Original Research Article

A Cross-Sectional Analysis of Urban Air Pollution and Modeled Health-Economic Burden

Seungyoon Shin

Department of Earth & Environmental Sciences, University of Rochester, 252 Elmwood Ave., Rochester, NY 14627, Unites States

ABSTRACT

This study performs a cross-sectional analysis of global urban air pollution and health-economic burden modeled by using publicly available World Air Quality Index Data that contain 16,695 records from 14,229 city-county pairs across 185 countries. EPA breakpoint interpolation was used to estimate PM2.5 concentrations from AQI values and summarize them at city and county levels. To estimate relative risk (RR), attributable fraction (AF), and normalized economic burden were estimated by applying a simplified concentration-response model (β corresponding to a 6% risk increase per 10μg/m^3) by using standardized population and value-of-statistical-life parameters. Mean inferred PM2.5 concentrations in many cities substantially exceeded the WHO guideline of 5μg/m^3 and varied by more than an order of magnitude across cities. Cities with higher PM2.5 concentrations showed proportionally higher modeled RR and AF, which were associated with markedly larger normalized economic burden estimates. Without estimating causal effects or absolute costs, this study demonstrates how a transparent and reproducible quantitative framework may reveal global disparities in pollution-related health burden by using open data. The quantitative framework proposed in this study provides a foundation for future studies to incorporate temporal, demographic, and policy variables.

Keywords: Air pollution; PM2.5; Urban air quality; Concentration-response modeling; Economic burden; Global cities

INTRODUCTION

Air pollution is one of the most serious environmental risk factors that affect global public health and economic stability. The World Health Organization has reported that increased risk of cardiovascular disease, respiratory illness, and premature mortality is associated with the exposure to ambient air pollution worldwide (1). Air quality challenges have been intensified by rapid urbanization, industrial activity, and increased vehicular emissions in many major cities, and this issue has become more serious especially in low- and middle-income regions (2). As a result, it has become an important interdisciplinary research objective at the intersection of environmental science, public health, and economics to understand spatial patterns of air pollution and their correlated economic burdens.

One of the most harmful air pollutants is fine particle matter with an aerodynamic diameter that is less than 2.5 micrometers (PM2.5) because of its ability to penetrate deep into the lungs and enter the bloodstream (3). Numerous epidemiological studies have shown strong associations between elevated PM2.5 concentrations and increased risk of mortality, ischemic heart disease,

Corresponding author: Seungyoon Shin, E-mail: syoonshin96@gmail.com.

Accepted February 27, 2026
https://doi.org/10.70251/HYJR2348.421925

stroke, and chronic respiratory conditions (4). These findings have particularly motivated researchers to develop concentration-response (C-R) functions, quantitatively relating changes in pollutant concentration to health risk changes, and also to estimate population-level impacts from environmental exposure (5).

Beyond the scope of health consequences, air pollution imposes substantial economic costs on societies. These costs are from increased healthcare expenditures, the economic valuation of premature mortality, and losses of productivity from illness (6). Many prior studies made an attempt to quantify such impacts by employing the concept of the Value of a Statistical Life (VSL), representing willingness of the society to pay to decrease death risk (7). In spite of varying estimates of VSL across countries and income levels, it has been applied widely to a standardized framework in the translation of health risks into economic terms (8). Seen in this perspective, prior studies in dealing with air pollution have increasingly integrated environmental measurements with economic modeling to estimate the societal burden of degraded air quality in a broad range.

Traditionally, global assessments of air pollution have relied on observations from satellite, chemical transport, or monitoring networks from governments (9). More recently, transparency and data availability have increased due to publicly accessible platforms, including the World Air Quality Index (WAQI) project, that has aggregated real-time and historical air quality data obtained from many monitoring stations worldwide (10). Based on these datasets, comparative analyses have become available across cities and regions, especially for pollutants, such as PM2.5, PM10, nitrogen dioxide, ozone, sulfur dioxide, and carbon monoxide (11). However, many prior studies examined national or regional averages, while only a few studies focused on city-level variation in a standardized global framework.

Despite extensive global assessments of air pollution, many studies relied on complicated epidemiological or proprietary health datasets. Hence, they had limited transparency and reproducibility. Few studies have combined publicly available air quality data at the city level through a simplified and transparent health-economic framework (12, 13). In the perspective of educational and exploratory research contexts, this gap is especially important. Therefore, a simplified concentration-response quantitative framework combined with standardized economic valuation may provide an accessible approach to compare relative pollution burden across cities with clearly communicated modeling assumptions and limitations.

To achieve this goal, this study aims to address the literature gap by analyzing a publicly available global air quality dataset obtained from the World Air Quality Index, while estimating modeled health-related economic costs correlated with PM2.5 exposure in major cities. Instead of deriving causal inference or precise mortality estimation, the analysis in this study targets to emphasize comparative patterns and modeled burden by specifically using concentration-response relationships that are reported in the literature. To be more specific, this study aims to answer the research question about how PM2.5 pollution levels differ across major global cities, and how the modeled health-related economic costs are correlated with these differences if applying standard concentration-response functions. This study hypothesizes that cities with higher observed PM2.5 concentrations will show significantly greater modeled health-related economic costs than cities with lower PM2.5 concentrations.

METHODS AND MATERIALS

Data Source

Air quality data were obtained from the World Air Quality Index (WAQI) that is a publicly available global dataset accessed through the Kaggle data repository. Data were downloaded from Kaggle on January 8, 2026. Air quality observations for thousands of city-country pairs worldwide were contained in this dataset, including geographic coordinates and pollutants-related air quality index (AQI) values. Overall AQI, PM2.5 AQI, PM10 AQI, ozone AQI, nitrogen dioxide AQI, sulfur dioxide AQI, carbon monoxide AQI, city name, country name, latitude, and longitude were used as main variables in this study.

A time variable was not included in this dataset that was cross-sectional. Each record represents a single observation snapshot rather than daily averages. Therefore, spatial comparisons across cities and countries were the focus in this study instead of temporal trend estimation. All records with missing or incomplete pollutant values were excluded from pollutant-specific calculations. Country and city names were standardized for the purpose of consistent grouping. Duplicate city-county records were removed by using exact matching, while retaining one record per unique city-county pair.

Although the dataset contained 16,695 records, only 14,229 unique city-country pairs were present. This shows that some cities were represented by partial pollutant records or multiple monitoring stations. Since

there was no time variable provided, repeated city entries were interpreted as pollutant-specific observations or station-level rather than temporal measurements. Duplicate city-country records were removed using exact matching to avoid overweighting cities with multiple entries. Therefore, one representative record per city-country pair was retained for city-level analyses.

Data Processing

PM2.5 AQI values were converted into estimated concentration of PM2.5 mass (μg/m³) by using standard Environmental Protection Agency AQI breakpoint ranges in the U.S. AQI values converted to PM2.5 concentration using EPA breakpoint interpolation as follows.

$$C = \frac{(I - I_{low})\,(C_{high} - C_{low})}{(I_{high} - I_{low})} + C_{low}$$

Where I represents the AQI value, C is PM2.5 concentration (μg/m³), and I_{low}, I_{high}, C_{low}, C_{high} are AQI breakpoints defined by the U.S. EPA. AQI values exceeding 500 were capped at the maximum breakpoint to avoid extrapolation beyond validated ranges.

With this conversion, AQI-based measurements were expressed in concentration units that were commonly used in epidemiological concentration-response modeling. Observations that exceeded AQI bounds were capped at standard maximum values for avoiding unrealistic extrapolation.

After this process of conversion, city-level summary statistics were calculated by averaging PM2.5 concentrations and AQI values in all records in each city. Country-level summaries were also computed to support broader geographic comparisons. With descriptive statistics generated, the overall distribution of air pollution levels across the dataset was characterized.

Health Risk Modeling

A simplified concentration-response (C-R) model was applied to estimate health-related impacts associated with PM2.5 exposure. With this approach that is commonly used for educational and comparative studies, this study followed prior established epidemiological literature. Relative risk (RR) of adverse health outcomes was estimated by the model as an exponential function of PM2.5 concentration above a reference level.

The relative risk was computed as follows.

$$RR = e^{\beta (C - C_0)}$$

Where C represents the estimated PM2.5 concentration (μg/m³), C_0 shows the reference concentration set to 5μg/m³ according to World Health Organization guidelines, and β was the concentration-response coefficient that was derived from prior literature and was 0.0058 per μg/m³ (equivalent to a 6% increase per 10 μg/m³) based on prior literature (5).

Representing the proportion of risk attributable to PM2.5 exposure, the attributable fraction (AF) was calculated as follows.

$$AF = \frac{RR - 1}{RR}$$

Economic Cost Estimation

Health-related economic burden was estimated by using a simplified framework for valuation. Normalized economic burden was calculated as follows:

$$\text{Economic Cost} = AF \times P \times VSL$$

Where P=1 (normalized population unit) and VSL=1 (normalized value of statistical life). Normalization enabled relative comparison across cities without implying absolute monetary estimates.

All data processing, calculations, and visualizations were conducted by using Python program. Results in this study were exhibited in descriptive summaries, ranked comparisons, and graphical distributions. However, this study did not claim any causal inference. Instead, this study provided modeled estimates only for comparative and educational purposes by using reproducible methods based on transparent assumptions.

RESULTS

Dataset Overview and Descriptive Statistics

The final dataset used for the analysis in this study consisted of 16,695 air quality records that represented 14,229 unique city-country pairs across 185 countries worldwide. For each record, overall Air Quality Index (AQI) values were contained, along with pollutant-specific AQI measures, including PM2.5, PM10, ozone, nitrogen dioxide, sulfur dioxide, and carbon monoxide. For all included observations, geographic coordinates were available, allowing for consistent grouping at the city and country levels.

According to descriptive statistics, substantial variability was revealed in air quality across the sampled cities. Overall AQI values were in a range from "Good" to "Hazardous" categorized in the levels. The majority of observations fell within the "Moderate" to "Unhealthy"

ranges. PM2.5 AQI values suggested a right-skewed distribution, showing how a subset of cities reported a substantially higher PM2.5 exposure, even though many cities experienced moderate pollution levels (Figure 1). After converting PM2.5 AQI to estimated mass concentration (μg/m³), a wide range was reported in inferred PM2.5 concentrations, with several cities exceeding the levels from international guideline by a large margin.

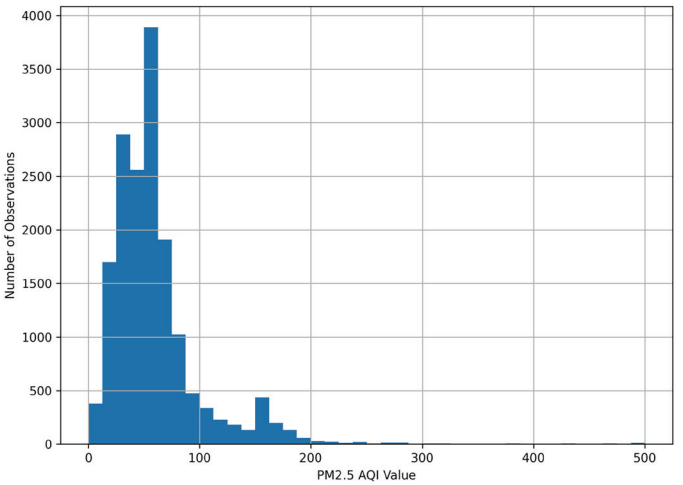

Figure 1. *Histogram of estimated PM2.5 mass concentrations (μg/m³) derived from 16,695 WAQI observations by using EPA AQI breakpoint interpolation. Substantial variability in urban air pollution exposure worldwide is shown in the right-skewed distribution.*

City-Level Air Pollution Patterns

According to city-level aggregation, there were pronounced differences in PM2.5 exposure across global urban centers. Mean inferred PM2.5 concentrations turned out to vary by more than an order of magnitude between the least and most polluted cities. Average values in the highest-ranked cities consistently exceeded the World Health Organization reference level of 5μg/m³. However, cities at the lower end of the distribution were clustered around or slightly above this threshold.

According to the ranked comparison of the top 20 cities by mean inferred PM2.5 concentrations, extreme pollution was not evenly distributed geographically (Figure 2). Instead, it turned out that high PM2.5 concentrations were concentrated in a limited number of urban regions. This suggested a strong regional clustering of air quality conditions. In contrast, cities with lower PM2.5 concentrations were distributed across multiple continents. This suggested how low pollution levels were available under both diverse socioeconomic and geographic contexts.

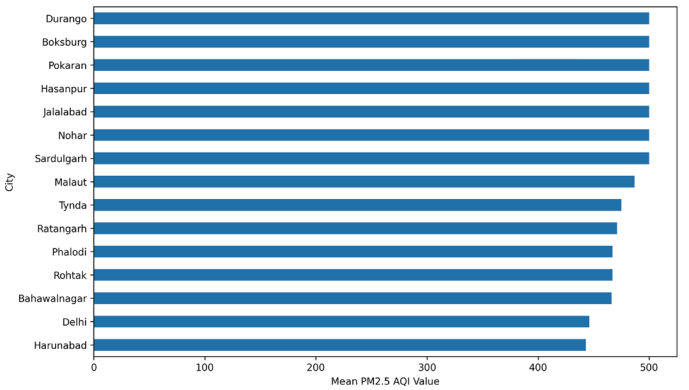

Figure 2. *Top 15 cities ranked by mean inferred PM2.5 concentration. With AQI-to-PM2.5 conversion, city-level average (μg/m³) was calculated by using EPA breakpoints and duplicate-record removal. Results showed strong geographic inequality in urban pollution exposure across major global cities.*

Country-Level Comparisons

In an attempt to reduce instability from small sample sizes, this study has restricted country-level summaries to countries with records of at least 50 cities. Among these countries, mean inferred PM2.5 concentrations turned out to vary significantly (Figure 3). This reflected

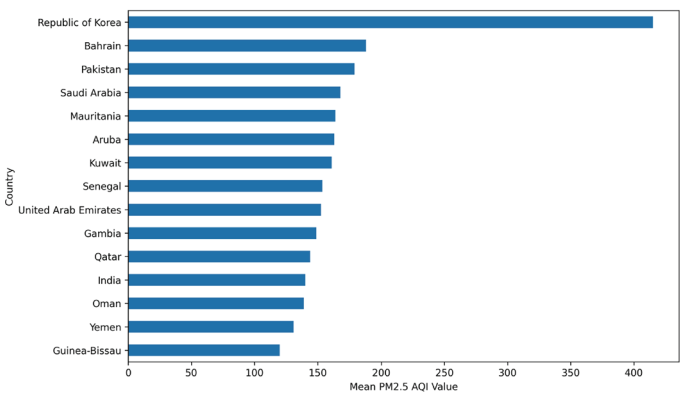

Figure 3. *Top 15 countries ranked by mean inferred PM2.5 concentration. Country means were calculated from cities with more than 50 observations to reduce instability in sampling. Larger modeled relative risk and normalized economic burden corresponded to higher national averages.*

differences in regional pollution profiles. Specifically, countries with higher mean PM2.5 concentrations also suggested higher overall AQI values. This showed consistency between PM2.5-specific and composite air quality measures.

The distribution of country-level PM2.5 concentrations suggested how there was global inequality in air quality exposure. Even if some countries exhibited relatively low average PM2.5 levels among their monitored cities, others suggested consistently elevated concentrations. Due to these differences, the basis has been formed for subsequent modeled health impact and also economic burden estimation.

Modeled Health Risk Estimates

With the simplified concentration-response model, relative risk (RR) values were calculated according to inferred PM2.5 concentration above the reference level for each city. As anticipated, RR turned out to monotonically increase as PM2.5 concentration increased. Cities with the highest inferred PM2.5 levels turned out to exhibit substantially greater RR values than one. This suggested an elevated modeled health risk compared to the reference concentration.

Representing the proportion of health risk attributable to PM2.5 exposure, the attributable fraction (AF) also suggested strong variation among cities. To be more specific, cities with low pollution levels had AF values close to zero, indicating minimal modeled excess risk. In contrast, cities with high PM2.5 concentrations exhibited AF values that approached or exceeded moderate levels. This suggested that there was a significant fraction of modeled health risk that may be attributed to PM2.5 exposure under the assumptions of the model.

Modeled Economic Burden

The resulting modeled economic costs suggested large disparities across cities when AF values were combined with normalized population and Value of Statistical Life (VSL) parameters. Specifically, cities with higher PM2.5 concentrations consistently showed greater modeled economic burden compared to cities with lower PM2.5 concentrations. Variation in inferred PM2.5 exposure drove these differences rather than changes in model parameters that were held constant across locations.

Importantly, the economic cost estimated modeled in this study were intended for relative comparison rather than precise valuation. When applying even conservative concentration-response assumptions, cities in the highest pollution categories suggested disproportionately large modeled economic burdens compared to the cities in lower pollution categories. According to this pattern, the hypothesis established in this study has been supported that air pollution disparities across cities were associated with significant differences in potential health-related economic outcomes.

Pearson correlation analysis showed a strong positive correlation between inferred PM2.5 concentration and modeled economic burden ($r = 0.82$). According to Bootstrapped 95% confidence intervals (1,000 resamples), stable estimates of mean PM2.5 concentration were reported across cities. Proportional changes in relative risk and economic burden were reported by a sensitivity analysis varying β by $\pm 20\%$, while preserving the ranking of cities. These results supported the robustness of comparative conclusions.

Summary of Findings

Overall, the findings in this study suggested substantial global variation in urban air quality and also corresponding modeled health-economic burden. Cities with higher PM2.5 concentrations turned out to consistently experience higher relative risk, greater attributable fraction, and also increased modeled economic costs. These findings underscore the potential societal implications of uneven air pollution exposure among major cities, while providing a quantitative framework for comparative analysis conducted through publicly available data.

DISCUSSION

This study examined global variation in urban air quality, while estimating modeled health-related economic burden by using air pollution data from publicly available source and also a simplified concentration-response quantitative framework. The results in this study showed substantial differences in PM2.5 exposure among cities and countries where a subset of urban areas showed significantly higher inferred PM2.5 concentrations than others. These disparities led directly to higher modeled relative risk, attributable fraction, and economic burden with consistent modeling assumptions that were consistent with hypothesis in this study.

The observed high concentration levels of PM2.5 in particular regions were consistent with prior global assessments of air pollution to factors, including industrial activity, population density, and regulatory enforcement. Even though this study did not particularly apply socioeconomic or policy variables, structural and

environmental determinants may play an important role in exhibiting urban air quality outcomes according to the spatial clustering of elevated pollution. More importantly, poor air quality may not be inevitable and can be mitigated through effective environmental management.

In this study, transparent estimation of health-related economic burden was enabled by the application of a simplified concentration-response model without facing the complexity of large-scale epidemiological simulations. With this approach, reproducibility and interpretation of assumptions were clearly possible. Since spatial comparisons across cities and countries were of the focus in this study, the modeled economic cost estimates shall not be interpreted as precise monetary values. However, they provide meaningful comparative insight into how disparities in pollution levels may translate into disproportionate societal impacts.

However, there are limitations in this study to be considered when interpreting the findings. The dataset was cross-sectional without temporal information. This prevented analysis of long-term trends or causal inference. In addition, PM2.5 concentrations were inferred from AQI values instead of applying direct measurement of mass concentrations. This may have introduced estimation error. Furthermore, since population size and Value of Statistical Life parameters were normalized for comparative purposes, the precision of absolute cost estimates was limited.

Despite these limitations, this study exhibits that publicly accessible air quality data combined with established modeling techniques may yield valuable insights in the perspective of global pollution disparities and their potential health-economic implications. Future research may take this framework as a foundation for incorporating temporal data, demographic variation, or policy evaluation to refine understanding of air pollution impact.

CONCLUSION

The analysis in this study did not establish causal relationships or precise monetary costs. However, it generated a reproducible framework for the purpose of comparative assessment of air pollution impacts. The findings in this study emphasize the potential societal implications of disparities in air quality exposure, while highlighting the importance of continued efforts for continued monitoring and mitigation. It is recommended for future research to extend the approach taken in this study by applying temporal data, population-specific parameters, or policy indicators to further refine estimates, while exploring drivers of observed disparities.

CONFLICT OF INTEREST

The author declares no conflicts of interest related to this work.

REFERENCES

1. World Health Organization. Ambient air pollution: A global assessment of exposure and burden of disease. Geneva: WHO; 2016. Available from: https://www.who.int/publications/i/item/9789241511353 (accessed on 2026-01-01).
2. Cohen AJ, Brauer M, Burnett R, Anderson HR, *et al*. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet*. 2017; 389 (10082): 1907-1918. https://doi.org/10.1016/S0140-6736(17)30505-6
3. Pope CA III, Dockery DW. Health effects of fine particulate air pollution: lines that connect. *J Air Waste Manag Assoc*. 2006; 56 (6): 709-742. https://doi.org/10.1080/10473289.2006.10464485
4. Brook RD, Rajagopalan S, Pope CA III, Brook JR, *et al*. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. *Circulation*. 2010; 121 (21): 2331-2378. https://doi.org/10.1161/CIR.0b013e3181dbece1
5. Burnett R, Chen H, Szyszkowicz M, Fann N, *et al*. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc Natl Acad Sci USA*. 2018; 115 (38): 9592-9597. https://doi.org/10.1073/pnas.1803222115
6. Landrigan PJ, Fuller R, Acosta NJR, Adeyi O, Arnold R, Baldé AB, *et al*. The Lancet Commission on pollution and health. *Lancet*. 2018; 391 (10119): 462-512. https://doi.org/10.1016/S0140-6736(17)32345-0
7. Viscusi WK, Aldy JE. The value of a statistical life: a critical review of market estimates throughout the world. *J Risk Uncertainty*. 2003; 27 (1): 5-76. https://doi.org/10.1023/A:1025598106257
8. Organisation for Economic Co-operation and Development. Mortality risk valuation in environment, health and transport policies. Paris: OECD Publishing; 2012.
9. Van Donkelaar A, Martin RV, Brauer M, Boys BL, *et al*. Global estimates of fine particulate matter using a combined geophysical-statistical method with satellite,

model, and monitor data. *Environ Sci Technol*. 2016; 50 (7): 3762-3772. https://doi.org/10.1021/acs.est.5b05833

10. World Air Quality Index Project. World Air Quality Index (WAQI) Project: real-time air quality information. Available from: https://aqicn.org (accessed on 2025-12-04).
11. OpenAQ. Open air quality data platform. Available from: https://openaq.org (accessed on 2025-12-02).
12. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis. *Lancet.* 2020; 396 (10258): 1223-1249.
13. Apte JS, Brauer M, Cohen AJ, Ezzati M, Pope CA III. Ambient PM2.5 reduces global and regional life expectancy. *Proc Natl Acad Sci USA*. 2018; 115 (38): 9592-9597. https://doi.org/10.1021/acs.estlett.8b00360