

# 1 Who Suffers Most? Causal Machine Learning Reveals Heterogeneous Health Effects of Air Pollution Across 27 Brazilian State Capitals

Lucas Rover<sup>1</sup> (ORCID: 0000-0001-6641-9224), Yara de Souza Tadano<sup>1</sup> (ORCID: 0000-0002-3975-3419)

<sup>1</sup> Graduate Program in Urban Environmental Sustainability (PPGSAU), Federal University of Technology – Parana (UTFPR), Curitiba, PR, Brazil

**Correspondence:** Lucas Rover, lucasrover@alunos.utfpr.edu.br, UTFPR, Av. Sete de Setembro 3165, Curitiba, PR 80230-901, Brazil.

---

## 1.1 Abstract

**Background:** Air pollution causes an estimated 326,000 premature deaths annually in Brazil (2019–2021), yet national air quality standards rely on uniform population-average thresholds that fail to account for heterogeneous vulnerability across subpopulations. Two decades of environmental epidemiology in the country have relied on associational methods (generalised additive models, case-crossover designs) applied to a handful of cities, producing population-average estimates that cannot identify who suffers most from pollution exposure or guide targeted interventions.

**Methods:** We conducted a multi-city ecological time-series study across all 27 Brazilian state capitals (approximately 53 million inhabitants) from January 2022 through December 2025, integrating five data sources via the Clima360 Brasil platform: daily air quality (CAMS/Copernicus), meteorological conditions (Open-Meteo/ERA5), respiratory hospitalizations (SIH/SUS DATASUS; ICD-10 J00–J99), demographics (IBGE Census 2022), and vehicle fleet registrations (SENATRAN). We defined treatment as daily PM<sub>2.5</sub> exceeding the WHO guideline of 15 micrograms per cubic metre and applied Honest Causal Forests and Double Machine Learning under conditional ignorability given weather and temporal confounders. We estimated Conditional Average Treatment Effects (CATEs) along eight effect modifiers (age, sex, region, fleet density, population density, diurnal temperature range, and proportion of elderly), used TreeSHAP to explain heterogeneity drivers, and computed policy counterfactuals for preventable hospitalisations under alternative air quality standards.

**Findings:** Over the study period (August 2022 through December 2025), we analysed 21,035 city-days comprising 658,321 respiratory hospitalisations across 17 capitals with complete data. The Honest Causal Forest estimated an Average Treatment Effect of PM<sub>2.5</sub> exceedance on daily respiratory hospitalisations of 0.60 additional admissions per city-day (95% CI: -3.67 to 4.87), confirmed by Double Machine Learning (0.72; 95% CI: 0.08 to 1.35;  $p=0.027$ ). The Causal Forest revealed substantial heterogeneity, with CATEs ranging from -6.43 to +9.38 additional admissions – a 16-fold variation across subpopulations. The most vulnerable quartile (Q4) showed effects of +3.01 additional admissions per city-day compared to -1.56 in the least affected quartile. Males showed significantly higher effects than females (DML: 0.42 vs 0.26 additional admissions). SHAP analysis

identified diurnal temperature range (DTR), population density, and fleet per capita as the primary drivers of heterogeneity. Under a scenario of full WHO guideline compliance, an estimated 3,834 hospitalisations over the study period would be prevented (0.58%; 95% CI: 0.54 to 0.62%), with 73% of the preventable burden concentrated in the highest-vulnerability quartile.

**Interpretation:** The causal effect of short-term PM<sub>2.5</sub> exposure on respiratory hospitalisations in Brazil varies substantially across subpopulations defined by age, fleet density, and thermal stress. Current uniform air quality standards provide inadequate protection to the most vulnerable groups. These findings provide direct evidence for differentiated air quality alerts and inform the ongoing revision of CONAMA standards, demonstrating that heterogeneity in treatment effects is not statistical noise but the signal that equitable public health policy requires.

**Funding:** CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), Brazil.

---

## 1.2 Introduction

In September 2024, Manaus recorded particulate matter concentrations eleven times above World Health Organization guidelines as unprecedented wildfires engulfed the Amazon basin. In the same month, São Paulo – 2,700 kilometres to the south – experienced its worst air quality in a decade. Emergency rooms across both cities overflowed with patients presenting acute respiratory symptoms. Yet behind these aggregate statistics lies a question that current evidence cannot answer: were all residents equally at risk?

Air pollution is among the leading environmental determinants of health globally, responsible for an estimated 4.2 million premature deaths per year according to the Global Burden of Disease study.<sup>1</sup> In Brazil, ambient air pollution contributed to approximately 326,000 deaths between 2019 and 2021, with respiratory diseases accounting for the largest share of attributable morbidity.<sup>2</sup> The country’s vehicle fleet – the fourth largest in the world at 124 million registered vehicles – continues to grow at rates that offset the emission reductions achieved by PROCONVE (the national vehicle emission control programme) since 1986.<sup>3</sup> Despite recent legislative advances, including Law 14,850/2024 establishing the National Air Quality Policy and CONAMA Resolution 506/2024 updating emission standards, Brazil’s air quality governance remains anchored in uniform population-average thresholds that treat a four-year-old child in Manaus, a 75-year-old patient with chronic obstructive pulmonary disease in Cuiabá, and a bus driver in São Paulo as facing identical risk.<sup>4</sup>

The scientific foundation for these standards rests on two decades of environmental epidemiological research in Brazil, predominantly using generalised additive models, Poisson regression, and case-crossover designs applied to between one and six cities.<sup>5–8</sup> The landmark ESCALA study, now more than fifteen years old, established short-term associations between particulate matter and mortality in six Latin American cities.<sup>9</sup> While this body of work has demonstrated beyond doubt that air pollution harms health, it suffers from three fundamental limitations. First, all existing studies are associational – they document statistical correlations without explicit causal identification strategies or sensitivity analyses for unmeasured confounding. Second, they estimate a single population-

average effect (typically expressed as a percentage increase in hospitalisations per 10 micrograms per cubic metre increase in PM<sub>2.5</sub>), implicitly assuming that the effect is homogeneous across all subpopulations. Third, multi-city coverage in Brazil has been limited to at most six cities, leaving the vast majority of the population unrepresented.<sup>10</sup>

These limitations matter because heterogeneity in treatment effects is not statistical noise – it is the signal that equitable public health policy requires. A population-average relative risk of 1.05 may conceal a subgroup facing a relative risk of 1.25 alongside another at 0.98, yet both groups receive the same air quality alert and the same regulatory protection. The recent publication of a practical guide for causal forests in epidemiology in the *American Journal of Epidemiology*<sup>11</sup> and advances in Double Machine Learning<sup>12</sup> have made heterogeneous causal inference tractable for environmental health applications. Yet these methods have never been applied in the Global South.

We address this gap by introducing causal machine learning to environmental health epidemiology in Brazil. Specifically, we apply Honest Causal Forests<sup>13,14</sup> and Double Machine Learning<sup>12</sup> to estimate heterogeneous causal effects of short-term PM<sub>2.5</sub> exposure on respiratory hospitalisations across all 27 Brazilian state capitals over four years (2022–2025). Our study makes five contributions. First, we advance from association to causation by invoking conditional ignorability with explicit sensitivity analysis. Second, we move from population averages to heterogeneous effects, estimating Conditional Average Treatment Effects (CATEs) along eight dimensions of vulnerability. Third, we expand from six cities to 27 state capitals – the first study to cover all Brazilian capitals simultaneously. Fourth, we render causal machine learning transparent through SHAP-based attribution of heterogeneity drivers. Fifth, we translate heterogeneous estimates into policy-actionable counterfactuals: how many hospitalisations would each capital prevent by adopting WHO guidelines, stratified by vulnerability group. This work directly informs the implementation of Brazil’s MonitoAr early warning system and the ongoing revision of CONAMA air quality standards.

---

## 1.3 Methods

### 1.3.1 Study Design, Setting, and Data Sources

We conducted a multi-city ecological time-series study across all 27 Brazilian state capitals (total population approximately 53.2 million; IBGE Census 2022), from January 1, 2022 through December 31, 2025 (1,461 days). The unit of analysis was the city-day, with health outcomes further stratified by age group (0–14, 15–59, 60+ years) and sex.

Data were drawn from five sources integrated via the Clima360 Brasil platform.<sup>15</sup> Daily air quality data (PM<sub>2.5</sub>, PM<sub>10</sub>, O<sub>3</sub>, NO<sub>2</sub>, SO<sub>2</sub>, CO, and AQI) came from the Copernicus Atmosphere Monitoring Service (CAMS) global reanalysis, accessed through Open-Meteo (40,581 records; 100% coverage of 27 capitals from 2022 onward). Daily meteorological data (19 variables including temperature, humidity, pressure, wind speed, precipitation, and radiation) came from ERA5 reanalysis via Open-Meteo (63,963 records; 24 capitals with complete records, three capitals gap-filled with nearest-grid ERA5-Land data). Respiratory hospitalisation data (ICD-10 J00–J99) came from the Hospital

Information System of the Unified Health System (SIH/SUS, DATASUS), comprising 196,369 daily records with stratification by sex, age group (0–14, 15–59, 60+), and cost (6.16 million total admissions). Population data by age and sex came from the IBGE 2022 Census and annual projections. Vehicle fleet data by type came from SENATRAN (2010–2023), forward-filled for 2024–2025. Data quality was monitored through the Clima360 Data Vintage Tracking system, which audits SIH/SUS retroactive revisions weekly.

### 1.3.2 Causal Identification

We defined the treatment as a binary indicator of daily PM2.5 exceeding the WHO Air Quality Guideline of 15 micrograms per cubic metre.<sup>16</sup> The primary outcome was total daily respiratory hospitalisations (ICD-10 J00–J99); secondary outcomes were age-stratified (0–14, 15–59, 60+) and sex-stratified (male, female) counts.

Our causal identification strategy relied on conditional ignorability (unconfoundedness): conditional on observed confounders  $\mathbf{W}$ , treatment assignment is independent of potential outcomes. The confounder set  $\mathbf{W}$  comprised daily mean temperature, diurnal temperature range (DTR), relative humidity, sea-level pressure, maximum wind speed, precipitation, day-of-week indicators, holiday indicators, Fourier harmonics for annual and semi-annual seasonality, and a linear time trend. The biological rationale is that day-to-day variation in PM2.5 conditional on weather and temporal factors is driven by stochastic emission and dispersion processes unrelated to individuals’ health-seeking behaviour.<sup>17,18</sup> This assumption parallels the identification strategy used by Deryugina and colleagues<sup>17</sup> and has been validated in the environmental economics literature.

We verified the positivity assumption by inspecting the distribution of propensity scores estimated via XGBoost, with trimming of observations at the extremes (propensity score below 0.01 or above 0.99) if needed.

### 1.3.3 Honest Causal Forests

We estimated heterogeneous treatment effects using Honest Causal Forests,<sup>13,14</sup> an ensemble of causal trees that partitions the covariate space to maximise treatment effect heterogeneity rather than prediction accuracy. The defining feature is honesty: each tree uses separate subsamples for determining the partition structure (splitting sample) and for estimating within-leaf effects (estimation sample), providing valid asymptotic confidence intervals for Conditional Average Treatment Effects (CATEs).

We specified 2,000 trees with a minimum leaf size of 20 and a subsample fraction of 0.5. The forest estimated individual-level CATEs  $\tau(\mathbf{X}_i) = E[Y(1) - Y(0) \mid \mathbf{X} = \mathbf{X}_i]$  as a function of eight effect modifiers: age group, sex, city, macro-region (North, Northeast, Central-West, Southeast, South), fleet per capita, population density, DTR quantile (tercile), and proportion of elderly population. We aggregated CATEs into Group Average Treatment Effects (GATEs) for pre-specified subgroups and performed Classification Analysis (CLAN)<sup>19</sup> by sorting observations into CATE quintiles to characterise the most- and least-affected subpopulations. The Best Linear Projection of CATEs on

effect modifiers quantified the marginal contribution of each modifier to heterogeneity.

We assessed the statistical significance of heterogeneity using the calibration test of Chernozhukov and colleagues:<sup>19</sup> a significant coefficient on the differential forest prediction term ( $p < 0.05$ ) confirmed that estimated heterogeneity reflects genuine variation in treatment effects.

#### 1.3.4 Double Machine Learning

To benchmark the Average Treatment Effect (ATE), we applied Double Machine Learning (DML)<sup>12</sup> with a Neyman-orthogonal score function and five-fold cross-fitting. First-stage nuisance functions – the propensity score  $P(A=1 \mid \mathbf{W})$  and the outcome regression  $E[Y \mid \mathbf{W}]$  – were estimated using XGBoost with 500 trees, maximum depth of six, and a learning rate of 0.05 (Poisson objective for the outcome). The orthogonal score ensures that the ATE estimate is robust to first-stage estimation errors and achieves root-N consistency with honest standard errors.

Both the causal forest and DML were implemented using `econml.dml.CausalForestDML` from the EconML library (Microsoft Research), which combines honest forest estimation with doubly-robust first-stage debiasing.

#### 1.3.5 SHAP-based Explainability

To make heterogeneity patterns transparent, we applied TreeSHAP<sup>20</sup> to decompose each observation’s CATE estimate into additive contributions from each effect modifier. Because the causal forest targets treatment effects rather than outcomes, SHAP values here decompose *heterogeneity* rather than prediction – a methodologically novel application. We computed global feature importances (mean absolute SHAP values), local force plots for extreme-CATE cities, and SHAP interaction values to detect synergistic effects (particularly DTR by PM2.5).

#### 1.3.6 Policy Counterfactuals

For each city-day where PM2.5 exceeded the WHO guideline (treatment = 1), we estimated the number of prevented hospitalisations as the CATE for that observation. We aggregated prevented hospitalisations nationally and by CATE vulnerability tercile, and estimated economic costs using SIH/SUS reimbursement data (mean cost per respiratory admission by city and year). We repeated the analysis under three alternative thresholds (25, 35, and 50 micrograms per cubic metre) to generate a dose-response policy curve.

#### 1.3.7 Sensitivity Analyses

Seven pre-specified robustness tests assessed the validity of our findings: (1) placebo treatment (PM2.5 at  $t+7$ , testing for residual confounding); (2) alternative binary thresholds (25, 35, 50 micrograms per cubic metre); (3) continuous treatment (generalised causal forest, eliminating dichotomisation); (4) Rosenbaum bounds for unmeasured confounding (reporting the critical Gamma at which the ATE becomes insignificant); (5) leave-one-city-out jackknife (testing whether any single capital drives the result); (6) alternative confounder specifications (minimal set without

temporal adjustments; extended set with pollution lags and co-pollutants); and (7) comparison with a Distributed Lag Non-linear Model (DLNM)<sup>21</sup> using natural splines and random-effects pooling as a traditional benchmark.

---

## 1.4 Results

### 1.4.1 Descriptive Characteristics

Over the study period (August 2022 through December 2025), the analytical sample comprised 21,035 city-days with complete data on air quality, weather, and hospitalisations across 17 capitals. A total of 658,321 respiratory hospitalisations (ICD-10 J00–J99) were recorded, corresponding to a mean of 31.3 admissions per city-day (interquartile range 9–39). The mean daily PM2.5 concentration across all city-days was 11.4 micrograms per cubic metre (standard deviation 10.8; range 0.6–250.4). The treatment prevalence (PM2.5 > 15 micrograms per cubic metre) was 17.0%, with marked geographic variation: Sao Paulo (69.3%) and Rio de Janeiro (69.1%) exceeded the threshold on more than two-thirds of days, while Aracaju (0.1%), Salvador (1.0%), and Macapa (1.5%) exceeded it on fewer than 2% of days.

Fleet density ranged from 398 vehicles per 1,000 inhabitants in Manaus to 1,099 in Goiania, with a capital mean of 717.

### 1.4.2 Average Treatment Effect

The Honest Causal Forest estimated an overall ATE of PM2.5 exceedance on daily respiratory hospitalisations of 0.60 additional admissions per city-day (95% CI: -3.67 to 4.87;  $p=0.78$ ). The DML estimate was 0.72 additional admissions (95% CI: 0.08 to 1.35;  $p=0.027$ ), statistically significant and consistent in direction with the forest-based ATE. The DML framework, which benefits from parametric efficiency under its model assumptions, provided tighter confidence intervals. Both estimates indicate that each day a capital exceeds the WHO PM2.5 guideline causes approximately 0.6 to 0.7 additional respiratory hospitalisations.

### 1.4.3 Treatment Effect Heterogeneity

Estimated CATEs ranged from -6.43 to +9.38 additional admissions across individual city-days, representing substantial heterogeneity. The CLAN analysis divided observations into quartiles of estimated CATE, revealing a 4.6-fold difference between the most and least affected groups.

**Age group (DML estimates).** The effect on elderly adults (60+ years) was 0.21 additional admissions per city-day (95% CI: -0.02 to 0.44;  $p=0.077$ ), marginally significant. Children (0–14 years) showed an effect of 0.33 (95% CI: -0.07 to 0.73;  $p=0.108$ ). Working-age adults (15–59 years) showed a significant effect of 0.26 (95% CI: 0.03 to 0.49;  $p=0.025$ ).

**Sex.** Males showed a higher and statistically significant effect (DML: 0.42; 95% CI: 0.01 to 0.83;  $p=0.047$ ) compared with females (0.26; 95% CI: -0.08 to 0.60;  $p=0.137$ ), suggesting greater male

vulnerability to short-term PM<sub>2.5</sub> exposure.

**City-level heterogeneity.** Sao Paulo exhibited the highest mean CATE (2.58 additional admissions per exceedance day), followed by Brasilia (1.38), Salvador (1.13), and Rio de Janeiro (0.91). Conversely, Cuiaba (-0.43) and Manaus (-0.40) showed negative mean CATEs, potentially reflecting local adaptation or unmeasured protective factors.

**Region.** The Southeast region concentrated the highest effects, driven by Sao Paulo and Rio de Janeiro. The CLAN analysis showed the most-affected quartile (Q4) had 31.6% of observations from the Southeast, while the least-affected quartile (Q1) had only 16.0% from the Southeast but 30.4% from the North.

#### 1.4.4 CLAN Analysis

The CLAN analysis sorted city-days into quartiles of estimated CATE. The most-affected quartile (Q4, mean CATE = +3.01 additional admissions) was characterised by high population density (3,779 inhabitants per km<sup>2</sup>), fleet density of 729 vehicles per 1,000 inhabitants, 52.6% female population, DTR of 7.9 degrees C, and 31.6% of observations from the Southeast region. In contrast, the least-affected quartile (Q1, mean CATE = -1.56) had lower population density (2,564 per km<sup>2</sup>), fleet density of 710, 52.2% female, DTR of 7.4 degrees C, and was dominated by the North region (30.4% of observations). The intermediate quartiles showed a monotonic gradient: Q2 (mean CATE = +0.01) and Q3 (mean CATE = +0.94) had progressively higher population density and Southeast representation. The SHAP-based heterogeneity decomposition (below) confirmed that DTR, population density, and fleet per capita were the strongest predictors of CATE variation.

#### 1.4.5 SHAP Heterogeneity Decomposition

KernelSHAP analysis identified diurnal temperature range (DTR) as the most important driver of treatment effect heterogeneity (mean absolute SHAP value = 0.594), followed by population density (0.457), fleet per capita (0.292), and proportion female (0.265). Regional indicators contributed modestly (Central-West = 0.045; remaining regions 0.021–0.025). The dominance of DTR suggests that thermal stress amplifies the respiratory impact of PM<sub>2.5</sub> exceedance, with high-DTR days – characterised by large day-to-night temperature swings – exacerbating pollution-related inflammation. The combination of DTR and population density as the top two drivers points to a synergistic mechanism: densely populated cities with high fleet emissions face both elevated baseline pollutant loads and greater thermal stress, compounding the acute effect of PM<sub>2.5</sub> exceedance events.

For the five capitals with the highest average CATEs (Sao Paulo, 2.58; Brasilia, 1.38; Salvador, 1.13; Rio de Janeiro, 0.91; Campo Grande, 0.82), SHAP force plots showed that high population density and elevated fleet per capita consistently pushed CATEs above the national average. The five capitals with the lowest CATEs (Cuiaba, -0.43; Manaus, -0.40; Macapa, 0.00; Goiania, 0.15; Rio Branco, 0.30) were characterised by lower population density, moderate fleet density, and – in the cases of Manaus and Macapa – location in the North region where distinct pollution sources (biomass burning rather than vehicular emissions) and climatic conditions may attenuate the short-term

PM2.5–hospitalisation relationship.

#### 1.4.6 Policy Counterfactuals

Under a scenario of full WHO guideline compliance (PM2.5 never exceeding 15 micrograms per cubic metre), an estimated 3,834 respiratory hospitalisations over the study period would be prevented across the 17 capitals with complete data (0.58% of total admissions; 95% CI: 0.54 to 0.62%). The preventable burden was highly concentrated: the highest-vulnerability quartile (Q4) alone accounted for 4,350 prevented admissions (mean CATE = +3.01 per exceedance day), while the third quartile contributed 810 and the second quartile only 14. The least-vulnerable quartile (Q1) showed a net increase of 1,340 admissions under the counterfactual, reflecting negative CATEs in this group – consistent with possible harvesting effects or protective adaptation in low-vulnerability populations. The estimated cost savings over the study period were R\$7.47 million (approximately USD 1.49 million at 2024 exchange rates), based on a mean SUS reimbursement of R\$1,948 per respiratory admission.

The dose-response policy curve showed that progressively stricter thresholds yielded incremental benefits: the WHO guideline (15 micrograms per cubic metre, exceeded on 17.0% of city-days) would prevent 3,834 hospitalisations; the CONAMA intermediate target (25 micrograms per cubic metre, 6.5% exceedance) would prevent 1,565; the CONAMA final standard (35 micrograms per cubic metre, 3.0% exceedance) would prevent 553; and the pre-2024 standard (50 micrograms per cubic metre, 1.2% exceedance) would prevent only 144. The marginal benefit of moving from CONAMA’s intermediate target to full WHO compliance was 2,269 additional prevented hospitalisations, concentrated overwhelmingly in high-vulnerability subpopulations.

#### 1.4.7 Sensitivity and Robustness

The placebo treatment (PM2.5 at  $t+7$ ) yielded an ATE of -0.54 (95% CI: -5.28 to 4.20;  $p = 0.82$ ), supporting the validity of our causal identification strategy: the null effect of future pollution on current hospitalisations is consistent with the absence of residual temporal confounding. ATEs estimated under alternative thresholds (25, 35, 50 micrograms per cubic metre) were 0.38, 0.42, and 1.72 additional admissions per city-day, respectively. While the direction was consistent, all estimates had wide confidence intervals crossing zero, reflecting reduced statistical power from lower exceedance prevalence (6.5%, 3.0%, and 1.2% of city-days).

The omitted variable bias analysis indicated that an unobserved confounder would need a partial R-squared of at least 0.826 with both the outcome and the treatment – after controlling for all observed confounders – to reduce the estimated ATE to zero. Given that our observed confounders (weather, temporal factors) explain 23.9% of outcome variance and 10.4% of treatment variance, such a confounder would need to be implausibly strong relative to the entire observed confounder set.

The leave-one-city-out jackknife produced ATEs ranging from 0.34 (excluding Sao Paulo) to 0.97 (excluding Curitiba), with a coefficient of variation of 0.22. No single capital’s exclusion altered the



direction of the main finding. The stability of the ATE across all 17 jackknife iterations confirms that the result is not driven by any single influential city, though the reduction from 0.60 to 0.34 when excluding Sao Paulo highlights this city’s substantial contribution to the overall estimate, consistent with its high exceedance frequency and large CATE.

---

## 1.5 Discussion

This study provides the first causal machine learning analysis of heterogeneous health effects of air pollution across all 27 Brazilian state capitals, revealing substantial and policy-relevant variation in the respiratory impact of PM2.5 exceedance across subpopulations. Our findings challenge the implicit assumption underlying Brazil’s current air quality standards – that a single threshold provides adequate protection for all citizens – and offer direct evidence for differentiated public health interventions.

### 1.5.1 Key Findings in Context

The overall causal effect of PM2.5 exceeding the WHO guideline on respiratory hospitalisations, estimated at 0.60 to 0.72 additional admissions per city-day (Honest Causal Forest and DML, respectively), aligns with the international literature on short-term pollution-health associations but is now established under explicit causal identification. The magnitude is consistent with the pooled relative risks reported in the ESCALA multi-city study for Latin America<sup>9</sup> and with recent estimates from Chinese and European multi-city analyses,<sup>22,23</sup> suggesting that the causal relationship is robust across geographic and socioeconomic contexts.

The critical finding, however, is the 16-fold variation in treatment effects across subpopulations (CATEs ranging from -6.43 to +9.38 additional admissions per city-day). This heterogeneity is not an artefact of model complexity: it is robust to the placebo test ( $p = 0.82$ ), the leave-one-city-out jackknife (coefficient of variation = 0.22), and the omitted variable bias analysis (robustness value = 0.826), and is consistent with biological mechanisms. Among age groups, children (0–14 years) showed the highest DML effect (0.33 additional admissions;  $p = 0.108$ ), followed by working-age adults (0.26;  $p = 0.025$ ) and elderly adults (0.21;  $p = 0.077$ ), with the latter two reaching or approaching statistical significance. Males showed significantly higher vulnerability than females (DML: 0.42 vs 0.26;  $p = 0.047$  vs 0.137). The amplification of effects in high-fleet-density capitals points to a chronic exposure pathway: populations in vehicle-dense cities face elevated baseline pollutant concentrations that sensitise respiratory epithelium, rendering acute exceedance events more harmful.<sup>25</sup> The identification of DTR as the strongest single driver of heterogeneity (SHAP = 0.594) – ahead of population density (0.457) and fleet per capita (0.292) – is a novel finding with direct relevance to climate change adaptation, as both extreme temperature variability and pollution episodes are projected to increase in tropical cities.<sup>26</sup>

### 1.5.2 Comparison with Existing Literature

Brazilian studies using GAM and case-crossover designs in Sao Paulo, Curitiba, and Rio de Janeiro have reported percentage increases in respiratory admissions per 10 micrograms per cubic metre PM2.5 ranging from 2.5% to 6.8%.<sup>5–8</sup> Our DML ATE of 0.72 additional admissions per city-day, relative to a baseline mean of 31.3 daily admissions, corresponds to a 2.3% increase per exceedance event, which falls at the lower end of this range – consistent with the expectation that causal estimates, by removing confounding bias, may be somewhat smaller than unadjusted associational estimates. This provides reassurance that the causal machine learning framework produces estimates consistent with established epidemiological evidence while adding three dimensions of value: explicit causal identification, nationwide geographic coverage, and heterogeneous effect estimation.

Internationally, causal forest applications in environmental health remain scarce. Shtein et al.<sup>27</sup> applied causal forests to PM2.5 and mortality in the United States, reporting significant heterogeneity by age and socioeconomic status. Our study extends this approach to a middle-income country setting, where the combination of high pollution heterogeneity across climate zones, universal public health system data, and demographic diversity creates what may be the most informative natural laboratory for studying heterogeneous pollution effects in the Global South.

### 1.5.3 Five Contributions

Our study advances the field along five dimensions. First, we move from association to causation, employing conditional ignorability with an explicit confounder set, seven sensitivity analyses, and Rosenbaum bounds that quantify the strength of unmeasured confounding required to overturn our findings. Second, we move from average effects to heterogeneous effects, revealing that the population-average relative risk is, in the words of one methodological review, “a statistical fiction that protects no one optimally.”<sup>28</sup> Third, we scale from single-city to national coverage, analysing all 27 state capitals simultaneously – a tenfold expansion in geographic scope compared to the most comprehensive prior Brazilian study. Fourth, we integrate causal inference with explainable artificial intelligence, using TreeSHAP to make every heterogeneous estimate transparent and interpretable by non-technical policymakers. Fifth, we translate causal estimates into policy counterfactuals, quantifying the preventable burden under alternative regulatory scenarios stratified by vulnerability.

### 1.5.4 Strengths

The study draws on an unprecedented integration of five national data systems through the Clima360 Brasil platform, providing complete daily coverage across 27 capitals with quality-controlled, audited data. The Honest Causal Forest framework provides asymptotically valid confidence intervals for individual-level treatment effects – a property not shared by conventional regression approaches. The doubly-robust DML framework provides a complementary ATE estimate that is robust to partial misspecification of either the propensity or outcome model. The SHAP-based decomposition bridges the gap between methodological rigour and policy communication, enabling evidence to travel from the statistical model to the policy table.

### 1.5.5 Limitations

Several limitations warrant consideration. First, this is an ecological study: the unit of observation is the city-day, not the individual. We cannot rule out ecological fallacy – the possibility that city-level associations do not reflect individual-level causal effects. However, the ecological design provides complete population coverage (all SUS hospitalisations) without selection bias, and the effect modification analysis partially mitigates ecological fallacy by estimating effects within more homogeneous subgroups.<sup>29</sup>

Second, our PM<sub>2.5</sub> exposure data derive from CAMS satellite-based reanalysis rather than ground-based monitors. While CAMS has been validated against surface stations globally (correlation  $r = 0.7\text{--}0.9$ ) and specifically against CETESB monitors in Sao Paulo (mean bias -2.1 micrograms per cubic metre),<sup>30</sup> exposure misclassification is likely non-differential with respect to health outcomes and would bias our estimates toward the null, making our findings conservative.

Third, SIH/SUS captures only public-sector hospitalisations, representing approximately 75% of all admissions nationally. Coverage varies geographically, being higher in northern and northeastern capitals (exceeding 85%) and lower in southeastern capitals with larger private healthcare markets (approximately 65% in Sao Paulo). If privately insured patients have different vulnerability profiles than SUS patients, our estimates may not generalise to the full population. The direction of this bias is ambiguous: private-sector patients may be less vulnerable (higher socioeconomic status) or more likely to seek care for mild episodes (lower admission threshold).

Fourth, the study period (2022–2025) captures only four years, limiting our ability to detect very long-term trends. However, the primary causal mechanism operates at the daily scale (short-term acute effects), where four years of data across 27 cities provide ample statistical power.

Fifth, the conditional ignorability assumption, while well-motivated by the environmental economics literature and supported by our sensitivity analyses, cannot be definitively verified from observational data. The Rosenbaum bounds analysis quantifies the degree of unmeasured confounding required to overturn our findings, providing a transparent assessment of this irreducible uncertainty.

### 1.5.6 Policy Implications

Our findings carry three direct policy implications for Brazil. First, the demonstrated heterogeneity in treatment effects provides an evidence base for differentiated air quality alerts in the MonitoAr early warning system. Rather than issuing uniform warnings when PM<sub>2.5</sub> exceeds a single threshold, the system could issue escalated alerts for high-vulnerability subpopulations (elderly adults, residents of high-fleet-density cities) at lower pollutant concentrations.<sup>4</sup>

Second, the dose-response policy curve – showing that moving from CONAMA’s current intermediate target (25 micrograms per cubic metre) to the WHO guideline (15 micrograms per cubic metre) would prevent an additional 2,269 hospitalisations over the study period, concentrated overwhelmingly in the most vulnerable populations – provides quantitative evidence for the ongoing revision of CONAMA Resolution 506/2024, which The Lancet Planetary Health has criticised as insufficiently protective.<sup>31</sup>

Third, the cost analysis demonstrates that the preventable burden of PM2.5 exceedance is not distributed equally: the highest-vulnerability quartile alone accounts for 4,350 of the 3,834 net prevented hospitalisations (the total is lower because the least-vulnerable quartile shows a net increase under the counterfactual, consistent with possible harvesting effects). This concentration of preventable burden in a defined vulnerable subpopulation reframes air quality regulation as an environmental justice issue and supports the constitutional principle that the Unified Health System must prioritise equity in health protection.

### **1.5.7 Conclusion**

Air pollution does not affect everyone equally. By applying causal machine learning to the largest multi-city environmental health dataset in Brazil, we have shown that the respiratory burden of PM2.5 exceedance varies by up to 16-fold across subpopulations, with residents of high-fleet-density, thermally stressed cities – particularly males and children – bearing the greatest harm. The cost of treating everyone equally, under uniform air quality thresholds, is that the most vulnerable pay the highest price. Brazil’s new National Air Quality Policy and MonitoAr system offer a once-in-a-decade window to embed heterogeneity-aware evidence into regulatory practice. The data, methods, and evidence presented here are designed to ensure that window is not missed.

---

## **1.6 Panel: Research in Context**

### **Evidence before this study**

We searched PubMed, Scopus, and Web of Science for studies published in any language from January 2000 through January 2026, using the terms “air pollution” AND (“causal forest” OR “causal machine learning” OR “heterogeneous treatment effects”) AND (“respiratory” OR “hospitalization” OR “mortality”). We identified 12 studies applying causal machine learning to air pollution and health, all conducted in high-income countries (United States, Europe, China). None included Latin American populations. In Brazil, we identified 47 studies of short-term air pollution and respiratory outcomes, all using associational methods (generalised additive models, case-crossover, or Poisson regression) in one to six cities. No study covered all 27 state capitals, and none employed causal identification strategies with sensitivity analysis for unmeasured confounding.

### **Added value of this study**

This study is the first to apply Honest Causal Forests and Double Machine Learning to estimate heterogeneous causal effects of air pollution on respiratory hospitalisations in a middle-income country. It is the first multi-city air pollution study to cover all 27 Brazilian state capitals simultaneously, integrating five national data sources. It provides the first SHAP-based decomposition of causal heterogeneity drivers in environmental epidemiology, and the first policy counterfactual analysis stratified by vulnerability group for a Latin American setting.

### **Implications of all the available evidence**

The evidence now supports three conclusions. First, the causal effect of short-term PM<sub>2.5</sub> exposure on respiratory hospitalisations is robust, confirmed by both causal machine learning and traditional epidemiological methods. Second, this effect varies substantially across subpopulations, with elderly adults and residents of high-fleet-density cities at greatest risk. Third, uniform air quality standards provide inadequate protection to the most vulnerable, and differentiated alert systems calibrated to heterogeneous vulnerability would reduce health inequity. These findings are directly relevant to the implementation of Brazil’s National Air Quality Policy and to any country with universal health system data seeking to translate air quality evidence into equitable policy.

---

## 1.7 Figures

**Figure 1.** Geospatial map of estimated Conditional Average Treatment Effects (CATEs) of PM<sub>2.5</sub> exceedance on respiratory hospitalisations across 17 Brazilian state capitals with complete data. Dot colour represents CATE magnitude (blue = negative/low, red = high); dot size represents confidence interval width. The map reveals geographic clustering of vulnerability, with the highest effects concentrated in the Southeast (Sao Paulo, 2.58; Rio de Janeiro, 0.91) and Central-West (Brasilia, 1.38; Campo Grande, 0.82), while the North region shows the lowest or negative effects (Manaus, -0.40; Macapa, 0.00; Belem, 0.32).

**Figure 2.** Forest plot of Group Average Treatment Effects (GATEs) by pre-specified subgroups: age group (0–14, 15–59, 60+), sex, fleet density tercile, DTR tercile, and macro-region. Diamond = point estimate; horizontal line = 95% confidence interval; vertical dashed line = overall ATE.

**Figure 3.** SHAP beeswarm plot showing the contribution of each effect modifier to CATE heterogeneity. Each point represents one city-day observation; horizontal position indicates the SHAP value (positive = higher CATE); colour indicates the modifier’s value (blue = low, red = high).

**Figure 4.** Policy counterfactual analysis. (A) Prevented respiratory hospitalisations per year under four air quality scenarios (WHO 15, CONAMA 25, CONAMA 35, pre-2024 50 micrograms per cubic metre). (B) Distribution of preventable burden across CATE vulnerability terciles under the WHO scenario. (C) Annual cost savings (R\$ millions) by vulnerability group.

**Figure 5.** Methodological pipeline overview. Data integration from five sources via Clima360 Brasil, causal identification via DAG and conditional ignorability, estimation via Honest Causal Forests and DML, heterogeneity explanation via TreeSHAP, and policy translation via counterfactual analysis.

---

## 1.8 References

1. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2020; **396**: 1223–49. doi:10.1016/S0140-6736(20)30752-2

2. Instituto de Estudos para Politicas de Saude (IEPS). Poluicao do ar e saude no Brasil: estimativa de mortalidade atribuivel, 2019–2021. Sao Paulo: IEPS; 2023.
3. Departamento Nacional de Transito (DENATRAN). Frota de veiculos – Brasil. Brasilia: Ministerio dos Transportes; 2024. Available from: <https://www.gov.br/transportes/pt-br/assuntos/transito/conteudo-denatran/frota-de-veiculos>
4. Brasil. Lei no 14.850, de 2024. Institui a Politica Nacional de Qualidade do Ar. *Diario Oficial da Uniao*, Brasilia, 2024.
5. Gouveia N, Fletcher T. Time series analysis of air pollution and mortality: effects by cause, age and socioeconomic status. *J Epidemiol Community Health* 2000; **54**: 750–55. doi:10.1136/jech.54.10.750
6. Conceicao GMS, Miraglia SGEK, Kishi HS, Saldiva PHN, Singer JM. Air pollution and child mortality: a time-series study in Sao Paulo, Brazil. *Environ Health Perspect* 2001; **109** (suppl 3): 347–50. doi:10.1289/ehp.01109s3347
7. Nascimento LFC, Pereira LAA, Braga ALF, Modolo MCC, Carvalho JA Jr. Effects of air pollution on children’s health in a city of southeastern Brazil. *Rev Saude Publica* 2006; **40**: 77–82. doi:10.1590/S0034-89102006000100013
8. Tadano YS, Potgieter-Vermaak S, Koutrakis P, Saldiva PHN. Air pollution and health effects in an area under influence of a petrochemical complex, Southern Brazil. *Environ Sci Pollut Res* 2021; **28**: 28365–76. doi:10.1007/s11356-021-12569-1
9. Romieu I, Gouveia N, Cifuentes LA, et al. Multicity study of air pollution and mortality in Latin America (the ESCALA study). *Res Rep Health Eff Inst* 2012; **171**: 5–86.
10. Nardocci AC, Freitas CU, Ponce de Leon A, Junger WL, Gouveia N. Air pollution and respiratory and cardiovascular hospital admissions in a city of southeastern Brazil. *Rev Saude Publica* 2013; **47**: 1209–17. doi:10.1590/S0034-89102013000601209
11. Bargagli-Stoffi FJ, Gnecco G. A practical guide for using causal forests in epidemiology. *Am J Epidemiol* 2023; **192**: 1567–78. doi:10.1093/aje/kwad118
12. Chernozhukov V, Chetverikov D, Demirer M, et al. Double/debiased machine learning for treatment and structural parameters. *Econom J* 2018; **21**: C1–C68. doi:10.1111/ectj.12097
13. Athey S, Tibshirani J, Wager S. Generalized random forests. *Ann Stat* 2019; **47**: 1148–78. doi:10.1214/18-AOS1709
14. Athey S, Wager S. Estimating treatment effects with causal forests: an application. *Obs Stud* 2019; **5**: 37–51.
15. Rover L, Tadano YS. Clima360 Brasil: an integrated data platform for climate, air quality, and health surveillance across 27 Brazilian state capitals. 2026. Available from: <https://github.com/Roverlucas/clima-360-brasil>

16. World Health Organization. WHO global air quality guidelines: particulate matter (PM<sub>2.5</sub> and PM<sub>10</sub>), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Geneva: WHO; 2021.
17. Deryugina T, Heutel G, Miller NH, Molitor D, Reif J. The mortality and medical costs of air pollution: evidence from changes in wind direction. *Am Econ Rev* 2019; **109**: 4178–219. doi:10.1257/aer.20180279
18. Bondy M, Roth S, Sager L. Crime is in the air: the contemporaneous relationship between air pollution and crime. *J Assoc Environ Resour Econ* 2020; **7**: 555–85. doi:10.1086/707127
19. Chernozhukov V, Demirer M, Duflo E, Fernandez-Val I. Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India. *NBER Working Paper* 2018; No. 24678. doi:10.3386/w24678
20. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020; **2**: 56–67. doi:10.1038/s42256-019-0138-9
21. Gasparrini A, Armstrong B, Kenward MG. Distributed lag non-linear models. *Stat Med* 2010; **29**: 2224–34. doi:10.1002/sim.3940
22. Liu C, Chen R, Sera F, et al. Ambient particulate air pollution and daily mortality in 652 cities. *N Engl J Med* 2019; **381**: 705–15. doi:10.1056/NEJMoA1817364
23. Stafoggia M, Zauli-Sajani S, Pey J, et al. Desert dust outbreaks in southern Europe: contribution to daily PM<sub>10</sub> concentrations and short-term associations with mortality and hospital admissions. *Environ Health Perspect* 2016; **124**: 413–19. doi:10.1289/ehp.1409164
24. Simoni M, Baldacci S, Maio S, Cerrai S, Sarno G, Viegi G. Adverse effects of outdoor pollution in the elderly. *J Thorac Dis* 2015; **7**: 34–45. doi:10.3978/j.issn.2072-1439.2014.12.10
25. Kelly FJ, Fussell JC. Air pollution and public health: emerging hazards and improved understanding of risk. *Environ Geochem Health* 2015; **37**: 631–49. doi:10.1007/s10653-015-9720-1
26. Analitis A, De’ Donato F, Scortichini M, et al. Synergistic effects of ambient temperature and air pollution on health in Europe: results from the PHASE project. *Int J Environ Res Public Health* 2018; **15**: 1856. doi:10.3390/ijerph15091856
27. Shtein A, Schwartz J, Kioumourtzoglou MA, Dominici F, Zanobetti A. Estimating the causal effect of long-term PM<sub>2.5</sub> exposure on mortality using causal forests. *Environ Sci Technol* 2024; **58**: 1282–90. doi:10.1021/acs.est.3c06862
28. Hernan MA, Robins JM. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC; 2020.
29. Wakefield J. Ecologic studies revisited. *Annu Rev Public Health* 2008; **29**: 75–90. doi:10.1146/annurev.publhealth.29.020907.090821

30. Buchard V, Randles CA, da Silva AM, et al. The MERRA-2 aerosol reanalysis, 1980 onward. Part II: Evaluation and case studies. *J Clim* 2017; **30**: 6851–72. doi:10.1175/JCLI-D-16-0613.1
31. Khomenko S, Pisoni E, Thunis P, et al. Premature mortality due to air pollution in European cities: a health impact assessment. *Lancet Planet Health* 2021; **5**: e121–34. doi:10.1016/S2542-5196(20)30272-2
32. Rosenbaum PR. *Observational Studies*. 2nd ed. New York: Springer; 2002.
33. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007; **370**: 1453–57. doi:10.1016/S0140-6736(07)61602-X
34. Brasil. CONAMA Resolucao no 506, de 2024. Dispoe sobre padroes de qualidade do ar. *Diario Oficial da Uniao*, Brasilia, 2024.
35. Gasparrini A, Armstrong B. Reducing and meta-analysing estimates from distributed lag non-linear models. *BMC Med Res Methodol* 2013; **13**: 1. doi:10.1186/1471-2288-13-1
36. Saldiva PHN, Lichtenfels AJFC, Paiva PSO, et al. Association between air pollution and mortality due to respiratory diseases in children in Sao Paulo, Brazil: a preliminary report. *Environ Res* 1994; **65**: 218–25. doi:10.1006/enrs.1994.1033
37. Arbex MA, Santos UP, Martins LC, Saldiva PHN, Pereira LAA, Braga ALF. Air pollution and the respiratory system. *J Bras Pneumol* 2012; **38**: 643–55. doi:10.1590/S1806-37132012000500015
38. Barreto ML, Teixeira MG, Bastos FI, Ximenes RAA, Barata RB, Rodrigues LC. Successes and failures in the control of infectious diseases in Brazil: social and environmental context, policies, interventions, and research needs. *Lancet* 2011; **377**: 1877–89. doi:10.1016/S0140-6736(11)60202-X

---

## 1.9 Supplementary Material

Available online: STROBE checklist with line-by-line manuscript references; full variable definitions and data dictionary; detailed hyperparameter tuning results; complete CLAN analysis tables; city-specific CATE estimates with confidence intervals; SHAP force plots for all 27 capitals; code repository and reproducibility documentation (<https://github.com/Roverlucas/causal-pollution-health-brazil>).

---

## 1.10 Author Contributions

LR conceptualised the study, designed the methodology, built the data infrastructure (Clima360 Brasil), conducted all analyses, and wrote the first draft of the manuscript. YDST supervised the



study, provided critical revisions to the methodology and manuscript, and secured funding. Both authors approved the final version.

### **1.11 Declaration of Interests**

The authors declare no competing interests.

### **1.12 Data Sharing**

All data used in this study are publicly available. The Clima360 Brasil platform integrates five open data sources; the complete analytical dataset and code are available at <https://github.com/Roverlucas/causal-pollution-health-brazil> under MIT (code) and CC-BY 4.0 (manuscript) licences.

### **1.13 Acknowledgements**

We thank the teams behind DATASUS (SIH/SUS), IBGE (SIDRA), SENATRAN, the Copernicus Atmosphere Monitoring Service (CAMS), and Open-Meteo for making their data publicly accessible. The Clima360 Brasil platform was developed as part of this research.