**BACKGROUND**

San Jose Diabetes Research Institute (SDRI) has a 75-year tenure conducting care, education, and research in diabetes. Incidence of diabetes on Latino families and recognizing that effective diabetes prevention and treatment must rest on understanding the determinants of diabetes as well as accurate risk calculation. SDRI launched Mi Familias in 2019.

California is home to 6 of the 10 most air polluted cities in the U.S.

Diabetes costs California more than $2 billion annually.

Diabetes has a higher prevalence among minorities.

One specific pollutant, particulate matter (PM$_{2.5}$), has been linked to a range of negative health impacts including diabetes. There has been no study exploring this association in California. This project aims to fill this crucial gap by assessing the relationship between particulate matter 2.5 (PM$_{2.5}$) and diabetes prevalence in California, USA, using a cross-sectional and panel data approach.

**RESEARCH QUESTIONS**

1. What are the yearly average PM$_{2.5}$ concentrations for each census tract in California?
2. What is the relationship between PM$_{2.5}$ concentration and diabetes prevalence in California?
3. How do these relationships change if we incorporate threshold values for PM$_{2.5}$ levels or demographic subgroups?

**DATA**

In this project we combine publicly available datasets for the years 2014-2017:

1. PM$_{2.5}$ concentrations from the California Air Resource Board
2. Diabetes prevalence from the Centers for Disease Control
3. Socio-demographic variables from the Census Bureau’s American Community Survey.

The PM$_{2.5}$ data was recorded as daily or hourly measurements from air quality monitors throughout the state, but for our statistical tests we need to convert these values to an annual PM$_{2.5}$ value for each census tract. Diabetes prevalence and socioeconomic variables were already provided in this format. We first averaged annual PM$_{2.5}$ concentrations for each monitoring location and interpolated these concentrations across the state. Annual PM$_{2.5}$ levels for each census tract are shown in the map below. After wrangling all datasets, we retained observations for 5,064 census tracts across California.

**WHAT IS DIABETES?**

**TYPE 1**

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Why</th>
<th>Prevention</th>
<th>Type 2</th>
<th>Why</th>
<th>Prevention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic</td>
<td>Rare</td>
<td>Diabetes</td>
<td>Genetics</td>
<td>Rare</td>
<td>Diabetes</td>
</tr>
<tr>
<td>Obesity</td>
<td>Common</td>
<td>Lifestyle</td>
<td>Obese</td>
<td>Common</td>
<td>Lifestyle</td>
</tr>
<tr>
<td>Stress</td>
<td>Unknown</td>
<td>Management</td>
<td>Stress</td>
<td>Unknown</td>
<td>Management</td>
</tr>
<tr>
<td>Excessive</td>
<td>Unknown</td>
<td>Prevention</td>
<td>Excessive</td>
<td>Unknown</td>
<td>Prevention</td>
</tr>
<tr>
<td>Smoking</td>
<td>Unknown</td>
<td>Cessation</td>
<td>Smoking</td>
<td>Unknown</td>
<td>Cessation</td>
</tr>
</tbody>
</table>

**KEY FINDINGS**

1. We find suggestive evidence of a positive association between average PM$_{2.5}$ concentration and diabetes prevalence that is sensitive to time and cross-sectional units.
2. We find stronger evidence of an association between non-attainment days of PM$_{2.5}$ and diabetes prevalence.
3. We see these same trends reflected in Latino subgroups. We see a larger positive association between non-attainment days of PM$_{2.5}$ and diabetes prevalence in high Latino census tracts compared to low-Latino tracts.

**FUTURE RESEARCH**

We used two different model types to assess possible relationships between PM$_{2.5}$, diabetes prevalence, and socio-demographic indicators.

**METHODS:**

The cross sectional model assesses the relationship between diabetes, PM$_{2.5}$ and socio-demographic variables within each year of study.

**RESULTS:**

We also used the CAPR dataset to calculate the proportion of days a census tract is in non-attainment of the National Ambient Air Quality Standard (NAAQS), which is 12μg/m$^3$. The same two model types as above were run on this dataset. The coefficient association between PM$_{2.5}$ and diabetes prevalence is smaller but still significant positive association between PM$_{2.5}$ and diabetes prevalence is both the cross-sectional and fixed effect models.

**ACKNOWLEDGMENTS**

We would like to extend our deepest thanks to our project partners at San Jose Diabetes Research Institute, especially Nirmala Ditta, David Lin and Mark Smilow. Additionally, we would like to thank our faculty advisors, Ashley Lamer, Olivier Deschesnes and Kyle Meng, as well as our Ph.D. advisor Patrick Lynch (without whom this project would not have been possible). Finally, we would like to thank the James S. Bowker Foundation.