# Leveraging the Sociome Data Commons in Interstitial Lung Diseases

Sociome Data Commons

PRESENTER:
**Anthony Solomonides, PhD MSc FAMIA FACMI**

Neha Kaushik, MD[1]; Anthony Solomonides, PhD, MSc[2]; Sandra Tilmon, MS MPH[3]; Sarah Gray, MD[1]; Mary E. Strek, MD[3]; Shashi Bellam, MD[1]; Ayodeji Adegunsoye, MD, PhD[3]; Julian Solway, MD[31]; and the Sociome Data Commons team (*)

[1] Division of Pulmonary and Critical Care, Endeavor Health, Evanston, IL [2] Endeavor Health Research Institute, Evanston, IL [3] Section of Pulmonary and Critical Care, University of Chicago Medicine, Chicago, IL

## Background

- Tobacco smoking: Poor prognosis in Interstitial Lung Disease (ILD)
- Factors associated with worse prognosis in never smoker patients: Unknown
- Residing near major highway: Known environmental hazard
- We investigated the prognostic implications of distance from a major highway in ILD patients
- Hypothesis: Never-smoker patients with ILD living closer to highways have poor survival

## Data

- Never Smoker ILD patients ≥ 18 years diagnosed between 2006-2016
- Demographics
- Baseline pulmonary function
- ILD subtypes
- Oxygen requirement
- NASA: PM 2.5, NO2, Ozone), 2000-2016 zip code average exposure
- Area Deprivation Index (ADI): 1 (least) to 100 (most deprived)
- Distance of patient's residence from major highways

*Highway distance*

- US census feature class distance code:
  - A1: Primary highway with limited access–only accessible via ramps
  - A2: Primary highway without limited access
- Highways and Patient Address Geocoding:
  - Topologically integrated geographic encoding and referencing (TIGER)
  - 2000 U.S. census data
- ArcGIS (v.10, ESRI Inc., Redlands, CA):
  - Nearest possible straight-line distance

## Methods

- Prognosis defined by Overall Survival (OS): Kaplan Meier
- Factors associated with (OS): Cox Proportional Hazards

## Results

| Variables | Highway Distance <200m (n=60) | Highway Distance ≥200m (n=230) | P-Value |
|---|---|---|---|
| Age (Median, IQR) | 75 (68-81) | 73 (65-82) | 0.67 |
| Sex (Females, n (%)) | 36 (60%) | 126 (54.8%) | 0.47 |
| Race (n, %) | | | 0.28 |
| White | 43 (71.7%) | 180 (78.3%) | |
| Non-White | 17 (28.3%) | 50 (21.7%) | |
| BMI (Median, IQR) | 28 (23-31) | 27 (23-31) | 0.64 |
| ILD Subtypes (n,%) | | | 0.48 |
| IPF | 20 (33.3%) | 88 (38.2%) | |
| Non IPF | 40 (66.7%) | 142 (61.8%) | |
| FVC < 80% (n,%) | 42 (70%) | 137 (59.6%) | 0.14 |
| DLCO < 80% (n,%) | 47 (78.3%) | 182 (79.1%) | 0.89 |
| Oxygen Requirement (n,%) | 13 (21.7%) | 67 (29.2%) | 0.24 |
| ILD Gap | 2 (1.5-4) | 2(2-3) | 0.93 |
| PM 2.5 (Median, IQR) | 12.2 (11.8-12.4) | 12.2 (11.9-13.4) | 0.33 |
| NO2 (Median, IQR) | 29.6 (22.2-32.5) | 31.1 (28.2-32.6) | 0.14 |
| O3 (Median, IQR) | 35.2 (34.6-36.2) | 35.1 (34.5-35.3) | 0.2 |
| ADI (Median, IQR) | 26 (16-49) | 22 (9-37) | 0.09 |

**Table 1**: Baseline characteristics by highway distance

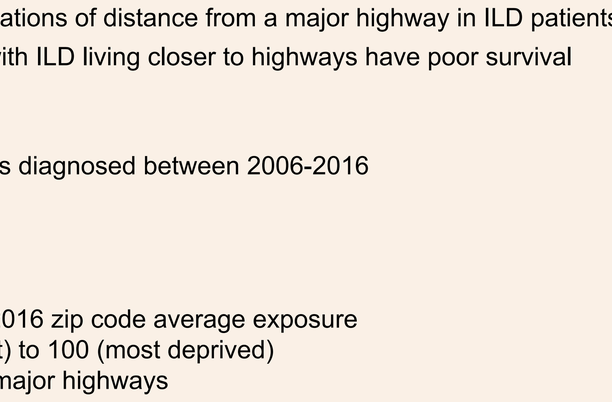


**Figure 1**: Overall survival from diagnosis by highway distance

| Variables | Univariate HR | Univariate 95% CI | Univariate P-value | Multivariate HR | Multivariate 95% CI | Multivariate P-value |
|---|---|---|---|---|---|---|
| Age | 1.05 | 1.03-1.06 | **<0.001** | 1.05 | 1.03-1.06 | **<0.001** |
| Sex (Female vs. Male) | 0.81 | 0.59-1.11 | 0.19 | 0.79 | 0.58-1.09 | 0.16 |
| Race (Non-White vs. White) | 0.70 | 0.46-1.05 | 0.08 | 0.89 | 0.59-1.36 | 0.61 |
| BMI | 0.98 | 0.95-1.01 | 0.22 | | | |
| FVC | 1.19 | 0.86-1.64 | 0.28 | | | |
| DLCO | 1.27 | 0.84-1.91 | 0.24 | | | |
| Oxygen Requirement | 1.53 | 1.09-2.13 | **0.01** | 1.96 | 1.39-2.76 | **<0.001** |
| Highway Distance < 200m | 1.36 | 0.97-1.89 | 0.06 | 1.46 | 1.04-2.05 | **0.03** |
| ILD Subtype (IPF vs. Non-IPF) | 1.71 | 1.24-2.36 | **0.001** | 1.53 | 1.08-2.16 | **0.02** |
| PM 2.5 | 1.08 | 0.77-1.51 | 0.64 | | | |
| NO2 | 1.01 | 0.98-1.04 | 0.43 | | | |
| O3 | .96 | 0.81-1.15 | 0.7 | | | |
| ADI | 1.002 | 0.99-1.01 | 0.66 | | | |

**Table 2**: Univariate and Multivariate Cox Proportional Hazards results

## Discussion

- First study to demonstrate elevated mortality in ILD patients residing near major highways
- Findings hold implications for policymakers and environmental health
- Limitations
  - Small Sample Size
  - Retrospective Design
  - Single Center Study
  - Findings need future validation and inquiries into causality

## Definitions

- **Interstitial Lung Disease** (ILD): a collective term for a spectrum of severe pulmonary disorders, many causing progressive fibrosing of lung tissue, affecting breathing and blood oxygenation
  - Fibrosis and scarring largely irreversible.
  - Causes include toxic exposures and autoimmune conditions
- Idiopathic Pulmonary Fibrosis (IPF): a significant fraction of ILD where the cause or causes remain unknown ( "idiopathic.")
  - Most severe and least tractable form of ILD
  - IPF accounts for 35-40% of ILD
  - The principal target in this study of factors that promote progression.

## Principal hypothesis

Non-clinical determinants of health—including social, environmental, behavioral, psychological, and economic factors, collectively, "the sociome"—play critical roles in shaping patient outcomes. In patients with ILD – and especially those with (IPF) – both higher neighborhood-level disadvantage as well as closer proximity to highways, are associated with worse baseline lung function, more rapid lung function decline, reduced odds of lung transplantation, and increased mortality.

## Data

- Endeavor Health and University of Chicago Medicine data to be combined to increase sample size and geographic representation.
- Common data dictionary established
  - Logical phenotypes for IPF
- IRB approval
- Data combined by a third party honest broker
  - Anonymization and deidentification
    - All PII replaced by one-way hash function-created "de-identifiers"—irreversible codes uncorrelated with the real identifiers
  - After calculation of patient distances to pollution and highway sites, all location information to be removed

## Community participation

- Pulmonary Fibrosis community interaction to be guided by the PCORI Dissemination & Implementation Framework.6
- An advocacy organization (e.g., Pulmonary Fibrosis Foundation) wukk consult on and refine the goals of the study.
- Institute for Translational Medicine (ITM) Community Navigators will facilitate these and other community links

## Methods

- Validation of the IPF phenotype via random sample chart review
- Rigorous statistical analysis of disease progression (including increasing fatigue, decline in forced vital capacity (FVC), need for oxygen, mental health issues) and exposure to SDC factors.
  - Primary outcomes: morbidity, mortality and survival times by the selected SDC factor.
  - Secondary outcomes: association of baseline lung function (FVC and DLCO) with SDC factor.
- Survival analyses
  - Survival time: time from diagnosis to death from any cause. Censoring at the date of the last follow-up for surviving patients
  - Visualization of survival distributions across groups via, Kaplan-Meier estimates and differences assessed via log-rank test
  - Hazard ratios: Cox Proportional Hazards (Cox PH), which also permits multivariate analysis
    - If the proportional hazards assumption is violated, alternatives include: Aalen additive regression model, Cox with time-varying covariates, and restricted mean survival time (RMST)
- Causal analyses
  - G-formula adjusts for the time-varying metrics and uses fitted Cox PH models to calculate the counterfactual of a hypothetical outcome under a different treatment scenario
  - Estimates causal effect by comparing simulated outcomes under different interventions via estimating the change in the hazard under a hypothetical intervention

## Dissemination

- Community dissemination of lessons learned
- Publication

## BACKGROUND

Sociome: Non-clinical aspects of life affecting health - social, environmental, behavioral, psychological, and economic factors

The Sociome Data Commons (SDC)

- Developed to facilitate integrating sociome factors into researchers' work by reducing data discovery and management
- Collects and curates high-quality and generalizable data ranging from precise location (industrial and health care sites) to census tract level estimates (PM 2.5)
  - Traditional SDoH (e.g., food insecurity) but also aims to encompasses data about the full social context of disease
- Curated census-tract level datasets organized by theme (social, environmental, behavioral, psychological, and economic)
  - Can be linked to patient's home address census tract
- Extends beyond averaged estimates, enabling high-resolution analyses such as:
  - Calculating precise distances to pollution sources
  - Processing satellite imagery at fine geographic scales, including individual residences.
- Preliminary analyses of the distribution of asthma exacerbations in a UChicago Medicine population; this work continues.
- With ILD, described above.

(*) The **SOCIOME DATA COMMONS** team
Sandra Tilmon [1], Sharmilee Nyenhuis[1 2], Anthony Solomonides [3], Bruno Barbarioli [4], Ankur Bhargava[5], Suzi Birz [1], Kathryn Bouzein [1], Celine Cardenas [6], Bradley Carlson [7], Ellen Cohen [1], Emily Dillon [8], Brian Furner [1], Zhong Huang [7], Julie Johnson [9], Nivedha Krishnan [10], Kevin Lazenby [7], Kaitlyn Li [11], Sonya Makhni [5], Doriane Miler [2], Jonathan Ozik [12], Carlos Santos [13], Marc Sleiman [7], Julian Solway [2], Sanjay Krishnan [4], Samuel Volchenboum [1]

[1]Pediatrics, University of Chicago, Chicago, IL, USA; [2]Medicine, University of Chicago, Chicago, IL, USA; [3]Endeavor Health, Research Institute, Evanston, IL, USA; [4]Computer Science, University of Chicago, Chicago, IL, USA; [5]Chicago Medicine, Chicago, IL, USA; [6]Wake Forest University, Winston-Salem, NC, USA; [7]Pritzker School of Medicine, University of Chicago, Chicago, IL, USA; [8]Psychiatry and Behavioral Sciences, Rush University Medical Center; Chicago, IL, USA; [9]Clinical Research Informatics, University of Chicago, Chicago, IL, USA; [10]University of Illinois at Chicago, Chicago, IL, USA; [11]University of Chicago, Chicago, IL, USA; [12]Decision and Infrastructure Sciences Division, Argonne National Laboratory, Lemont, IL, USA; [13]Internal Medicine, Rush University Medical Center, Chicago, IL, USA.

**Reference**
Tilmon S, Nyenhuis S, Solomonides A, et al. Sociome Data Commons: A scalable and sustainable platform for investigating the full social context and determinants of health. *J Clin Transl Sci*. 2023;7(1):e255. Published 2023 Nov 7. doi:10.1017/cts.2023.670

Grant support: NIH UL1TR002389-07