Rethinking PM2.5 Exposure:

Chronic Disease Trends in the U.S. (2015 - 2022)

Bo Wudhinitikornkij

Introduction

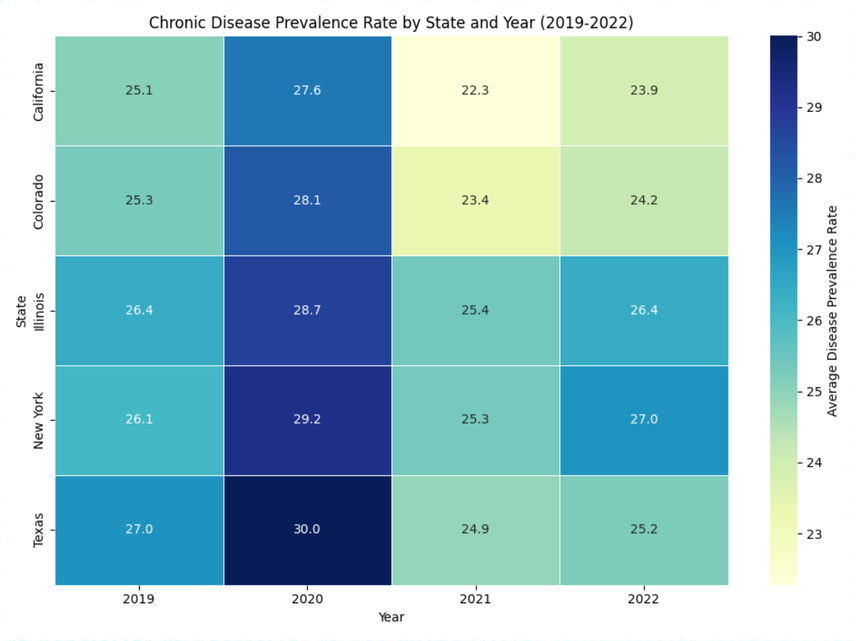
- Chronic diseases such as cardiovascular disease, respiratory disorders, and cancer remain major health burdens in the United States. At the same time, PM2.5 air pollution continues to raise global concerns due to its proven links to respiratory and cardiovascular outcomes.

- This project aims to analyze U.S. chronic disease prevalence alongside national PM2.5 trends and to compare U.S. air quality patterns with global PM2.5 levels. Together, these trends help contextualize how air pollution may relate to population health in the United States.

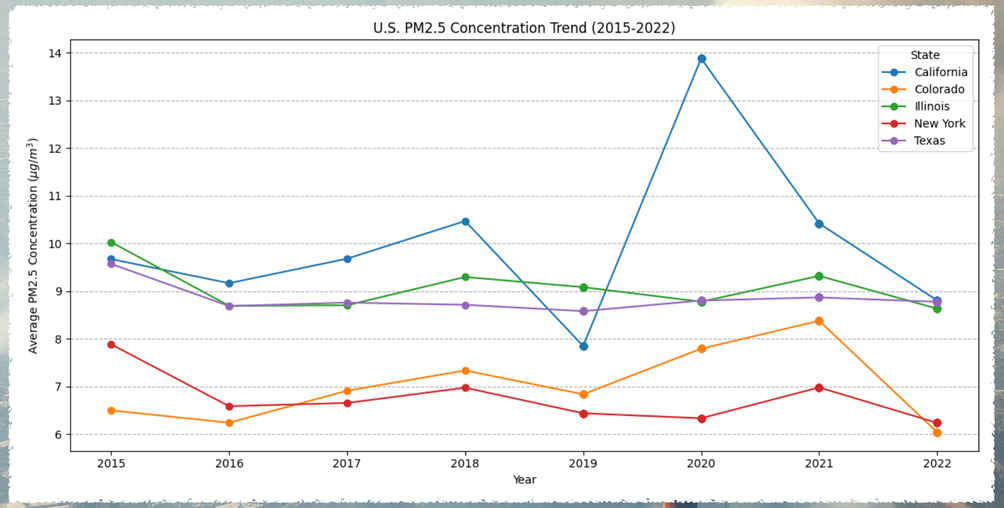
Data Sources	Description	Type	Format	Data Size
EPA U.S. Environmental Protection Agency	Air quality system (AQS) API https://aqs.epa.gov/aqsweb/documents/data_api.html	API	json	Retrieve: 25,270 Processed: 40 rows x 3 columns
CDC Centers for Disease Control and Prevention	U.S. Chronic Disease Indicators https://catalog.data.gov/dataset/u-s-chronic-disease-indicators	Web request	json	Retrieve: 309,215 Processed: 281 rows x 5 columns
WHO World Health Organization	Air pollution: concentrations of fine particular matter (PM2.5) https://www.who.int/data/gho/data/indicators/indicator-details/GHO/concentrations-of-fine-particulate-matter-%28pm2-5%29	Web request from Google Drive	json	Retrieve: 4,725

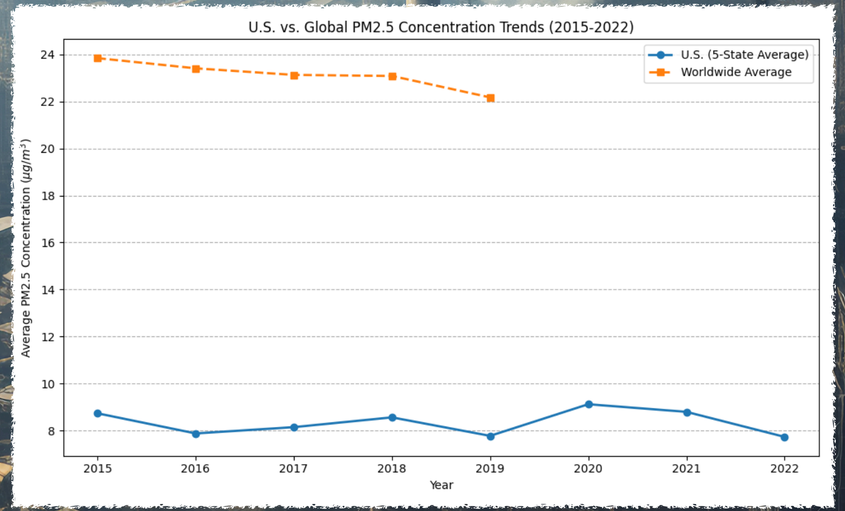
Summary of the results

Environmental Exposure & Health Landscape (2015–2022)



- 2020 stands out with the highest prevalence values, especially in **Texas**, which shows the **darkest** shade.
- **California** and **Colorado** show consistently lower prevalence from 2021–2022, visible as **lighter** colors.
- **Highest states**:
  - 2019–2020: Texas
  - 2021–2022: New York and Illinois with similar values
- **Lowest states**: California, especially in 2021, appears lowest overall.

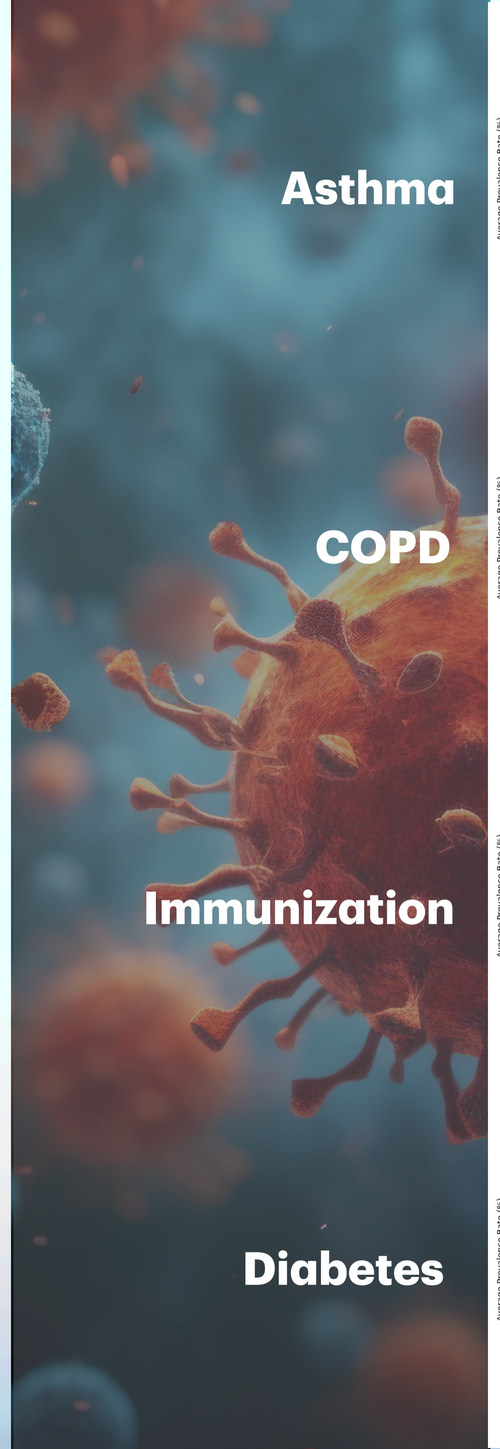


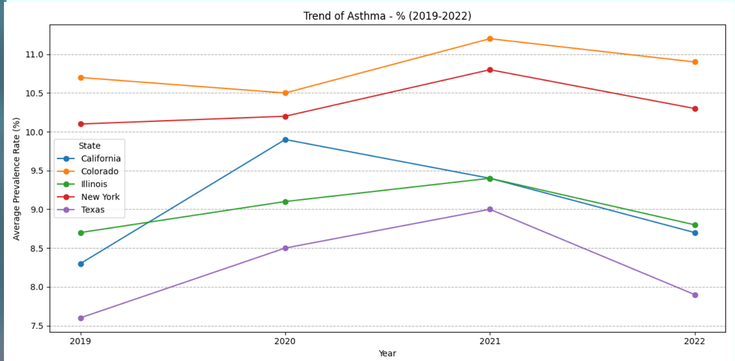


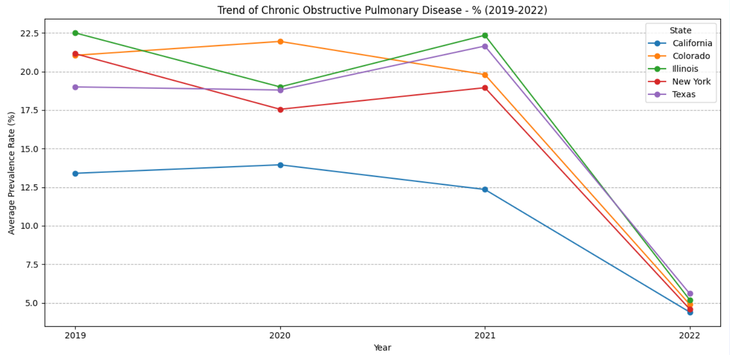
Summary of the results

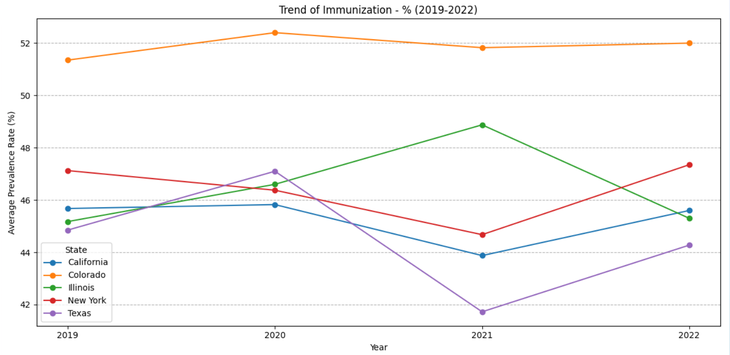
U.S. Chronic Disease Trends (2019 - 2022)

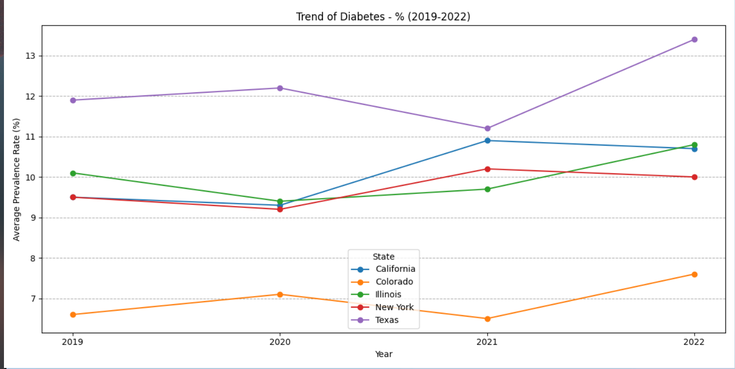
- **Overall**
  - Chronic disease rates change only slightly over time → suggests long-term structural health patterns rather than short-term fluctuations.
- **Respiratory diseases (Asthma & COPD):**
  - Clear separation between states — CO & NY trend higher, CA & TX remain lower.
- **Interesting mismatch:**
  - California shows low respiratory prevalence despite having one of the highest PM2.5 levels among the 5 states.
- **Diabetes:**
  - Displays a totally different state ranking, showing that metabolic conditions follow different drivers than respiratory diseases.
- **Immunization:**
  - Behaves differently from disease indicators — trends appear more influenced by policy and access than environmental exposure.







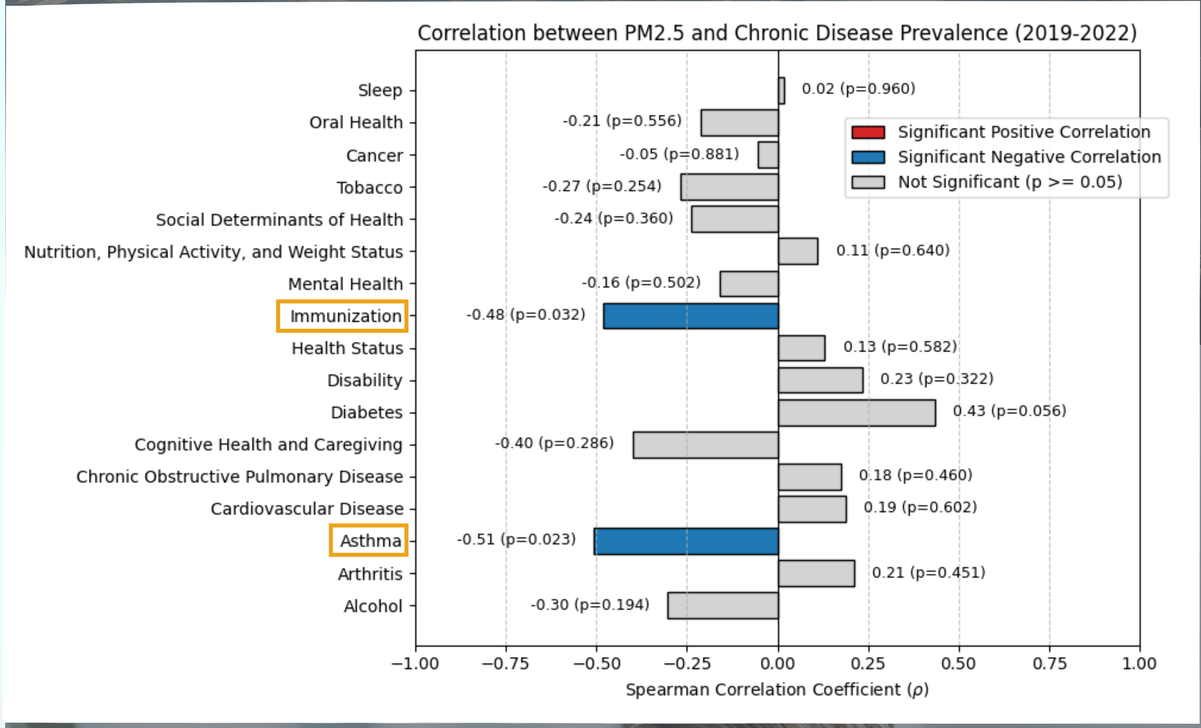




Statistical Correlation Analysis

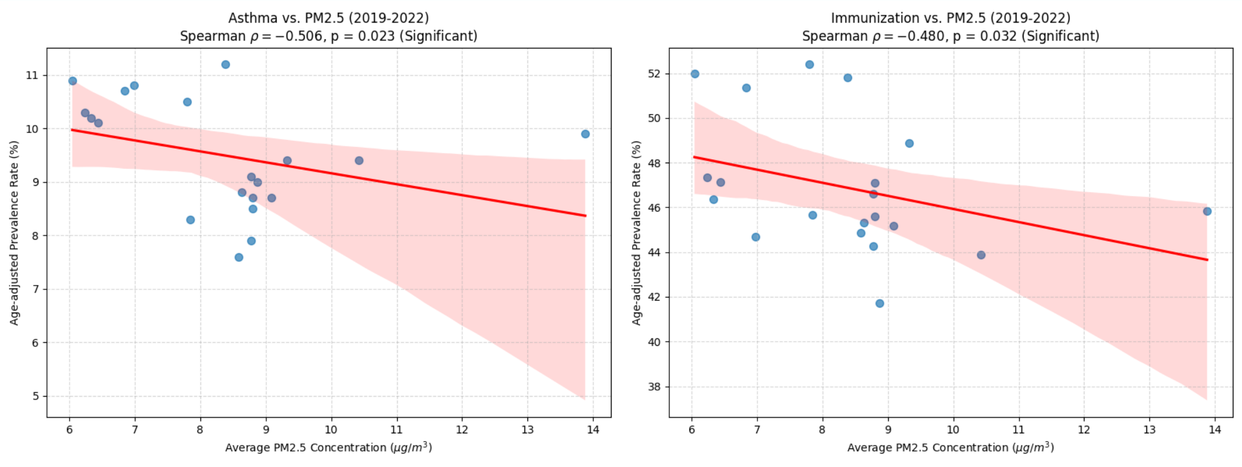
U.S. PM2.5 vs. 17 Chronic Diseases

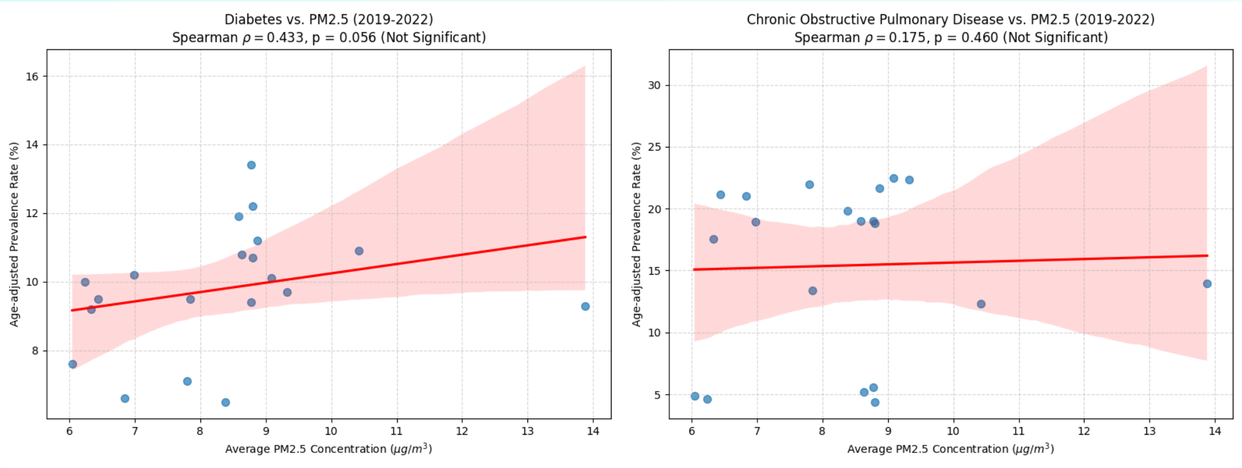
- **Only 2 relationships are statistically significant:**
  - **Asthma** (ρ = –0.51, p = 0.022) → higher PM2.5 = lower reported asthma
  - **Immunization** (ρ = –0.48, p = 0.032) → higher PM2.5 = lower immunization rates
- Most diseases show weak or non-significant correlations, suggesting PM2.5 is not a strong direct driver of chronic disease prevalence within this limited 5-state dataset.
- Diabetes is the only condition showing a moderate positive trend (ρ = 0.43, $p \approx$ 0.056), but not statistically significant.
- Cardiovascular and COPD correlations are very weak, showing no meaningful association in this sample.
- **Overall pattern:**
  - Environmental exposure varies more than chronic disease does.
  - Chronic diseases appear driven by long-term structural, behavioral, or socioeconomic factors, not year-to-year air quality differences at the state level.



Visual Confirmation

Scatter Plots of PM2.5 Relationship





- **Significant Relationships : (Asthma & Immunization)**
  - Both plots exhibit a **clear, tight downward slope**, visually confirming the strong negative correlations (ρ≈−0.50).

- **Non-Significant Relationships : (COPD & Diabetes)**
  - Both plots show **wide scattering** of data points around a nearly flat or weak regression line.

Conclusion

U.S. Chronic Disease and PM2.5 Analysis

I. Statistical Findings

- **Significant Negative Link:** Only **Asthma** ($\rho=-0.51$) and the **Immunization** control metric ($\rho=-0.48$) show a statistically significant relationship with PM2.5.
- **Biological Hypothesis Rejected**: We found no significant correlation for COPD or Diabetes ($P>0.05$).

II. Ecological Factors

- The negative correlation means that cleaner states have higher Asthma rates. This is driven by Ecological Confounding.
- The identical pattern in Immunization confirms a powerful third factor—likely Socioeconomic Status or strong public health policy—that influences both air quality and disease rates.
- For Asthma, factors like population migration (Avoidance Effect) and local policy are stronger determinants than direct environmental exposure in this dataset.

Challenges

U.S. Chronic Disease and PM2.5 Analysis

- **Data Sparsity:** CDC data was insufficient, forcing us to restrict the study window to 2019–2022 and use only the Age-adjusted Prevalence metric.
- **Small Sample Size:** Final analysis was limited to only n≈20 data points (5 states x 4 years), severely reducing statistical power and masking weaker effects (COPD, Diabetes).
- **Correlation Method:** Had to use Spearman Rank Correlation over Pearson due to high variability and extreme outliers in the PM2.5 data (e.g., wildfire spikes).
- **Analytical Interpretation:** Interpreting the significant negative correlation for Asthma required moving beyond biology to consider Ecological Confounding (e.g., Avoidance Effect).
- **The Confounder:** The significant result for the Immunization control metric confirmed that a powerful, unmeasured socioeconomic variable was influencing the entire dataset. This must be addressed with external data in future work.

“Q & A”

Thank you for your attention