prevention of brain or metabolic disease at this time. The main policy-relevant implication is the need for further longitudinal and mechanistic research to clarify whether reducing UPF intake would improve health outcomes.



de_Souza_2026 - Audience Evidence Briefs

Experimental AI-generated audience briefs from ASP reports. Inspect before use.

Briefs

- [Scientist Brief](#)
- [Legal Brief](#)
- [Judge Brief](#)
- [Reporter Brief](#)
- [Policy Brief](#)

Prompt bundles

- [Scientist Prompt](#)
- [Legal Prompt](#)
- [Judge Prompt](#)
- [Reporter Prompt](#)
- [Policy Prompt](#)

https://moirai.shinyapps.io/asp_monograph_de_souza_2026/

Evidence Brief: Ultraprocessed Food Intake and Health Outcomes in Middle-Aged Canadian Adults (de_Souza_2026)

Audience

Legal professionals (litigation, regulatory, or policy context)

Key Points for Legal Audiences

- **Potential Legal Claim Supported:** A party could reasonably use this paper to argue that higher intake of ultraprocessed foods (UPFs) is statistically associated with greater adiposity (body fat and visceral adipose tissue) and, less robustly, with subclinical vascular brain injury (VBI) in middle-aged Canadian adults. The study does *not* support claims that UPF intake causes these outcomes or that reducing UPF intake would prevent them.
- **Type of Evidence:** The study provides cross-sectional observational evidence of statistical associations only. It does not establish general or specific causation, nor does it support interventional or mechanistic conclusions. The findings are best used for hypothesis generation.
- **Major Limitations and Cross-Examination Points:**
 - Cross-sectional design: Cannot establish temporality or rule out reverse causation (i.e., whether higher adiposity or VBI leads to higher UPF intake, or vice versa).
 - Potential for residual confounding: Unmeasured or inadequately controlled factors (e.g., socioeconomic status, cultural dietary patterns) may explain observed associations.
 - Exposure misclassification: UPF intake was measured by food frequency questionnaire and classified post hoc, with no calibration or biomarker validation; this could bias results.
 - Sensitivity of findings: Associations with VBI were not robust across all definitions of UPF exposure and analytic models.
 - No mechanistic or experimental evidence: The study does not include biological, toxicological, or interventional data to support causality.
- **Causal Language and Overstatement:** While the paper sometimes uses language implying modifiability or prevention, these statements are not justified by the study design. The authors themselves caution against causal interpretation and recommend further longitudinal research.
- **Assumptions Required for Causal Argument:** To argue causation, one would have to assume: (1) that the observed associations are not due to reverse causation or confounding; (2) that the dietary exposure measurement is accurate and temporally relevant; (3) that the analytic models fully account for all relevant confounders and mediators. None of these assumptions are directly tested or fully supported in the paper.
- **What Would Strengthen or Rebut Reliance on This Paper:**
 - Strengthen: Longitudinal or experimental studies showing temporality and effect of UPF reduction on outcomes; formal bias analysis; biomarker validation of exposure.
 - Rebut: Evidence of reverse causation, strong confounding by unmeasured variables, or inconsistent findings in other populations or with different UPF definitions.
- **Recommended Legal Wording:** Lawyers should state: "This study reports a statistical association between higher ultraprocessed food intake and greater adiposity, and a less robust association with subclinical vascular brain injury, in a cross-sectional Canadian cohort. The study does not establish that UPF intake causes these outcomes, nor does it support policy or clinical recommendations to reduce UPF intake for prevention based on this evidence alone."

Bottom Line

The de_Souza_2026 study provides credible cross-sectional evidence that higher ultraprocessed food intake is statistically associated with greater body fat and, less consistently, with subclinical vascular brain injury in middle-aged Canadian adults. However, the study's design does not allow conclusions about causation, temporality, or the effectiveness of reducing UPF intake for prevention. Observed associations may be influenced by confounding, measurement error, or reverse causation. Legal arguments or policy recommendations based solely on this paper should be made with caution and clear acknowledgment of these limitations.

- **Causal Language and Overstatement:** While the paper sometimes uses language implying modifiability or prevention, these statements are not justified by the study design. The authors themselves caution against causal interpretation and recommend further longitudinal research.
- **Assumptions Required for Causal Argument:** To argue causation, one would have to assume: (1) that the observed associations are not due to reverse causation or confounding; (2) that the dietary exposure measurement is accurate and temporally relevant; (3) that the analytic models fully account for all relevant confounders and mediators. None of these assumptions are directly tested or fully supported in the paper.
- **What Would Strengthen or Rebut Reliance on This Paper:**
 - Strengthen: Longitudinal or experimental studies showing temporality and effect of UPF reduction on outcomes; formal bias analysis; biomarker validation of exposure.
 - Rebut: Evidence of reverse causation, strong confounding by unmeasured variables, or inconsistent findings in other populations or with different UPF definitions.
- **Recommended Legal Wording:** Lawyers should state: “This study reports a statistical association between higher ultraprocessed food intake and greater adiposity, and a less robust association with subclinical vascular brain injury, in a cross-sectional Canadian cohort. The study does not establish that UPF intake causes these outcomes, nor does it support policy or clinical recommendations to reduce UPF intake for prevention based on this evidence alone.”

Bottom Line

The de_Souza_2026 study provides credible cross-sectional evidence that higher ultraprocessed food intake is statistically associated with greater body fat and, less consistently, with subclinical vascular brain injury in middle-aged Canadian adults. However, the study’s design does not allow conclusions about causation, temporality, or the effectiveness of reducing UPF intake for prevention. Observed associations may be influenced by confounding, measurement error, or reverse causation. Legal arguments or policy recommendations based solely on this paper should be made with caution and clear acknowledgment of these limitations.

Why evaluating causal claims is difficult

Ambiguous definitions, competing explanations, design assumptions, model uncertainty, and transportability

“Cause” has multiple meanings

Traditions model causal pathways, predict outcomes, or evaluate causal evidence.

Structural / path modeling

Wright · Simon · Blalock · Kenny (1921–79)

Direct, indirect, and total effects depend on specified causal structures and assumptions. (“Effects depend on their causes.”)

INUS / causal pies

Mackie 1965 · Rothman 1976

Different **combinations of component causes** can be sufficient for the same outcome.

Granger causality

Granger 1969

Past X improves **prediction** of future Y beyond the other available past information. (“Causes predict their effects”)

Probabilistic causation

Suppes 1970

Causes precede and **raise the probability** of effects, if non-spurious. (“Causes make their effects more likely”)

Refutationist epidemiology

Maclure 1985, 1990

Seek evidence capable of **refuting** rival noncausal explanations for observed associations. (“Causes can’t be explained away.”)

Bradford Hill / WoE

Hill 1965 · later IARC / EPA applications

Consider strength, consistency, temporality, plausibility, etc. of **associations** not due to chance, confounding, or bias

Selected milestones; these approaches are complementary, and not all define causation in the same sense.

Causation as the effect of interventions: $P(Y \mid \text{do}(x)) \neq P(Y \mid x)$

From hypothetical fixing to potential outcomes, do-calculus, and manipulation.

Structural interventions

Haavelmo 1943–44

Hypothetically fix X in a structural system and derive the consequences; fixing differs from conditioning.

Potential outcomes

Neyman 1923 · Rubin 1974

Compare potential outcomes under alternative interventions: $Y(x)$ versus $Y(x')$.

Longitudinal effects

Robins 1986

Estimate effects of time-varying treatment strategies when treatment and confounders evolve together.

Structural causal models

Pearl 1995–2000

Use structural assumptions and causal graphs to identify distributions such as $P(Y \mid \text{do}(X = x))$.

Manipulationist causation

Woodward 2003

X causes Y when an appropriate hypothetical intervention on X would change the distribution of Y . (“Seeing is not doing.”)

Risk-assessment question: What would change under an intervention on X ?

Many causal statements lack conceptual clarity

- “Evidence is sufficient to conclude that there is a **causal relationship** with relevant pollutant exposures (e.g., doses or exposures generally within one to two orders of magnitude of recent concentrations). **That is**, the pollutant has been **shown to result in** health effects in studies in which **chance, confounding, and other biases could be ruled out with reasonable confidence.**” (US EPA, 2015)
- Evidence is sufficient to conclude that a causal relationship is **likely to exist** with relevant pollutant exposures. **That is**, the pollutant has been **shown to result in** health effects in studies where **results are not explained by chance, confounding, and other biases**, but uncertainties remain in the evidence overall.

How?

Many causal statements lack conceptual clarity

- “Evidence is sufficient to conclude that there is a **causal relationship** with relevant pollutant exposures (e.g., doses or exposures generally within one to two orders of magnitude of recent concentrations). **That is**, the pollutant has been shown to **result in** health effects in studies in which **chance, confounding, and other biases could be ruled out with reasonable confidence.**” (US EPA, 2015)



Four ways an observed association can arise

Exposure → Outcome

Intervening on X changes Y

$$X \rightarrow Y$$

Reverse causation

Y influences exposure X

$$Y \rightarrow X$$

Common cause

U influences both X and Y

$$X \leftarrow U \rightarrow Y$$

Collider / selection bias

Conditioning on S can link X and Y

$$X \rightarrow S \leftarrow Y$$

Condition on S (selection / adjustment)

Four more ways associations can mislead

Spurious trends

Independent series can co-move

Coincident trends $\neq$ causal effect
(Example: Dublin)

Misspecified models

Wrong models can invent slopes

Example: $Y = kX$; intercept fixed at zero
(Example: enrofloxacin)

Aggregation / ecological bias

Pooling can reverse patterns

Simpson reversal; ecological inference risk

Internal validity threats

Other changes can mimic effects

History · maturation · regression to mean

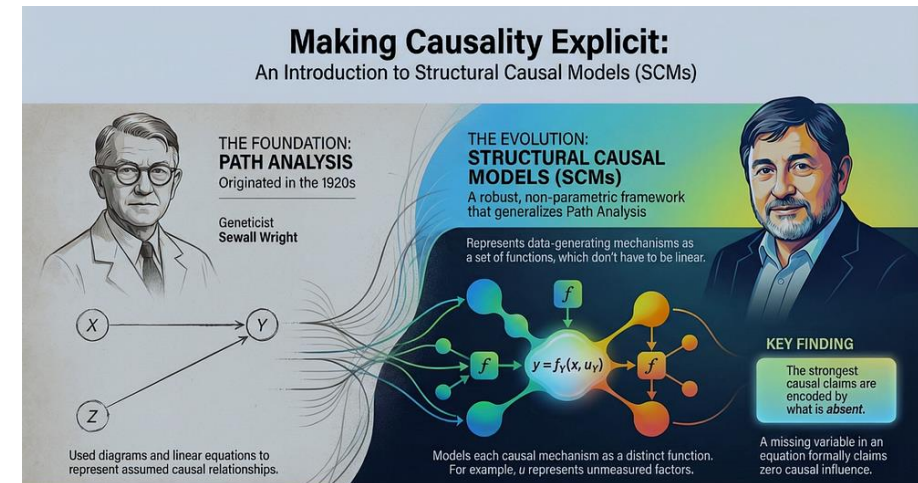


Opportunities to do better

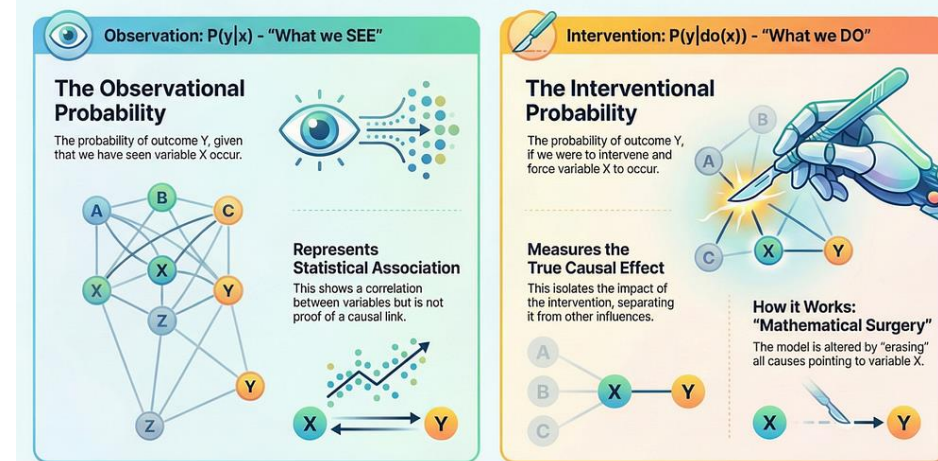
Five domains: conceptual & definitional clarity, study design, causal modeling & analysis, interpretation, applicability

Clarity about causality is available

- **Interventional causation:**
- “X is a **direct cause** of Y” *means* that changing X changes Y (or its probability distribution)...
- ...even when other direct causes of Y are held fixed
- This is what decision-makers need to know
- *Changing* is different from *observing*: “Seeing is not doing”
 - Not the same as predictive (Granger) or attributional (Bradford Hill, IARC, EPA) causation
 - Implied by mechanistic causation
 - “How does it work?”



Seeing vs. Doing: How the 'do-operator' Reveals True Cause & Effect

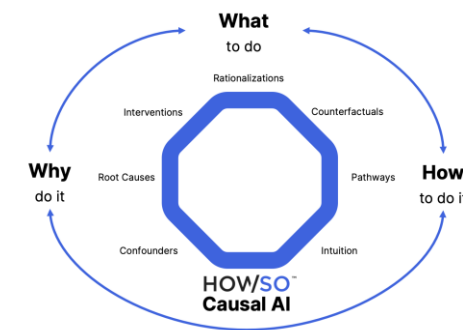
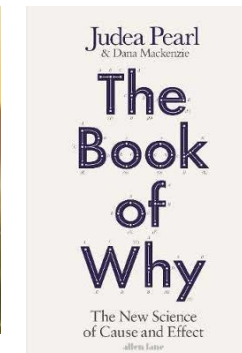


Common causal interpretation errors are now avoidable

Common error	How to avoid it
Treating association as causation	Define the intervention ; estimate $P(Y \mid \text{do}(X))$, not just $P(Y \mid X)$.
Ignoring rival causal explanations	Draw competing DAGs; test implications and seek discriminating evidence .
Inadequate confounding control	Identify common causes; use justified adjustment sets and sensitivity analyses.
Adjusting for the wrong variables	Use justified adjustment sets : avoid conditioning on colliders, mediators, post-treatment variables.
Mistaking before–after differences for effects	Model trends and co-interventions; use credible comparison groups .
Overstating what the evidence establishes	Report identification assumptions , uncertainty, limits, and applicability.

Wrap-up on current situation

- Regulatory and legal determinations of causation today are largely Rorschach tests
 - Strong, consistent, plausible exposure-response associations that are not readily explained by chance, confounding, or bias... are often *not* indicators of causality
 - Rooster's crowing does not cause sunrise
 - CARET trial, formaldehyde, etc.
- The “Causal Revolution” provides tools for doing better... but has a steep learning curve
- To what extent can AI help bridge the gap?
 - Clearer, more transparent, and often more accurate causal reasoning and conclusions
 - Less judgment, more science
 - Reduce causal overreach in conclusions, interpretations, and policy recommendations
 - Augment rather than replace human experts



Aspirations for applying AI

- Improved AI assistant for causal evidence assessment, multi-document synthesis, active seeking of relevant information
 - Enable human experts to query and improve evidence bundles and reasoning
- High-impact applications
 - Formaldehyde, EtO, others
 - Use AI to identify highest-impact evidence
- Supports new synthesis articles, AI software, repositories of AI/human evaluations of influential articles

Principles for valid causal inference

Five domains: conceptual & definitional clarity, study design, causal modeling & analysis, interpretation, applicability

What is needed for a thorough, constructive, review?

5 domains of necessary conditions for valid causal inference

A • Conceptual clarity	Counterfactual contrast, estimand, intervention, population
B • Study design	Time zero, comparison groups, exchangeability, positivity
C • Analysis	Model validity, measurement error, missingness, sensitivity
D • Interpretation	Causal language, assumptions, alternatives, uncertainty
E • Applicability	Transportability, effect modifiers, target interventions

Design and analysis prerequisites

- Define time zero, eligibility, exposure, outcome, and follow-up.
- Examine exchangeability, positivity, consistency, and temporality.
- Avoid conditioning on colliders or inappropriate post-treatment variables.
- Probe error, missing data, latent variables, and model uncertainty.

Interpretation and applicability

- Distinguish estimates from identification assumptions and untested explanations.
- Check robustness across defensible specifications and samples.
- Assess external validity, effect modifiers, and target-population differences.
- Policy predictions require effects of specified feasible interventions.

From principles to auditable causal questions

Define

Explicit intervention,
estimand,
population, outcome,
timing

Identify

Rival DAGs,
assumptions,
comparison groups,
bias pathways

Probe

Negative controls,
alternate models,
sensitivity,
transportability

Document

Quote → question →
judgment →
assumptions →
proposed check

The crucial test: Does the review reveal errors or uncertainties that matter to the causal conclusion?

Statistical Methods Review

Measurement error	Validation, calibration, SIMEX, Bayesian models
Missing data	MCAR / MAR / MNAR; imputation, weighting, sensitivity
Latent variables	Unmeasured confounding, controls, IVs, bias analysis
Model uncertainty	Alternative specifications, multiverse, specification curves
Validation	Diagnostics, calibration, falsification, robustness

What does ASP1 ask about a paper? 6 core modules

Six specialist modules · five prespecified questions each · screened for applicability

01

Grouped STROBE Review

Reporting transparency

02

Causal Methods Review

Causal inference

03

Statistical Methods Review

Quantitative analysis

04

Systematic Review / Meta-analysis Review

Evidence synthesis

05

Risk-of-Bias Review

Risk-of-bias appraisal

06

Mechanistic Review

Mechanistic evidence

Screened for applicability; AI-generated appraisals require human verification.

Grouped STROBE Review

Applicable: when observational epidemiologic evidence is reported

A	Research question	Are the study question, objectives, rationale, and key concepts clearly framed?
B	Study design	Are the setting, dates, eligibility, participant selection, design, and follow-up described?
C	Variables & methods	Are variables, measures, size, methods, confounding, missingness, bias, and sensitivity reported?
D	Results & interpretation	Are participant flow, results, uncertainty, limitations, and interpretation reported fairly?
E	External validity	Are generalizability, funding, and limits on applying the findings beyond this setting discussed?

Causal Methods Review

Applicable: when module 0 identifies causal claims

A	Causal question	Are exposure, outcome, intervention, population, and estimand defined? Are association and causation distinguished?
B	Design & bias	Does the study design support causal inference, addressing selection, eligibility, follow-up, and missingness?
C	Measurement & analysis	Are variables measured well, with suitable confounding control and sensitivity analysis?
D	Causal interpretation	Are alternatives, uncertainty, limitations, and transportability addressed without causal overreach?
E	Policy implications	Are recommendations for action justified by the evidence, assumptions, and likely effects of intervention?

Statistical Methods Review

Applicable: when statistical analyses are presented

A	Measurement error	Are exposure errors addressed through validation, calibration, SIMEX, Bayesian methods, or sensitivity?
B	Missing data	Are MCAR/MAR/MNAR assumptions and relevant imputation, weighting, or sensitivity checks addressed?
C	Latent variables	Are unmeasured variables probed with controls, IVs, quantitative bias analysis, or sensitivity checks?
D	Model uncertainty	Are model and selection uncertainties probed with alternative models, multiverse or specification curves?
E	Validation & robustness	Are model assumptions and robustness checked with diagnostics, falsification, controls, or sensitivity?

Systematic Review / Meta-analysis Review

Applicable: when systematic reviews or meta-analyses are presented

A	Review plan	Are the objectives, eligibility criteria, protocol, and methods clearly stated and sufficiently specific?
B	Search & screening	Is the search comprehensive and reproducible, with transparent screening and selection?
C	Evidence extraction	Are data extraction, appraisal methods, and synthesis inputs defined so they can be reproduced?
D	Synthesis methods	Are synthesis methods suitable, and are heterogeneity, bias, uncertainty, and limitations handled responsibly?
E	Auditability	Are sources, flow, exclusions, judgments, and implications clear enough for reuse and decisions?

Risk-of-Bias Review

Applicable: when risk-of-bias results are presented

A	Framework & object	Is the risk-of-bias framework named, and is it clear what is being evaluated for bias?
B	Bias domains	Are selection bias, information bias, confounding, and other important domain-specific biases assessed?
C	Judgment procedure	Are bias judgments transparent, reproducible, and grounded in study features and evidence?
D	Interpretive use	Are risk-of-bias judgments used cautiously and appropriately in synthesis and conclusions?
E	Evidence trail	Are domain judgments, reasons, disagreements, and supporting evidence documented for external review?

Mechanistic Review

Applicable: when mechanistic or toxicological evidence is presented

A	Biological pathway	Is a biologically meaningful pathway from exposure to a precursor or adverse outcome identified?
B	Dose relevance	Are experimental concentrations, doses, or modeled exposures relevant to plausible in vivo or human exposure?
C	Mechanistic robustness	Are the reported mechanisms coherent, specific, and resistant to obvious alternative explanations?
D	Translation	Does the study address extrapolation across systems, doses, time scales, or species?
E	Human relevance	Are findings informative for human hazard, causal interpretation, or risk assessment, with limits clearly stated?

Implementation

Formaldehyde case study

AI software to help boost review and synthesis productivity

https://moirai.shinyapps.io/asp_monograph_formaldehyde_leukemia/

ASP Monograph Builder

Create, extend, process, inspect, and package multi-paper ASP evidence reviews.

Monograph Collection

Open existing Monograph

PFOS/PFOA/PFCs (0 papers; 20261016)

New Monograph topic

PFOS/PFOA/PFCs

Create New Monograph

Add Papers

1. Select PDFs. 2. Add them to the collection. 3. Process all newly staged papers when ready.

Choose one or more PDFs

Browse... Grandjean 2012.pdf

Upload complete

Add Selected PDFs

Processing Options

Model

gpt-4.1

- ☒ Key Assumptions
- ☒ Key References
- ☒ Reference Support Audit
- ☒ Overall Assessment
- ☒ Audience Briefs

Process New Papers

Critical Synthesis

Current Monograph

Topic: PFOS/PFOA/PFCs

Run root: C:/Users/Christine/Documents/ASP1/hosted_builder_workspace/runs/20261016_100758

PDF intake complete.

Added: 1

Skipped: 0

Papers

Paper	Status	Core reports available	Core reports expected	Error message
Grandjean 2012	staged	0	9	NA

Open Outputs

Paper

Open Evidence Bundle

Open Briefs

Rebuild Monograph

Open Monograph Portal

Open Critical Synthesis

Download Monograph ZIP

Edit...

ASP Monograph Evidence Portal

Review topic: Formaldehyde and leukemia

Automated paper-level methodological evidence review.

15

Papers staged

15

Papers completed

15

Complete 9/9 core reviews

0

Failed papers

Individual Paper Evidence Bundles

Paper	Status	Core Reports	Outputs	
Beane Freeman JNCI 101 751 2009	completed	9/9	Open Evidence Bundle	Open Briefs
Blair Am J Ind Med 40 3 2001	completed	9/9	Open Evidence Bundle	Open Briefs
Checkoway et al 2015	completed	9/9	Open Evidence Bundle	Open Briefs
Coggon et al 2014	completed	9/9	Open Evidence Bundle	Open Briefs
Hauptmann J Natl Cancer Inst 101 1696 2009	completed	9/9	Open Evidence Bundle	Open Briefs
Hayes Am J Ind Med 18 641 1990	completed	9/9	Open Evidence Bundle	Open Briefs

Checkoway et al 2015	completed	9/9	Open Evidence Bundle	Open Briefs
Coggon et al 2014	completed	9/9	Open Evidence Bundle	Open Briefs
Hauptmann J Natl Cancer Inst 101 1696 2009	completed	9/9	Open Evidence Bundle	Open Briefs
Hayes Am J Ind Med 18 641 1990	completed	9/9	Open Evidence Bundle	Open Briefs
Lauer et al 2024 triangulation with supplements	completed	9/9	Open Evidence Bundle	Open Briefs
Meyers Am J Ind Med 56 1027 2013	completed	9/9	Open Evidence Bundle	Open Briefs
Ott Am J Ind Med 16 631 1989	completed	9/9	Open Evidence Bundle	Open Briefs
Pira Cancer Causes Control 25 1343 2014	completed	9/9	Open Evidence Bundle	Open Briefs
Saberi Hosnijeh Occup Env Med 70 464 2013	completed	9/9	Open Evidence Bundle	Open Briefs
Stroup JNCI 77 1217 1986	completed	9/9	Open Evidence Bundle	Open Briefs
Talibov et al 2014	completed	9/9	Open Evidence Bundle	Open Briefs
Walrath Cancer Res 44 4638 1984	completed	9/9	Open Evidence Bundle	Open Briefs
Walrath Int J Cancer 31 407 1983	completed	9/9	Open Evidence Bundle	Open Briefs

ASP Monograph Evidence Portal

Review topic: Formaldehyde and leukemia

Automated paper-level methodological evidence review.

15

Papers staged

15

Papers completed

15

Complete 9/9 core reviews

0

Failed papers

Individual Paper Evidence Bundles

Paper	Status	Core Reports	Outputs	
Beane Freeman JNCI 101 751 2009	completed	9/9	Open Evidence Bundle	Open Briefs
Blair Am J Ind Med 40 3 2001	completed	9/9	Open Evidence Bundle	Open Briefs
Checkoway et al 2015	completed	9/9	Open Evidence Bundle	Open Briefs
Coggon et al 2014	completed	9/9	Open Evidence Bundle	Open Briefs
Hauptmann J Natl Cancer Inst 101 1696 2009	completed	9/9	Open Evidence Bundle	Open Briefs
Hayes Am J Ind Med 18 641 1990	completed	9/9	Open Evidence Bundle	Open Briefs

ASP Monograph Evidence Portal

Review topic: Formaldehyde and leukemia

Automated paper-level methodological evidence review.

15

Papers staged

15

Papers completed

15

Complete 9/9 core reviews

0

Failed papers

Individual Paper Evidence Bundles

Paper	Status	Core Reports	Outputs	
Beane Freeman JNCI 101 751 2009	completed	9/9	Open Evidence Bundle	Open Briefs
Blair Am J Ind Med 40 3 2001	completed	9/9	Open Evidence Bundle	Open Briefs
Checkoway et al 2015	completed	9/9	Open Evidence Bundle	Open Briefs
Coggon et al 2014	completed	9/9	Open Evidence Bundle	Open Briefs
Hauptmann J Natl Cancer Inst 101 1696 2009	completed	9/9	Open Evidence Bundle	Open Briefs
Hayes Am J Ind Med 18 641 1990	completed	9/9	Open Evidence Bundle	Open Briefs

Beane_Freeman_JNCI_101_751_2009 ASP Evidence Bundle

Self-contained paper-level ASP evidence review bundle.

Generated from a single selected paper.

Paper: Beane_Freeman_JNCI_101_751_2009

Available items: 16

Reusable single-paper output folder

Evidence Bundle and Core Reports

Source PDF

Open

Open ASP Bundle

Open

Study Summary

Open

Extracted Causal Claims

Open

Grouped STROBE Review

Open

Causal Critique

Open

Statistical Methods Review

Open

SR/MA Review

Open

Risk-of-Bias Review

Open

Mechanistic Review

Open

Cross-Module Synthesis

Open

Key Assumptions

Open

Key References

Open

Reference Support Audit

Open

Integrated Assessment

Open

Audience Briefs

Open

This folder is self-contained. Copy or rename the whole folder to save it permanently.

ASP Evidence Review Bundle

Paper: Beane_Freeman_JNCI_101_751_2009.pdf

Generated: 2026-08-18 14:59:29 MDT

Model: gpt-4.1

Available Reports

Module	Status	Link	Notes
Study Summary	Completed	Open report	summary
Extracted Causal Claims	Completed	Open report	Verbatim causal and policy-relevant claims extracted from the paper
Grouped STROBE Review	Completed	Open report	full_review
Causal Critique	Completed	Open report	claim_level_review
Statistical Methods Review	Completed	Open report	methods_review
SR/MA Review	Screening only	Open report	screening_only
Risk-of-Bias Review	Screening only	Open report	screening_only
Mechanistic Review	Screening only	Open report	screening_only
Cross-Module Synthesis	Completed	Open report	synthesis
Overall Assessment	Completed	Open report	integrated_assessment

Modules run in this order: aspsum, aspstroke, aspcrit, aspmethods, aspprima, asproba, aspmech.

ASP Evidence Review Bundle

Paper: Beane_Freeman_JNCI_101_751_2009.pdf

Generated: 2026-08-18 14:59:29 MDT

Model: gpt-4.1

Available Reports

Module	Status	Link	Notes
Study Summary	Completed	Open report	summary
Extracted Causal Claims	Completed	Open report	Verbatim causal and policy-relevant claims extracted from the paper
Grouped STROBE Review	Completed	Open report	full_review
Causal Critique	Completed	Open report	claim_level_review
Statistical Methods Review	Completed	Open report	methods_review
SR/MA Review	Screening only	Open report	screening_only
Risk-of-Bias Review	Screening only	Open report	screening_only
Mechanistic Review	Screening only	Open report	screening_only
Cross-Module Synthesis	Completed	Open report	synthesis
Overall Assessment	Completed	Open report	integrated_assessment

Modules run in this order: aspsum, aspstroke, aspcrit, aspmethods, aspprima, asproba, aspmech.

Summary Table

Paper: Beane_Freeman_JNCI_101_751_2009.pdf
Generated: 2026-08-18 14:56:32 MDT **Model:** gpt-4.1 **Mode:** summary

Main Study Goals / Research Questions	Study Design and Data Used to Pursue Each Goal	Main Data Analyses Performed	Main Findings and Conclusions, Including Limitations
Does occupational exposure to formaldehyde increase mortality from lymphohematopoietic malignancies, including leukemia, myeloid leukemia, Hodgkin lymphoma, and multiple myeloma?	Prospective cohort study of 25,619 U.S. workers employed in formaldehyde-producing/using plants before 1966; Follow-up through December 31, 2004 (median 42 years); Vital status and cause of death from National Death Index (NDI) Plus and other records; Exposure assessment from job histories, expert estimation, and historical measurement data; Exposure metrics: peak, average intensity, cumulative exposure.	Poisson regression to estimate relative risk (RR) and 95% CIs for associations between formaldehyde exposure (peak, avg. intensity, cumulative) and selected causes of death; Standardized Mortality Ratios (SMRs) against US population; Sensitivity analyses for potential confounders and post-1980 exposure; Trend tests and analyses by time since first exposure and by calendar period.	Statistically significant increased risk for lymphohematopoietic malignancies and Hodgkin lymphoma with highest peak exposure (≥ 4 ppm); Non-significant RR elevations observed for leukemia, myeloid leukemia, and multiple myeloma for highest peak exposure; No consistent associations with average intensity or cumulative exposure; Risk patterns strongest in earlier follow-up (pre-1995), declining after; Associations not explained by confounders examined; Limitations: Mortality (not incidence), exposure assessment stopped in 1980, use of death certificates which may lack disease specificity.
How do the associations between formaldehyde exposure and cancer risk change over time and with time since first (peak) exposure?	Cohort follow-up by increments and analysis by calendar time and years since first exposure using same cohort and exposure data as above; Time-stratified analyses.	Cumulative RR estimates calculated by moving end of follow-up forward in yearly increments; Time period-specific analyses (pre-1981, 1981-94, 1995-2004); Analyses stratified by time since first exposure and time since first high peak exposure.	Highest relative risks for myeloid leukemia and other outcomes were observed closest in time to relevant exposure (<25 years since first peak exposure), with risks declining in later follow-up; Risk estimates for myeloid leukemia highest through mid-1990s and then declined; Pattern is consistent with a potential causal link within a relatively short induction/incubation period; Alternative explanation: pattern may reflect chance variations over time or increased precision of risk estimates.
Are the observed associations influenced by confounding exposures or exposure misclassification (e.g., benzene or unknown post-1980 formaldehyde exposures)?	Sensitivity and stratified analyses controlling for co-exposures (e.g. benzene, wood dust, antioxidants, plastics-related chemicals); Alternative scenarios for exposure post-1980 (assumed zero, assumed continued exposure, censored).	Poisson regression analyses stratified or adjusted for pay category, plant, co-exposures; Analyses excluding those with potential benzene exposure; Sensitivity analyses for post-1980 exposure assumptions.	Results robust to adjustment for co-exposures and pay category; Exclusion of those with possible benzene exposure did not materially change results; Results also robust to alternative assumptions on post-1980 exposure (though risk for myeloid leukemia slightly stronger with censored follow-up); Possible nondifferential misclassification would likely bias results toward the null; limitations of exposure assessment noted.

Brief Study Summary

This prospective cohort study extended the follow-up of 25,619 U.S. workers employed in formaldehyde-using or -producing industries before 1966, examining mortality from lymphohematopoietic malignancies through December 31, 2004. Exposure was assessed using historical work records and expert estimation, resulting in quantitative metrics for peak, average intensity, and cumulative formaldehyde exposure. Statistical analyses evaluated associations between these exposures and death from lymphatic and hematopoietic cancers, with special focus on changes in risk over time, time since exposure, and robustness to confounding and exposure classification assumptions.

Main Contributions

- Provided extended long-term prospective data (>40 years) for a large cohort exposed to formaldehyde.
- Quantitatively estimated occupational exposure and analyzed risk for multiple lymphohematopoietic cancer subtypes.
- Found significant associations between peak formaldehyde exposure and mortality from lymphohematopoietic cancers and Hodgkin lymphoma; and observed time-dependent risk patterns, especially for myeloid leukemia.
- Systematically addressed possible confounding and exposure misclassification through sensitivity analyses.
- Integrated findings with evolving evidence and IARC carcinogenicity evaluations, highlighting the need for further research.

Main Conclusions and Limitations

- Results are consistent with a possible link between high-level peak occupational formaldehyde exposure and increased mortality from myeloid leukemia, Hodgkin lymphoma, and perhaps multiple myeloma, with strongest associations appearing in earlier follow-up periods and within 25 years of exposure.
- Patterns over time suggest a short induction/incubation period, but declining risks in later years may indicate chance, survivor bias, or exposure misclassification.
- Limitations include reliance on mortality data (not incidence), possible misclassification of exposure (especially post-1980), and limited specificity of death certificate diagnoses.
- Further epidemiological and mechanistic studies are warranted to clarify associations and causal mechanisms.

ASP Evidence Review Bundle

Paper: Beane_Freeman_JNCI_101_751_2009.pdf

Generated: 2026-08-18 14:59:29 MDT

Model: gpt-4.1

Available Reports

Module	Status	Link	Notes
Study Summary	Completed	Open report	summary
Extracted Causal Claims	Completed	Open report	Verbatim causal and policy-relevant claims extracted from the paper
Grouped STROBE Review	Completed	Open report	full_review
Causal Critique	Completed	Open report	claim_level_review
Statistical Methods Review	Completed	Open report	methods_review
SR/MA Review	Screening only	Open report	screening_only
Risk-of-Bias Review	Screening only	Open report	screening_only
Mechanistic Review	Screening only	Open report	screening_only
Cross-Module Synthesis	Completed	Open report	synthesis
Overall Assessment	Completed	Open report	integrated_assessment

Modules run in this order: aspsum, aspstroke, aspcrit, aspmethods, aspprima, asproba, aspmech.

Paper: Beane_Freeman_JNCI_101_751_2009.pdf

Generated: 2026-08-18 14:58 MDT **Model:** gpt-4.1

ASP Statistical/Methodological Validity Review

Domain	Strengths	Concerns	Evidence/Quotes	Overall Appraisal
M1. Exposure measurement and measurement error	- Comprehensive exposure assessment using work histories, job/department/task coding, and historical measurement data. - Exposure indices developed: peak exposure (categorical), average intensity, cumulative exposure (quantitative, time-dependent). - Some recognition of measurement error and potential for non-differential misclassification. - Sensitivity analyses conducted regarding post-1980 exposure assumptions.	- No direct quantitative correction for measurement error or errors-in-variables bias (e.g., regression calibration not used). - Exposure after 1980 not measured: primary analysis assumed zero exposure; sensitivity analyses only explored best/worst-case scenarios. - Validation of exposure estimates limited; only modest comparison to available measurement data (r=0.50, small sample).	"Exposure to formaldehyde was estimated for each job from individual work histories...using current and past measurement data." (p. 753) "Although validation was not possible, effects of misclassification may be gauged by comparing full-shift exposure estimates with the monitoring data. A sample... showed a correlation coefficient of 0.50 (8)." (p. 759) "We assumed that zero exposure occurred after [1980]...carried out two sensitivity analyses..." (p. 754)	Reasonable for historical cohort data, but could underestimate error and bias; measurement error likely biases against detection of associations.
M2. Missing data and selection mechanisms	- Extensive tracing process to ascertain deaths and causes (including linking to NDI, Social Security, etc.). - Discussion of unexplained missed deaths in earlier tracing, with transparent reporting regarding inclusions.	- No quantitative analysis or modeling of missing data patterns. - Missing post-1980 exposure data handled only via deterministic sensitivity analyses (not via multiple imputation or robust quantitative methods). - No test or discussion of MCAR/MAR/MNAR assumptions.	"We were unable to access the full file provided by NDI after the initial 1994 linkage...we have included them in all analyses for this update." (p. 752) "We may have underascertained cases because we considered an individual alive if NDI Plus did not identify the subject as deceased." (p. 759)	Some transparency about limitations, but limited implementation of advanced missing data methods.
M3. Omitted, latent, and unmeasured variables	- Extensive efforts to control for potential confounding by other occupational exposures (benzene, wood dust, etc) and plant, pay category, demographics. - Sensitivity analyses including and excluding subjects with benzene exposure show	- No use of negative or positive controls, instrumental variables, or quantitative bias analysis. - No formal modeling of potential latent variable structure in exposure or outcome. - Possible unmeasured confounding remains,	"Controlling for duration of exposure to 11 potentially confounding substances...did not meaningfully change results...Excluding individuals with possible benzene exposure...did not change the relative risk..." (p. 755)	Plausibly important omitted variable bias is acknowledged but not fully ruled out; mainstream confounders addressed carefully, other bias less so.

Most Important Methodological Strengths

- Comprehensive, quantitative historical exposure assessment with multiple indices (peak, average, cumulative).
- Detailed, transparent reporting of cohort construction, death ascertainment, and exposure estimation processes.
- Thoughtful and extensive sensitivity analyses for exposure assumptions, lag periods, confounder control, and time-windows.
- Statistical modeling (time-dependent Poisson regression) appropriate to cohort data structure.
- Pragmatic acknowledgment of limitations and their likely bias direction.

Most Important Methodological Concerns

- Exposure misclassification is likely, with only partial validation and mostly qualitative treatment of measurement error—no correction for error-induced bias.
- No formal quantitative modeling or imputation for missing data; unknown post-1980 exposures only addressed with deterministic scenarios.
- Confounding by unmeasured variables (e.g., socioeconomic status, unmeasured co-exposures) cannot be entirely ruled out; negative/positive controls not used.
- No use of more advanced model uncertainty methods (e.g., model averaging) or bias quantification using external data.

Implications for Interpreting the Paper

- The methods used are generally appropriate for a historical occupational cohort with the data available, and the sensitivity analyses provide some reassurance on key design choices.
- The principal threat to validity is likely to be non-differential exposure misclassification and some potential for unmeasured confounding, both of which would, if anything, tend to bias estimates towards the null rather than create spurious positive associations.
- The lack of formal quantification of measurement error/missing data uncertainty means that effect sizes should be regarded as minimum estimates: more sophisticated modern approaches could potentially adjust these upwards, but would require additional data.

ASP Evidence Review Bundle

Paper: Beane_Freeman_JNCI_101_751_2009.pdf

Generated: 2026-08-18 14:59:29 MDT

Model: gpt-4.1

Available Reports

Module	Status	Link	Notes
Study Summary	Completed	Open report	summary
Extracted Causal Claims	Completed	Open report	Verbatim causal and policy-relevant claims extracted from the paper
Grouped STROBE Review	Completed	Open report	full_review
Causal Critique	Completed	Open report	claim_level_review
Statistical Methods Review	Completed	Open report	methods_review
SR/MA Review	Screening only	Open report	screening_only
Risk-of-Bias Review	Screening only	Open report	screening_only
Mechanistic Review	Screening only	Open report	screening_only
Cross-Module Synthesis	Completed	Open report	synthesis
Overall Assessment	Completed	Open report	integrated_assessment

Modules run in this order: aspsum, aspstroke, aspcrit, aspmethods, aspprisma, asprob, aspmech.

Document: Beane_Freeman_JNCI_101_751_2009
Generated: 2026-08-18 15:01:49 MDT **Model:** gpt-4.1
Inputs: synthesis, causal critique, key assumptions, references, and reference support audit where available.

Overall Assessment

The Beane Freeman et al. (2009) study presents a well-conducted, large-scale, prospective cohort analysis of occupational formaldehyde exposure and mortality from lymphohematopoietic malignancies among US industrial workers. The evidence is carefully and transparently reported. Statistical associations—particularly with high peak formaldehyde exposures and mortality from Hodgkin lymphoma (statistically significant) and myeloid leukemia (non-significant elevation)—are robustly observed. However, the study is explicit about substantial uncertainties due to exposure misclassification (especially post-1980), measurement reliance on mortality rather than incidence, and incomplete control for confounding. Importantly, conclusions and recommendations are appropriately cautious, largely hypothesis-generating, and align with the strengths and limitations of the findings. Calls for regulatory action are avoided, with further research highlighted as the appropriate path forward.

Key Conclusions and Support

Conclusion or Recommendation	Support Level	Best Supporting Evidence	Main Caveats	Assessment
Evaluation of risks over time suggests a possible link between formaldehyde exposure and lymphohematopoietic malignancies, particularly myeloid leukemia but also perhaps Hodgkin lymphoma and multiple myeloma.	Partial	Beane Freeman et al. (2009) cohort data; significant association for Hodgkin lymphoma with high peak exposure; non-significant elevation for myeloid leukemia.	Exposure misclassification (post-1980), outcome ascertainment limited to mortality data, small case numbers for rare cancers, possible residual confounding.	Supported for association, especially for Hodgkin lymphoma at high peak exposures; less so for myeloid leukemia and multiple myeloma; causality not established.
Observed patterns could be due to chance but are also consistent with a causal association within the relatively short induction–incubation periods characteristic of leukemogenesis.	Partial	Empirical time-trend and latency analysis (declining risks with increasing time since exposure).	Temporal pattern is suggestive but alternative explanations possible (chance, bias, exposure/outcome misclassification).	Interpretation is appropriate—pattern compatible with causality but not definitive; alternative explanations not excluded.
The long-term follow-up of industrial workers in this study suggested a possible association between formaldehyde exposure and mortality due to lymphohematopoietic malignancies, particularly myeloid leukemia.	Partial to Moderate	Main cohort data, robust statistical methods, replication of earlier findings for Hodgkin lymphoma.	Limited by outcome data based on mortality, incomplete exposure data after 1980, and confounding.	Findings justify describing the association as "possible" rather than established; reflects the preponderance of evidence.



Document: Beane_Freeman_JNCI_101_751_2009
Generated: 2026-08-18 15:01:49 MDT **Model:** gpt-4.1
Inputs: synthesis, causal critique, key assumptions, references, and reference support audit where available.

Overall Assessment

The Beane Freeman et al. (2009) study presents a well-conducted, large-scale, prospective cohort analysis of occupational formaldehyde exposure and mortality from lymphohematopoietic malignancies among US industrial workers. The evidence is carefully and transparently reported. Statistical associations—particularly with high peak formaldehyde exposures and mortality from Hodgkin lymphoma (statistically significant) and myeloid leukemia (non-significant elevation)—are robustly observed. However, the study is explicit about substantial uncertainties due to exposure misclassification (especially post-1980), measurement reliance on mortality rather than incidence, and incomplete control for confounding. Importantly, conclusions and recommendations are appropriately cautious, largely hypothesis-generating, and align with the strengths and limitations of the findings. Calls for regulatory action are avoided, with further research highlighted as the appropriate path forward.

Key Conclusions and Support

Conclusion or Recommendation	Support Level	Best Supporting Evidence	Main Caveats	Assessment
Evaluation of risks over time suggests a possible link between formaldehyde exposure and mortality	Partial	Beane Freeman et al. (2009) cohort data; significant association for Hodgkin lymphoma	Exposure misclassification (post-1980), outcome uncertainty limited to mortality	Supported for association, especially for Hodgkin lymphoma at high exposure levels



Key Conclusions and Support

Conclusion or Recommendation	Support Level	Best Supporting Evidence	Main Caveats	Assessment
Evaluation of risks over time suggests a possible link between formaldehyde exposure and lymphohematopoietic malignancies, particularly myeloid leukemia but also perhaps Hodgkin lymphoma and multiple myeloma.	Partial	Beane Freeman et al. (2009) cohort data; significant association for Hodgkin lymphoma with high peak exposure; non-significant elevation for myeloid leukemia.	Exposure misclassification (post-1980), outcome ascertainment limited to mortality data, small case numbers for rare cancers, possible residual confounding.	Supported for association, especially for Hodgkin lymphoma at high peak exposures; less so for myeloid leukemia and multiple myeloma; causality not established.
Observed patterns could be due to chance but are also consistent with a causal association within the relatively short induction-incubation periods characteristic of leukemogenesis.	Partial	Empirical time-trend and latency analysis (declining risks with increasing time since exposure).	Temporal pattern is suggestive but alternative explanations possible (chance, bias, exposure/outcome misclassification).	Interpretation is appropriate—pattern compatible with causality but not definitive; alternative explanations not excluded.
The long-term follow-up of industrial workers in this study suggested a possible association between formaldehyde exposure and mortality due to lymphohematopoietic malignancies, particularly myeloid leukemia.	Partial to Moderate	Main cohort data, robust statistical methods, replication of earlier findings for Hodgkin lymphoma.	Limited by outcome data based on mortality, incomplete exposure data after 1980, and confounding.	Findings justify describing the association as "possible" rather than established; reflects the preponderance of evidence.
Statistically significant associations between formaldehyde exposure and Hodgkin lymphoma as well as a possible association with multiple myeloma are observed, but the evidence is limited and based on small numbers.	Partial	Statistical significance for Hodgkin lymphoma (peak exposure), non-significant elevation for multiple myeloma.	Very low case numbers increase chance findings; outcome misclassification.	The claim is accurately limited by the rarity of outcomes and methodologic challenges.
There is little or no evidence of association between formaldehyde exposure and lymphohematopoietic malignancies when exposure is measured as average intensity or cumulative exposure.	Strong	Cohort analysis of different exposure metrics; null/statistically insignificant results for average/cumulative exposure.	Same outcome misclassification and exposure assessment issues, but consistent null findings here.	Clear and well-supported within the cohort data provided.
Further epidemiological study and exploration of potential molecular mechanisms are warranted.	Strong	Cohort's empirical results, statistical uncertainties, mechanistic gaps, and literature (esp. IARC evaluation).	None—recommendation is proportional to the strength and limitations of evidence.	Strongly justified and appropriately cautious.
The IARC recently classified formaldehyde	Strong (for context)	Direct IARC monograph citation.	N/A—contextual summary of agency review.	Accurate use of authoritative summ

Key Recommendations and Support

- **Recommendation:** Further epidemiological and mechanistic research should be conducted to clarify the possible association between formaldehyde exposure and lymphohematopoietic malignancy risks.
Support Level: Strong.
Comment: The main body of evidence, uncertainties, and gaps support this as a fully justified, proportional recommendation that does not go beyond the presented evidence.
- **Recommendation:** No direct regulatory or policy action is recommended based solely on this study.
Support Level: Strong.
Comment: The cautious interpretation and explicit lack of causal certainty correctly constrain the implications for policy.

Where the Evidence Is Strongest

- Null associations: The absence of associations for average or cumulative exposure metrics is consistently observed and robust to sensitivity analyses.
- Statistical association between high peak exposure and Hodgkin lymphoma: This finding is statistically significant, internally consistent, and repeatedly observed.
- Appropriateness of recommendation for further research: The methodological care and identified uncertainties combine to make the call for further study fully evidence-based.
- Reporting discipline and accurate context statements (e.g., referencing IARC determinations).

Where Support Is Partial, Contextual, or Unclear

- The core association between formaldehyde exposure and myeloid leukemia is only partly supported: the elevation is not statistically significant for the full period, and effect sizes are unstable over time.
- Results regarding multiple myeloma and other rare outcomes are fundamentally limited by very small case numbers, producing instability and risk of chance associations.
- Causal inference is explicitly hedged and cannot be confirmed due to exposure/outcome misclassification, possible residual/unmeasured confounding, and incomplete mechanistic evidence.
- Underlying assumptions about post-1980 exposure scenarios, completeness of death certificate data, and referent group suitability are only partly addressed or remain open to bias.

Key Assumptions Affecting Interpretation



Key Assumptions Affecting Interpretation

- **Exposure assessment accuracy:** Assumed to reliably assign peak, average, and cumulative exposure, yet direct validation was not done and errors could bias associations (likely toward the null, per internal discussion).
- **Confounding control:** Adjustment for major occupational and demographic confounders, but cannot fully rule out unmeasured or residual confounding.
- **Mortality-based outcome ascertainment:** Assumes reliable coding of lymphohematopoietic malignancies and subtypes; acknowledged limitations could impact results, especially for rare cancer types.
- **Post-1980 exposure estimation:** Requires assumption of little or no new exposure, addressed only with sensitivity analyses.
- **Pooling of data:** Assumes no major effect modification by sex, race, or facility, which was empirically checked and found absent in this cohort.
- **Peak exposure as key etiologic metric:** Chosen due to stronger associations but with only limited empirical justification.

Reference-Support Issues

- **Direct support:** The core document itself is the main source of evidence for all association claims. Its own analyses and sensitivity checks drive the conclusions.
- **Context and corroboration:** Prior cohort analyses (Hauptmann et al. 2003; Blair et al. 1986) offer relevant precedents and design continuity, while Pinkerton et al. 2004 and Coggon et al. 2003 provide contextual benchmarks showing weaker or negative associations in other cohorts.
- **Mechanistic evidence:** IARC 2006 is referenced for overall plausibility, but no new mechanistic data is generated or critically appraised by the present study.
- **Supporting statements on misclassification and bias:** Methodological texts (Checkoway et al. 2004) and cancer classification accuracy studies (Percy et al. 1990) are used to justify bias discussions.
- **Limitations:** None of the cited references can fill the critical gaps in exposure adjudication post-1980 or incidence-based outcome ascertainment, nor do they enable firmer causal inference in this population.

Bottom-Line Appraisal

This study provides useful, robust, but limited evidence for an association—especially for Hodgkin lymphoma—with high peak occupational formaldehyde exposure. It is exemplary in methodological transparency, risk acknowledgement, and interpretive restraint. No claims of established causality or immediate policy implications are made, and none would be warranted on the basis of these results. Rather, the study's strength lies in precise characterization of remaining scientific uncertainties and in providing a careful, evidence-informed rationale for further research. Policymakers and health risk assessors should weigh these findings in the context of an evolving, still-inconclusive evidence base, using the results to prioritize new studies and better-designed epidemiologic and mechanistic investigations.



Critical Synthesis: Formaldehyde Exposure and Leukemia Risk

Evidence-calibrated, source-traceable cross-document critical synthesis based on 70 normalized propositions from 15 papers.

Traceability rule: Representative quotations are inserted verbatim from the standardized ASP source-evidence registry. Author statements are distinguished from the evidence needed to establish them.

Overall Assessment

Across the critical corpus, the strongest defensible conclusion is that current epidemiological and mechanistic evidence does not support a robust or causal relationship between occupational formaldehyde exposure and risk of leukemia (specifically myeloid leukemia, including AML and CML) in working populations, as would be required for regulatory classification or direct intervention. This conclusion is observed consistently in methodologically robust studies that adjust for major confounders and are most resistant to bias, and is further constrained by a lack of mechanistic (biological or toxicological) plausibility. While some studies, especially those with greater limits in design and exposure assessment, report positive statistical associations, these signals either disappear or are greatly attenuated in analyses accounting for risk of bias or confounding. Null findings are observed across multiple large cohorts and confirmatory analyses, but limited power, exposure misclassification, outcome ascertainment uncertainties, and incomplete confounder control mean that small or moderate effects cannot be definitively excluded. Statistically significant associations for certain rare outcomes or in weakly controlled or selected cohorts should be viewed as hypothesis-generating only, not causal. The evidence does not justify immediate regulatory or policy action specifically targeting formaldehyde as a leukemogen, and the strongest recommended actions are for further high-quality research and ongoing surveillance, not population-level interventions. Strong prior regulatory causal claims are not substantiated by current evidence. The corpus robustly supports restraint in policy or risk management decisions based solely on present data and recommends that the research community prioritize improved mechanistic, exposure, and bias analysis approaches.



Proposition-Level Evidence Overview

Use the links in this table to jump directly to detailed proposition families or representative source quotations.

Canonical proposition	Causal support	Papers	Representative source evidence	Main limiting evidence
Occupational formaldehyde exposure does not clearly increase the risk of myeloid leukemia (including AML and CML) in working populations, and strong regulatory causal claims are not justified by the available epidemiologic or mechanistic evidence.	Causal support is weak for a formaldehyde–leukemia relationship and moderate-to-strong against it; does not achieve high robustness. No evidence for intervention effect.	Lauer_et_al_2024_triangulation_with_supplements , Coggon_et_al_2014 , Checkoway_et_al_2015 , Pira_Cancer_Causes_Control_25_1343_2014 , Meyers_Am_J_Ind_Med_56_1027_2013 , Hauptmann_J_Natl_Cancer_Inst_101_1696_2009	Lauer_et_al_2024_triangulation_with_supplements — empirical result or observation Pira_Cancer_Causes_Control_25_1343_2014 — empirical result or observation Checkoway_et_al_2015 — author interpretation or causal conclusion	Some case-control/PMR/occupational studies with weak exposure assessment (e.g. Hauptmann 2009, Meyers 2013, Walrath studies) report positive associations, but methodological reviews attribute these to confounding, exposure misclassification, or selection bias. All null findings limited by inability to rule out small effect sizes due to lack of power and residual bias.
Observed positive associations between occupational formaldehyde exposure and leukemia or other lymphohematopoietic cancers in select studies are likely artefacts of methodological limitations and do not provide robust causal evidence.	Support for causality is weak or negative; associations are best interpreted as artefacts, not actionable causal evidence.	Lauer_et_al_2024_triangulation_with_supplements , Hauptmann_J_Natl_Cancer_Inst_101_1696_2009 , Blair_Am_J_Ind_Med_40_3_2001 , Walrath_Cancer_Res_44_4638_1984 , Meyers_Am_J_Ind_Med_56_1027_2013	Lauer_et_al_2024_triangulation_with_supplements — empirical result or observation Blair_Am_J_Ind_Med_40_3_2001 — recommendation Meyers_Am_J_Ind_Med_56_1027_2013 — limitation or caveat	Persistence of positive effect estimates in some lower-quality studies; possibility, not ruled out, that a small minority of cases reflect true effects. Interpretive uncertainty remains due to lack of

Current evidence cannot rigorously exclude small or moderate effects of occupational formaldehyde exposure on leukemia risk; absence of observed effect is not definitive proof of safety.	Cannot claim robust safety (i.e., non-carcinogenicity) for all exposure scenarios—evidence limits rather than proves absence.	Lauer_et_al_2024_trianguation_with_supplements, Coggon_et_al_2014, Checkoway_et_al_2015	Lauer_et_al_2024_trianguation_with_supplements — limitation or caveat	No single study or meta-analysis offers sufficient certainty to definitively exclude all possible effects. Lack of statistical power to detect small or rare effects in any design.
Current evidence does not support immediate or strong regulatory, policy, or workplace intervention regarding formaldehyde as a cause of leukemia; research and surveillance should be prioritized.	No robust causal support; interventional estimates are not supported. Major policy action would outpace available evidence.	Lauer_et_al_2024_trianguation_with_supplements, Coggon_et_al_2014, Checkoway_et_al_2015, Beane_Freeman_JNCI_101_751_2009, Meyers_Am_J_Ind_Med_56_1027_2013, Walrath_Cancer_Res_44_4638_1984, Walrath_Int_J_Cancer_31_407_1983	Lauer_et_al_2024_trianguation_with_supplements — recommendation Coggon_et_al_2014 — author interpretation or causal conclusion	Findings of increased risk in historically uncontrolled PMR and case-control studies (see F2)—but these are specifically reviewed and judged not actionable.
There is a lack of mechanistic (biological, toxicological, experimental) evidence supporting a plausible pathway by which inhaled formaldehyde causes myeloid leukemia in humans at exposure levels observed in occupational settings.	No support for mechanistic plausibility; further weakens any causal claim from epidemiology alone.	Checkoway_et_al_2015, Lauer_et_al_2024_trianguation_with_supplements, Hauptmann_J_Natl_Cancer_Inst_101_1696_2009	Checkoway_et_al_2015 — empirical result or observation Lauer_et_al_2024_trianguation_with_supplements — limitation or caveat	Novel, as-yet-undescribed mechanisms could be discovered. Current experimental and toxicological approaches may not detect some relevant modes of action.

Detailed Proposition-Level Synthesis

F1: Occupational formaldehyde exposure does not clearly increase the risk of myeloid leukemia (including AML and CML) in working populations, and strong regulatory causal claims are not justified by the available epidemiologic or mechanistic evidence.

Directly supported empirical finding

Most methodologically robust studies (Lauer 2024, Coggon 2014, Checkoway 2015, Pira 2014) observe no statistically significant association between occupational formaldehyde exposure and myeloid leukemia risk after accounting for bias. Meta-analytic risk estimates attenuate to null when high-risk-of-bias studies are excluded.

Assumption-dependent stronger interpretation

If future studies could quantify and remove all sources of exposure misclassification, confounding, and selection bias, and if small but real effects are not being obscured, then formaldehyde would not be a leukemogen in typical workplace conditions.

Scope and boundaries

Supported for occupational exposures in US/European industrial and embalming settings studied to date; not directly generalizable to extremely high exposures or unstudied populations. Effect cannot be rigorously excluded for small or rare subgroups.

Papers

- Lauer_et_al_2024_trianguation_with_supplements
- Coggon_et_al_2014
- Checkoway_et_al_2015
- Pira_Cancer_Causes_Control_25_1343_2014
- Meyers_Am_J_Ind_Med_56_1027_2013
- Hauptmann_J_Natl_Cancer_Inst_101_1696_2009

Agreement pattern

Strong empirical and interpretive agreement among post-2010 cohort studies, critical syntheses, and meta-analyses; qualified only by limitations in exposure, confounding, and outcome ascertainment which limit ability to exclude small effects.

Evidence for

- Strong agreement among the highest-quality, lowest bias studies and meta-analyses.
- Attenuation of risk estimates after bias-based exclusion.
- Null findings for myeloid leukemia, even at highest exposure strata.

Evidence against or limiting

- Some case-control/PMR/occupational studies with weak exposure assessment (e.g. Hauptmann 2009, Meyers 2013, Walrath studies) report positive associations, but methodological reviews attribute these to confounding, exposure misclassification, or selection bias.
- All null findings limited by inability to rule out small effect sizes due to lack of power and residual bias.
- No formal quantitative bias analysis performed to definitively bound bias-induced null results.

Alternative interpretations

- Residual or unmeasured confounding, or misclassification, could mask a small true risk.
- Low event numbers or exposure prevalence limit power to detect rare or modest effects.
- Over-adjustment, healthy worker effect, or exposure proxies may bias toward the null.

Representative Source Evidence

Lauer_et_al_2024_triangulation_with_supplements — empirical result or observation
reports/integrated_assessment.txt
Lauer_et_al_2024_triangulation_with_supplements::Q1

“Most methodologically robust epidemiological studies do not support a clear association between occupational formaldehyde exposure and risk of ML, AML, or CML; isolated studies reporting high risks are likely influenced by multiple sources of bias.”

Checkoway_et_al_2015 — author interpretation or causal conclusion

reports/causal_claims.txt

Checkoway_et_al_2015::Q1

“Findings from this re-analysis do not support the hypothesis that formaldehyde is a cause of AML.”

Why this matters: Re-analysis does not support AML causal relationship.

ASP assessment: Key null finding for AML, cornerstone for F1.

[↑ Back to overview](#)

Methodological assessment

Convergent null results are most credible (triangulation/meta-analyses, large cohorts), but no study achieves full confounder adjustment or direct measurement. Exposure misclassification and unmeasured confounders are recurring threats. Residual bias cannot be formally excluded, but null evidence is consistently present across independent studies. Some studies are hypothesis-generating only.

Causal support assessment

Causal support is weak for a formaldehyde–leukemia relationship and moderate-to-strong against it; does not achieve high robustness. No evidence for intervention effect.

Key inferential assumptions and validation status

- Exposure proxies reflect cumulative/peak exposure (validation: only partial).
- No major unmeasured confounding remains (validation: not achieved).
- Mortality data adequately reflect incidence for leukemia outcomes (validation: assumed, not demonstrated).
- No effect-masking from healthy worker bias or under-ascertainment (validation: not directly tested).

Unresolved uncertainties

- Presence and magnitude of any small true formaldehyde effect on leukemia risk.
- Residual confounding by co-exposures (e.g., benzene), lifestyle factors.
- Accuracy of exposure categorization, especially at high ends of exposure and post-1980.
- Effect of long latency or rare subtypes (e.g., CML, HL, rare solid tumors).

What evidence would discriminate?

- Prospective cohort or nested case-control study with direct, repeated quantitative exposure measurement (biomarkers, personal monitoring), robust confounder and co-exposure adjustment, and formal quantitative bias analysis.
- Evidence from randomized or quasi-experimental exposure reductions with sufficient sample size and outcome ascertainment.
- Mechanistic validation of a plausible biological pathway from inhaled/formaldehyde exposure to leukemogenesis at observed exposure levels.

Synthesis conclusion

The risk of myeloid leukemia due to occupational formaldehyde exposure is most credibly null or very small in the general working populations studied to date; current high-confidence regulatory, legal, or causal claims are not justified.

[↑ Back to overview](#)

F2: Observed positive associations between occupational formaldehyde exposure and leukemia or other lymphohematopoietic cancers in select studies are likely artefacts of methodological limitations and do not provide robust causal evidence.

Directly supported empirical finding

Meta-analysis and triangulation methods (Lauer 2024) show that positive risk estimates in some studies become null when high-risk-of-bias studies (those with proxy exposure assessment or uncontrolled confounding) are excluded. Studies with positive findings (Hauptmann 2009, Blair 2001, Meyers 2013, Walrath) have inconsistent exposure assessment, reliance on occupation as proxy, and lack of confounder control.

Assumption-dependent stronger interpretation

Major Agreements Across Papers

- Current epidemiological evidence does not support a robust or causal association between occupational formaldehyde exposure and myeloid leukemia risk.
- Positive associations in less methodologically robust studies are likely artefacts of bias or confounding.
- Mechanistic (biological/toxicological) evidence for formaldehyde-induced leukemogenesis in humans is lacking.
- All current high-quality reviews and syntheses recommend further research and surveillance rather than regulatory policy or intervention.

Major Disagreements and Tensions

- Some older studies or studies with limited design suggest positive statistical associations, but these are largely dismissed as artefactual upon critical synthesis and triangulation.
- Unresolved debate about the degree to which null findings can be interpreted as evidence of safety (absence of evidence vs. evidence of absence).

Cross-Paper Methodological Pattern

Null or near-null findings are consistent across higher quality, larger, and more modern cohort studies, but all are methodologically limited by: (1) exposure misclassification—common use of proxy metrics (ordinal job categories, duration of employment); (2) incomplete confounder adjustment—often lacking smoking, alcohol, and co-exposure data; (3) insufficient statistical power to detect small effect sizes for rare outcomes; (4) lack of mechanistic support; and (5) absence of advanced quantitative bias analysis to rule out alternative explanations for observed findings.



Evidence Gaps

- Direct, validated, and quantitative individual-level exposure assessment.
- Comprehensive confounder and co-exposure measurement and adjustment.
- Mechanistic evidence bridging exogenous formaldehyde exposure and human leukemogenesis.
- Large-scale, prospective designs or natural/quasi-experimental reductions in workplace exposure with pre-registered bias analysis.
- Studies with sufficient power to detect moderate-to-small risk increases.

Conclusions Supported by the Current Evidence

- Occupational formaldehyde exposure does not robustly increase risk of myeloid leukemia in working populations studied to date.
- Policy, regulatory, or interventional action is not justified by the present evidence and should await further research.
- Further high-quality research, including improved exposure assessment, mechanistic studies, and ongoing occupational surveillance, is warranted.

Conclusions Not Supported by the Current Evidence

- Regulatory or authoritative causal declarations that occupational formaldehyde exposure is a cause of leukemia (as of the evidence reviewed).
- Population-level interventions to reduce formaldehyde exposure on the basis of leukemia risk.
- Claims that occupational or environmental exposures to formaldehyde have definitively been proven safe (i.e., that all risk is excluded).



Priority Future Evidence

- Prospective, quantitative, and highly controlled studies with direct exposure biomonitoring and rigorous confounder measurement.
- Experimental and mechanistic research that demonstrates or rules out a plausible pathway from inhaled formaldehyde to leukemogenesis.
- Integrated, pre-registered bias analyses to formally quantify the possible effects of residual confounding and measurement error.
- Replication of findings in high-exposure subgroups or contexts (if present).

Bottom Line

The corpus robustly supports the conclusion that occupational formaldehyde exposure does not clearly or causally increase the risk of myeloid leukemia, and strong regulatory or intervention claims are not justified by the available evidence. Statistical associations reported in older and weaker studies are best explained by methodological limitations rather than by a true causal relationship. While small risks cannot be fully excluded, the lack of both epidemiologic and mechanistic evidence argues for a policy and scientific stance of restraint, prioritizing improved research and surveillance over regulatory action at this time.



Visualizations & Interactive Evidence Browser

How to assimilate key findings in less than 5 seconds (after learning how)?

Retrieved and Reviewed Papers

Evidence node: Ultra-Processed Foods

Study	Title	Local PDF	Evidence Bundle	Evidence Browser & Scores	Briefs	Methodology Score	Domain Scores
Bonaccio, 2023	UPF consumption, all-cause mortality, and cardiovascular mortality in type 2 diabetes	Open PDF	Open Bundle	Open EB	Open Briefs	0.591 ●	A 0.500; B 0.826; C 0.532; D 0.682; E 0.560
Dai, 2024	Umbrella review and updated meta-analyses of UPFs and human health	Open PDF	Open Bundle	Open EB	Open Briefs	0.414 ●	A 0.321; B 0.800; C 0.288; D 0.655; E 0.349
Gauci, 2025	UPF cardiovascular mortality	Open PDF	Open Bundle	Open EB	Open Briefs	0.557 ●	A 0.500; B 0.941; C 0.466; D 0.701; E 0.467
Fu, 2024	UPFs and hearing loss	Open PDF	Open Bundle	Open EB	Open Briefs	0.416 ●	A 0.417; B 0.398; C 0.357; D 0.510; E 0.510
Henney, 2024	UPFs and dementia	Open PDF	Open Bundle	Open EB	Open Briefs	0.442 ●	A 0.333; B 0.623; C 0.407; D 0.510; E 0.441
Lane, 2024	UPF health outcomes	Open PDF	Open Bundle	Open EB	Open Briefs	0.321 ●	A 0.364; B 0.408; C 0.192; D 0.712; E 0.038
Levy, 2023	How and why ultra-processed foods harm human health	Open PDF	Open Bundle	Open EB	Open Briefs	0.035 ●	A 0.062; B 0.000; C 0.000; D 0.197; E 0.000
Liang, 2025	Ultra-processed foods and risk of all-cause mortality: updated systematic review and dose-response meta-analysis	Open PDF	Open Bundle	Open EB	Open Briefs	0.549 ●	A 0.429; B 0.771; C 0.365; D 0.702; E 0.754
Lutz, 2025	UPF and neuropsychiatric outcomes	Open PDF	Open Bundle	Open EB	Open Briefs	0.084 ●	A 0.125; B 0.009; C 0.052; D 0.304; E 0.000
Menegassi, 2025	Ultraprocessed food and risk of cancer: mechanistic pathways and public health implications	Open PDF	Open Bundle	Create EB	Open Briefs	—	—
Mendoza, 2024	UPF cardiovascular outcomes	Open PDF	Open Bundle	Open EB	Open Briefs	0.643 ●	A 0.500; B 0.861; C 0.586; D 0.625; E 0.739
Monteiro, 2025	Ultra-processed foods and human health: main thesis and evidence	Open PDF	Open Bundle	Open EB	Open Briefs	0.521 ●	A 0.500; B 0.394; C 0.500; D 0.806; E 0.517
Prescott, 2024	UPF and crime	Open PDF	Open Bundle	Open EB	Open Briefs	0.098 ●	A 0.156; B 0.009; C 0.000; D 0.392; E 0.165
Solans, 2021	Ultra-Processed Food and Chronic Lymphocytic Leukemia in the MCC-Spain Study	Open PDF	Open Bundle	Open EB	—	0.552 ●	A 0.500; B 0.733; C 0.429; D 0.727; E 0.556

Generated from current ASP bundle inventory; local PDFs are linked when available.

ASP Causal Evidence Browser

Interactive, read-only browser for hierarchical ASP causal scoring, drill-down evidence review, and gap diagnostics.

Navigation

Top

Up

Path

Overall

Jump to subsection

Select a subsection...

Load CSV

Browse...

No file selected

Loaded file

CAUSAL_EB_export.csv

Path:

/srv/connect/apps/asp_eb_henney_2024/CAUSAL_EB_export.csv | Working CSV:
CAUSAL_working_expanded_Henney_2024.csv

Selected Evidence

Click a row in the Gap, Strength, or Evidence

TOP

Overall Summary

Score 0.442 | LB 0.269 | UB 0.616 | D=391 Y=105 P=136 NO=150

SCORE

0.442

Primary score

LOWER BOUND

0.269

Yes / denominator

UPPER BOUND

0.616

(Yes + Partly) / denominator

COUNTS

D=391

Y=105 P=136 NO=150

Overview

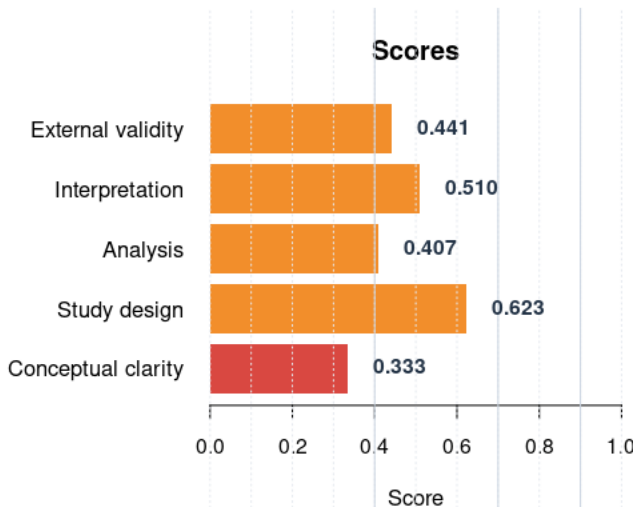
Domain Dashboard

Top Gaps

Evidence

Components

Direct subsections of the current node. Click a bar or table row to drill down.



Current Node Summary

Node type TOP

Label Top

Domain

Item

Children 5

Leaf rows in subtree 391

Section Snapshot

CHILDREN

5

LEAVES

391



ASP Causal Evidence Browser

Interactive, read-only browser for hierarchical ASP causal scoring, drill-down evidence review, and gap diagnostics.

Navigation

Top

Up

Path

Overall > Domain E: External validity

Jump to subsection

Select a subsection...

Load CSV

Browse... No file selected

Loaded file

CAUSAL_EB_export.csv

Path:
/srv/connect/apps/asp_eb_henney_2024/CAUSAL_EB_export.csv | Working CSV:
CAUSAL_working_expanded_Henney_2024.csv

Selected Evidence

Click a row in the Gap, Strength, or Evidence

DOMAIN

DOMAIN: External validity

Score 0.441 | LB 0.276 | UB 0.605 | D=76 Y=21 P=25 NO=30

SCORE

0.441

Primary score

LOWER BOUND

0.276

Yes / denominator

UPPER BOUND

0.605

(Yes + Partly) / denominator

COUNTS

D=76

Y=21 P=25 NO=30

Overview

Domain Dashboard

Top Gaps

Evidence

Components

Direct subsections of the current node. Click a bar or table row to drill down.



Current Node Summary

Node type	DOMAIN
Label	External validity
Domain	Robustness / generalizability
Item	
Children	4
Leaf rows in subtree	76

Section Snapshot

CHILDREN
4

LEAVES
76



ASP Causal Evidence Browser

Interactive, read-only browser for hierarchical ASP causal scoring, drill-down evidence review, and gap diagnostics.

Navigation

Top

Up

Path

Overall > Domain C: Analysis

Jump to subsection

Select a subsection...

Load CSV

Browse... No file selected

Loaded file

CAUSAL_EB_export.csv

Path:
/srv/connect/apps/asp_eb_henney_2024/CAUSAL_EB_export.csv | Working CSV:
CAUSAL_working_expanded_Henney_2024.csv

Selected Evidence

Click a row in the Gap, Strength, or Evidence

DOMAIN

DOMAIN: Analysis

Score 0.407 | LB 0.259 | UB 0.556 | D=135 Y=35 P=40 NO=60

SCORE

0.407

Primary score

LOWER BOUND

0.259

Yes / denominator

UPPER BOUND

0.556

(Yes + Partly) / denominator

COUNTS

D=135

Y=35 P=40 NO=60

Overview

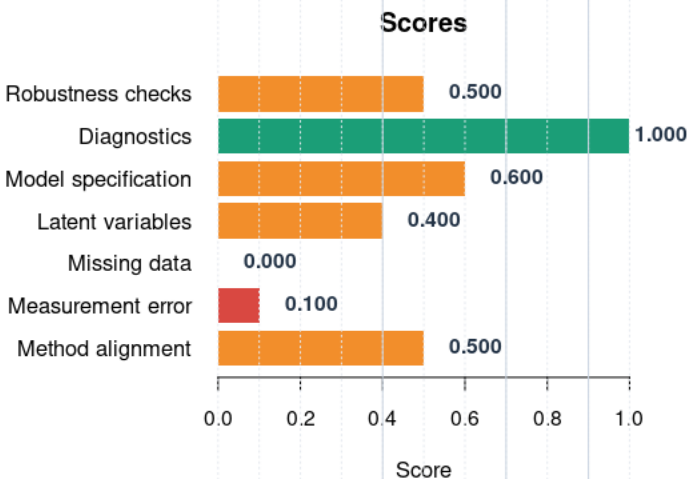
Domain Dashboard

Top Gaps

Evidence

Components

Direct subsections of the current node. Click a bar or table row to drill down.



Current Node Summary

Node type	DOMAIN
Label	Analysis
Domain	Analytic/internal validity
Item	
Children	7
Leaf rows in subtree	135

Section Snapshot

CHILDREN
7

LEAVES
135



ASP Causal Evidence Browser

Interactive, read-only browser for hierarchical ASP causal scoring, drill-down evidence review, and gap diagnostics.

Navigation

Top

Up

Path

Overall > Domain C: Analysis

Jump to subsection

Select a subsection...

Load CSV

Browse...

No file selected

Loaded file

CAUSAL_EB_export.csv

Path:
/srv/connect/apps/asp_eb_henney_2024/CAUSAL_EB_export.csv | Working CSV:
CAUSAL_working_expanded_Henney_2024.csv

Selected Evidence

Click a row in the Gap, Strength, or Evidence

DOMAIN

DOMAIN: Analysis

Score 0.407 | LB 0.259 | UB 0.556 | D=135 Y=35 P=40 NO=60

SCORE

0.407

Primary score

LOWER BOUND

0.259

Yes / denominator

UPPER BOUND

0.556

(Yes + Partly) / denominator

COUNTS

D=135

Y=35 P=40 NO=60

Overview Domain Dashboard Top Gaps Evidence

Text-first view of the largest reporting gaps and strengths in this section. Hover over truncated wording to see the full leaf text. Click any row to load the corresponding evidence into the Selected Evidence panel.

Largest Reporting Gaps

Show 25 entries Search:

Leaf	Question	Rating
C[C1]::C2.2	Effect of measurement error on identification assessed	NO
C[C1]::C2.3	Assumptions about measurement error or misclassification structure are stated and justified, including whether errors are differential, non-differential, classical, Berkson, or proxy-based	NO
C[C1]::C2.4	Validation, replicate measurement, calibration, or other correction information is used when key variables are	NO

Largest Reporting Strengths

Show 25 entries Search:

Leaf	Question	Rating
C[C1]::C4.2	Formal sensitivity analysis conducted	YES
C[C1]::C5.4	Uncertainty from plausible alternative model specifications is transparently disclosed	YES
C[C1]::C7.3	Results are compared across multiple plausible valid adjustment sets or covariate-selection strategies	YES
C[C2]::C4.2	Formal sensitivity analysis conducted	YES
C[C2]::C5.4	Uncertainty from plausible	YES

ASP Causal Evidence Browser

Interactive, read-only browser for hierarchical ASP causal scoring, drill-down evidence review, and gap diagnostics.

Navigation

Top

Up

Path

Overall > Domain C: Analysis

Jump to subsection

Select a subsection...

Load CSV

Browse... No file selected

Loaded file

CAUSAL_EB_export.csv

Path:
/srv/connect/apps/asp_eb_henney_2024/CAUSAL_EB_export.csv | Working CSV:
CAUSAL_working_expanded_Henney_2024.csv

Selected Evidence

Item: C[C1]::C5.4

DOMAIN

DOMAIN: Analysis

Score 0.407 | LB 0.259 | UB 0.556 | D=135 Y=35 P=40 NO=60

SCORE

0.407

Primary score

LOWER BOUND

0.259

Yes / denominator

UPPER BOUND

0.556

(Yes + Partly) / denominator

COUNTS

D=135

Y=35 P=40 NO=60

Overview

Domain Dashboard

Top Gaps

Evidence

Leaf-Level Evidence

Ratings, quotes, and explanations for leaves in the current section. Click any row to load the corresponding evidence into the Selected Evidence panel.

Show 50 entries

Search:

Leaf	Question	Rating	Evidence
All	All	All	All
C[C1]::C2.1	Measurement error evaluated for key variables	PARTIAL	Quote Explanation
C[C1]::C3.1	Missing data mechanism evaluated	NO	Quote Explanation
C[C1]::C4.1	Potential unmeasured confounding evaluated	PARTIAL	Quote Explanation
CG[CG1]::C1.1	Analytic methods align with causal estimand	PARTIAL	Quote

ASP Causal Evidence Browser

Interactive, read-only browser for hierarchical ASP causal scoring, drill-down evidence review, and gap diagnostics.

Navigation

Top

Up

Path

Overall > Domain C: Analysis

Jump to subsection

Select a subsection...

Load CSV

Browse... No file selected

Loaded file

CAUSAL_EB_export.csv

Path:
/srv/connect/apps/asp_eb_henney_2024/CAUSAL_EB_export.csv | Working CSV:
CAUSAL_working_expanded_Henney_2024.csv

Selected Evidence

Item: C[C1]::C2.1

DOMAIN

DOMAIN: Analysis

Score 0.407 | LB 0.259 | UB 0.556 | D=135 Y=35 P=40 NO=60

SCORE

0.407

Primary score

LOWER BOUND

0.259

Yes / denominator

UPPER BOUND

0.556

(Yes + Partly) / denominator

COUNTS

D=135

Y=35 P=40 NO=60

Overview

Domain Dashboard

Top Gaps

Evidence

Show 50 entries

Search:

Leaf

Question

Rating

Evidence

All

All

All

All

C[C1]::C2.1

Measurement error evaluated for key variables

PARTIAL

Quote

this meant there was no pre-determined cut-off to quantify exposure. Moderate intake in one population may exceed high intake in another. | whereas no studies with MCI as the outcome did so despite codes existing, and all included studies assessed diet using FFQs which increases recall bias and underreporting of true intake [54].

Explanation

ASP Causal Evidence Browser

Interactive, read-only browser for hierarchical ASP causal scoring, drill-down evidence review, and gap diagnostics.

Navigation

Top

Up

Path

Overall > Domain C: Analysis

Jump to subsection

Select a subsection...

Load CSV

Browse... No file selected

Loaded file

CAUSAL_EB_export.csv

Path:
/srv/connect/apps/asp_eb_henney_2024/CAUSAL_EB_export.csv | Working CSV:
CAUSAL_working_expanded_Henney_2024.csv

Selected Evidence

Item: C[C1]::C2.1

DOMAIN

DOMAIN: Analysis

Score 0.407 | LB 0.259 | UB 0.556 | D=135 Y=35 P=40 NO=60

SCORE

0.407

Primary score

LOWER BOUND

0.259

Yes / denominator

UPPER BOUND

0.556

(Yes + Partly) / denominator

COUNTS

D=135

Y=35 P=40 NO=60

Overview

Domain Dashboard

Top Gaps

Evidence

Show 50 entries

Search:

Leaf

Question

Rating

Evidence

All

All

All

All

C[C1]::C2.1

Measurement error evaluated for key variables

PARTIAL

Quote

Explanation

The paper qualitatively evaluates measurement error for key variables: it notes exposure classification/cut-off inconsistencies and acknowledges FFQ-related recall bias and underreporting, and coding differences for outcomes. However, there is no quantitative evaluation or correction, so support is incomplete.



Explanation

The paper qualitatively evaluates measurement error for key variables: it notes exposure classification/cut-off inconsistencies and acknowledges FFQ-related recall bias and underreporting, and coding differences for outcomes. However, there is no quantitative evaluation or correction, so support is incomplete.

Navigation

Top

Up

Path

Overall > Domain C: Analysis

Jump to subsection

Select a subsection...

Load CSV

Browse... No file selected

Loaded file

CAUSAL_EB_export.csv

Path:
/srv/connect/apps/asp_eb_henney_2024/CAUSAL_EB_export.csv | Working CSV:
CAUSAL_working_expanded_Henney_2024.csv

Selected Evidence

Item: C[C1]::C2.1

RATING

PARTIAL

DOMAIN

DOMAIN: Analysis

Score 0.407 | LB 0.259 | UB 0.556 | D=135 Y=35 P=40 NO=60

SCORE

0.407

Primary score

LOWER BOUND

0.259

Yes / denominator

UPPER BOUND

0.556

(Yes + Partly) / denominator

COUNTS

D=135

Y=35 P=40 NO=60

Overview Domain Dashboard Top Gaps Evidence

C[C1]::C3.1	Missing data mechanism evaluated	NO	Quote Explanation
C[C1]::C4.1	Potential unmeasured confounding evaluated	PARTIAL	Quote Explanation
CG[CG1]::C1.1	Analytic methods align with causal estimand	PARTIAL	Quote Explanation
CG[CG1]::C5.1	Risk of model misspecification evaluated	YES	Quote Explanation
CG[CG1]::C6.1	Model diagnostics conducted	YES	Quote Explanation
CG[CG1]::C7.1	Sensitivity analyses performed	YES	Quote Explanation
C[C1]::C2.2	Effect of measurement error on identification assessed	NO	Quote

Visually assessing a large literature

Interactive heat maps sorted by methodology scores within “lenses”

https://moirai.shinyapps.io/asp_methodology_audit/

ASP Methodology Audit Builder v1.4.1

Projects

Audit Matrix

Overview

Paper View

Frameworks

User Guide

Data

Color-blind colors

Project: NO2 and Asthma Methodology Audit

Cross-lens Meta View

A compact paper-by-lens summary. The dot is a conservative rule-based synthesis of the scorable criterion ratings within each lens; it is not an arithmetic quality score. Click a paper heading to sort lenses for that paper; click a lens heading to sort papers for that lens; click any summary cell to inspect the rating distribution and criteria driving it.

Default: alphabetical papers and standard lens order.

Reset sorting

Methodological lens	Kyrmánidou 2026 review	Johannessen 2026 Iceland	Lee 2024 meta-analysis	Ohyama 2022 HONO	Permaul 2020 SICAS	So 2025/26 Denmark	Yue 2017 mice	Clark 2010 BC
Causal Methodology Audit	G0 Y2 O5 R5	G0 Y4 O7 R1	G4 Y2 O5 R1	G6 Y4 O1 R1	G0 Y5 O6 R1	G1 Y6 O4 R1	G5 Y5 O1 R1	G0 Y7 O4 R1
Evidence-Synthesis Quality	G1 Y3 O5 R5	G0 Y0 O0 R0	G10 Y3 O2 R0	G0 Y0 O0 R0	G0 Y0 O0 R0	G0 Y0 O0 R0	G0 Y0 O0 R0	G0 Y0 O0 R0
Diagnostic Red Flags for Interventional Causal Claims	G0 Y1 O5 R4	G0 Y2 O8 R0	G4 Y1 O5 R0	G4 Y4 O1 R1	G0 Y5 O5 R0	G1 Y5 O4 R0	G2 Y6 O1 R1	G0 Y8 O2 R0
Statistical Analysis Quality	G0 Y2 O9 R2	G3 Y2 O9 R1	G7 Y3 O4 R0	G3 Y7 O3 R0	G4 Y5 O4 R2	G8 Y5 O2 R0	G2 Y8 O3 R0	G2 Y10 O2 R1
Confounding & Identification Audit	G0 Y0 O14 R1	G3 Y4 O3 R4	G0 Y7 O7 R0	G3 Y1 O0 R0	G6 Y1 O5 R2	G4 Y4 O5 R0	G4 Y0 O0 R0	G3 Y7 O2 R2

► How the meta-view summary dot is derived

Type here to search

64°F Mostly cloudy

10:36 PM 10/22/2026

ASP Methodology Audit Builder v1.4.1

Projects

Audit

Project: NO2 and Asthma Methodology Audit

Cross-lens Meta View

A compact paper-by-lens summary. The dot is a conservative heading to sort papers for that lens; click any summary cell to

Default: alphabetical papers and standard lens order.

Methodological lens

Kyrmanidou 2026

Causal Methodology Audit

Evidence-Synthesis Quality

Diagnostic Red Flags for Interventional Causal Claims

Statistical Analysis Quality

Confounding & Identification Audit

How the meta-view summary dot is derived

Kyrmanidou 2026 review — Causal Methodology Audit

Meta-view summary: Not addressed / omitted

Scorable criteria: 12 | Green 0 | Yellow 2 | Orange 5 | Red 5

Rule triggered: At least 30% of scorable criteria are red.

Criteria are shown from weakest to strongest; gray, N/A, and unrated items follow the scorable ratings.

Defined causal estimand

no common causal estimand

Well-defined intervention versions

policy recommendations not linked to intervention regimes

Explicit identification strategy

association synthesis without causal identification audit

Positivity / support assessed

support/overlap not assessed

Testable implications examined where available

no causal-model implications tested

Adjustment strategy matches causal structure

quality appraisal but no confounder/adjustment audit

Estimator matches estimand

heterogeneous measures summarized narratively

Reasonable robustness analyses examined

quality assessment and heterogeneity discussion, limited sensitivity

Rival explanations seriously examined

residual confounding acknowledged but not systematically evaluated

o 2025/26 Denmark

Yue 2017 mice

Clark 2010 BC

G1 Y6 O4 R1

G5 Y5 O1 R1

G0 Y7 O4 R1

G0 Y0 O0 R0

G0 Y0 O0 R0

G0 Y0 O0 R0

G1 Y5 O4 R0

G2 Y6 O1 R1

G0 Y8 O2 R0

G8 Y5 O2 R0

G2 Y8 O3 R0

G2 Y10 O2 R1

G4 Y4 O5 R0

G4 Y0 O0 R0

G3 Y7 O2 R2

Type here to search

64°F Mostly cloudy

10:37 PM 10/22/2026

ASP Methodology Audit Builder v1.4.1

Projects

Audit Matrix

Overview

Paper View

Frameworks

User Guide

Data

Color-blind colors

Project: NO2 and Asthma Methodology Audit

Cross-lens Meta View

A compact paper-by-lens summary. The dot is a conservative rule-based synthesis of the scorable criterion ratings within each lens; it is not an arithmetic quality score. Click a paper heading to sort lenses for that paper; click a lens heading to sort papers for that lens; click any summary cell to inspect the rating distribution and criteria driving it.

Default: alphabetical papers and standard lens order.

Reset sorting

Methodological lens	Kyrmánidou 2026 review	Johannessen 2026 Iceland	Lee 2024 meta-analysis	Ohyama 2022 HONO	Permaul 2020 SICAS	So 2025/26 Denmark	Yue 2017 mice	Clark 2010 BC	Gehring 2020 PIAMA	Lavigne 2018 Ontario	Pedersen 2023 Denmark	Shukla 2026 UK	Wang 2025 ECHO
Causal Methodology Audit	G0 Y2 O5 R5	G0 Y4 O7 R1	G4 Y2 O5 R1	G6 Y4 O1 R1	G0 Y5 O6 R1	G1 Y6 O4 R1	G5 Y5 O1 R1	G0 Y7 O4 R1	G0 Y6 O3 R1	G1 Y7 O3 R1	G2 Y7 O2 R1	G3 Y7 O2 R0	G0 Y8 O3 R1
Evidence-Synthesis Quality	G1 Y3 O5 R5	G0 Y0 O0 R0	G10 Y3 O2 R0	G0 Y0 O0 R0	G0 Y0 O0 R0	G0 Y0 O0 R0	G0 Y0 O0 R0	G0 Y0 O0 R0	G0 Y0 O0 R0	G0 Y0 O0 R0	G0 Y0 O0 R0	G0 Y0 O0 R0	G0 Y0 O0 R0
Diagnostic Red Flags for Interventional Causal Claims	G0 Y1 O5 R4	G0 Y2 O6 R0	G4 Y1 O5 R0	G4 Y4 O1 R1	G0 Y5 O5 R0	G1 Y5 O4 R0	G2 Y6 O1 R1	G0 Y6 O2 R0	G0 Y6 O2 R0	G2 Y7 O1 R0	G3 Y6 O1 R0	G3 Y5 O2 R0	G0 Y7 O3 R0
Statistical Analysis Quality	G0 Y2 O9 R2	G3 Y2 O9 R1	G7 Y3 O4 R0	G3 Y7 O3 R0	G4 Y5 O4 R2	G8 Y5 O2 R0	G2 Y8 O3 R0	G2 Y10 O2 R1	G2 Y10 O2 R1	G5 Y7 O2 R1	G5 Y7 O2 R1	G8 Y5 O2 R0	G2 Y12 O1 R0
Confounding & Identification Audit	G0 Y0 O14 R1	G3 Y4 O3 R4	G0 Y7 O7 R0	G3 Y1 O0 R0	G6 Y1 O5 R2	G4 Y4 O5 R0	G4 Y0 O0 R0	G3 Y7 O2 R2	G6 Y6 O2 R0	G6 Y4 O3 R1	G9 Y3 O0 R2	G3 Y2 O5 R1	G6 Y4 O3 R1
Assumptions & Identification Requirements	G0 Y3 O7 R2	G2 Y4 O6 R0	G2 Y6 O4 R0	G5 Y6 O0 R2	G1 Y6 O6 R0	G2 Y7 O3 R0	G5 Y6 O0 R2	G1 Y8 O4 R0	G1 Y8 O4 R0	G1 Y8 O4 R0	G3 Y6 O4 R0	G4 Y2 O5 R0	G1 Y8 O3 R0
Measurement & Exposure Assessment	G0 Y1 O9 R1	G5 Y4 O2 R1	G1 Y5 O5 R0	G6 Y5 O1 R0	G5 Y5 O3 R0	G2 Y5 O4 R1	G7 Y4 O1 R0	G3 Y8 O2 R0	G2 Y7 O4 R0	G4 Y7 O2 R0	G7 Y5 O1 R0	G2 Y3 O5 R2	G5 Y4 O2 R1
Outcome Validity & Ascertainment	G0 Y0 O9 R0	G6 Y4 O1 R0	G2 Y3 O5 R0	G3 Y2 O0 R0	G7 Y2 O0 R0	G8 Y3 O0 R0	G3 Y2 O0 R0	G4 Y6 O0 R0	G5 Y5 O0 R0	G5 Y5 O0 R0	G4 Y6 O0 R0	G4 Y5 O2 R0	G3 Y5 O3 R0
Transportability & External Validity	G1 Y7 O3 R2	G5 Y3 O4 R1	G6 Y6 O1 R0	G0 Y1 O3 R5	G2 Y5 O4 R1	G3 Y5 O4 R1	G0 Y2 O2 R5	G2 Y7 O2 R1	G2 Y7 O2 R1	G3 Y6 O2 R1	G4 Y5 O2 R1	G3 Y6 O6 R1	G3 Y8 O2 R0
Citation & Evidentiary Support	G1 Y5 O2 R0	G4 Y3 O1 R0	G8 Y0 O0 R0	G4 Y3 O2 R0	G4 Y5 O0 R0	G7 Y1 O0 R0	G4 Y4 O1 R0	G4 Y5 O0 R0	G4 Y5 O0 R0	G4 Y5 O0 R0	G4 Y5 O0 R0	G6 Y2 O0 R0	G4 Y4 O0 R0

How the meta-view summary dot is derived

Type here to search

64°F Mostly cloudy

10:34 PM 10/22/2026

ASP Methodology Audit Builder v1.4.1

Projects

Audit Matrix

Overview

Paper View

Frameworks

User Guide

Data

Color-blind colors

Project: NO2 and Asthma Methodology Audit

Methodological lens

Confounding & Identification Audit

Requirements

Causal Methodology Audit

Citation & Evidentiary Support

Confounding & Identification Audit

Well addressed

Meaningfully but incompletely addressed

Weak/indirect control or substantial residual concern

Not addressed / omitted

Not assessable from available material

Not applicable (does not count against the study)

Not yet rated

Confounding & Identification Audit

Confounding, co-exposure identification, exposure measurement, and outcome/selection issues.

Question this lens answers: Which plausible confounders, co-exposures, measurement problems, and selection mechanisms remain capable of explaining the observed association?

Applicability: Primary observational studies; adapt cautiously to syntheses when primary-study control is explicitly audited.

Potential confounding / identification issue	Clark 2010 BC	Gehring 2020 PIAMA	Johannessen 2026 Iceland	Kyrmánidou 2026 review	Lavigne 2018 Ontario	Lee 2024 meta-analysis
POTENTIAL CONFOUNDERS Baseline/familial asthma susceptibility	familial asthma not adjusted	familial asthma addressed	severe maternal asthma only	not systematically audited	familial asthma addressed	varies by study
CO-EXPOSURE / POLLUTANT-SPECIFIC IDENTIFICATION PM2.5 as co-pollutant	multipollutant adjustment infeasible	two-pollutant PM2.5 adjustment performed	single-pollutant models only	not systematically audited	two-pollutant PM2.5 adjustment performed	few NO2 studies adjusted
POTENTIAL CONFOUNDERS Individual/household SES	individual ses incompletely controlled	individual ses addressed		not systematically audited	individual SES unavailable	varies by study
CO-EXPOSURE / POLLUTANT-SPECIFIC IDENTIFICATION NO2-specific identification from TRAP	NO2 remains a source/mixture marker	some NO2-specific robustness, mixture ambiguity remains	highly correlated exposures	review does not isolate NO2	some NO2-specific robustness, mixture ambiguity remains	pooled estimate not necessarily NO2-specific
POTENTIAL CONFOUNDERS Seasonality / calendar time	calendar structure partially addressed	calendar structure partially addressed	birth year/month	not systematically audited	calendar structure partially addressed	varies by study
CO-EXPOSURE / POLLUTANT-SPECIFIC IDENTIFICATION Black carbon / EC	black/elemental carbon	black/elemental carbon	measured; not jointly modeled	not systematically audited	black/elemental carbon not	largely unaddressed

Type here to search

64°F Mostly cloudy 10:39 PM 10/22/2026

ASP Methodology Audit Builder v1.4.1

Projects

Audit Matrix

Overview

Paper View

Frameworks

User Guide

Data

Color-blind colors

Project: NO2 and Asthma Methodology Audit

Methodological lens

Confounding & Identification Audit

Show short notes under dots

Reload CSVs

Reset sorting

Well addressed

Meaningfully but incompletely addressed

Weak/indirect control or substantial residual concern

Not addressed / omitted

Not assessable from available material

Not applicable (does not count against the study)

Not yet rated

Confounding & Identification Audit

Confounding, co-exposure identification, exposure measurement, and outcome/selection issues.

Question this lens answers: Which plausible confounders, co-exposures, measurement problems, and selection mechanisms remain capable of explaining the observed association?

Applicability: Primary observational studies; adapt cautiously to syntheses when primary-study control is explicitly audited.

Potential confounding / identification issue	Clark 2010 BC	Gehring 2020 PIAMA	Johannessen 2026 Iceland	Kyrmánidou 2026 review	Lavigne 2018 Ontario	Lee 2024 meta-analysis
POTENTIAL CONFOUNDERS						
Baseline/familial asthma susceptibility	familial asthma not adjusted	familial asthma addressed	severe maternal asthma only	not systematically audited	familial asthma addressed	varies by study
CO-EXPOSURE / POLLUTANT-SPECIFIC IDENTIFICATION						
PM2.5 as co-pollutant	multipollutant adjustment infeasible	two-pollutant PM2.5 adjustment performed	single-pollutant models only	not systematically audited	two-pollutant PM2.5 adjustment performed	few NO2 studies adjusted
POTENTIAL CONFOUNDERS						
Individual/household SES	individual ses incompletely controlled	individual ses addressed		not systematically audited	individual SES unavailable	varies by study
CO-EXPOSURE / POLLUTANT-SPECIFIC IDENTIFICATION						
NO2-specific identification from TRAP	NO2 remains a source/mixture marker	some NO2-specific robustness, mixture ambiguity remains	highly correlated exposures	review does not isolate NO2	some NO2-specific robustness, mixture ambiguity remains	pooled estimate not necessarily NO2-specific
POTENTIAL CONFOUNDERS						
Seasonality / calendar time	calendar structure partially addressed	calendar structure partially addressed	birth year/month	not systematically audited	calendar structure partially addressed	varies by study
CO-EXPOSURE / POLLUTANT-SPECIFIC IDENTIFICATION						
Black carbon / EC	black/elemental carbon	black/elemental carbon	measured; not jointly modeled	not systematically audited	black/elemental carbon not	largely unaddressed

Type here to search

64°F Mostly cloudy 10:38 PM 10/22/2026

ASP Methodology Audit Builder v1.4.1

Projects

Audit Matrix

Overview

Paper View

Frameworks

User Guide

Data

Color-blind colors

Project: NO2 and Asthma Methodology Audit

Methodological lens

Diagnostic Red Flags for Interventional Causal Claims

Show short notes under dots

Reload CSVs

Reset sorting

Well addressed

Meaningfully but incompletely addressed

Weak/indirect control or substantial residual concern

Not addressed / omitted

Not assessable from available material

Not applicable (does not count against the study)

Not yet rated

SAVE / EXPORT

No unexported session edits.

Download edited audit CSV

Download

Diagnostic Red Flags for Interventional Causal Claims

Screening-oriented lens based on diagnostic red flags for asserted interventional causal claims.

Question this lens answers: Are there methodological warning signs that materially weaken an interventional causal interpretation?

Applicability: Studies or syntheses used to support interventional causal claims.

Potential confounding / identification issue	Clark 2010 BC	Gehring 2020 PIAMA	Johannessen 2026 Iceland	Kyrmanidou 2026 review	Lavigne 2018 Ontario	Lee 2024 meta-analysis	Ohyama 2022 HONO	Pedersen 2023 Denmark	Perrin 2023
RED FLAGS Identification assumptions remain implicit or unjustified	key causal assumptions mostly implicit	key causal assumptions mostly implicit	key identifying assumptions largely implicit	key identifying assumptions largely implicit	key causal assumptions mostly implicit	key identifying assumptions largely implicit	key causal assumptions partly explicit	key causal assumptions mostly implicit	key causal assumptions mostly implicit
RED FLAGS The causal explanation is not seriously tested against the evidence	causal explanation not strongly tested	causal explanation not strongly tested	pregnancy vs early-life and age>3 comparisons	no causal-model implications tested	causal explanation partly tested	heterogeneity/meta-regression explored, not causal constraints	causal explanation partly tested	causal explanation partly tested	causal explanation partly tested
RED FLAGS Undefined exposure contrast	exposure contrast well defined	exposure contrast well defined	statistical contrast defined; intervention version not specified	statistical contrast defined; intervention version not specified	exposure contrast well defined	statistical contrast defined; intervention version not specified	exposure contrast well defined	exposure contrast well defined	exposure contrast well defined
RED FLAGS No credible basis for identifying the claimed effect	identification basis present but assumption-dependent	identification basis present but assumption-dependent	cohort adjustment undermined by single-pollutant design	association synthesis without causal identification audit	identification basis present but assumption-dependent	meta-analysis of heterogeneous observational studies	identification basis present but assumption-dependent	identification basis present but assumption-dependent	identification basis present but assumption-dependent
RED FLAGS Causal model or data analysis does not match the causal question	analysis-to-question alignment reasonable	analysis-to-question alignment reasonable	association model not tied to formal causal estimand	association model not tied to formal causal estimand	analysis-to-question alignment reasonable	association model not tied to formal causal estimand	analysis-to-question alignment reasonable	analysis-to-question alignment reasonable	analysis-to-question alignment reasonable
RED FLAGS Important analytic choices are unexplained or inadequately examined	analytic choices partly examined	analytic choices partly examined	age-threshold and exposure-window comparisons only	quality assessment and heterogeneity discussion, limited sensitivity	analytic choices stress-tested	random effects, meta-regression, publication-bias and scale sensitivity	analytic choices stress-tested	analytic choices stress-tested	analytic choices stress-tested
RED FLAGS Important alternative explanations are not seriously examined	alternative explanations partly examined	alternative explanations partly examined	smoking, diagnosis, healthcare, maternal asthma discussed	residual confounding acknowledged but not systematically evaluated	alternative explanations partly examined	heterogeneity, missingness, selection and confounding explicitly examined	alternative explanations partly examined	alternative explanations partly examined	alternative explanations partly examined

Type here to search

64°F Mostly cloudy

10:40 PM 10/22/2026

ASP Methodology Audit Builder v1.4.1

Projects

Audit Matrix

Overview

Paper

Methodological lens

Diagnostic Red Flags for Interventional Causal Claims

Show short notes under dots

Reload CSVs

Reset sorting

Well addressed

Meaningfully but incompletely addressed

Weak/indirect control or substantial residual concern

Not addressed / omitted

Not assessable from available material

Not applicable (does not count against the study)

Not yet rated

SAVE / EXPORT

No unexported session edits.

Download edited audit CSV

Download

Diagnostic Red Flags for Interventional Causal Claims

Screening-oriented lens based on diagnostic red flags

Question this lens answers: Are there methodological

Applicability: Studies or syntheses used to support interventi

Potential confounding / identification issue

RED FLAGS

Identification assumptions remain implicit or unjustified

RED FLAGS

The causal explanation is not seriously tested against the evidence

RED FLAGS

Undefined exposure contrast

RED FLAGS

No credible basis for identifying the claimed effect

RED FLAGS

Causal model or data analysis does not match the causal question

RED FLAGS

Important analytic choices are unexplained or inadequately examined

RED FLAGS

Important alternative explanations are not seriously examined

Kyrmanidou 2026 review — No credible basis for identifying the claimed effect

Screening question: What comparison, design feature, or other evidence supports the conclusion about what would have happened had exposure been different?

The evidence may show association without a defensible basis for determining what would have happened under a different exposure.

AI PROPOSAL

Reviewer: ChatGPT initial audit

Confidence: high

Rating

Not addressed / omitted

Confidence

high

Review status

AI proposal

Short note shown under dot

association synthesis without causal id

Why this rating? (rationale)

Assessment mirrors the identification-strategy criterion: The review uses PRISMA and quality tools but does not systematically

Supporting quotation

Page

3-4,6-7

Source location

Methods/Discussion

Reviewer

ChatGPT initial audit

Cancel

Save assessment

Project: NO2 and Asthma Methodology Audit

Lee 2024 meta-analysis

Ohyama 2022 HONO

Pedersen 2023 Denmark

Perr

key identifying assumptions largely implicit

key causal assumptions partly explicit

key causal assumptions mostly implicit

ke

heterogeneity/meta-regression explored, not causal constraints

causal explanation partly tested

causal explanation partly tested

ci

statistical contrast defined; intervention version not specified

exposure contrast well defined

exposure contrast well defined

es:

meta-analysis of heterogeneous observational studies

identification basis present but assumption-dependent

identification basis present but assumption-dependent

ider but

association model not tied to formal causal estimand

analysis-to-question alignment reasonable

analysis-to-question alignment reasonable

i a

random effects, meta-regression, publication-bias and scale sensitivity

analytic choices stress-tested

analytic choices stress-tested

a

heterogeneity, missingness, selection and confounding explicitly examined

alternative explanations partly examined

alternative explanations partly examined

alt in

Type here to search

64°F Mostly cloudy

10:42 PM 10/22/2026

ASP Methodology Audit Builder v1.4.1

Projects

Audit Matrix

Overview

Paper View

Frameworks

User Guide

Data

Color-blind colors

Project: NO2 and Asthma Methodology Audit

Methodological lens

Statistical Analysis Quality

Show short notes under dots

Reload CSVs

Reset sorting

Well addressed

Meaningfully but incompletely addressed

Weak/indirect control or substantial residual concern

Not addressed / omitted

Not assessable from available material

Not applicable (does not count against the study)

Not yet rated

SAVE / EXPORT

No unexported session edits.

Download edited audit CSV

Download current matrix

Statistical Analysis Quality

Audit of statistical analysis quality for estimating and characterizing associations, distinct from whether a causal effect is identified.

Question this lens answers: How credible, stable, transparent, and reproducible is the reported statistical association, conditional on the study design?

Applicability: Quantitative empirical studies and quantitative evidence syntheses, with design-appropriate interpretation.

Potential confounding / identification issue	Clark 2010 BC	Gehring 2020 PIAMA	Johannessen 2026 Iceland	Kyrmanidou 2026 review	Lavigne 2018 Ontario	Lee 2024 meta-analysis	Ohyama 2022 HONO	Pedersen 2023 Denmark	Perrin 2023 US
MODEL SPECIFICATION AND DEPENDENCE									
Model specification	model family and major time/cluster structure are appropriate to the design	model family and major time/cluster structure are appropriate to the design	Cox with PH checks; single-pollutant models forced by collinearity	narrative synthesis because formal meta-analysis not performed	model family and major time/cluster structure are appropriate to the design	prespecified pooling rules, random effects, heterogeneity analysis	simple group-comparison analysis	model family and major time/cluster structure are appropriate to the design	model family and major time/cluster structure are appropriate to the design
ROBUSTNESS AND UNCERTAINTY									
Uncertainty quantification	effect estimates are accompanied by confidence intervals and inferential uncertainty	effect estimates are accompanied by confidence intervals and inferential uncertainty	HRs with 95% CIs and PH checks	narrative effect ranges without pooled uncertainty	effect estimates are accompanied by confidence intervals and inferential uncertainty	random-effects CIs and I ² heterogeneity	SE/p-values reported; limited interval emphasis	effect estimates are accompanied by confidence intervals and inferential uncertainty	effect estimates are accompanied by confidence intervals and inferential uncertainty
DATA QUALITY AND MEASUREMENT									
Missing data handling	missingness quantified but assumptions/handling remain imperfect	missingness quantified but assumptions/handling remain imperfect	high registry/address completeness; key smoking covariate unavailable	primary-study missing-data handling not systematically synthesized	missingness quantified but assumptions/handling remain imperfect	missingness and loss-to-follow-up audited across primary studies	attrition/exclusions not fully itemized	missingness quantified but assumptions/handling remain imperfect	missingness quantified but assumptions/handling remain imperfect
DATA QUALITY AND MEASUREMENT									
Outcome misclassification	outcome definition is explicit; residual misclassification remains	outcome definition is explicit; residual misclassification remains	asthma before age 3 poorly validated	heterogeneous asthma incidence/exacerbation definitions	outcome definition is explicit; residual misclassification remains	substantial heterogeneity in asthma definitions	objective physiologic/mechanistic outcomes	outcome definition is explicit; residual misclassification remains	outcome definition is explicit; residual misclassification remains
DATA QUALITY AND MEASUREMENT									
Covariate measurement quality	key covariates measured with mixed individual/area or questionnaire/registry precision	key covariates measured with mixed individual/area or questionnaire/registry precision	smoking unavailable; maternal asthma captures severe cases only	confounder/covariate measurement not systematically audited	key covariates measured with mixed individual/area or questionnaire/registry precision	confounder sets vary materially across studies	observational covariate measurement not applicable	key covariates measured with mixed individual/area or questionnaire/registry precision	key covariates measured with mixed individual/area or questionnaire/registry precision
MODEL SPECIFICATION AND DEPENDENCE									
Functional form / lag structure	exposure windows or functional form are examined to a useful but incomplete degree	exposure windows or functional form are examined to a useful but incomplete degree	IQR-linear contrasts; limited shape/specification analysis	exposure windows summarized broadly, not harmonized	exposure windows or functional form are examined to a useful but incomplete degree	standardized increments; heterogeneous primary exposure windows	time/dose patterns examined	exposure windows or functional form are examined to a useful but incomplete degree	exposure windows or functional form are examined to a useful but incomplete degree
MODEL SPECIFICATION AND DEPENDENCE									
Dependence / clustering / autocorrelation	clustered/dependence is	clustered/dependence is	family clusters not clearly	not assessable at review level	within person-level	clustered/dependence is	clustered/dependence is	clustered/dependence is	clustered/dependence is

Type here to search64°F Mostly cloudy10:45 PM 10/22/2026

ASP Methodology Audit Builder v1.4.1

Projects

Audit Matrix

Overview

Paper

Methodological lens

Statistical Analysis Quality

Show short notes under dots

Reload CSVs

Reset sorting

Well addressed

Meaningfully but incompletely addressed

Weak/indirect control or substantial residual concern

Not addressed / omitted

Not assessable from available material

Not applicable (does not count against the study)

Not yet rated

SAVE / EXPORT

No unexported session edits.

Download edited audit CSV

Download current matrix

Statistical Analysis Quality

Audit of statistical analysis quality for estimating and controlling confounding

Question this lens answers: How credible, stable, transparent, and applicable are the results?

Applicability: Quantitative empirical studies and quantitative synthesis

Potential confounding / identification issue

MODEL SPECIFICATION AND DEPENDENCE

Model specification

ROBUSTNESS AND UNCERTAINTY

Uncertainty quantification

DATA QUALITY AND MEASUREMENT

Missing data handling

DATA QUALITY AND MEASUREMENT

Outcome misclassification

DATA QUALITY AND MEASUREMENT

Covariate measurement quality

MODEL SPECIFICATION AND DEPENDENCE

Functional form / lag structure

MODEL SPECIFICATION AND DEPENDENCE

Dependence / clustering / autocorrelation

Project: NO2 and Asthma Methodology Audit

Lee 2024 meta-analysis

Ohyama 2022 HONO

Pedersen 2023 Denmark

Perrin 2023

prespecified pooling rules, random effects, heterogeneity analysis

simple group-comparison analysis

model family and major time/cluster structure are appropriate to the design

random-effects CIs and I² heterogeneity

SE/p-values reported; limited interval emphasis

effect estimates are accompanied by confidence intervals and inferential uncertainty

missingness and loss-to-follow-up audited across primary studies

attrition/exclusions not fully itemized

missingness quantified but assumptions/handling remain imperfect

substantial heterogeneity in asthma definitions

objective physiologic/mechanistic outcomes

outcome definition is explicit; residual misclassification remains

confounder sets vary materially across studies

observational covariate measurement not applicable

key covariates measured with mixed individual/area or questionnaire/registry precision

standardized increments; heterogeneous primary exposure windows

time/dose patterns examined

exposure windows or functional form are examined to a useful but incomplete degree

Lee 2024 meta-analysis — Outcome misclassification

Screening question: Could outcome misclassification materially distort the reported association?

Validity, sensitivity/specificity, and consistency of the outcome definition.

AI PROPOSAL Reviewer: ChatGPT Confidence: high

Rating

Confidence

Review status

Weak/indirect control or substantial residual concern

high

AI proposal

Short note shown under dot

substantial heterogeneity in asthma de

Why this rating? (rationale)

Asthma definitions vary across self-report, physician diagnosis, medication use, and administrative ICD criteria, limiting comparability.

Supporting quotation

Page

Source location

Reviewer

7-8

Results

ChatGPT

Cancel

Save assessment

Type here to search

64°F Mostly cloudy

10:45 PM 10/22/2026

Break / discussion

Questions on causal estimands, rival models, and what evidence would discriminate among them



How does it work?

ASP architecture and inspectable outputs

Six core review modules, claim extraction, evidence bundles, and five-domain Evidence Browser

ASP1: the proposed middle ground

Current top-down AI

“Read and critique this article.”

Flexible, potentially insightful, but less reproducible or auditable.

ASP1: Bottom-up AI

Prespecified, high-level questions + task decomposition + source quotations + synthesis.

More trustworthy & transparent AI

Atomic checklist

Numerous low-level items can be inspectable but rigid, complex, and narrow.

STRUCTURE + FLEXIBILITY

ASP1 is a methodological reviewer of a supplied document—not an independent arbiter of truth.

Two design commitments

1 Prespecify the questions

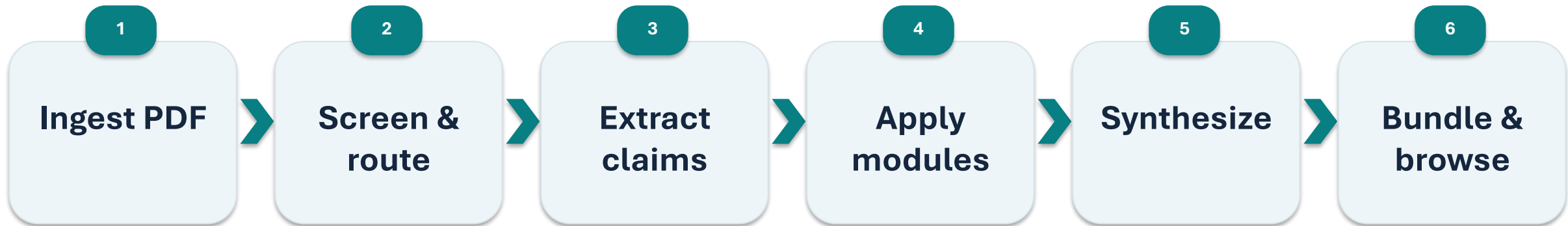
Set evaluative frameworks before seeing the target document—a “veil of ignorance” that discourages post hoc tailoring.

2 Ground the judgments

Link evaluations to source quotations or identified locations; keep direct textual evidence distinct from reviewer inference.

The result is a review that a human can inspect, challenge, correct, and reuse.

From uploaded PDF to Evidence Review Bundle



Persistent artifacts: summaries • claim tables • module appraisals • quotations • cross-module assessment

Applicable modules are selected by document content; multiple modules may apply.

Six methodological review modules

01

Observational epidemiology / STROBE

Design, variables, analyses, reporting

02

Causal inference / policy

Identification, confounding, causal claims

03

Statistical methods

Models, estimation, uncertainty, sensitivity

04

Systematic review / meta-analysis

Search, eligibility, synthesis, heterogeneity

05

Risk of bias

Bias domains, judgments, rationale

06

Mechanistic / toxicological

Pathways, dosimetry, human relevance

The causal module: five prespecified questions

- 1 What is the causal question—exposure, outcome, population, intervention?
- 2 Does the design support causal inference; are confounding and selection addressed?
- 3 Are exposures, outcomes, confounders and modifiers measured and analyzed appropriately?
- 4 Are alternative explanations and limitations considered without overstating causality?
- 5 Do policy or intervention implications follow from the presented evidence?

Module 0: claims before critique

- A: descriptive/associational, without intervention implication.
- B: stronger directional language that remains observational.
- C: claim implying an effect of changing exposure.
- D: explicit intervention or policy recommendation.
- Store quotations, source locations, and stable identifiers C1, C2, ...

Statistical Methods Review

Measurement error	Validation, calibration, SIMEX, Bayesian models
Missing data	MCAR / MAR / MNAR; imputation, weighting, sensitivity
Latent variables	Unmeasured confounding, controls, IVs, bias analysis
Model uncertainty	Alternative specifications, multiverse, specification curves
Validation	Diagnostics, calibration, falsification, robustness

Two complementary review layers

Core modules

High-level specialist prompts, selected by document content.

Evidence Browser

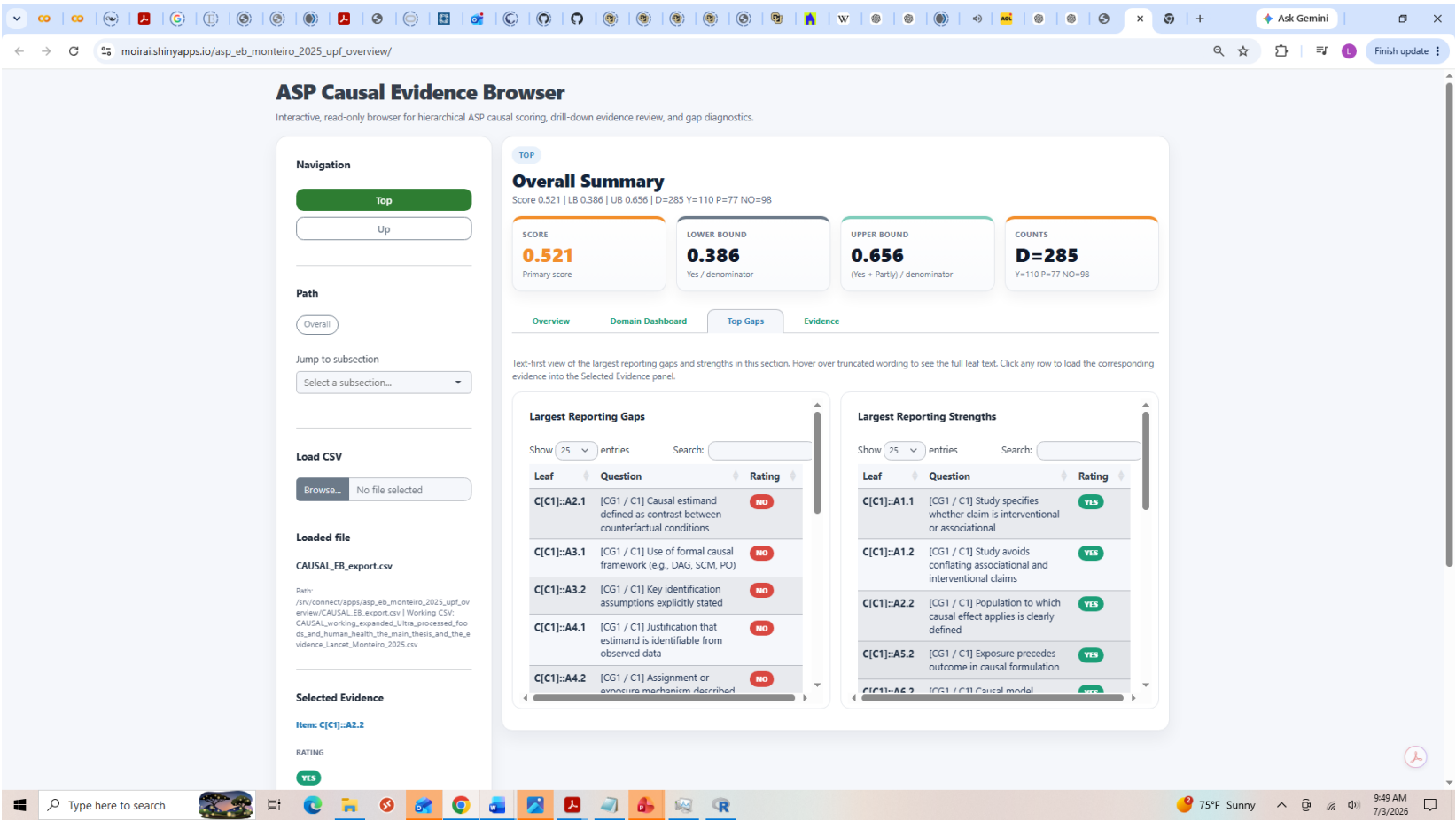
Detailed five-domain items rated Yes / Partly / No / N/A, with explanations and quotations.

Evidence Browser scores are heuristic

Yes = 1 Partly = 0.5 No = 0

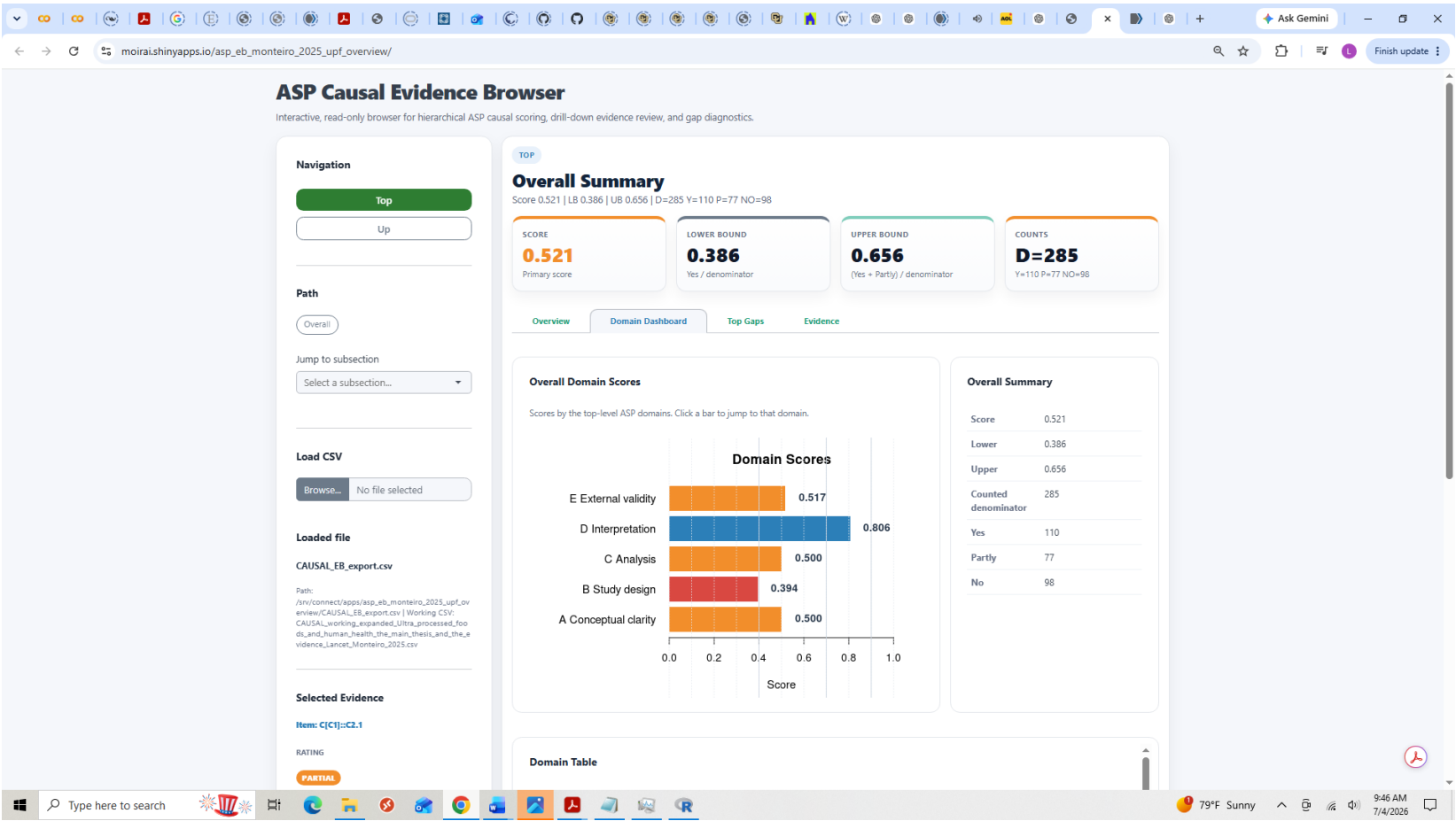
Scores are not probabilities of causal truth or validated measures of methodological quality.

Evidence Browser: gaps and strengths



Inspect item-level evidence rather than relying on the summary score.

Evidence Browser: domain dashboard



Example overall score 0.521; sensitivity range 0.386–0.656.

From AI assessment to defensible scientific judgment

Prespecify

Run reusable methodological questions without tailoring standards to the desired conclusion.

Ground

Link claims, judgments, and limitations to original quotations and document locations.

Challenge

Identify false positives, omitted analyses, and criticisms that are not consequential.

Verify

Experts check empirical details, causal relevance, and independent evidence.

ASP outputs are hypotheses and review aids—not scientific verdicts

Claim linked to source

What did the paper explicitly report?
Check the quote, its scope, and the population.

AI's interpretation

Does the inferential criticism apply? Is it material? Does other evidence address it?

For every consequential concern: inspect source → compare alternatives → verify significance → revise the assessment.

Cross-document synthesis

A formaldehyde–leukemia case and a PowerPoint-simulated walk-through of the ASP portal

Suggested segment: 30 minutes

ASP1, monographs, and the next scale of synthesis

ASP1: within a document

Evaluate one supplied paper through applicable methodological modules and cross-module synthesis.

Monograph Builder: across documents

Experimental extension ingests paper-level Evidence Review Bundles and synthesizes evidence, claims, and tensions across papers.

Separate but composable tasks: document appraisal → literature synthesis → decision analysis.

Formaldehyde–leukemia: three rival causal models

Direct genotoxicity

Inhaled formaldehyde → target-tissue DNA damage → leukemia

- test tissue delivery.

Indirect pathway

Airway injury → systemic signaling → hematopoietic effects

- test mediating mechanisms.

Confounding

leukemia risk ← Job factors and co-exposures → exposure

- test adjustment and negative evidence.

Rival models are hypotheses, not established mechanisms.

Different studies can discriminate among rival models

Beane Freeman et al. 2009 Peak-exposure associations differed by metric; reconstruction and reference choices matter.

Checkoway et al. 2015 AML pattern did not increase at the highest peak-exposure levels; consider subtype and latency.

Coggon et al. 2014 Highest exposure category did not show elevated myeloid leukemia risk; consider power and comparability.

Mechanistic tests Target-tissue adducts, dose relevance, assay specificity, and alternative pathways.

Independent study verification remains necessary.

The challenge: discriminate rival models, not tally positive studies

Direct mechanism

What evidence supports exposure delivery to relevant target tissue and an exposure-specific mutational pathway?

Indirect pathway

What reproducible intermediate mechanisms connect inhalation to marrow outcomes?

Confounding / bias

Which co-exposures and measurement errors can account for apparently associated risk patterns?

Specific rival hypotheses yield different empirical predictions; an ASP synthesis must preserve contrary evidence.

Watch the audit trail—not the apparent confidence of the prose

Document-level

Open a paper, review claims, statistical critique, quotations and caveats.

Proposition-level

Compare support for a specific claim across papers and evidentiary modalities.

Scientific verification

Challenge the AI's grouping, omissions, study descriptions, and inference.

The displayed formaldehyde results are AI-generated illustrative assessments, not independently verified conclusions.

Choose the evidence collection

ASP Monograph Evidence Portal

Review topic: Formaldehyde and leukemia

Automated paper-level methodological evidence review.

15
Papers staged

15
Papers completed

15
Complete 9/9 core reviews

0
Failed papers

Individual Paper Evidence Bundles

Paper	Status	Core Reports	Outputs
Beane Freeman JNCI 101 751 2009	completed	9/9	Open Evidence Bundle Open Briefs
Blair Am J Ind Med 40 3 2001	completed	9/9	Open Evidence Bundle Open Briefs
Checkoway et al 2015	completed	9/9	Open Evidence Bundle Open Briefs
Coggon et al 2014	completed	9/9	Open Evidence Bundle Open Briefs
Hauptmann J Natl Cancer Inst 101 1696 2009	completed	9/9	Open Evidence Bundle Open Briefs
Hayes Am J Ind Med 18 641 1990	completed	9/9	Open Evidence Bundle Open Briefs

Entry point

- One topic can contain multiple paper bundles.
- Examine completion and failed-paper status.
- Start with claims, not a global score.

Open the selected paper's evidence bundle

ASP Evidence Review Bundle

Paper: Beane_Freeman_JNCI_101_751_2009.pdf

Generated: 2026-08-18 14:59:29 MDT

Model: gpt-4.1

Available Reports

Module	Status	Link	Notes
Study Summary	Completed	Open report	summary
Extracted Causal Claims	Completed	Open report	Verbatim causal and policy-relevant claims extracted from the paper
Grouped STROBE Review	Completed	Open report	full_review
Causal Critique	Completed	Open report	claim_level_review
Statistical Methods Review	Completed	Open report	methods_review
SR/MA Review	Screening only	Open report	screening_only
Risk-of-Bias Review	Screening only	Open report	screening_only
Mechanistic Review	Screening only	Open report	screening_only
Cross-Module Synthesis	Completed	Open report	synthesis
Overall Assessment	Completed	Open report	integrated_assessment

Modules run in this order: aspsum, aspstroke, aspcrit, aspmethods, asprisma, asprob, aspmech.

Outputs

- Study summary and verbatim claims.
- Causal and statistical reviews.
- Synthesis, assumptions, and integrated assessment.

Examine the statistical-methods critique

Logical Validity Review			
	Strengths	Concerns	Evidence/Quotes
Measurement	- Comprehensive exposure assessment using work histories, job/department/task coding, and historical measurement data. - Exposure indices developed: peak exposure (categorical), average intensity, cumulative exposure (quantitative, time-dependent). - Some recognition of measurement error and potential for non-differential misclassification. - Sensitivity analyses conducted regarding post-1980 exposure assumptions.	- No direct quantitative correction for measurement error or errors-in-variables bias (e.g., regression calibration not used). - Exposure after 1980 not measured: primary analysis assumed zero exposure; sensitivity analyses only explored best/worst-case scenarios. - Validation of exposure estimates limited; only modest comparison to available measurement data (r=0.50, small sample).	"Exposure to formaldehyde was estimated for each job from individual work histories...using current and past measurement data." (p. 753) "Although validation was not possible, effects of misclassification may be gauged by comparing full-shift exposure estimates with the monitoring data. A sample... showed a correlation coefficient of 0.50 (8)." (p. 759) "We assumed that zero exposure occurred after [1980]...carried out two sensitivity analyses..." (p. 754)
Mechanisms	- Extensive tracing process to ascertain deaths and causes (including linking to NDI, Social Security, etc.). - Discussion of unexplained missed deaths in earlier tracing, with transparent reporting regarding inclusions.	- No quantitative analysis or modeling of missing data patterns. - Missing post-1980 exposure data handled only via deterministic sensitivity analyses (not via multiple imputation or robust quantitative methods). - No test or discussion of MCAR/MAR/MNAR assumptions.	"We were unable to access the full file provided by NDI after the initial 1994 linkage...we have included them in all analyses for this update." (p. 752) "We may have underascertained cases because we considered an individual alive if NDI Plus did not identify the subject as deceased." (p. 759)
Univariable			

Review focus

- The source reports historical exposure assessment.
- AI flags possible exposure error and incomplete correction.
- Human reviewer must assess direction and materiality.

Inspect claim-specific support and caveats

Document: Beane_Freeman_JNCI_101_751_2009
Generated: 2026-08-18 15:01:49 MDT **Model:** gpt-4.1
Inputs: synthesis, causal critique, key assumptions, references, and reference support audit where available.

Overall Assessment

The Beane Freeman et al. (2009) study presents a well-conducted, large-scale, prospective cohort analysis of occupational formaldehyde exposure and mortality from lymphohematopoietic malignancies among US industrial workers. The evidence is carefully and transparently reported. Statistical associations—particularly with high peak formaldehyde exposures and mortality from Hodgkin lymphoma (statistically significant) and myeloid leukemia (non-significant elevation)—are robustly observed. However, the study is explicit about substantial uncertainties due to exposure misclassification (especially post-1980), measurement reliance on mortality rather than incidence, and incomplete control for confounding. Importantly, conclusions and recommendations are appropriately cautious, largely hypothesis-generating, and align with the strengths and limitations of the findings. Calls for regulatory action are avoided, with further research highlighted as the appropriate path forward.

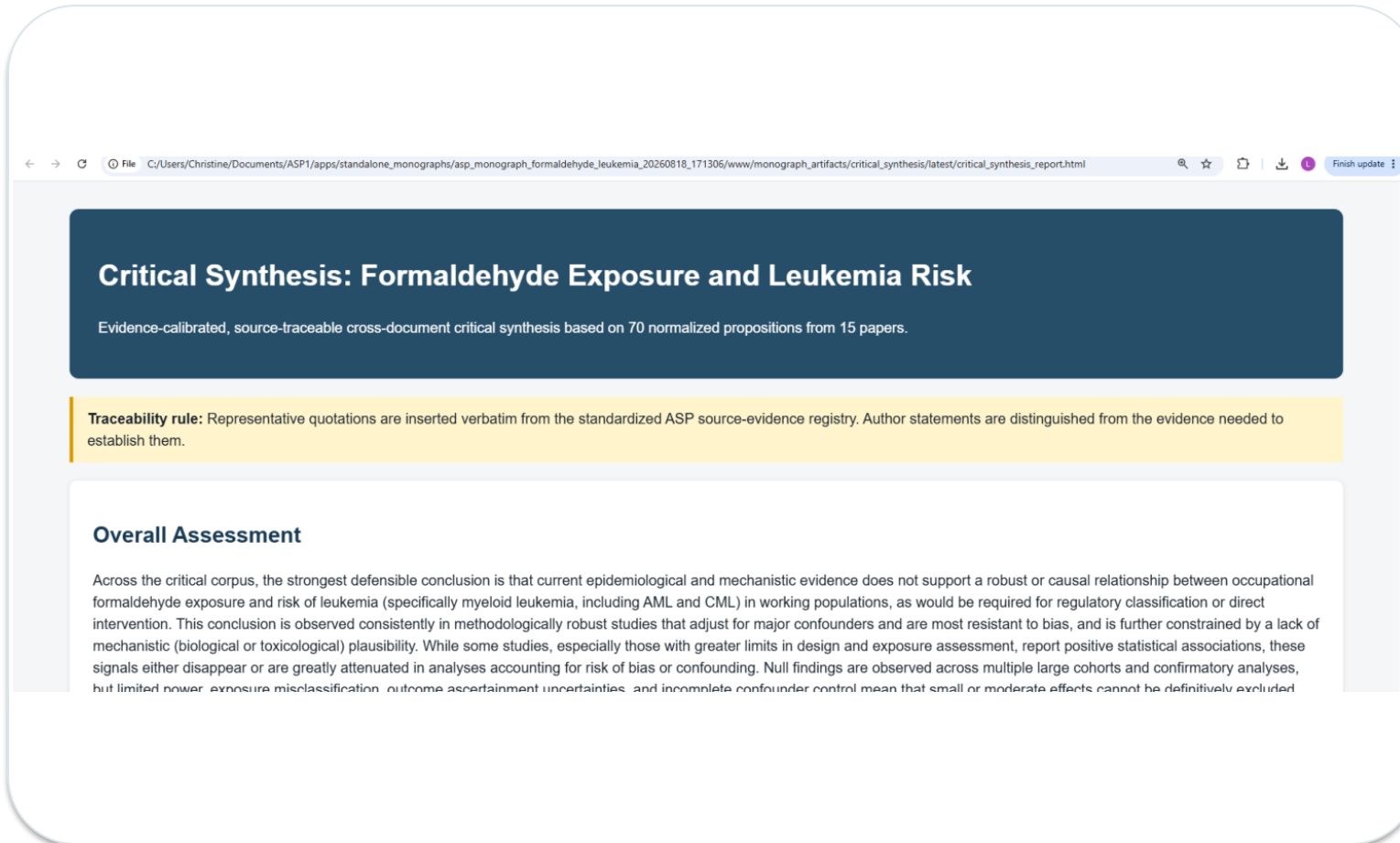
Key Conclusions and Support

--	--	--	--	--

Check support

- Separate statistical association from causal interpretation.
- Identify key dependence on exposure reconstruction.
- Verify adverse and null findings in the source.

Move from paper appraisals to critical synthesis



Cross-paper layer

- Statements are cross-linked across the corpus.
- Synthesis can expose contradictions and assumptions.
- AI-generated verdicts require independent validation.

Organize the evidence by causal proposition

Canonical proposition	Causal support	Papers	Representative source evidence
Occupational formaldehyde exposure does not clearly increase the risk of myeloid leukemia (including AML and CML) in working populations, and strong regulatory causal claims are not justified by the available epidemiologic or mechanistic evidence.	Causal support is weak for a formaldehyde–leukemia relationship and moderate-to-strong against it; does not achieve high robustness. No evidence for intervention effect.	Lauer_et_al_2024_triangulation_with_supplements, Coggon_et_al_2014, Checkoway_et_al_2015, Pira_Cancer_Causes_Control_25_1343_2014, Meyers_Am_J_Ind_Med_56_1027_2013, Hauptmann_J_Natl_Cancer_Inst_101_1696_2009	Lauer_et_al_2024_triangulation — empirical result or observation Pira_Cancer_Causes_Control_25_1343_2014 — empirical result or observation Checkoway_et_al_2015 — causal conclusion
Observed positive associations between occupational formaldehyde exposure and leukemia or other lymphohematopoietic cancers in select studies are likely artefacts of methodological limitations and do not provide robust causal evidence.	Support for causality is weak or negative; associations are best interpreted as artefacts, not actionable causal evidence.	Lauer_et_al_2024_triangulation_with_supplements, Hauptmann_J_Natl_Cancer_Inst_101_1696_2009, Blair_Am_J_Ind_Med_40_3_2001, Walrath_Cancer_Res_44_4638_1984, Meyers_Am_J_Ind_Med_56_1027_2013	Lauer_et_al_2024_triangulation — empirical result or observation Blair_Am_J_Ind_Med_40_3_2001 — recommendation Meyers_Am_J_Ind_Med_56_1027_2013 — limitation or caveat

Claims first

- Read the exact proposition.
- See which papers are used for and against it.
- Inspect how the system weights limitations and studies.

Trace a proposition back to source evidence

Detailed Proposition-Level Synthesis

F1: Occupational formaldehyde exposure does not clearly increase the risk of myeloid leukemia (including AML and CML) in working populations, and strong regulatory causal claims are not justified by the available epidemiologic or mechanistic evidence.

Directly supported empirical finding

Most methodologically robust studies (Lauer 2024, Coggon 2014, Checkoway 2015, Pira 2014) observe no statistically significant association between occupational formaldehyde exposure and myeloid leukemia risk after accounting for bias. Meta-analytic risk estimates attenuate to null when high-risk-of-bias studies are excluded.

Assumption-dependent stronger interpretation

If future studies could quantify and remove all sources of exposure misclassification, confounding, and selection bias, and if small but real effects are not being obscured, then formaldehyde would not be a leukemogen in typical workplace conditions.

Scope and boundaries

Supported for occupational exposures in US/European industrial and embalming settings studied to date; not directly generalizable to extremely high exposures or unstudied populations. Effect cannot be rigorously excluded for small or rare subgroups.

Papers

- Lauer_et_al_2024_triangular_with_supplements
- Coggon_et_al_2014
- Checkoway_et_al_2015

Audit trail

- Separate empirical finding from causal extrapolation.
- Inspect alternative explanations.
- Check cited quotations and evidence coverage.

Compare agreements and unresolved tensions

Major Agreements Across Papers

- Current epidemiological evidence does not support a robust or causal association between occupational formaldehyde exposure and myeloid leukemia risk.
- Positive associations in less methodologically robust studies are likely artefacts of bias or confounding.
- Mechanistic (biological/toxicological) evidence for formaldehyde-induced leukemogenesis in humans is lacking.
- All current high-quality reviews and syntheses recommend further research and surveillance rather than regulatory policy or intervention.

Major Disagreements and Tensions

- Some older studies or studies with limited design suggest positive statistical associations, but these are largely dismissed as artefactual upon critical synthesis and triangulation.
- Unresolved debate about the degree to which null findings can be interpreted as evidence of safety (absence of evidence vs. evidence of absence).

Cross-Paper Methodological Pattern

Triangulation

- Convergence can reflect shared biases.
- Null results may be uninformative if underpowered.
- Focus on predictions that discriminate models.

Turn unresolved gaps into next-study questions

Evidence Gaps

- Direct, validated, and quantitative individual-level exposure assessment.
- Comprehensive confounder and co-exposure measurement and adjustment.
- Mechanistic evidence bridging exogenous formaldehyde exposure and human leukemogenesis.
- Large-scale, prospective designs or natural/quasi-experimental reductions in workplace exposure with pre-registered bias analysis.
- Studies with sufficient power to detect moderate-to-small risk increases.

Conclusions Supported by the Current Evidence

- Occupational formaldehyde exposure does not robustly increase risk of myeloid leukemia in working populations studied to date.
- Policy, regulatory, or interventional action is not justified by the present evidence and should await further research.
- Further high-quality research, including improved exposure assessment, mechanistic studies, and ongoing occupational surveillance, is warranted.

Conclusions Not Supported by the Current Evidence

- Regulatory or authoritative causal declarations that occupational formaldehyde exposure is a cause of leukemia (as of the evidence reviewed).
- Population-level interventions to reduce formaldehyde exposure on the basis of leukemia risk.

Research needs

- Name the missing causal link.
- Propose reanalysis or experiment with high discrimination.
- Distinguish research priorities from policy prescriptions.

Causal support, decision relevance, and action are separate

Evidentiary question

Does changing exposure X alter outcome Y in the specified population and conditions? What is identified?

Decision question

Which feasible action is preferable given benefits, costs, uncertainty, distribution, and error consequences?

Weak causal evidence does not prove that inaction is best; strong causal evidence alone does not select a policy.

Human verification is indispensable

- Verify quotations, estimates, and methods against the original document.
- Check whether each review question is relevant to the specific design.
- Consider supporting, contrary, and omitted evidence.
- Document material corrections and unresolved disagreements.

Potential value and unresolved validation

Potential contribution

Systematic coverage, auditable evidence trails, early detection of methodological omissions.

Not established

Sensitivity, specificity, inter-rater reliability, performance versus expert reviewers or alternative models.

What would establish practical value?

- Benchmark against multiple expert panels and much broader, diverse corpora.
- Compare ASP1 directly with simple frontier-model + well-written rubric prompting.
- Stress-test subtle identification failures, omitted source details, conflicting evidence, and justifiable action under uncertainty.
- Evaluate sensitivity to rubrics, model versions, and alternative synthesis approaches.
- Develop and validate multi-document synthesis while maintaining audit trails.

Main conclusion

Make methodological expertise accessible and auditable

The greatest benefit may come during study design, analysis, and manuscript preparation—while weaknesses can still be corrected.

What would make this approach scientifically credible?

This is an invitation to test and improve ASP—not a claim that its performance is already validated.

Reproducibility

How stable are findings across documents, prompts, models, and expert auditors?

Error analysis

Which false positives and false negatives are consequential?

Scientific impact

Which missing analyses, rival models, or new measurements would change conclusions?

Paper, implementation, and demonstration examples

Published ASP1 article

doi.org/10.1111/risa.70356

Single-paper reviewer

moirai.shinyapps.io/asp_single_paper/

Evidence Browser / UPF examples

moirai.shinyapps.io/asp_upf_standalone/

Formaldehyde monograph

moirai.shinyapps.io/asp_monograph_formaldehyde_leukemia/

Backup Slides

Other worked examples and failure modes

Gas stoves, Grandjean PFCs, vinyl chloride, alcohol, glyphosate, and UPFs

Suggested segment: 30 minutes

Gas-stove worked example

PAF 12.7% (95% CI 6.3%–19.3%)

Derived from gas-stove prevalence and pooled observational OR 1.34 (1.12–1.57), not an identified intervention effect.

Extracted claims C1–C5

C1: Theoretical prevention

PAF interpreted as asthma theoretically prevented if stoves absent.

C2: Guidance

PAF may guide public-health interventions.

C3: Mitigation

Stove replacement and ventilation discussed.

C4: Predicted benefit

Mitigation “will lessen” asthma burden.

C5: Further studies

Call for intervention research.

What the gas-stove review found

Descriptive strengths

Pooled OR, national prevalence, PAF formula, 10,000 Monte Carlo draws.

Causal gap

Observed association and model assumptions do not identify cases prevented by intervention.

Robustness opportunity

Recalculate PAF using only North American studies.

A consequential sensitivity question

North American OR 1.36 (0.76–2.43)

Versus pooled North American–European OR 1.34 (1.12–1.57). Test sensitivity of PAF to this choice.

Human verification: keep and correct

Keep the central critique

Conditional PAF is distinct from the effect of stove removal or ventilation.

Correct ancillary overstatements

Authors acknowledged assumptions; nine states had separate estimates; secondary modeling need not collect new data.

Grandjean: a false-positive AI critique (using 2025 LLM)

Claim misclassification

“Were associated with reduced humoral immune response” was treated as implying intervention effects.

Overlooked qualifications

“If the associations are causal” and reported model checks were not fully credited.

Applications across evidence types

- Vinyl chloride: well-established liver-cancer causal conclusion was supported.
- Alcohol and CHD: biomarker effects did not directly establish fewer coronary events.
- Glyphosate: opposing meta-analyses yielded qualified assessments.
- UPFs: cross-paper comparisons are possible, but heuristic scores are not validity rankings.

Test corpus: six publications, diverse evidence

Mendoza et al. (2024)	Prospective cohorts + meta-analysis	Cardiovascular disease
Solans et al. (2021)	Population case–control	Chronic lymphocytic leukemia
Henney et al. (2024)	Systematic review + meta-analysis	Dementia
Lane et al. (2024)	Umbrella review	Multiple adverse outcomes
Menegassi & Vinciguerra (2025)	Mechanistic narrative review	Cancer
Monteiro et al. (2025)	Broad evidence + policy synthesis	Chronic disease / policy

Purposefully selected illustration—not a systematic UPF review or formal validation cohort.

Recurring result: support changes by claim type

ASSOCIATION

Usually relatively well supported

CAUSAL INTERPRETATION

Often partly supported or uncertain

POLICY EFFECTIVENESS

Frequently exceeds direct evidence

Qualitative pattern; no numeric performance scores or causal effect estimates were established.

Mendoza: an inferential tension in plain sight

Study interpretation

Authors expressly acknowledge that hazard ratios cannot support causal interpretations.

Policy-facing language

Yet they suggest soft drinks and processed meats “should be particularly discouraged.”

ASP1 extracted both passages and flagged the additional causal assumption underlying advice based on observational associations.

Monteiro: the most demanding test case

Recognized strengths

- Synthesis of 104 prospective studies
- Broad international coverage
- Plausible, diverse mechanisms
- Short-term randomized feeding evidence

Identified limitations

- Predominantly observational base
- Residual confounding and measurement error
- Heterogeneity of the UPF category
- Short-term trials vs chronic endpoints

ASP1 classified some broad causal theses as partly supported/uncertain and urgent policy claims as having questionable support.

ASP1 versus expert commentators

ASP1 contributed

- Systematic inferential decomposition
- Explicit claim-by-claim support assessments
- Consistent treatment of evidence limitations
- Inspectable quotation and report trails

Human experts contributed

- Detailed socioeconomic context
- Sharper identification of internal tensions
- Specific flaws in dietary measurement
- Vivid mechanistic and historical critiques

Major methodological concerns overlapped substantially; this was a qualitative comparison, not a formal benchmark.

Where the humans sharpened the analysis

Exposure classification

Kuhnle: historical dietary questionnaires predated Nova and were not designed to distinguish some UPFs from counterparts.

Historical generalization

McConway: trends for the UK and data-history limits challenged sweeping displacement claims.

Mechanistic breadth

Griffin: a large catalogue of possible mechanisms risks a diffuse causal story.

Human review added precision and judgment about which limitations were consequential.

UPF evidence collection

Retrieved and Reviewed Papers

Evidence node: Ultra-Processed Foods

Study	Title	Local PDF	Evidence Bundle	Evidence Browser & Scores	Briefs	Methodology Score	Domain Scores
Bonaccio, 2023	UPF consumption, all-cause mortality, and cardiovascular mortality in type 2 diabetes	Open PDF	Open Bundle	Open EB	Open Briefs	0.591 ●	A 0.500; B 0.826; C 0.532; D 0.682; E 0.560
Dai, 2024	Umbrella review and updated meta-analyses of UPFs and human health	Open PDF	Open Bundle	Open EB	Open Briefs	0.414 ●	A 0.321; B 0.800; C 0.288; D 0.655; E 0.349
Gauci, 2025	UPF cardiovascular mortality	Open PDF	Open Bundle	Open EB	Open Briefs	0.557 ●	A 0.500; B 0.941; C 0.466; D 0.701; E 0.467
Fu, 2024	UPFs and hearing loss	Open PDF	Open Bundle	Open EB	Open Briefs	0.416 ●	A 0.417; B 0.398; C 0.357; D 0.510; E 0.510
Henney, 2024	UPFs and dementia	Open PDF	Open Bundle	Open EB	Open Briefs	0.442 ●	A 0.333; B 0.623; C 0.407; D 0.510; E 0.441
Lane, 2024	UPF health outcomes	Open PDF	Open Bundle	Open EB	Open Briefs	0.321 ●	A 0.364; B 0.408; C 0.192; D 0.712; E 0.038
Levy, 2023	How and why ultra-processed foods harm human health	Open PDF	Open Bundle	Open EB	Open Briefs	0.035 ●	A 0.062; B 0.000; C 0.000; D 0.197; E 0.000
Liang, 2025	Ultra-processed foods and risk of all-cause mortality: updated systematic review and dose-response meta-analysis	Open PDF	Open Bundle	Open EB	Open Briefs	0.549 ●	A 0.429; B 0.771; C 0.365; D 0.702; E 0.754
Lutz, 2025	UPF and neuropsychiatric outcomes	Open PDF	Open Bundle	Open EB	Open Briefs	0.084 ●	A 0.125; B 0.009; C 0.052; D 0.304; E 0.000
Menegassi, 2025	Ultraprocessed food and risk of cancer: mechanistic pathways and public health implications	Open PDF	Open Bundle	Create EB	Open Briefs	—	—
Mendoza, 2024	UPF cardiovascular outcomes	Open PDF	Open Bundle	Open EB	Open Briefs	0.643 ●	A 0.500; B 0.861; C 0.586; D 0.625; E 0.739
Monteiro, 2025	Ultra-processed foods and human health: main thesis and evidence	Open PDF	Open Bundle	Open EB	Open Briefs	0.521 ●	A 0.500; B 0.394; C 0.500; D 0.806; E 0.517
Prescott, 2024	UPF and crime	Open PDF	Open Bundle	Open EB	Open Briefs	0.098 ●	A 0.156; B 0.009; C 0.000; D 0.392; E 0.165
Solans, 2021	Ultra-Processed Food and Chronic Lymphocytic Leukemia in the MCC-Spain Study	Open PDF	Open Bundle	Open EB	—	0.552 ●	A 0.500; B 0.733; C 0.429; D 0.727; E 0.556

Generated from current ASP bundle inventory; local PDFs are linked when available.

Heuristic scores support navigation and comparison, not authoritative ranking.

Figure 3, p. 27

Three complementary tests of scientific value

Coverage

Are important methodological flaws omitted? How often does ASP find consequential ones?

Precision

Does ASP mistakenly flag correct statements, irrelevant omissions, or already-performed checks?

Usefulness

Does AI assistance change expert conclusions, uncover better analyses, or improve reporting?

Compare with expert panels, direct frontier-model + rubric prompting, and blinded follow-up verification.

Alcohol biomarkers versus coronary disease outcomes

Mechanistic / biomarker evidence

Alcohol changed HDL and haemostatic biomarkers in short-term studies.

Intervention-level claim

A projected 24.7% lower CHD risk depends on linking biomarker changes to clinical outcomes.

Critical question: Does a biomarker response identify the long-term effect of changing alcohol consumption on CHD events?

Gas stoves: population-attributable fraction $\neq$ prevented cases

Reported association

Pooled observational odds ratio: 1.34 (95% CI 1.12–1.57).

Reported attribution

Estimated PAF: 12.7% (95% CI 6.3%–19.3%).

To predict prevented asthma cases, we also need a defined intervention and justified causal, transportability, and implementation assumptions.

How would ASP1 audit the Dublin coal-ban claim?

Proposed review questions; no ASP1 evaluation of this particular study is claimed here.

A · Causal estimand

Effect of which intervention on which mortality outcome, population and time scale?

B · Design

Does a defensible counterfactual trend or control group identify the ban's effect?

C · Analysis

Do seasonality, co-pollutants, long-run trend, autocorrelation, and lag choice alter inference?

D–E · Interpretation

What conclusions survive competing models, and what transports beyond Dublin?

A model can manufacture a positive exposure coefficient

Assume no causal effect

X and Y are independent and both have positive means.

Force regression through zero

Fitting $Y = \beta X$ without an intercept yields $\hat{\beta} = \Sigma XY / \Sigma X^2 > 0$ for positive data.

Question for the AI reviewer: Were functional form, intercept, diagnostics, and plausible alternative models examined?

Break / discussion

What is a valid criticism? What would establish that the AI improves expert review?