

GHEAS 双轨健康数据分析方法学

以 PM2.5 与人口健康为例：从观察性关联到因果识别

项目	登记内容
文档类型	方法学研究报告与可复用模板
版本	v2.0 - Dual Track
基线	GHEAS Research Release 1.0 (冻结保存)
状态	轨道 A 已运行；轨道 B 引擎已验证，真实政策数据待登记
日期	2026-07-20

核心结论 当前国家级面板不能识别 PM2.5 对期望寿命的因果方向和大小。最可靠的结果是：横截面负相关主要混合了国家发展差异，纵向结果又对趋势、差分、暴露窗口和时间安慰剂高度敏感。

本报告的成果不是寻找一个更显著的系数，而是建立一套能够阻止错误确定性的分析模式：轨道 A 系统暴露观察性结果的不稳定性；轨道 B 只有在政策或自然实验的识别前提通过后才允许使用因果语言。

1. 研究定位与方法学目标

GHEAS v1
已完成从官方数据、清洗、国家匹配、横截面分析、多因素回归、双向固定效应到论文发布的完整流程。v2 不覆盖这一成果，而是将其冻结为可复核范本，并重新设计从研究问题到结论的通用方法。

层次	回答的问题	允许的结论
基线 v1	原有流程能否重复计算	生态学关联范本
轨道 A	关联是否跨合理模型稳定	条件关联及其不确定性
轨道 B	是否存在可辩护的外生变化	限定前提下的因果效应
综合层	两条证据路线是否一致	证据边界、冲突原因和下一数据门槛

2. 统一研究协议

每项研究先登记目标人群、暴露对照、结果尺度、时间窗口和权重，随后登记变量角色与数据版本。协议通过 JSON Schema 校验并生成 SHA-256；模型输出携带协议哈希、输入哈希和软件版本。

变量角色	PM2.5 案例中的处理	原因
预先混杂因素	滞后 GDP、滞后城市化	在暴露和结果之前，可能共同影响两者
中介	同期医疗支出	污染负担或发展可能通过医疗支出影响健康
结果后代/碰撞风险	65 岁以上人口比例	生存改善本身会改变老年人口比例
仅描述或分层	地区、当前收入组	用于异质性和年度冲击，不机械解释为因果控制

3. 数据起点与测量边界

现有轨道 A 数据包含 4,406 个国家-年份、186 个国家或经济体、2000-2023 年共 24 年。PM2.5 是基于 GBD 的模型化国家暴露估计；期望寿命综合联合国人口估计和国家统计来源。两者均有测量与修订不确定性，但现有下载不包含逐观测不确定区间。

重要边界 期望寿命是期间生命表指标，不是对当年出生婴儿实际未来寿命的直接观察；国家平均暴露也不能转写成个人暴露效应。

4. 轨道 A：强化观察性分析

主模型为国家固定效应、年份固定效应、当年 PM2.5，以及至少滞后一年的人均 GDP 和城市化。总效应主模型排除 65 岁以上人口比例与同期医疗支出。预设稳健性包括国家独立趋势、一年与五年变化模型、3/5/10 年暴露均值、非线性样条、未来暴露安慰剂和 Driscoll-Kraay 不确定性。

ID	规格	N	估计	95% 区间
A0	Two-way fixed effects, exposure only	4,406	+0.391	+0.107 至 +0.674
A1	Primary: two-way fixed effects with lagged confounders	4,220	+0.401	+0.150 至 +0.651
A2	Primary plus country-specific linear trends	4,220	-0.038	-0.183 至 +0.107

ID	规格	N	估计	95% 区间
A3	1-year change model	4,220	-0.006	-0.086 至 +0.074
A4	5-year change model	3,476	+0.134	-0.029 至 +0.297
A5	Three-year mean exposure	4,034	+0.561	+0.199 至 +0.922
A6	Five-year mean exposure	3,662	+0.609	+0.202 至 +1.017
A7	Ten-year mean exposure	2,732	+1.179	+0.575 至 +1.783
A8	Future-exposure placebo	4,034	+0.364	+0.121 至 +0.607
A9	Nonlinear 5-year exposure-response spline	3,662	+1.207	+0.051 至 +2.364
A10	Primary model with Driscoll-Kraay uncertainty	4,220	+0.401	+0.107 至 +0.694

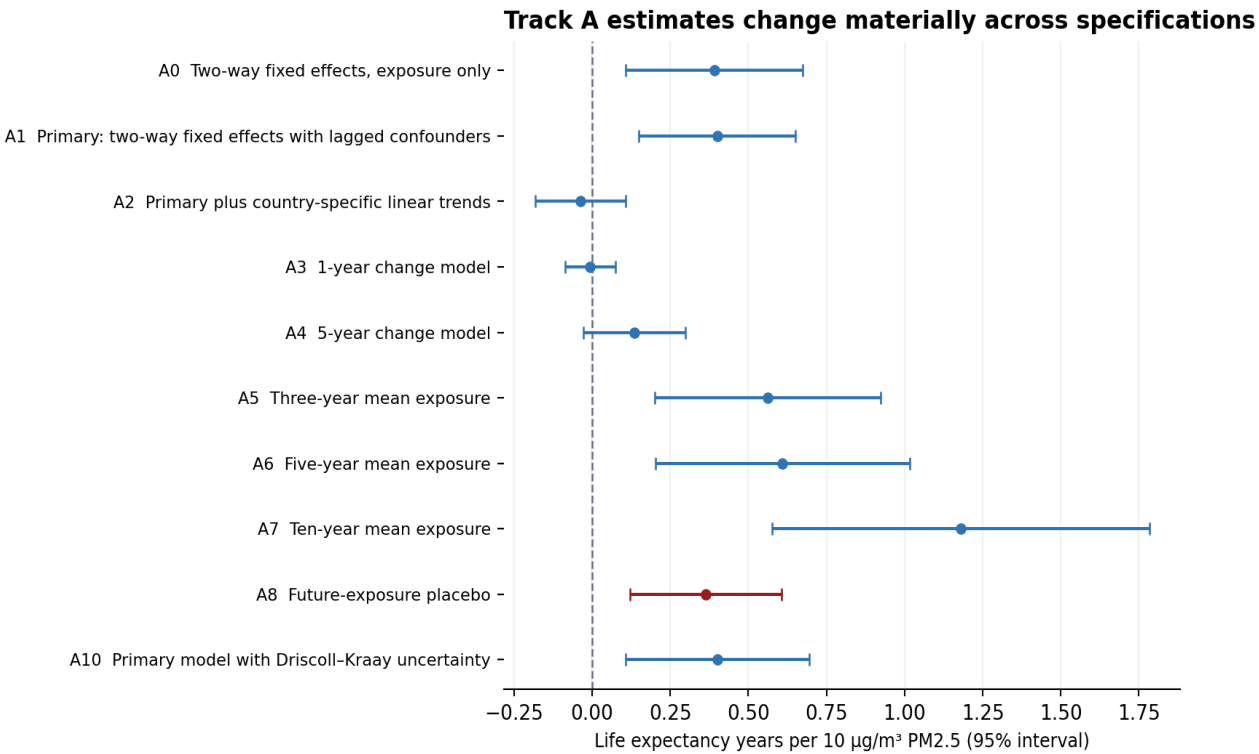


图 1. 轨道 A 的估计对模型规格高度敏感。A8 为未来暴露安慰剂。

5. 轨道 A 的判定

A1 主模型估计为 +0.401 年/10 $\mu\text{g}/\text{m}^3$ (95% 区间 +0.150 至 +0.651) , 但加入国家独立线性趋势后变为 -0.038 (-0.183 至 +0.107) 。一年变化模型同样接近零。更关键的是, 下一年的 PM2.5 仍与当前期望寿命显著相关 (A8 : +0.364 , 95% 区间 +0.121 至 +0.607) 。

轨道 A 结论 轨道 A 未能识别稳定的独立关联：趋势、变化模型和暴露窗口会改变方向或区间结论。这说明当前结果对研究设计敏感，并不表示 PM2.5 对健康无害。

不能把 A1 的正系数解释为污染具有保护作用；它更可能反映残余趋势、暴露测量误差或时间结构未被充分建模。轨道 A 的价值正是阻止研究者从多个规格中挑出一个符合叙事的结果。

6. 轨道 B：政策或自然实验

轨道 B 需要明确的处理分配机制，例如法规分期实施、燃煤设施关闭、行政阈值、地理边界或外生输送变化。参考实现采用分组-时间双重差分：每个政策批次与当时尚未处理或从未处理的地区比较，再按处理批次规模汇总事件时间效应。

识别检查	通过标准	失败后的处理
暴露第一阶段	政策确实改变 PM 暴露	不估计健康因果效应
处理前趋势	预先事件期与零相容	降级为设计不充分
负对照结果	不应受污染影响的结果接近零	调查共同政策或趋势
预期与反转	处理时间可定义且无未建模反转	更换设计或重新编码
干扰与溢出	对照地区未被输送或政策波及	重建比较组或空间模型

7. 轨道 B 的方法验证与现实数据门槛

确定性模拟数据包含 80 个地区、16 年、3 个分期政策批次和 20 个从未处理地区。引擎恢复出污染下降 -5.71 个模拟单位、健康结果改善 0.51 个单位，负对照结果为 -0.030，且处理前区间与零相容。

解释限制 这些数字只验证程序能恢复已知模拟效应，不构成真实 PM2.5 证据。真实轨道 B 仍为空值，直到政策名称、法律来源、执行时间、地理范围、分配规则和对照地区全部登记。

8. 双轨综合结论

当前国家级面板不能识别 PM2.5 对期间期望寿命的因果方向和大小。

判断	当前结论
可以支持	国家横截面的较大负相关主要受发展条件混杂；纵向估计对趋势、变化规格、暴露窗口和时间安慰剂敏感。
不能支持	当前数据既不支持“没有影响”，也不支持一个数值化的因果寿命效应。
方法学成果	轨道 A 用于暴露关联敏感性；轨道 B 用明确分配机制和证伪检查约束因果语言。
下一证据门槛	登记真实政策或自然实验面板，并获得年度年龄标准化心肺疾病结果。

9. 可迁移到其他研究主题的模板

阶段	动作	必须留下的记录
01	定义问题	人群、暴露对照、结果、时间和权重
02	建立变量角色	混杂、中介、碰撞/后代、描述变量
03	登记数据	来源、版本、许可、测量类型和不确定性
04	轨道 A	关联主模型、趋势、滞后、变化和安慰剂
05	轨道 B	外生分配、第一阶段、事件研究和负对照
06	综合证据	解释一致与冲突，不机械平均系数
07	发布结论	同时报告支持、反面证据、限制和下一门槛

这个模板适用于温度与死亡、饮水污染与疾病、食品暴露与代谢结果、职业暴露与健康、政策变化与医疗利用等问题。可迁移的是协议和停止规则，不是 PM2.5 的具体变量或系数。

10. 数据升级路线

第一优先级是获得年度年龄标准化全因和心肺疾病死亡率，并记录估计区间；第二优先级是人口、吸烟、气候、冲突和教育等具有明确时间顺序的数据；第三优先级是登记一个可公开复核的政策或自然实验。WHO GHE 2021 已被选为优先结果体系，但公开摘要下载主要提供若干基准年份，旧 OData 中发现的相关 IHD 和脑血管指标只有 2004 年，不能冒充完整年度面板。

如果轨道 B 使用分期政策，必须同时验证政策是否真正改变污染暴露；如果第一阶段不存在，即使健康结果显著，也不能归因于污染。对于阈值或边界分配，应改用回归不连续；对于外生输送或监管工具，应使用工具变量并单独论证排除限制。

11. 可重复运行与文件结构

交付物	作用
protocol.json	研究问题、变量角色、数据源和判定规则
study-protocol.schema.json	机器校验协议是否完整
track-a-results.json / report.md	观察性全规格结果和判定
track-b-demo-results.json / report.md	因果引擎的确定性验证
policy-events.template.csv	真实政策面板的标准字段
dual-track-reconciliation.json / .md	两轨证据的统一边界
archive/release-1.0/manifest.sha256	冻结基线的完整性校验

12. 参考来源

1. World Bank. PM2.5 air pollution, mean annual exposure (EN.ATM.PM25.MC.M3); underlying GBD 2023 exposure estimates. <https://data.worldbank.org/indicator/EN.ATM.PM25.MC.M3>
2. World Bank. Life expectancy at birth, total (SP.DYN.LE00.IN); UN World Population Prospects and national statistical sources. <https://data.worldbank.org/indicator/SP.DYN.LE00.IN>
3. World Health Organization. Global Health Estimates 2021: deaths by cause, age, sex, country and region, 2000-2021. <https://www.who.int/data/global-health-estimates/>
4. Callaway B, Sant'Anna PHC. Difference-in-Differences with multiple time periods. *Journal of Econometrics*. 2021;225(2):200-230. doi:10.1016/j.jeconom.2020.12.001.
5. Ebenstein A, Fan M, Greenstone M, He G, Zhou M. New evidence on sustained exposure to air pollution and life expectancy from China's Huai River Policy. *PNAS*. 2017;114(39):10384-10389. doi:10.1073/pnas.1616784114.
6. Gasparrini A, Armstrong B, Kenward MG. Distributed lag non-linear models. *Statistics in Medicine*. 2010;29(21):2224-2234. doi:10.1002/sim.3940.
7. Burnett R, et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *PNAS*. 2018;115(38):9592-9597. doi:10.1073/pnas.1803222115.
8. Stevens GA, et al. Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. *PLoS Medicine*. 2016;13(6):e1002056.

附录 A. 判定规则摘要

轨道 A 只有在主规格、趋势模型、变化模型、暴露窗口和不确定性估计在方向与实质大小上相容，且未来暴露安慰剂通过时，才能称为稳定关联。即使达到这一标准，也不升级为因果。

轨道 B 只有在政策分配机制、比较组重叠、处理前趋势、暴露第一阶段、预期效应、负对照和溢出风险全部被记录并通过时，才允许在限定前提下使用因果语言。显著性不是升级证据等级的充分条件。

协议 SHA-256 : 78b4a0c505bc948ca0b6e19ec0624a603e160850a745491dbc8368e5ac4c78fb

轨道 A 输入 SHA-256 : c7179fa8cbe584df7f0094b43b72a64e129a303a18bde64fabe3423eb5ddf46c