Eduvest – Journal of Universal Studies
Volume 6 Number 2, February, 2026
p- ISSN 2775-3735- e-ISSN 2775-3727

The Impact of NO_2 and SO_2 Pollutants on Respiratory Diseases: A Case Study of Indonesia

Reizka Asadelia Rafmawan
Universitas Indonesia
Email: r.asadeliaa@gmail.com

ABSTRACT

Air pollution remains a significant public health challenge in Indonesia, driven by rapid industrialization and urbanization. Key pollutants such as NO_2 and SO_2 are strongly linked to respiratory conditions, yet comprehensive national evidence on their age-specific impacts remains limited. This study aims to examine the causal effects of NO_2 and SO_2 exposure on the incidence of acute respiratory infections (ARI), pneumonia, and asthma in Indonesia, with a particular focus on differences across age groups. This study employs a quasi-experimental design using district-level pollution data and individual health data from Riskesdas 2018. Analysis was conducted via multiple linear regression, coefficient stability testing, and IPW to estimate robust causal associations. Age-stratified analysis was performed across five groups: 0–4, 5–17, 18–49, 50–74, and 75+ years. Results show that NO_2 exhibits a strong positive association with ARI incidence, particularly among children aged 0–4 and 5–17 years, and is linked to asthma in adults aged 18–49. SO_2 shows significant positive effects on ARI among older adults (50–74 years) and on asthma in those aged 75 and above. However, unexpected negative or non-significant relationships were found between NO_2 and pneumonia/asthma, and between SO_2 and certain outcomes, likely reflecting data constraints and unobserved confounders. In conclusion, this study reveals age-specific pollutant–health relationships and underscores the need for targeted air quality interventions. Recommendations include strengthening monitoring systems, implementing pollutant-specific warnings, and integrating environmental–health data to support evidence-based policies and protect vulnerable groups.

KEYWORDS *Pollution; Respiratory Diseases; Quasiexperimental Analysis; Case Study.*

INTRODUCTION

Air pollution remains one of the most pervasive and costly externalities in modern economies. Beyond ecological degradation, it threatens human capital formation—reducing health, productivity, and cognitive performance (Chang et al., 2019; Fu et al., 2021; Zivin & Neidell, 2012). Airborne pollutants such as nitrogen dioxide (NO_2) and sulphur dioxide (SO_2) present a critical public health concern, given the sensitivity of these pollutants to respiratory illnesses such as acute respiratory infections (ARIs), pneumonia, and asthma, and the potential of these substances to impose both immediate health impacts and long-term socioeconomic burdens. However, in Indonesia—a country marked by vast environmental and demographic diversity—nationally representative, pollutant-specific evidence remains limited.

While prior studies have examined Indonesia's haze events Jayachandran, (2009), elevated $PM_{2.5}$ mortality Siregar et al., (2024), and lung function decline (Kim et al., 2017), many of these studies often rely on aggregate exposure proxies (e.g., haze, aerosol indices) or use generalized or gravest health outcomes like self-rated health or all-cause mortality. City-level studies (Andriati et al., 2021; Anurogo et al., 2023; Arsyad & Priyana, 2023) have begun to address pollutant–health relationships but often lack individual- and/or district-level controls, rely solely on diagnosed cases, or ignore subgroup vulnerability. Moreover, these

studies predominantly concentrate on urban centers such as Bogor, Tangerang, and municipalities within Jakarta—despite air pollution being a widespread concern across Indonesia.

To understand the broader public health burden of air pollution in Indonesia, it is essential to move beyond urban-focused analyses and examine exposure risks at a national scale using disaggregated, representative data. According to USAID (2018), public perceptions of air pollution appear relatively consistent across demographic and regional groups. However, only 17% of respondents ranked air pollution as their top concern, and within this subset, 64% identified health impacts—particularly respiratory diseases—as the main reason. These findings underscore the need to approach air pollution as a national environmental and health issue.

Environmental data further reinforce this urgency. Ground-based pollutant measurements show that approximately 30% of Indonesian districts exceed 10 μg/m³ of NO_2 or SO_2. Moreover, satellite-based $PM_{2.5}$ data (IHME) indicate that all districts exceed the 5 μg/m³ threshold. However, existing studies rarely disaggregate pollutant effects across Indonesia or isolate the impacts of NO_2 and SO_2—pollutants commonly emitted by the energy and transport sectors. Few analyses assess these effects using individual-level, nationally representative data, and even fewer investigate age-specific vulnerabilities, despite well-established differences in physiological susceptibility between children, adults, and the elderly.

Consequently, this study aims to fill these knowledge gaps by examining how NO_2 and SO_2 exposure affects the incidence of ARI, pneumonia, and asthma in Indonesia and how these effects differ by age. Using the 2018 Riskesdas survey—a nationally representative health dataset linked with district-level pollution data—we estimate pollutant–disease effects for both all and diagnosed cases across five age groups (0–4, 5–17, 18–49, 50–74, and 75+). We apply a linear regression framework alongside coefficient stability testing and inverse probability weighting to identify robust causal associations while addressing potential bias due to sample imbalance and unobserved confounding.

This study contributes to the literature by providing new causal evidence that NO_2 and SO_2 are significantly associated with respiratory illness risk in Indonesia, with varying effect sizes across pollutants, disease types, and age groups. This study revealed strong and consistent effects of NO_2 on ARIs, pneumonia, and asthma among children and robust associations of SO_2 with ARIs and asthma in older adults. By applying coefficient stability testing and inverse probability weighting, this study accounts for potential attrition and unmeasured confounders, although anomalies—such as NO_2's negative links to asthma and pneumonia and SO_2's negative associations with respiratory outcomes—highlight residual bias from measurement error and omitted variable bias. This study offers three policy recommendations: (1) strengthening pollutant-specific alert systems with WHO threshold triggers and public accessibility; (2) expanding monitoring infrastructure with subdistrict-level automated devices; and (3) integrating health records into a national big data platform. These results support more granular, life-course-sensitive environmental health policies and lay the groundwork for future cost-of-illness research and targeted pollution control in Indonesia’s development agenda.

METHOD

This study adopted a quantitative research design to investigate the associations between NO_2 and SO_2 pollution and respiratory disease incidence in Indonesia. The research draws upon two primary data sources: pollutant exposure datasets and individual-level health survey data. This study covers approximately 400 Indonesian districts and municipalities.

The pollution exposure data for NO_2 and SO_2 come from ground-level measurements from the Ministry of Environment and Forestry (MoEF) (Direktorat Jenderal Pengendalian Pencemaran dan Kerusakan Lingkungan Kementerian Lingkungan Hidup dan Kehutanan RI, 2018) and satellite-based estimates from the MERRA-2 model (Global Modeling and Assimilation Office (GMAO), 2015a). Additional environmental covariates include $PM_{2.5}$ data from the Institute for Health Metrics and Evaluation (IHME) and meteorological variables (temperature, rainfall, humidity) sourced from NASA's Global Modeling and Assimilation Office (GMAO) (Global Burden of Disease Collaborative Network, 2021; Global Modeling and Assimilation Office (GMAO), 2015b).

The health data are derived from the 2018 Basic Health Research (Riskesdas), conducted by the Indonesian Ministry of Health (Kementerian Kesehatan RI, 2018). The survey uses a stratified, two-stage sampling design with probability proportional to size (PPS), which targets households across all census blocks. The final sample includes all respondents interviewed in the 2018 Riskesdas round.

Table 1. Definitions of the Variables

Categories	Variables
Pollution	NO_2 SO_2
Respiratory Diseases	ARI: the presence of ≥2 of the following symptoms—fever, cough (<2 weeks), runny nose, sore throat—or a self-reported formal diagnosis within the past month (0 = unaffected, 1 = affected) Pneumonia: report of diagnosis or ≥2 symptoms in past 12 months (≥5 years: high fever, productive cough, dyspnea; <5 years: symptom combinations including respiratory distress signs—rapid breathing, nasal flaring, or chest wall indrawing) (0 = unaffected, 1 = affected) Asthma: report of diagnosis and symptom recurrence in past 12 months. Newly diagnosed if diagnosis age equals or differs by one year from current age. (0 = unaffected, 1 = affected)
Demographic and Socioeconomic Controls	Gender (0: male, 1: female) Marital status (0: Never married, 1: Married, 2: Divorced, 3: Widowed) Education (0: Never attended school, 1: Incomplete primary school, 2: Completed primary school, 3: Completed junior high school, 4: Completed senior high school, 5: Completed diploma level education, 6: Completed university degree) Age, age^2 Urban–rural residence (0: urban areas, 1: rural areas)

Family-level exposure	Presence of other sick members (0: absence, 1: present)
Behavioral controls	Active smoking (0: never smoked, 1: former smoker who quit within the last month, 2: currently an occasional smoker, 3: former daily smoker now smoking occasionally, 4: former occasional smoker now smoking daily, 5: currently a daily smoker) Passive smoking a. DPS (Daily Passive Smoker) (1: exposed daily to second-hand smoke) b. OPS (Occasional Passive Smoker) (1: exposed occasionally) Open waste burning (0: otherwise, 1: burning)
District-level variables	Meteorological conditions (temperature, rainfall, humidity) Copollutants (e.g., NO_2 or SO_2, and $PM_{2.5}$) Seasonal effects Presence of steam-fired coal power plants (PLTU) (0: absence, 1: present) Healthcare facilities (number of doctors, number of hospitals, number of polindes (Poliklinik Desa), number of puskesmas and its line)

Source: Processed from Riskesdas 2018 and MoEF environmental data

Ordinary least squares

To examine the impact of air pollution on respiratory illness, this study begins with a set of ordinary least squares (OLS) regressions. The baseline model estimates individual respiratory outcomes—ARI, pneumonia, and asthma—as a function of NO_2 and SO_2 levels measured at the district level. While this linear framework is aligned with established practices in environmental health economics, it is prone to omitted variable bias (OVB) if confounders such as demographic, behavioral, or environmental factors are not adequately controlled.

To mitigate OVB, the models incorporate extensive individual-level covariates, including age, education, marital status, smoking behavior, and household health exposure. These are complemented by district-level controls such as meteorological conditions, copollutants, and health infrastructure indicators. This dual-level control structure aims to close backdoor paths from pollution to health outcomes. Nonetheless, several estimation challenges remain: unmeasured confounders, classical measurement error, and behavioral adaptations that obscure true exposure–response relationships.

Equation 1. *The fully controlled baseline model*

$$Y_{id} = \alpha + \widehat{\beta_1} \cdot Pollution_d + X'_{id}\gamma + Z'_d\delta + \epsilon_{id}$$

where:

1. Y_{id} denotes an indicator for respiratory illness (ARI, pneumonia, asthma) for individual i in district d
2. where $Pollution_d$ is the main exposure variable—either NO_2 or SO_2 concentrations at the district level—using measurements from MoEF or MERRA-2 satellite data.

3. X'_{id} is a vector of individual-level covariates, including demographic and socioeconomic controls, family-level exposure, and behavioral controls
4. Z'_d is a vector of district-level variables, including meteorological conditions, copollutants, seasonal effects, the presence of steam-fired coal power plants (PLTUs), and healthcare facilities.
5. ϵ_{id} is the individual-level error term

All regressions are estimated at the individual level. Furthermore, the fully controlled OLS model serves as a baseline for comparison with more robust specifications.

H_0 (null hypothesis):

There is no significant association between $Pollution$ and Y.

H_1 (alternative hypothesis):

There is a significant association between $Pollution$ and Y.

Despite the use of comprehensive controls, OLS estimates are still vulnerable to attenuation and selection biases. Pollution data coverage is incomplete and is absent in approximately 18% of districts—often those with lower industrial activity. This non-randomness may bias the results by overrepresenting high-pollution areas. To address this, inverse probability weighting is applied to reweight the sample on the basis of predicted pollution data availability, thereby correcting for potential attrition bias.

Finally, to assess heterogeneous effects and improve causal interpretation, the analysis is stratified by age group. This acknowledges the biological and epidemiological differences in vulnerability to pollution. In contrast to integrated cost-of-illness models, this study focuses strictly on estimating pollutant-specific marginal health effects. While rooted in the dose–response function literature, it stops short of full economic valuation—serving instead as a foundational step toward a more comprehensive environmental health assessment for Indonesia.

Robustness models

To strengthen the validity of the OLS estimates, we apply two complementary robustness strategies: the coefficient stability test and inverse probability weighting. The CST, which is based on the Oster, (2019) framework, assesses how much selection of unobserved variables would be required to nullify the observed treatment effect. By comparing changes in coefficients and R^2 across nested models, we estimate a δ parameter indicating the strength of selection needed on unobservable relative to observables. A large δ or a treatment effect that remains materially different from zero under δ = 1 is interpreted as evidence of robustness against omitted variable bias.

Equation 2. *The CST model*

$$Y_{id} = \alpha + \beta_1^* \cdot Pollution_d + X'_{id}\gamma + Z'_d\delta + \epsilon_{id}$$

To address potential selection bias caused by the nonrandom availability of pollution data—especially from ground-based monitors—we use IPW. We first estimate a logit model predicting the likelihood of a district being observed (i.e., "polluted") on the basis of industrial characteristics. The inverse of the predicted probability is then used to weight observations in the regression model. This adjustment rebalances the estimation sample to resemble the full population more closely, thus correcting for attrition and enhancing external validity.

Equation 3. *The IPW model*

$$Y_{id} = \alpha + \beta \cdot Pollution_d + X'_{id}\gamma + Z'_d\delta + \epsilon_{id}, \qquad weighted\ by\ \frac{1}{\widehat{P_d}}$$

Heterogeneity Model

We also examined heterogeneity in the pollution–health relationship by conducting age-based subgroup analyses. Given that susceptibility to pollution varies across the life course, we divided the sample into five biologically relevant age groups: 0–4, 5–17, 18–49, 50–74, and 75+. Fully controlled models are estimated separately for each group, allowing us to identify population-specific vulnerabilities and avoid misinterpretation from pooled estimates. This stratified design provides clearer causal interpretations and more targeted policy insights.

Equation 4. *The Age-Based Subgroup Estimation Model*

$$Y_{id}^{(g)} = \alpha + \beta^{(g)} \cdot Pollution_d + X'_{id}\gamma^{(g)} + Z'_d\delta^{(g)} + \epsilon_{id}^{(g)}$$

Each regression is conducted within subgroup g.

These robustness and heterogeneity checks enhance the credibility of the main findings. The CST addresses concerns about unobserved confounding, the IPW corrects for attrition bias due to missing pollution data, and the age-stratified analyses reveal life-course variation in pollution sensitivity. Together, these approaches ensure that the estimated associations between air pollution and respiratory health outcomes are not only statistically robust but also substantively meaningful for policymaking.

RESULT AND DISCUSSION

Understanding the distribution and characteristics of air pollution is essential given its implications for human capital and long-term economic growth. This study examines nitrogen dioxide (NO_2) and sulphur dioxide (SO_2) levels across approximately 400 Indonesian districts via 2018 data from the Ministry of Environment and Forestry (MoEF). NO_2 concentrations range from 1.31 to 37.91 µg/m^3, with 29.2% of districts exceeding the WHO's annual guideline (10 µg/m^3). SO_2 levels vary between 3.12 and 25.92 µg/m^3, and 26.7% of districts exceed the Canadian 2025 guideline (10.5 µg/m^3). MERRA-2 satellite data for SO_2 range from 0.032--25.03 µg/m^3; however, only 3.9% of the data exceed the Canadian standard. As Figure 1 illustrates, ground-based MoEF readings often record higher values than satellite-based MERRA-2 readings do.

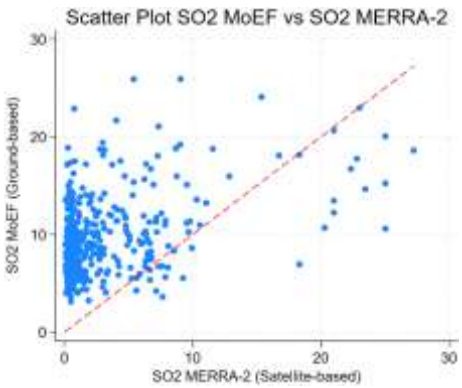

Figure 1. Scatter plot of local SO_2 ground-based and satellite-based measurements, 2018

Source: Author's computation

The discrepancy arises from methodological differences in data construction. The MoEF dataset represents ground-based measurements, combining limited real-time monitoring with passive samplers—typically one or two points—each designed to reflect four distinct environments: industrial areas, residential zones, transportation corridors, and office districts. These measurements are aggregated on a biannual basis.

In contrast, the SO_2 data from MERRA-2 represent satellite-based estimates with broader spatial coverage. However, such estimates are derived at a coarse spatial resolution (~50 × 70 km) and reflect pollutant concentrations from higher atmospheric layers. The accuracy of these methods may be compromised by cloud cover and atmospheric interference.

As a result, ground-based data often record higher pollutant concentrations than satellite-derived estimates do. Nonetheless, both sources remain susceptible to measurement error, potentially introducing classical measurement error that may attenuate estimated effects or bias results—particularly when the misalignment between measured and true exposure correlates with disease incidence.

On the health side, Table 2 presents prevalence estimates from the 2018 Riskesdas survey. ARIs affect 32.35% of the population (85.3 million), but only 4.37% report a diagnosis. Pneumonia affects 14.5% (symptomatic) and 2.01% (diagnosed) of patients. Asthma is least prevalent (1.45% symptomatic, 0.28% diagnosed), suggesting underdiagnosis or access barriers.

Table 2. Cases of ARI, pneumonia, and asthma in Indonesia, 2018

Types of Illnesses	Prevalence Rate (All Cases)	Total Population (All Cases)	Prevalence Rate (Self-reported Diagnosed Cases)	Total Population (Self-reported Diagnosed Cases)
ARI	32.35%	85,307,522	4.37%	11,534,516
Pneumonia	14.50%	38,242,086	2.01%	5,295,961
Asthma	1.45%	3,830,087	0.28%	745,533

Source: (Kementerian Kesehatan RI, 2018)

Note: Author's own processing

Estimates combine self-reported diagnoses with assessments based on self-reported symptoms; the total population of Indonesia is 263,722,841.

Ordinary least squares (OLS)

Table 3. Comparison of Estimation Results: Baseline, Model with Individual-level Controls, and Full Model with Individual-level and District-level Controls in OLS Regression

	ARI			Pneumonia			Asthma		
	(1) Baseline	(2) +Individual-level	(3) Full Model	(1) Baseline	(2) +Individual-level	(3) Full Model	(1) Baseline	(2) +Individual-level	(3) Full Model
NO_2	0.000455***	0.00114***	0.000849***	-0.000386***	0.0000491	-0.000409***	0.0000107	-0.0000437	-0.000100***

	(0.0000832)	(0.0000852)	(0.000110)	(0.0000623)	(0.0000643)	(0.0000831)	(0.0000219)	(0.0000230)	(0.0000297)
Obs.	882255	882255	860823	882255	882255	860823	882255	882255	860823
F test	29.95	2131.8	1457.6	38.46	1497.0	1065.6	0.238	326.5	219.8
R-sq.	0.0000339	0.0527	0.0544	0.0000436	0.0376	0.0404	0.000000270	0.00844	0.00861
SO_2	-0.0000610 (0.000119)	0.000442*** (0.000118)	-0.000902*** (0.000133)	-0.000253** (0.0000895)	0.000122 (0.0000889)	-0.000635*** (0.000100)	0.0000547 (0.0000315)	0.0000376 (0.0000318)	-0.00000953 (0.0000358)
Obs.	882255	882255	860823	882255	882255	860823	882255	882255	860823
F test	0.261	2124.3	1457.6	7.993	1497.0	1065.6	3.012	326.4	219.8
R-sq.	0.000000296	0.0525	0.0544	0.00000906	0.0376	0.0404	0.00000341	0.00844	0.00861
SO_2 (MERRA2)	0.000411*** (0.0000846)	0.00110*** (0.0000859)	0.000592*** (0.000140)	-0.000619*** (0.0000665)	-0.00000383 (0.0000680)	0.000117 (0.000114)	0.00000283 (0.0000226)	-0.00000358 (0.0000235)	-0.000262*** (0.0000405)
Obs.	1017290	1017290	860823	1017290	1017290	860823	1017290	1017290	860823
F test	23.61	2518.5	1456.7	86.62	1873.9	1064.4	1.567	372.6	221.1
R-sq.	0.0000232	0.0539	0.0544	0.0000851	0.0406	0.0403	0.00000154	0.00835	0.00866

Source: OLS regression results by the author

Note: Author's own processing. Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Each panel compares estimates across three specifications: (1) the baseline model without controls, (2) the model with individual-level controls (X), and (3) the fully controlled model including both individual-level (X) and district-level (Z) variables. The pollution variables include the district-level concentrations of NO_2 and SO_2, which are measured from both ground-based monitoring and satellite-derived data. All regressions are estimated at the individual level via OLS, and the results are shown for three health outcomes: acute respiratory infections (ARIs), pneumonia, and asthma. The fully controlled OLS model (3) is used as the primary baseline for subsequent causal inference tests, including CST and IPW.

Table 3 presents the evolution of regression estimates across three model specifications: (1) a baseline model including only NO_2 or SO_2 exposures, (2) a model incorporating individual-level controls, and (3) a full model that additionally includes district-level covariates. This stepwise structure facilitates assessment of how much bias is addressed by sequentially adjusting for observed confounders.

For NO_2, the coefficients and explanatory power of the models increase as more covariates are introduced. In the case of the ARI, the coefficient increases from 0.000455 at the baseline to 0.00114 with individual-level controls and then slightly decreases to 0.000849 in the full model, whereas R^2 improves substantially from 0.00003 to 0.0544. This finding indicates that most omitted variable bias stems from individual-level confounding. For pneumonia, the coefficient of NO_2 shifts from significantly negative to slightly positive before

returning to a significant negative estimate. In the asthma models, NO_2 is insignificant at first but becomes significantly negative (–0.000100) once full controls are applied, confirming model refinement through the inclusion of both individual and district-level variables.

For SO_2 (MoEF), the ARI coefficient begins as statistically insignificant, becomes significantly positive with individual-level controls, and becomes significantly negative in the full specification, coinciding with an increase in R^2 to 0.0544. A similar reversal occurs for pneumonia, reinforcing the importance of controlling for district-level environmental confounders. In contrast, SO_2 was not significantly related to asthma across the models. Satellite-derived SO_2 (MERRA-2) is consistently positively and significantly associated with the ARI across specifications, albeit with slight attenuation. For pneumonia, the coefficient shifts from significantly negative to statistically null and then to slightly positive. For asthma, the relationship switches from a weakly positive to a significantly negative association (–0.000262), with R^2 rising from near zero to 0.00866, indicating improved model fit with full adjustment.

Robustness

Coefficient Stability Test and Inverse Probability Weights

To assess the robustness of the estimated pollution–health associations, we employed both coefficient stability test (CST) and inverse probability weighting (IPW) methods, followed by Oster's (2019) coefficient stability method. In addition to reporting both the non and weighted fully controlled OLS estimates, we present bias-adjusted treatment effects (β if δ = 1), the corresponding δ values required to nullify the effect (δ for β = 0), and the assumed maximum explanatory power (R^2 max) of the full model—which is calculated as 1.3 times the R^2 from the weighted fully controlled model.

Table 4. Comparison of Estimation Results: Method Non-weighting vs. Reweighting via IPW

Variable	Non	Reweighting			
	Full Controlled Effect	Full Controlled Effect	β if $\delta = 1$	δ for $\beta = 0$	R^2max
		ARI			
NO_2	0.000849*** (0.000110)	0.000618*** (0.000166)	0.000682***	6.218	0.06864
R-squared	0.0544	0.0528			
SO_2 (Ground-based)	-0.000902*** (0.000133)	-0.0000160 (0.000212)	-0.000306	-0.0605	0.06864
R-squared	0.0544	0.0528			
SO_2 (Satellite-based)	0.000592*** (0.000140)	0.000235 (0.000161)	-0.000709	0.400	0.06864
R-squared	0.0544	0.0528			
		Pneumonia			
NO_2	-0.000409*** (0.0000831)	-0.000519*** (0.000126)	-0.000463***	3.136	0.05499
R-squared	0.0404	0.0423			
SO_2 (Ground-based)	-0.000635*** (0.000100)	-0.0000623 (0.000158)	-0.000247	-0.377	0.05499

R-squared	0.0404	0.0423			
SO_2 (Satellite-based)	0.0000117 (0.000114)	-0.000579*** (0.000132)	-0.000736***	2.358	0.05499
R-squared	0.0403	0.0423			
		Asthma			
NO_2	-0.000100*** (0.0000297)	-0.0000315 (0.0000441)	-0.0000749	-1.217	0.011297
R-squared	0.00861	0.00869			
SO_2 (Ground-based)	-0.00000953 (0.0000358)	-0.000124* (0.0000633)	-0.000142*	-20.92	0.011297
R-squared	0.00861	0.00869			
SO_2 (Satellite-based)	-0.000262*** (0.0000405)	-0.000316*** (0.0000465)	-0.000691***	-40.35	0.011336
R-squared	0.00866	0.00872			

Source: CST and IPW analysis results by the author

Note: Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Full Controlled Effect:

The coefficient from the regression with all observable controls. “Non” refers to non-weighted OLS; “IPW” is weighted by inverse probability. R-squared: Indicates the proportion of variance in the dependent variable explained by the model. Both non-weighted and IPW-adjusted models are provided. β if δ = 1: The bias-adjusted treatment effect assumes that the selection of unobserved confounders is equal in magnitude to that of observables (δ = 1), following (Oster, 2019). R2 max: Maximum R-squared value used to simulate the bias-adjusted treatment effect. We assume R2 max = 1.3 × R2 (IPW-adjusted models). δ for β = 0: The minimum strength of selection on unobservable (relative to observables) variables required to reduce the estimated effect to zero. A large δ implies robustness; a low or negative δ indicates sensitivity to omitted variables.

For the ARI (Table 4.), the estimated NO_2 treatment effect decreases to 0.000682 under IPW, with a strong δ value of 6.218, indicating high robustness to unobserved confounding. In contrast, SO_2 (ground-based) attenuates to –0.000306 with a weak δ (–0.0605), and the satellite-based SO_2 estimate becomes –0.000709 with a low δ of 0.400. The sign reversal and attenuation of SO_2 effects suggest possible attrition bias due to the underrepresentation of less industrialized areas. Additionally, the unusual negative value in the estimated SO_2 effect indicates a potential bias–measurement error and omitted variable bias–and heterogeneity pollution-health effects.

For pneumonia (Table 4.), the NO_2 effect becomes slightly more negative (–0.000463), with a strong δ value of 3.136, indicating moderate robustness. The SO_2 effect (ground-based) weakens to –0.000247 with a non-robust δ (–0.377), whereas satellite-based SO_2 has a stronger negative effect (–0.000736) with a δ of 2.358. These shifts again highlight the influence of attrition bias. Notably, the unusual negative values in the estimated effects of NO_2 and SO_2 may indicate potential sources of bias—such as measurement error, omitted variable bias, and heterogeneity in pollution–health effects.

For asthma patients (Table 4.), all the pollutant estimates turn negative. NO_2 attenuates to –0.0000749 with a non-robust δ (–1.217), whereas SO_2 becomes more negative both in ground-based (–0.000142, δ = –20.92) and satellite-based (–0.000691, δ = –40.35) models. These consistent unusual negative values and extreme δ values suggest strong susceptibility to

measurement error, unobserved bias, and heterogeneous pollution–health effects across subgroups.

Heterogeneity

Age-Based Subgroup Estimation

Having established the sensitivity of pooled models to attrition bias and unobserved confounding factors, we proceed to stratify the analysis across five age-defined subgroups to uncover heterogeneity in pollution–health responses and assess their robustness via Oster's coefficient stability test and reweighting via IPW. In addition to reporting the weighted fully controlled OLS estimates, we present bias-adjusted treatment effects (β if δ = 1), the corresponding δ values required to nullify the effect (δ for β = 0), and the assumed maximum explanatory power (R^2 max) of the full model—which is calculated as 1.3 times the R^2 from the weighted fully controlled model.

Across all the patients with ARIs (Table 5.), NO_2 shows strong, positive, and robust effects in children (0–4: 0.00321, δ = 3.422; 5–17: 0.00101, δ = 1.829), whereas effects in adults aged 50–74 are smaller and not robust. SO_2 appears negative and less robust in children but becomes positive and robust in adults aged 50–74 across data sources. In individuals diagnosed with ARIs, NO_2 remains robust in children, becomes negative in older adults, and is positive in adults aged 18–49 years. SO_2 from MoEF is positive and strong in adolescents (5–17, δ = 8.154), whereas satellite SO_2 is positive and robust across all ages for diagnosed ARIs. Unusual negative estimates for NO_2 and SO_2 suggest possible bias from measurement error and omitted variable bias in pollution–health effects.

Table 5. ARI, Pneumonia, and Asthma—Pollutant Age-Based Subgroup Estimation

Variable	0-4			5-17			18-49			50-74			75+		
	Full Controlled Effect	β if $\delta = 1$	δ for $\beta = 0$	Full Controlled Effect	β if $\delta = 1$	δ for $\beta = 0$	Full Controlled Effect	β if $\delta = 1$	δ for $\beta = 0$	Full Controlled Effect	β if $\delta = 1$	δ for $\beta = 0$	Full Controlled Effect	β if $\delta = 1$	δ for $\beta = 0$
							ARI								
NO_2	0.00338*** (0.00626)	0.00321***	3.422	0.00168*** (0.00349)	0.00101***	1.829	0.00895 (0.00238)	-0.00578	0.671	-0.000501 (0.000357)	0.000250	0.723	-0.00181 (0.00105)	-0.00130	2.284
R-squared	0.0625			0.0548			0.0376			0.0383			0.0325		
SO_2 (Ground-based)	-0.000315 (0.00785)	-0.00128	-0.364	-0.000266 (0.00446)	-0.00105	-0.378	0.00575 (0.00302)	-0.00207	0.233	0.000322 (0.000456)	0.000653	-1.139	-0.00121 (0.00126)	-0.000898	3.250
R-squ	0.0625			0.0548			0.0376			0.0383			0.0325		

ared															
SO_2 (Satellite-based)	0.00232*** (0.000574)	-0.000422***	0.915	0.00141*** (0.000329)	-0.00124***	0.686	-0.000283 (0.000232)	-0.00183	-0.459	-0.00108** (0.000376)	0.000124**	0.947	-0.00226 (0.00127)	-0.000599	1.161
R-squared	0.0627			0.0549			0.0376			0.0383			0.0326		
						Pneumonia									
NO_2	-0.000100 (0.000310)	0.0000579	0.694	0.000716** (0.000274)	0.000919***	41.06	-0.000713*** (0.000186)	-0.000811***	7.456	-0.00179*** (0.000281)	-0.000146***	2.629	-0.00402*** (0.000916)	-0.000371***	3.285
R-squared	0.0175			0.0524			0.0378			0.0415			0.0425		
SO_2 (Ground-based)	-0.000358 (0.000372)	-0.000496	-3.618	-0.000796* (0.000362)	-0.00131*	-1.933	0.000341 (0.000223)	0.000161	1.759	0.000439 (0.000352)	0.000752	-1.708	-0.00152 (0.00104)	-0.000923	2.303
R-squared	0.0175			0.0524			0.0378			0.0415			0.0425		
SO_2 (Satellite-based)	-0.000743* (0.000297)	-0.00106*	2.948	0.000169 (0.000286)	-0.000492	0.420	-0.000533** (0.000191)	-0.000103***	6.594	-0.00165*** (0.000315)	-0.000314***	1.092	-0.00235* (0.00114)	0.000130*	0.778
R-squared	0.0175			0.0524			0.0378			0.0417			0.0426		
						Asthma									
NO_2	0.000985 (0.000170)	0.000254	-0.997	-0.000313 (0.000733)	-0.000153	-0.364	0.000258 (0.000618)	0.000342	1.113	-0.000235* (0.000119)	-0.000176*	2.346	-0.000507 (0.000485)	-0.000377	2.431
R-squ	0.0175			0.00947			0.00513			0.00720			0.0184		

ared														
SO_2 (Ground-based)	-0.000459 (0.000254)	-0.000408	5.223	-0.000423 (0.000101)	-0.0000902	-1.032	-0.000193* (0.000951)	-0.000220*	-21.48	0.0000389 (0.000162)	0.000110	-0.620	0.000466 (0.000438)	0.000608	-4.364
R-squared	0.0175			0.00947			0.00513			0.00720			0.0184		
SO_2 (Satellite-based)	-0.000357* (0.000140)	-0.000225*	1.403	-0.000827 (0.000777)	-0.000366	-0.860	-0.000323*** (0.000659)	-0.000729***	-11.35	-0.0000651*** (0.000138)	-0.000946***	2.605	-0.000216 (0.000568)	0.0000458	0.477
R-squared	0.0174			0.00947			0.00516			0.00732			0.0183		

Source: Age-subgroup analysis with IPW and CST by the author

Note: Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The fully controlled effect refers to the coefficient estimated from the IPW-weighted OLS regression model including the full set of observed control variables. The model adjusts for sample imbalance via inverse probability weighting (IPW) to mitigate attrition bias across subgroups. R-squared: Indicates the proportion of variance in the dependent variable explained by the model. β if δ = 1: The bias-adjusted treatment effect assumes that the selection of unobserved confounders is equal in magnitude to that of observables (δ = 1), following (Oster, 2019). R2 max: Maximum R-squared value used to simulate the bias-adjusted treatment effect. We assume R2 max = 1.3 × R2 (IPW-adjusted models). δ for β = 0: The minimum strength of selection on unobservable (relative to observables) variables required to reduce the estimated effect to zero. A large δ implies robustness; a low or negative δ indicates sensitivity to omitted variables.

For pneumonia (Table 5.), NO_2 effects are positive in children (0–4: 0.0000579, δ = 0.694; 5–17: 0.000919, δ = 41.06), whereas SO_2 is negative in children but becomes positive and robust for adults aged 18–74 (MoEF) and for those aged 75+ (satellite). Diagnosed pneumonia shows robust NO_2 effects only in children aged 0–4 years and robust SO_2 effects across all age groups (MoEF) and all but the youngest (satellite). For asthma patients (Table 5.), NO_2 is positive in young children and robust in adults aged 18–49 years. SO_2 is generally negative in children but positive in older adults, with satellite data confirming significance only in the 75+ group. In individuals diagnosed with asthma, both NO_2 and satellite SO_2 effects are negative, whereas MoEF SO_2 remains positive and robust in adults aged 50–74 years. Again, negative estimates of NO_2 and SO_2 may reflect bias from measurement error, omitted confounders, and underlying heterogeneity in exposure–response relationships.

Air pollution is a classic negative externality—its costs are underpriced, diffuse, and deeply embedded in urban economies. Empirical research has shown that pollutants such as

$PM_{2.5}$, NO_2, and SO_2 significantly affect economic outcomes, ranging from asset prices and corporate valuations to national GDP. In urban housing markets, residents internalize air quality, as seen in Mexico City, where a 10% rise in AOD reduces home values by 1.3% and raises wages by 2.3% (Goodwin et al., 2021); in Shanghai, a 1 $\mu g/m^3$ reduction in $PM_{2.5}$ and NO_2 raises housing prices by 155.93 and 278.03 yuan/m^2, respectively (Zou et al., 2022), whereas $PM_{2.5}$ and SO_2 reductions from the Blue-Sky Protection Campaign increased home values by up to 6% in Changsha (Feng et al., 2024), underscoring the capitalization of cleaner air. Pollution also affects corporate finance and markets: in China, a 1 $\mu g/m^3$ increase in $PM_{2.5}$ leads to a 0.3-point drop in ESG scores for manufacturing firms, especially in polluting industries (Zhao et al., 2024); in New York, daily S&P 500 returns fall by nearly 12% on high-$PM_{2.5}$ days owing to cognitive strain on traders (Heyes et al., 2016); and NO_2 disrupts earnings forecasts after site visits in polluted cities, highlighting its role as a behavioral pollutant that alters mood, attention, and risk perception (Dong et al., 2021). Finally, air pollution undermines labor productivity across physical and cognitive tasks: a 10-ppb rise in ozone reduces agricultural output by 5.5% (Zivin & Neidell, 2012), a 10-point API increase cuts daily call center output by 0.35% Chang et al., (2019), and a 1 $\mu g/m^3$ drop in $PM_{2.5}$ increases firm productivity by 0.82%, whereas a 1% national $PM_{2.5}$ reduction could increase GDP by 0.039% (Fu et al., 2021). Collectively, these findings show that pollution control yields not only health but also substantial economic gains.

Air pollution not only leads to productivity losses but also imposes substantial health costs, particularly through its effects on respiratory illnesses. Using the IPW model to correct for sampling bias, this study revealed that NO_2 remains significantly and robustly associated with the ARI, indicating a stable causal effect even after adjusting for attrition. These findings are in line with the well-documented link between NO_2 and ARI, which is consistent with localized studies such as (Andriati et al., 2021; Surury et al., 2022), which show that higher NO_2 concentrations increase the incidence of ARI. However, NO_2 appears to have no significant effect on asthma and is even negatively associated with pneumonia—findings that are unusual and diverge from established biomolecular signaling literature and localized evidence from (Anurogo et al., 2023; Arsyad & Priyana, 2023), who reported that rising NO_2 levels increase asthma incidence.

Similarly, SO_2 had no significant effect on the incidence of ARIs or pneumonia on the basis of ground-based measurements and had a significantly negative effect on asthma incidence. When satellite-based data are used, the effect of SO2 on the incidence of ARIs remains insignificant and is significantly negatively associated with pneumonia and asthma. These results are uncommon and inconsistent with the known biomolecular mechanisms linking SO_2 to respiratory outcomes and contrast with findings from (Anurogo et al., 2023; Arsyad & Priyana, 2023), who reported positive associations between SO_2 concentrations and asthma incidence. (Surury et al., 2022), using annual SO_2 data in Jakarta, also reported that SO_2 increased the incidence of ARIs, although the relationship was nonlinear and inconsistent. (Lestari & Haryanto, 2022), using now-unavailable monthly SO_2 data from Jakarta's environmental agency, reported an atypical and inconsistent correlation between SO_2 and ARI among children.

Several factors may explain these inconsistencies. First, the quality of the NO_2 and SO_2 concentration data used. Although this study utilized ground-based pollutant data covering

82% of Indonesia, the data quality may vary regionally, as it largely relies on passive samplers, with only one or two sampling points estimated to represent each of the four distinct sites per district—aggregated biannually. This stems from limitations in automatic monitoring infrastructure: in 2018, Indonesia had only 13 cities equipped with automatic pollutant monitoring, covering just 2.53% of the country. To enhance coverage, the study complemented ground-based data with satellite-derived pollution estimates. However, these satellite measurements have low spatial resolution and capture pollutants from higher atmospheric layers, which are often affected by clouds and atmospheric interference, which may lead to measurement error and misrepresentation of true individual-level exposure.

These data limitations likely contribute to discrepancies with the established literature. For example, Liu & Ao, (2021), using hourly ground-based pollutant data from Taiwan, reported that a one-unit improvement in the AQI results in NT$2.3 billion (approximately US$74 million) savings in annual outpatient respiratory expenditures. Similarly, Sofwan et al., (2021) used hourly pollutant data in Malaysia and reported significant positive associations between pollutants (including NO_2 and SO_2) and respiratory illness risk.

Second, although this study employed coefficient stability testing to assess robustness to unobservable variables (assuming equal importance between observed and unobserved variables with $\delta = 1$), the adjusted results suggest limited correction and indicate remaining bias from unobserved confounders. Omitted variable bias at both the individual and regional levels may distort estimations. Community-level prevention and healthcare investment, regional air quality control policies, and local allergen levels are key unobserved variables. Individual behaviors—such as mask usage, outdoor exposure duration, and occupational proximity to pollution sources—also remain unobserved. Owing to these data limitations, heterogeneity across Indonesia's population and geography may have introduced bias in the study's estimates, particularly regarding the NO_2 pneumonia and NO_2 asthma results and the generally negative SO_2 findings, which deviate from well-established biomolecular logic.

Age-stratified analysis confirmed that children (0–4 and 5–17 years) are particularly vulnerable, with NO_2 significantly increasing the risks of ARI, pneumonia, and asthma, supporting the findings of the clinical literature on childhood immune immaturity (Azmi et al., 2016). Among adults, NO_2 increases the ARI from 50–74, whereas SO_2 increases the ARI from 50–74 (across both data sources). For asthma, NO_2 is robust in adults 18–49, and SO_2 is robust in the oldest adults 75+. These results suggest that cumulative exposure, lifestyle, and comorbidities heighten adult susceptibility, especially to SO_2, echoing prior links to late-life asthma Wu et al., (2014), and reinforce the urgency of quantifying pollution's long-term economic burden.

Health effects from air pollution, particularly SO_2, have economic implications. Ramakrishnan et al., (2016) reported that SO_2 significantly reduces GDP per capita by 0.12 percentage points per 1% increase in emissions across high-income countries. Combined with broader literature on productivity loss (Chang et al., 2019; Fu et al., 2021; Zivin & Neidell, 2012), housing value decline (Feng et al., 2024; Goodwin et al., 2021), and impaired firm valuation Zhao et al., (2024), the findings show that air quality is not just a health issue but also a pillar of economic resilience. Pollution control must be prioritized as an investment in human capital and long-term prosperity.

CONCLUSION

This study reveals that NO_2 and SO_2 pollutants significantly elevate respiratory disease incidence in Indonesia, with NO_2 showing robust positive associations with acute respiratory infections (ARIs) among children (0–4 and 5–17 years) and asthma in young adults (18–49 years), while SO_2 impacts ARIs in middle-aged adults (50–74 years) and asthma in the elderly (75+ years), underscoring life-course vulnerabilities. Unexpected negative or non-significant links (e.g., NO_2 with pneumonia/asthma) highlight data limitations like sparse spatial/temporal pollution coverage and measurement bias. Policy recommendations include pollutant-specific WHO-aligned alert systems with public real-time access, subdistrict-level automated monitoring, and integrated national health-environmental data platforms. Sensitivity analyses via inverse probability weighting also flag $PM_{2.5}$'s role in ARIs/asthma, suggesting future research employ granular national/regional data to explore nonlinear/threshold effects, economic cost-of-illness valuations, and multi-pollutant interactions for refined pollution control strategies.

REFERENCES

Andriati, R., Adha, M. Z., & Hadi, T. E. Z. (2021). The Analysis of Meteorological Factors and Ambient Air Quality (PM10, CO, SO2, NO2, and O3) with the Incidence of Acute Respiratory Infection (ARI) in Tangerang City, Indonesia During 2010-2019. *International Journal of Advancement in Life Sciences Research*, *4*(4), 7–14. https://doi.org/10.31632/ijalsr.2021.v04i04.002

Anurogo, D., Sulaeman, S., Yamtana, Y., & Andarmoyo, S. (2023). Assessing the Impact of Air Quality on Respiratory Health in Urban Environments: A Case Study of Tangerang. *West Science Interdisciplinary Studies*, *1*(10), 928–939. https://doi.org/10.58812/wsis.v1i10.278

Arsyad, K. A., & Priyana, Y. (2023). Studi Kausalitas antara Polusi Udara dan Kejadian Penyakit Saluran. *Jurnal Multidisiplin West Science*, *02*(06), 462–472. https://doi.org/https://doi.org/10.58812/jmws.v2i6.434

Azmi, S., Aljunid, S. M., Maimaiti, N., Ali, A. A., Muhammad Nur, A., De Rosas-Valera, M., Encluna, J., Mohamed, R., Wibowo, B., Komaryani, K., & Roberts, C. (2016). Assessing the burden of pneumonia using administrative data from Malaysia, Indonesia, and the Philippines. *International Journal of Infectious Diseases*, *49*, 87–93. https://doi.org/10.1016/j.ijid.2016.05.021

Chang, T. Y., Zivin, J. G., Gross, T., & Neidell, M. (2019). The effect of pollution on worker productivity: Evidence from call center workers in China. *American Economic Journal: Applied Economics*, *11*(1), 151–172. https://doi.org/10.1257/app.20160436

Direktorat Jenderal Pengendalian Pencemaran dan Kerusakan Lingkungan Kementerian Lingkungan Hidup dan Kehutanan RI. (2018). Laporan Tahunan Direktorat Pengendalian Pencemaran Udara 2018. In *Kementerian Lingkungan Hidup dan Kehutanan*.

Dong, R., Fisman, R., Wang, Y., & Xu, N. (2021). Air pollution, affect, and forecasting bias: Evidence from Chinese financial analysts. *Journal of Financial Economics*, *139*(3), 971–984. https://doi.org/10.1016/j.jfineco.2019.12.004

Feng, P., Zhou, Z., Cohen, J. P., & Yasar, M. (2024). Air Pollution and the Effects on House

Prices: A Push for Sustainability. *Journal of Real Estate Research*, *46*(3), 336–359. https://doi.org/10.1080/08965803.2024.2344999

Fu, S., Viard, V. B., & Zhang, P. (2021). *Air Pollution And Manufacturing Firm Productivity: Nationwide Estimates For China*. *131*(April), 3241–3273.

Global Burden of Disease Collaborative Network. (2021). *Global Burden of Disease Study 2019 (GBD 2019) Air Pollution Exposure Estimates 1990-2019*. https://doi.org/https://doi.org/10.6069/70JS-NC54

Global Modeling and Assimilation Office (GMAO). (2015a). *MERRA-2 tavgM_2d_aer_Nx: 2d,Monthly mean,Time-averaged,Single-Level,Assimilation,Aerosol Diagnostics V5.12.4*. https://doi.org/10.5067/FH9A0MLJPC7N

Global Modeling and Assimilation Office (GMAO). (2015b). *MERRA-2 tavgM_2d_flx_Nx: 2d,Monthly mean,Time-Averaged,Single-Level,Assimilation,Surface Flux Diagnostics V5.12.4*. https://doi.org/10.5067/0JRLVL8YV2Y4

Goodwin, M. B., Fontenla, M., & Gonzalez, F. (2021). Estimating the impact of pollution on wages and housing prices using satellite imagery. *Applied Economics Letters*, *28*(20), 1750–1753. https://doi.org/10.1080/13504851.2020.1853665

Heyes, A., Neidell, M., & Saberian, S. (2016). The Effect of Air Pollution on Investor Behavior: Evidence from the S&P 500. *National Bureau of Economic Research*, w22753.

Kementerian Kesehatan RI. (2018). *Basic Health Research (Riset Kesehatan Dasar) 2018*.

Lestari, P. A., & Haryanto, B. (2022). SO2 Concentration and the Occurrence of Acute Respiratory Infection in Children Under Five. *Kesmas: Jurnal Kesehatan Masyarakat Nasional (National Public Health Journal)*, *17*(4), 292–296. https://doi.org/10.21109/kesmas.v17i4.6290

Liu, Y. M., & Ao, C. K. (2021). Effect of air pollution on health care expenditure: Evidence from respiratory diseases. *Health Economics (United Kingdom)*, *30*(4), 858–875. https://doi.org/10.1002/hec.4221

Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business and Economic Statistics*, *37*(2), 187–204. https://doi.org/10.1080/07350015.2016.1227711

Ramakrishnan, S., Hishan, S. S., Nabi, A. A., Arshad, Z., Kanjanapathy, M., Zaman, K., & Khan, F. (2016). An interactive environmental model for economic growth: evidence from a panel of countries. *Environmental Science and Pollution Research*, *23*(14), 14567–14579. https://doi.org/10.1007/s11356-016-6647-8

Sofwan, N. M., Mahiyuddin, W. R. W., Latif, M. T., Ayub, N. A., Yatim, A. N. M., Mohtar, A. A. A., Othman, M., Aizuddin, A. N., & Sahani, M. (2021). Risks of exposure to ambient air pollutants on the admission of respiratory and cardiovascular diseases in Kuala Lumpur. *Sustainable Cities and Society*, *75*(2), 103390. https://doi.org/10.1016/j.scs.2021.103390

Surury, I., Azizah, M., & Prastiwi, N. D. (2022). Spatial Analysis of Acute Respiratory Infection (ARI) Based on the Air Pollution Standard Index (PSI) at DKI Jakarta Region in 2018-2019. *Jurnal Kesehatan Lingkungan*, *14*(2), 90–98. https://doi.org/10.20473/jkl.v14i2.2022.90-98

Wu, T. J., Wu, C. F., Lee, Y. L., Hsiue, T. R., & Guo, Y. L. (2014). Asthma incidence, remission, relapse and persistence: A population-based study in southern Taiwan.

Respiratory Research, *15*(1), 1–9. https://doi.org/10.1186/s12931-014-0135-9

Zhao, Y., Dong, L., Sun, Y., Ma, Y., & Zhang, N. (2024). Is air pollution the original sin of firms? The impact of air pollution on firms' ESG scores. *Energy Economics*, *136*(January), 107704. https://doi.org/10.1016/j.eneco.2024.107704

Zivin, J. G., & Neidell, M. (2012). The impact of pollution on worker productivity. *American Economic Review*, *102*(7), 3652–3673. https://doi.org/10.1257/aer.102.7.3652

Zou, G., Lai, Z., Li, Y., Liu, X., & Li, W. (2022). Exploring the nonlinear impact of air pollution on housing prices: A machine learning approach. *Economics of Transportation*, *31*(July). https://doi.org/10.1016/j.ecotra.2022.100272