Test file: Causal Inference in Environmental Epidemiology: Application of Targeted Maximum Likelihood Estimation to Assess the Effect of Air Pollution Reduction Policies on Cardiovascular Mortality

Authors: Mohammad Tarique[1], Lei Yang[2], Chao Luo[3]*
Affiliations:
[1] University of Missouri, Columbia 65211, USA
[2] Xi'an Jiaotong University, Xi'an 710049, China
[3] The Affiliated Huaian No.1 People's Hospital of Nanjing Medical University, Huaian 223300, China
***Corresponding Author:**
Chao Luo, Email: hayylch@njmu.edu.cn

Abstract

This study employs targeted maximum likelihood estimation to rigorously estimate the causal effect of comprehensive air pollution reduction policies on cardiovascular mortality across 156 urban areas from 2005 to 2022. By analyzing longitudinal mortality data, air quality monitoring, and policy implementation records, we found that sustained PM2.5 reductions of 10 $\mu g/m^3$ achieved through multi-sectoral policies were associated with an 11.2% decrease in cardiovascular mortality. The TMLE approach, which efficiently adjusts for time-varying confounding and model misspecification, provides robust causal evidence supporting the cardiovascular benefits of air quality regulations. Policy effects were modified by baseline pollution levels, population age structure, and concurrent public health interventions.

Keywords: Causal Inference; Environmental Epidemiology; Air Pollution; Cardiovascular Mortality; Targeted Maximum Likelihood Estimation; Policy Evaluation; Time-Varying Confounding

1. Introduction

Establishing causal relationships in environmental epidemiology presents substantial methodological challenges due to complex confounding structures, time-varying exposures, and frequent model misspecification. While numerous ecological studies have reported associations between air pollution and cardiovascular mortality, causal interpretation of these relationships remains limited by methodological constraints. Traditional regression approaches often fail to adequately address time-varying confounding and may produce biased effect estimates. Targeted maximum likelihood estimation represents an advanced causal inference methodology that combines machine learning with semiparametric theory to generate efficient, double-robust effect estimates. This study applies TMLE to evaluate the causal effect of air pollution reduction policies on cardiovascular mortality, addressing key methodological challenges in environmental epidemiology while providing policy-relevant evidence for public health decision-making. The analysis incorporates 18 years of longitudinal data from diverse urban settings, examining effect modification by demographic and environmental factors to identify populations most likely to benefit from intervention.

2. Methods

2.1 Study Population and Data Sources

We analyzed data from 156 urban areas across 15 countries from 2005 to 2022, representing diverse geographic, economic, and environmental contexts. Mortality data were obtained from national vital statistics systems, with cardiovascular mortality defined using ICD-10 codes I00-I99. Air quality data included daily PM2.5, NO2, and O3 concentrations from ground monitoring stations supplemented by satellite-based estimates. Policy implementation was documented through systematic review of environmental regulations, including industrial emission standards, vehicle fleet modernization, and clean energy initiatives. Covariate data encompassed demographic characteristics, socioeconomic indicators, healthcare access, meteorological factors, and co-occurring public health programs.

2.2 Causal Framework and Estimand Definition

We formalized the analysis using a causal directed acyclic graph identifying air pollution reduction policies as the intervention of interest, cardiovascular mortality as the outcome, and time-varying confounders including economic development, healthcare infrastructure, and behavioral factors. The target causal parameter was the population intervention effect of sustained PM2.5 reduction on cardiovascular mortality risk. We defined the intervention as achieving and maintaining a 10 μg/m³ reduction in annual average PM2.5 concentrations through policy implementation, corresponding to realistic pollution control targets in urban settings.

2.3 Targeted Maximum Likelihood Estimation Implementation

TMLE was implemented through a two-stage procedure: first, initial estimates of the outcome and exposure mechanisms were obtained using an ensemble machine learning approach combining generalized additive models, random forests, and gradient boosting. Second, these estimates were targeted through a fluctuation step to optimize the bias-variance tradeoff for the causal parameter of interest. We accounted for spatial and temporal correlation through geographic and calendar-time stratification. Sensitivity analyses assessed robustness to unmeasured confounding, selection bias, and exposure measurement error using quantitative bias analysis and E-values.

3. Results

3.1 Causal Effect Estimates

The TMLE analysis demonstrated a significant causal effect of air pollution reduction on cardiovascular mortality. A sustained 10 μg/m³ reduction in PM2.5 concentrations was associated with an 11.2% decrease in cardiovascular mortality. The effect was non-linear, with greater relative benefits observed at higher baseline pollution levels. The TMLE estimates showed improved efficiency and smaller confidence intervals compared to traditional regression approaches, particularly in the presence of complex confounding structures. The double-robust property of TMLE provided protection against model misspecification, with consistent results across different machine learning algorithm combinations.

3.2 Effect Modification and Heterogeneity

Significant effect modification was observed across population subgroups. Older adults demonstrated greater mortality reductions, with effects approximately 40% larger in populations over 65 years. Urban areas with higher baseline cardiovascular disease burden showed amplified benefits, suggesting that pollution reduction may address existing health disparities. The timing and comprehensiveness of policy implementation also modified effects, with integrated multi-sectoral approaches yielding larger benefits than isolated interventions. Geographic heterogeneity reflected regional differences in pollution composition, population susceptibility,

Zentime

and healthcare system capacity.

3.3 Sensitivity and Validation Analyses

Sensitivity analyses supported the robustness of the causal conclusions. The E-value analysis indicated that unmeasured confounding would need to be substantial to explain away the observed effects. Multiple imputation for missing exposure data produced consistent results, and negative control outcomes showed no association with the intervention, supporting the specificity of the findings. Cross-validation demonstrated good model performance across different geographic and temporal subsets, with minimal evidence of overfitting.

4. Discussion

4.1 Methodological Advancements and Causal Interpretation

This study demonstrates the value of advanced causal inference methods in environmental epidemiology. The TMLE approach successfully addressed time-varying confounding while maintaining statistical efficiency, providing stronger causal evidence than previous ecological studies. The machine learning integration enabled flexible modeling of complex exposure-response relationships without strict parametric assumptions. The findings substantially strengthen causal inference regarding air pollution effects on cardiovascular health, addressing a key limitation in the existing evidence base.

4.2 Public Health and Policy Implications

The results provide robust evidence supporting the cardiovascular benefits of air pollution control policies. The magnitude of mortality reduction suggests substantial public health gains achievable through sustained environmental regulations. The identified effect modifiers can guide targeted interventions for vulnerable populations and optimize resource allocation. Policy makers should consider these findings when setting air quality standards and evaluating regulatory impacts, particularly in highly polluted urban areas with aging populations.

4.3 Limitations and Future Directions

Despite methodological advancements, several limitations warrant consideration. Exposure misclassification remains possible, though the multi-source exposure assessment likely minimized this bias. The ecological study design prevents individual-level causal inference, and residual confounding cannot be entirely excluded. Future research should incorporate individual-level data, examine specific biological pathways, and evaluate interactive effects with climate change mitigation strategies.

5. Conclusion

The application of targeted maximum likelihood estimation provides robust causal evidence that air pollution reduction policies significantly decrease cardiovascular mortality. The methodological approach addresses key limitations in environmental epidemiology and can be adapted to other complex exposure-outcome relationships. These findings strengthen the scientific foundation for air quality regulations and highlight the importance of causal inference methods in environmental health research. Continued advancement and application of rigorous causal methods will enhance our ability to evaluate environmental interventions and protect population health in an increasingly urbanized world.

References

Zentime

[1] van der Laan, M. J., & Rose, S. (2018). Targeted Learning: Causal Inference for Observational and Experimental Data. Springer.
[2] Dominici, F., et al. (2022). Assessing the effect of air pollution on mortality using causal inference methods: a systematic review. Epidemiologic Reviews, 44(1), 1-15.
[3] Henneman, L. R., et al. (2023). Causal inference in environmental health: lessons from air pollution epidemiology. Environmental Health Perspectives, 131(2), 025001.
[4] Schwartz, J., et al. (2021). Estimating causal effects of air pollution exposure on mortality: new approaches to an old question. American Journal of Epidemiology, 190(8), 1579-1588.
[5] Zigler, C. M., & Dominici, F. (2022). The need for causal inference in air pollution epidemiology: balancing science and policy. Annual Review of Public Health, 43, 1-15.

Zentime