Sansum Diabetes Research Institute (SDRI) has a 75-year tenure conducting research, education, and care in diabetes. Incited by the disproportionate burden of diabetes on Latino families and recognizing that effective diabetes prevention and treatment rest on understanding the determinants of diabetes as well as accurate risk calculation, SDRI launched the Mil Familias project (www.milfamilias.com) in 2018.

A 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 relationships between particulate matter 2.5 (PM$_{2.5}$) and diabetes prevalence throughout the state 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?**

---

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 of Disease Control
3) Sociodemographic 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 needed to convert these values to an average 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 maps below. After wrangling all datasets, we retained observations for 5,084 census tracts across California.

Type 1 diabetes is hereditary, but type 2 is preventable. Well established risk factors contributing to type 2 diabetes include family history, sedentary lifestyle, unhealthy diet, and age.
We used two model types to assess relationships between PM$_{2.5}$, diabetes prevalence, and sociodemographic indicators. In all models, sociodemographic variables included educational attainment, poverty rate, unemployment rate, and race/ethnicity. Here, PM$_{2.5}$ was calculated as a yearly average.

**Cross Sectional Model**

Assesses the relationship between diabetes, PM$_{2.5}$, and sociodemographic variables within each year of study.

**Fixed Effects Model**

Uses the same variables as the cross sectional model, but also controls for time-invariant factors between census tracts.

### Results

There is a significantly positive association between PM$_{2.5}$ and diabetes prevalence in the years 2014, 2015, and 2016.

In models that include data from 2017, there is no association between PM$_{2.5}$ and diabetes.

### Thresholds and Subgroups

We conducted an alternative analysis to explore the possibility of non-linear relationships between PM$_{2.5}$ and diabetes. We calculated PM$_{2.5}$ as the proportion of days in non-attainment of the National Ambient Air Quality Standard (NAAQ), which is 12µg/m$^3$.

**Thresholds**

When we incorporate non-attainment days of PM$_{2.5}$ into the models, there is a significant positive relationship (p<0.001) between PM$_{2.5}$ and diabetes prevalence across all models and years.
Subgroups

Census tracts in California with a high Latino population (>33% Latino of any race) consistently have higher concentrations of PM$_{2.5}$ and diabetes prevalence compared to census tracts with smaller Latino populations.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes</td>
<td>6.63</td>
<td>11.53</td>
<td>8.22</td>
<td>11.19</td>
<td>8.49</td>
<td>11.15</td>
<td>8.4</td>
<td>11.06</td>
</tr>
<tr>
<td>PM$_{2.5}$</td>
<td>10.58</td>
<td>11.90</td>
<td>9.59</td>
<td>10.91</td>
<td>9.48</td>
<td>10.63</td>
<td>10.63</td>
<td>11.53</td>
</tr>
<tr>
<td>Unemployment</td>
<td>5.96</td>
<td>8.50</td>
<td>5.33</td>
<td>7.62</td>
<td>4.69</td>
<td>6.69</td>
<td>4.13</td>
<td>5.83</td>
</tr>
<tr>
<td>Education</td>
<td>8.84</td>
<td>31.71</td>
<td>8.68</td>
<td>31.17</td>
<td>8.57</td>
<td>30.70</td>
<td>8.38</td>
<td>29.88</td>
</tr>
</tbody>
</table>

Future Research

The area of public health that incorporates environmental conditions is a new and growing field of research. There are many opportunities for new development, such as the following:

- Literature studying the relationships between chronic health conditions and environmental factors is not conclusive on the timeline at which pollution can affect the body.
- Diabetes prevalence in the United States differs by race and ethnicity. Additional studies could incorporate interactions between pollution and demographics.

Conclusion

Subgroups

We ran the fixed effects model on both subgroups using PM$_{2.5}$ calculated as yearly averages and as non-attainment days. We see there is a larger positive relationship between PM$_{2.5}$ and diabetes prevalence in heavily Latino-populated census tracts compared to areas with small Latino populations only when we incorporate PM$_{2.5}$ as non-attainment days.

Key Results

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.