Rethinking PM2.5 Exposure:

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

Bo Wudhinitikornkij

Introduction

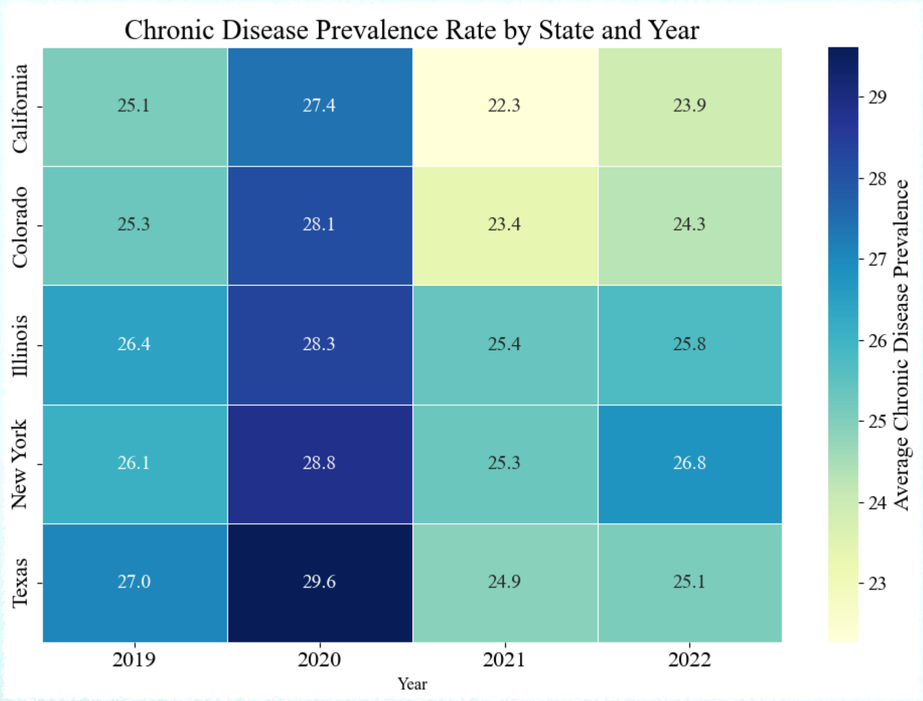
- Chronic diseases such as cardiovascular disease, respiratory disorders, and cancer remain major health burdens in the United States. At the same time, PM2.5 air pollution continues to raise global concerns due to its proven links to respiratory and cardiovascular outcomes.

- 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.

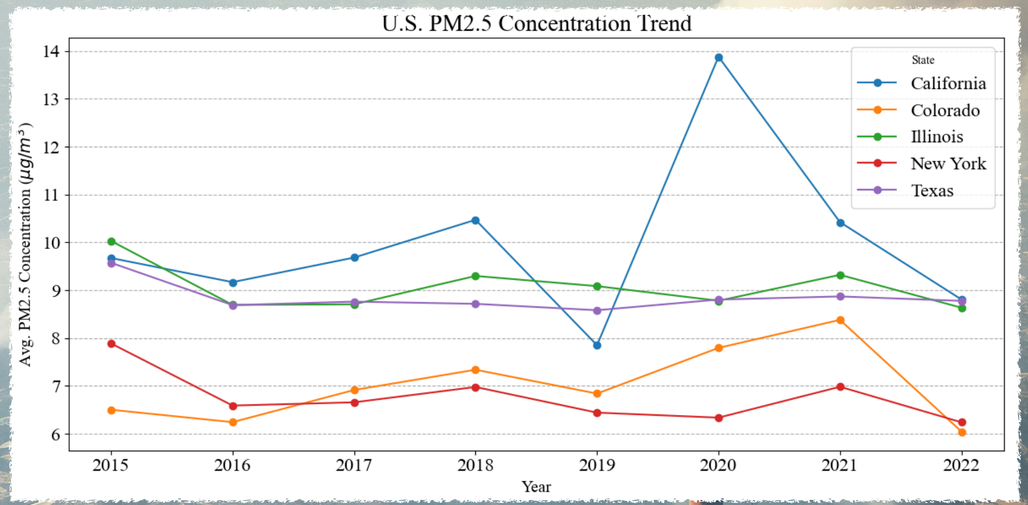
Data Sources	Description	Type	Format	Data Size
EPA U.S. Environmental Protection Agency	Air quality system (AQS) API https://aqs.epa.gov/aqsweb/documents/data_api.html	API	json	Retrieve: 25,270 Aggregated: 40 rows x 3 columns
CDC Centers for Disease Control and Prevention	U.S. Chronic Disease Indicators https://catalog.data.gov/dataset/u-s-chronic-disease-indicators	Web request	json	Retrieve: 309,215 Aggregated: 290 rows x 6 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	csv	Retrieve: 4,725

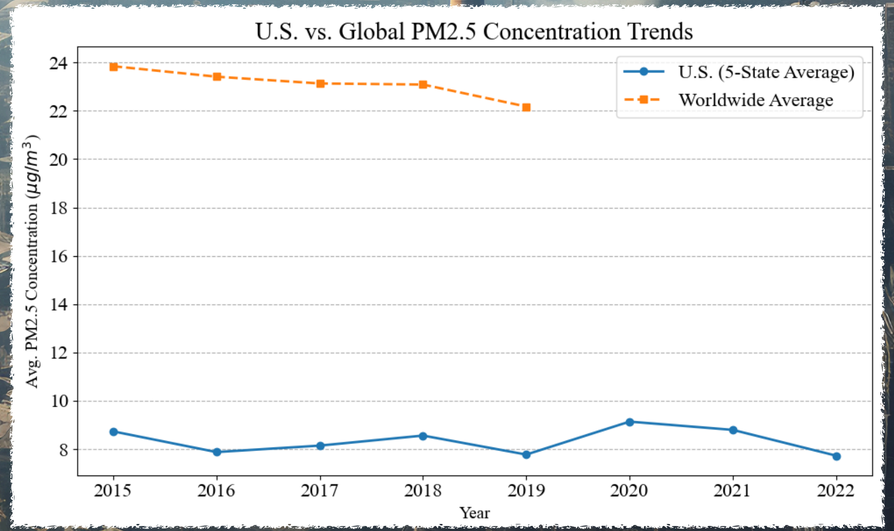
Summary of the results

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.

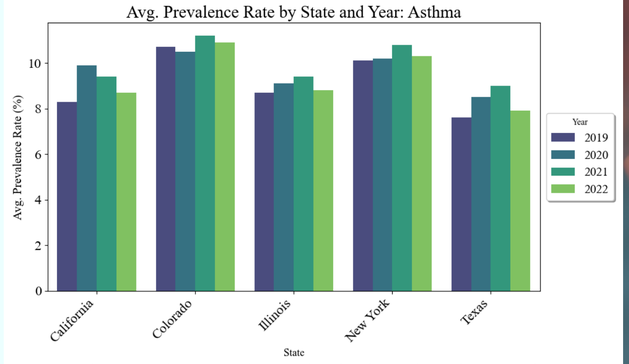


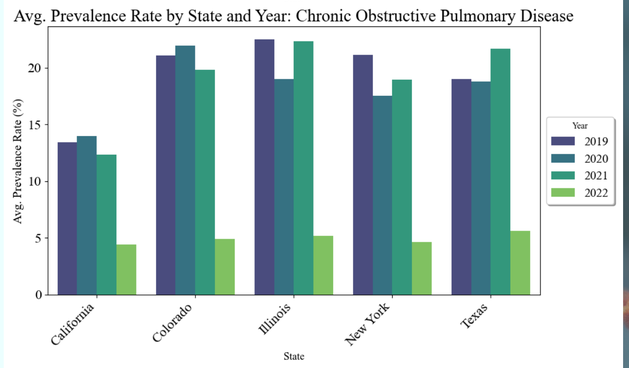


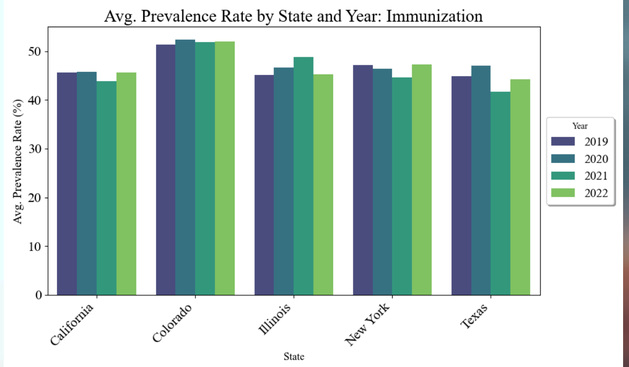
Summary

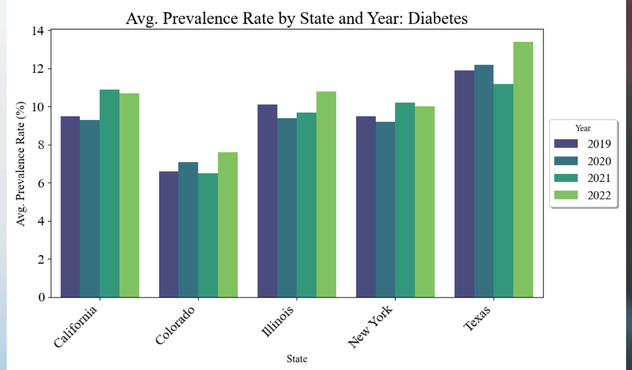
U.S. Chronic Disease Trends (2019 - 2022)

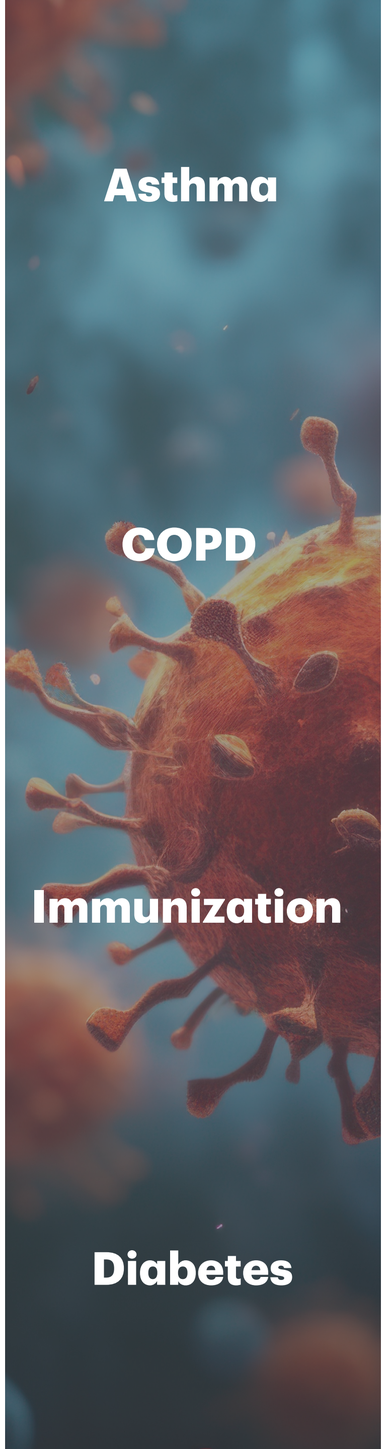
- **Overall**
  - Chronic disease rates change only slightly over time → suggests long-term structural health patterns rather than short-term fluctuations.
- **Respiratory diseases (Asthma & COPD):**
  - Clear separation between states — CO & NY trend higher, CA & TX remain lower.
- **Interesting mismatch:**
  - 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.

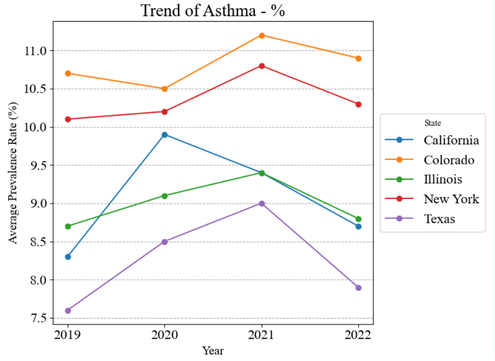


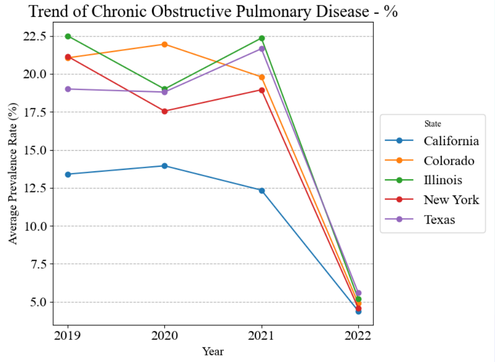


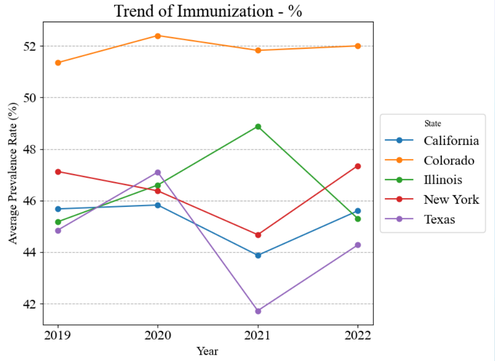


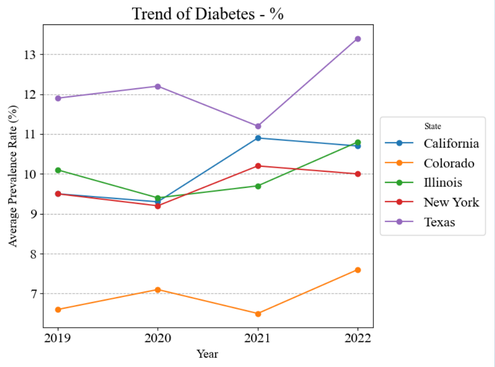








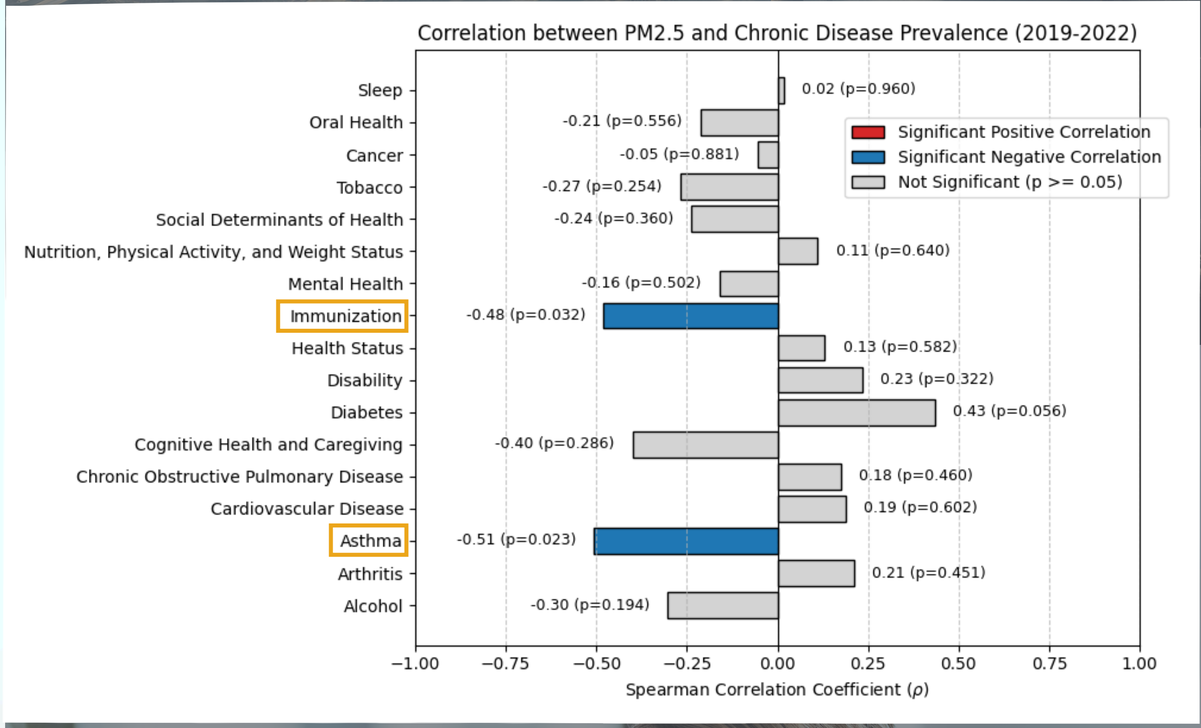




Statistical Correlation Analysis

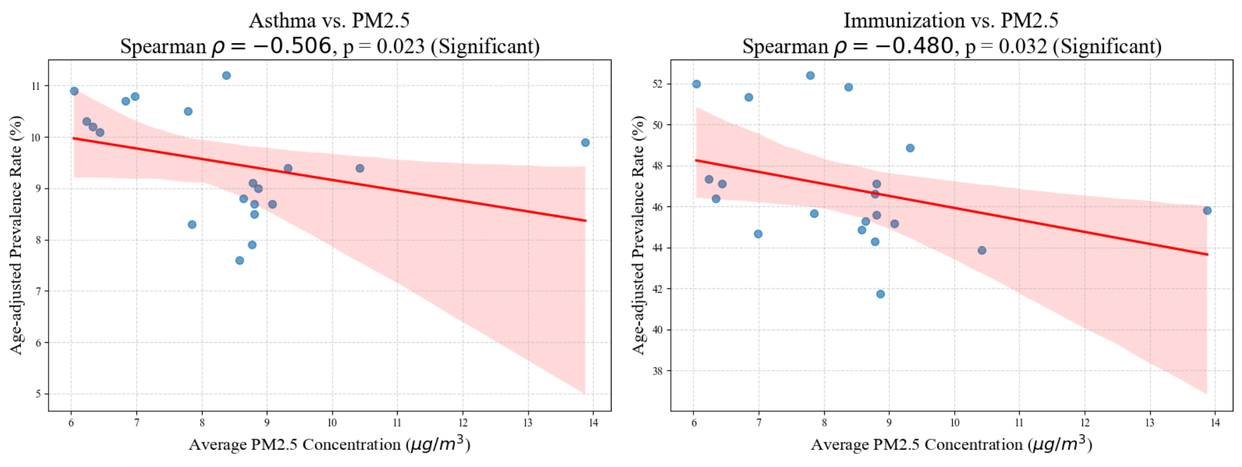
U.S. PM2.5 vs. 17 Chronic Diseases

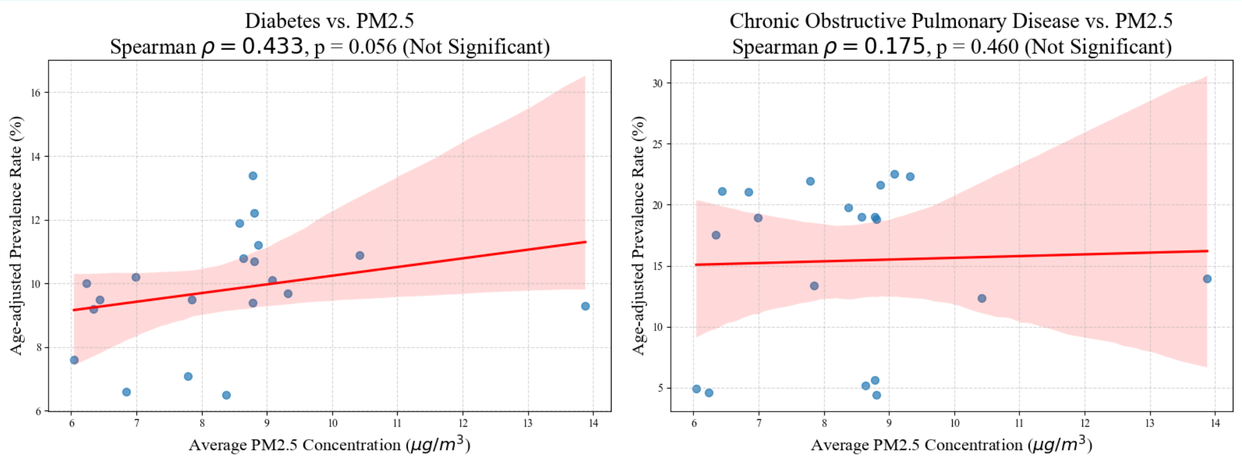
- **Only 2 relationships are statistically significant:**
  - **Asthma** (ρ = –0.51, *p* = 0.022) → higher PM2.5 = lower Asthma
  - **Immunization** (ρ = –0.48, *p* = 0.032) → higher PM2.5 = lower Immunization
- Most diseases show weak or non-significant correlations, suggesting PM2.5 is not a strong 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* ≈ 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





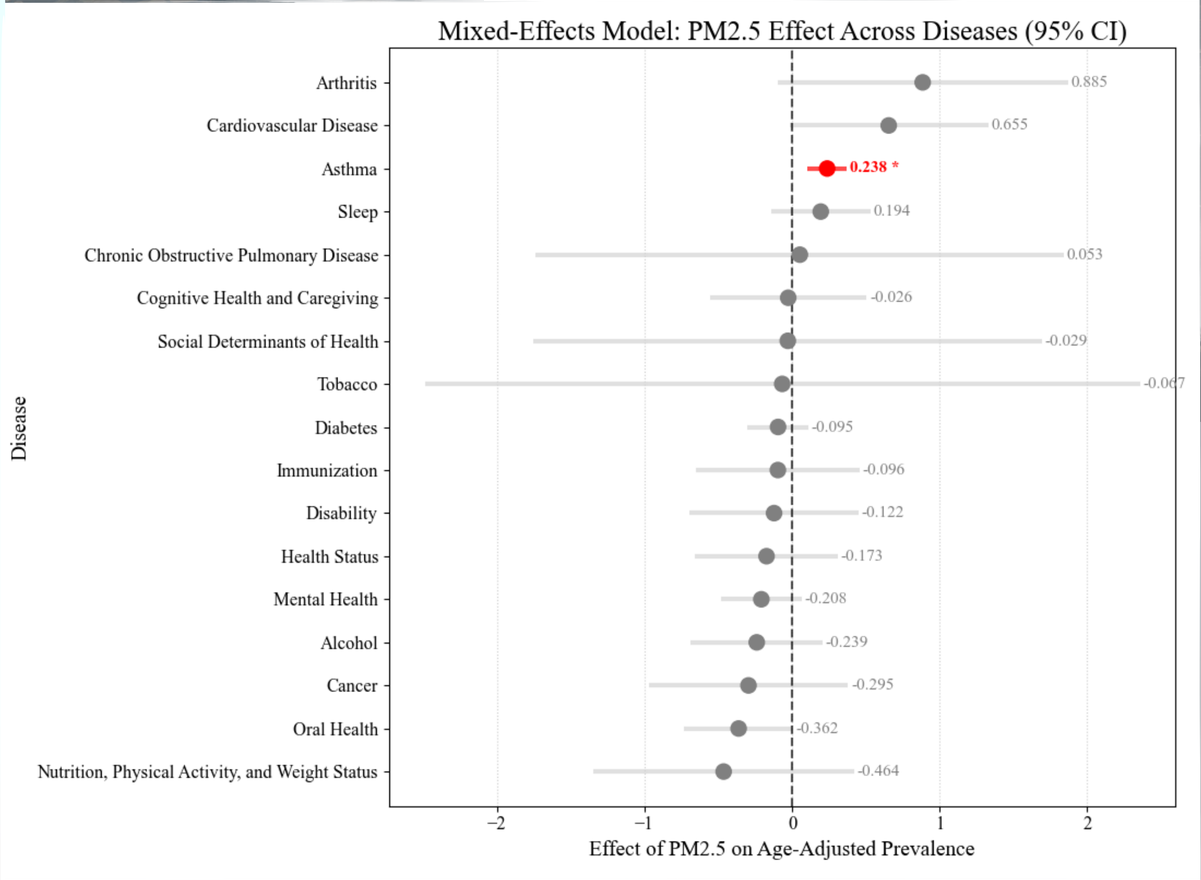
- **Significant Relationships : (Asthma & Immunization)**
  - Both plots exhibit a **clear, tight downward slope**, visually confirming the strong negative correlations (ρ≈−0.50).

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

Mixed-Effects Models Analysis

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

- **Key Independent Risk Factor**
  - The MEM identified **Asthma** as the only disease with a statistically significant positive association with PM2.5 concentration.
  - The confidence intervals for **all other diseases** (including Diabetes, COPD, etc.) cross the zero vertical line. This means the effect of PM2.5 is **not statistically different from zero** when controlling for the year trend.
- **Resolution of Contradiction**
  - **Simple Correlation (ρ=−0.51):** Showed a misleading negative link, confusing the overall decline in PM2.5 over time with the rise in Asthma rates.
  - **MEM (Coef. = +0.238):** Confirmed the expected **positive relationship** (higher PM2.5 = higher Asthma) after controlling for the strong, underlying societal trends that are causing Asthma rates to increase year-over-year.



Conclusion

U.S. Chronic Disease and PM2.5 Analysis

I. Initial Findings (Spearman Correlation)

- **Significant Negative Link:** Only **Asthma** ($\rho=-0.51$) and the **Immunization** ($\rho=-0.48$) show a statistically significant relationship with PM2.5.
- This negative link contradicted the biological hypothesis and suggested a major source of **Ecological Confounding**.

II. Mixed-Effects Model (MEM) Findings

- **Asthma Reversal (Biological Hypothesis Confirmed)**: the relationship with Asthma **reversed** and became **highly significant and positive** (Coef.=+0.238).
- PM2.5 is an **independent risk factor for Asthma**. The strong negative correlation was indeed an artifact of Ecological Confounding.
- The MEM found **no significant independent PM2.5 effect** for COPD or Diabetes, confirming that pollution is not the primary driver for these outcomes in this specific dataset.

III. Ecological Factors

- **Policy Insight:** For most chronic diseases in this dataset, PM2.5 is a **proxy** for larger determinants (Socioeconomic Status, healthcare access), which are far stronger drivers of prevalence than the air pollutant itself.

Challenges

U.S. Chronic Disease and PM2.5 Analysis

- **Data Constraints**
  - **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 5 states/areas, severely reducing statistical power and likely masking weaker, but biologically relevant, effects (e.g., for COPD or Diabetes).
- **Analytical & Confounding Issues:**
  - **Correlation Method Limitations:** We had to use Spearman Rank Correlation (due to PM2.5 outliers/variability), but its low statistical power dictated the need for a more robust method.
  - **Ecological Confounding:** Simple correlation produced a significant, but misleading, negative link for Asthma. Interpreting this required accounting for Ecological Confounding (e.g., the avoidance effect of migration).
- **Solution Through Modeling**
  - **MEM:** The Mixed-Effects Model (MEM) was critical for success, providing a hierarchical approach that was robust to the small sample size and allowed us to isolate the independent effect of PM2.5 by explicitly controlling for the powerful influence of the Time Trend (Year).

“Q & A”

Thank you for your attention