

Investigating the Impact of Social Determinants of Health on Diagnostic Delays and Access to Antifibrotic Treatment in Idiopathic Pulmonary Fibrosis

Rui Li Liwei Wang (presenter)



Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive lung disease characterized by the thickening and scarring (fibrosis) of lung tissue.

Challenges:

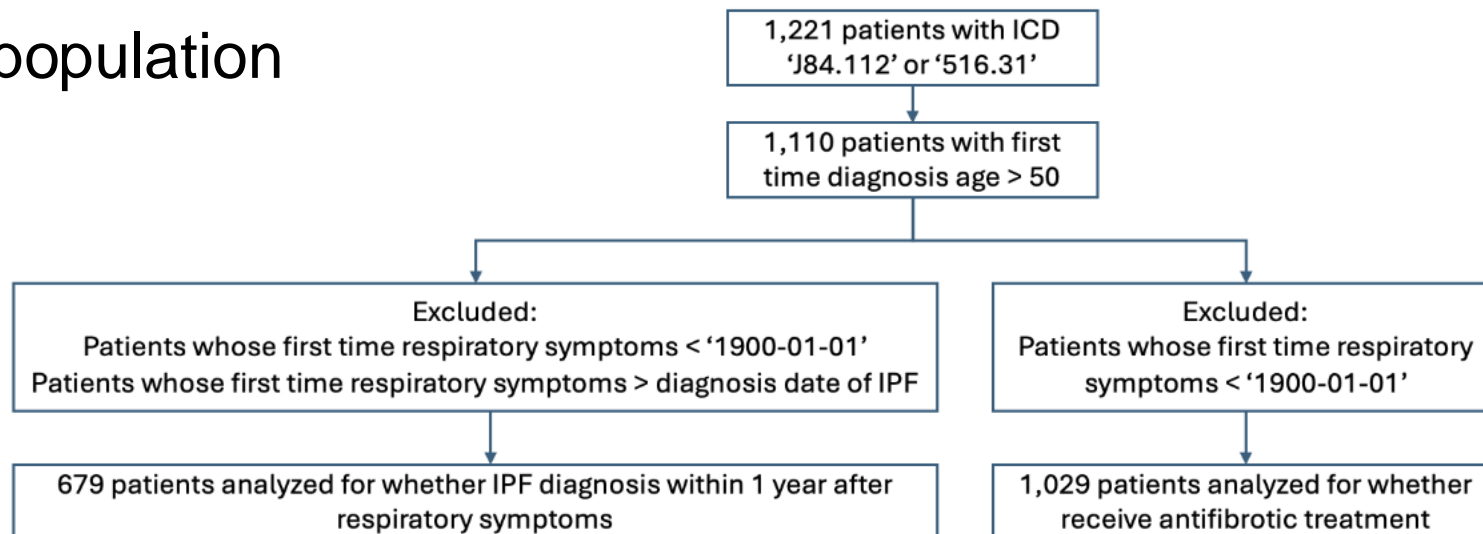
- Long diagnostic process. Its symptoms are similar to those of other lung diseases, its rarity and gradual progression make early detection difficult, and there are currently no reliable biomarkers.
- Hard to treatment. FDA has approved two drugs for the treatment of IPF nintedanib and pirfenidone. But researches show that only 10.3% of patients with IPF were treated with an antifibrotic treatment.

Method

- Data source

We use UTHealth OMOP CDM dataset. It contains EHR data from the outpatient practice of the University of Texas Health Sciences Center at Houston's McGovern Medical School.

- Study population



Method

- Ascertainment of clinical related outcomes
 - time-to-diagnosis: the time to IPF diagnosis after the initial onset of respiratory symptoms. Binary classification, 0 indicated the diagnosis took longer than one year; 1 indicated the diagnosis occurred within one year.
 - whether the patient received antifibrotic treatment. Binary classification, 1 indicated the patient received antifibrotic therapy; 0 indicated the patient did not.

Method

- classification models

- Logistic regression:

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \text{ where } z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_n X_n.$$

$\beta_0, \beta_1, \cdots \beta_n$ are model coefficients, and $X_1, X_2, \cdots X_n$ are input features.

- XGBoost:

It constructs an ensemble of decision trees in a sequential manner, where each subsequent tree focuses on correcting the errors made by the previous ones.

Method

- Ascertainment of SDoHs from HHS



Economic
stability



Healthcare access
and quality



Education access
and quality



Neighborhood and
build environment



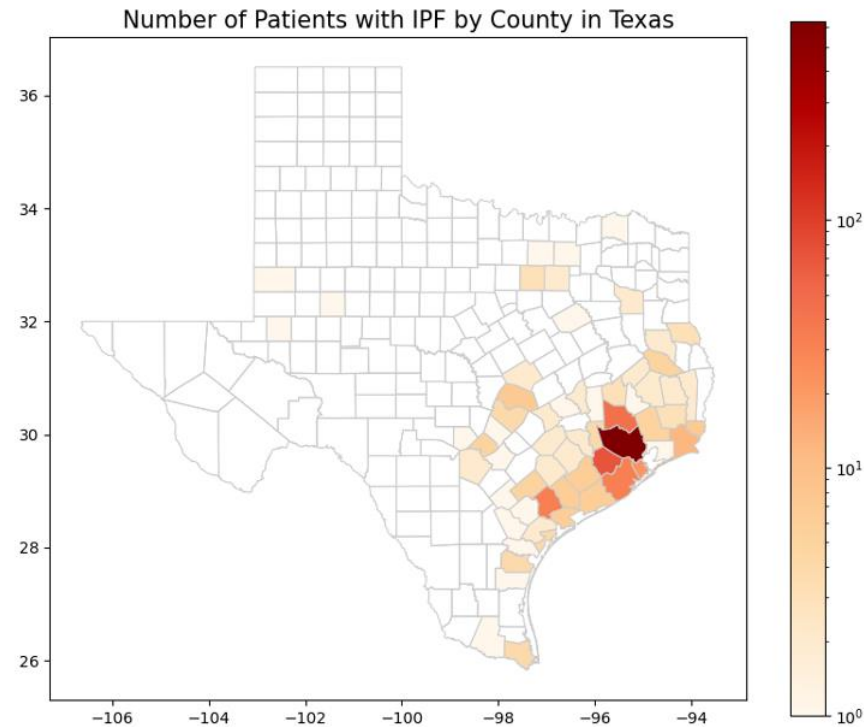
Social and comm-
unity context

Four SDoH features: household income, the percentage of individuals with a bachelor's degree or higher, health insurance status, PM2.5 levels.

The patient's zip code, age, gender and race are used to extract area-specific averages from the American Community Survey, accessed via the U.S. Census Bureau's API. These demographic-specific averages were then assigned to each patient as representative SDoH data.

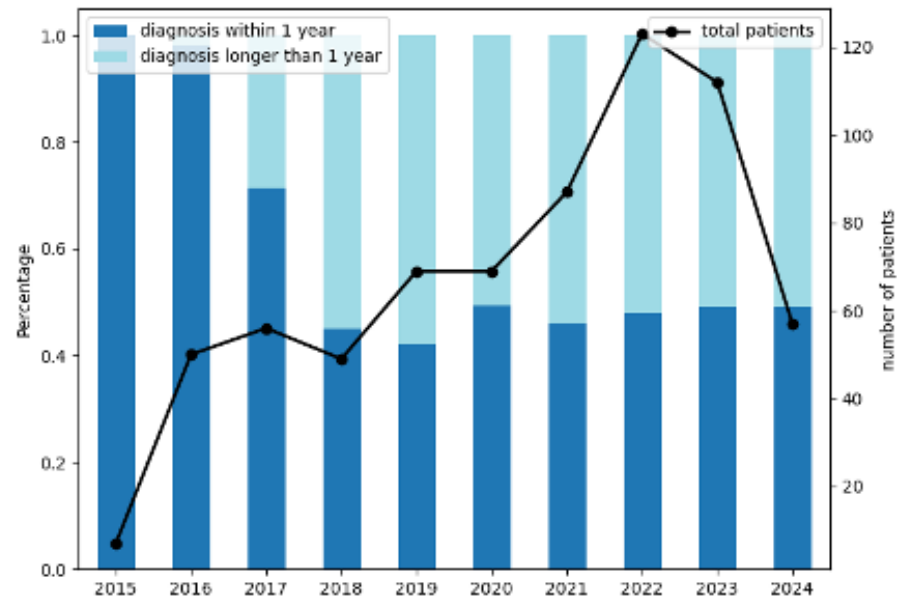
Statistics

- The county-level distribution of patients with IPF in Texas

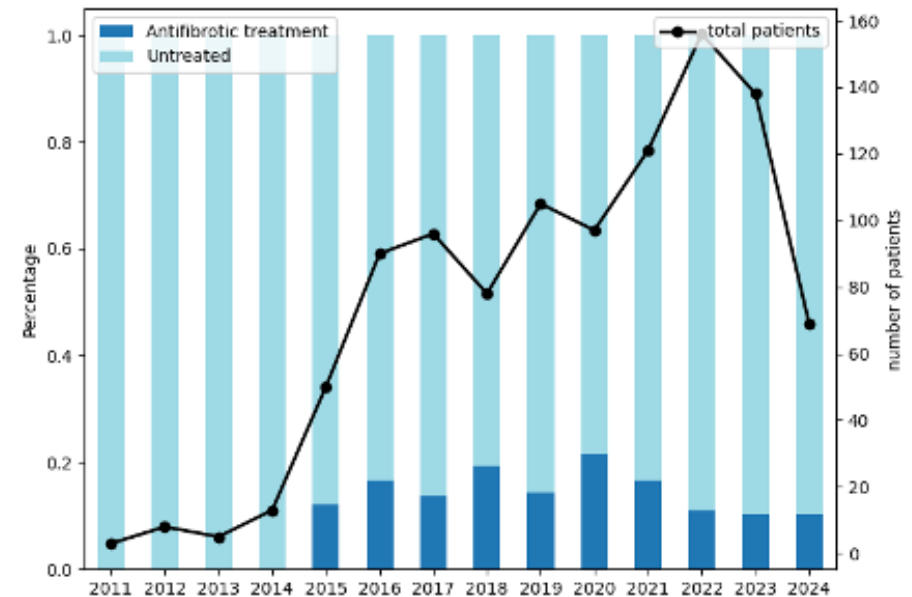


Statistics

- Trends of Treatment/Diagnosis Time



(a) Trend of time-to-diagnosis



(b) Trend of receiving antifibrotic treatment

Statistics

Table 1: Demographics and SDoH of Studied Patients

	whether IPF diagnosis within 1 year after respiratory symptoms				whether receive antifibrotic treatment			
	within 1 year <= 1 year (N=363)	after 1 year > 1 year (N=316)	Total (N=679)	p value	Treated with anti- fibrotic (N=143)	Untreated (N=886)	Total (N=1029)	p value
Age				0.598				0.892
Mean (SD)	70.5 (9.4)	70.2 (9.3)	70.4 (9.3)		72.0 (7.9)	70.1 (9.7)	70.4 (9.5)	
Median	70.7	70	70.3		71	70	70	
Age Group				0.965				0.002
Less than 55	16 (4.4)	13 (4.2)	29 (4.3)		3 (2.1)	46 (5.2)	49 (4.8)	
55-64	84 (23.2)	80 (25.3)	164 (24.2)		19 (13.3)	230 (25.9)	249 (24.2)	
65-74	148 (40.7)	129 (40.8)	277 (40.8)		71 (49.6)	345 (38.9)	416 (40.4)	
74-84	91 (25.1)	75 (23.7)	166 (24.4)		42 (29.4)	206 (23.3)	248 (24.1)	
Over 85	24 (6.6)	19 (6.0)	43 (6.3)		8 (5.6)	59 (6.7)	67 (6.5)	
Gender				<0.001				0.077
Female	140 (38.6)	166 (52.5)	306 (45.1)		53 (37.1)	402 (45.4)	455 (44.3)	
Male	223 (61.4)	150 (47.5)	373 (54.9)		90 (62.9)	484 (54.6)	574 (55.7)	
Race/Ethnicity				0.298				0.024
White	160 (44.1)	120 (38.0)	280 (41.3)		81 (56.6)	374 (42.2)	455 (44.3)	
Hispanic or Latino	82 (22.6)	84 (26.6)	166 (24.4)		21 (14.7)	185 (20.9)	206 (20.0)	
Black or African American	47 (12.9)	53 (16.8)	100 (14.7)		13 (9.1)	129 (14.5)	142 (13.8)	
Asian	11 (3.0)	11 (3.4)	22 (3.3)		4 (2.8)	29 (3.3)	33 (3.2)	
No matching	63 (17.4)	48 (15.2)	111 (16.3)		24 (16.8)	169 (19.1)	193 (18.7)	
Household Income				0.162				0.006
Mean (SD)	82,455 (37869)	78,339 (38513)	80,539 (38197)		89,187 (42042)	78,916 (35829)	80,344 (36905)	
Median	71,795	72,284	71,858		77,794	72,284	72,284	
Income Group				0.010				0.046
Low class (<30,000)	0 (0.0)	8 (2.5)	8 (1.2)		1 (0.7)	14 (1.6)	15 (1.4)	
Lower-middle class (30,000-58,020)	106 (29.2)	100 (31.6)	206 (30.3)		29 (20.3)	264 (29.8)	293 (28.5)	
Middle class (58,021-94,000)	149 (41.0)	137 (43.5)	286 (42.1)		66 (46.1)	389 (43.9)	455 (44.2)	
Upper-middle class (94,001-153,000)	88 (24.3)	57 (18.0)	145 (21.4)		35 (24.5)	180 (20.3)	215 (20.9)	
Upper class (≥153,000)	20 (5.5)	14 (4.4)	34 (5.0)		12 (8.4)	39 (4.4)	51 (5.0)	
Education				0.071				0.011
Bachelor's degree or higher percent	30.7	27.6	29.3		33.9	28.5	29.2	
Insurance status				0.181				<0.001
coverage percent	93.1	92.1	92.6		95.6	92.2	92.6	
PM 2.5 exposure				0.998				0.205
Mean (SD)	10.4 (0.4)	10.4 (0.4)	10.4 (0.4)		10.3 (0.6)	10.4 (0.4)	10.3 (0.4)	
Median	10.5	10.5	10.5		10.5	10.5	10.5	

Experiment

- Classification performance evaluation
 - F1 score, average precision, Cohen's kappa

Table 2: Classification Performance

	whether IPF diagnosis within 1 year after respiratory symptoms		whether receive anti- fibrotic treatment	
	logistic regression	XGBoost classification	logistic regression	XGBoost classification
F1	0.605	0.600	0.268	0.250
AP	0.510	0.619	0.177	0.262
Cohen's Kappa	0.107	0.109	0.064	0.187

Experiment

- Association between SDoH and clinical outcomes in logistic regression.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad , \text{ where } z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_n X_n .$$

$\beta_0, \beta_1, \cdots \beta_n$ are model coefficients, and $X_1, X_2, \cdots X_n$ are input features.

The **coefficients** quantify the impact of each feature on the log-odds of the outcome: a positive coefficient increases the log-odds, while a negative coefficient decreases it. The larger magnitude indicates stronger influence.

odds ratio $= e^{\beta}$ quantifies how the odds of the outcome change with a one-unit increase in the corresponding feature. If >1 , indicates a positive association between the input and the outcome.

Experiment

- Associations between SDoH and Clinical Outcomes in Logistic Regression

	whether IPF diagnosis within 1 year after respiratory symptoms		whether receive anti- fibrotic treatment	
	Coefficient	Odds Ratio	Coefficient	Odds Ratio
age	-0.089	0.915	0.021	1.021
income	-0.199	0.819	0.062	1.064
education	0.310	1.363	0.007	1.007
insurance	0.174	1.190	0.331	1.392
PM2.5_mean	-0.182	0.833	-0.1	0.905
Gender				
Female	-0.309	0.734	-0.169	0.844
Male	0.309	1.363	0.170	1.185
Race				
White	-0.075	0.927	0.285	1.330
Hispanic or Latino	-0.003	0.997	-0.240	0.787
Black or African American	-0.046	0.955	-0.049	0.952
Asian	-0.222	0.801	0.057	1.059
No matching concept	0.347	1.415	0.053	0.948

Experiment

- Association between SDoH and clinical outcomes in XGBoost classification.

XGBoost constructs an ensemble of decision trees in a sequential manner. It achieves better performance but lacks direct interpretability in terms of coefficients.

Relative risk: $RR = \frac{P(Y = 1|X)_{in\ compared\ group}}{P(Y = 1|X)_{in\ reference\ group}}$.

If $RR > 1$, the probability of predicting label 1 is higher in the comparison group than in the reference group.

Experiment

- Association between SDoH and clinical outcomes in XGBoost classification.

	whether IPF diagnosis within 1 year after respiratory symptoms	whether receive anti- fibrotic treatment
	risk ratio (95% CI)	risk ratio (95% CI)
Age Group		
Less than 55	0.628 (0.169-1.349)	0.136 (0.028-0.326)
55-64	0.709 (0.529-0.934)	0.553 (0.111-1.365)
65-74	1.000 (reference)	1.000 (reference)
74-84	1.091 (0.887-1.331)	1.314 (0.465-2.846)
Over 85	0.706 (0.372-1.096)	0.762 (0.048-2.451)
Gender		
Female	1.000 (reference)	1.000 (reference)
Male	1.477 (1.205-1.832)	1.766 (0.701-4.576)
Race/Ethnicity		
White	1.054 (0.818-1.399)	4.849 (1.028-22.948)
Hispanic or Latino	0.795 (0.545-1.093)	1.281 (0.187-5.644)
Black or African American	1.000 (reference)	1.000 (reference)
Asian	0.648 (0.515-0.851)	2.102 (0.0167-10.626)
no matching	0.942 (0.651-1.289)	1.749 (0.089-10.039)
Income Group		
Low class (<30,000)	0.121 (0.091-0.156)	0.078 (0.022-0.182)
Lower-middle class (30,000-58,020)	0.858 (0.681-1.065)	0.809 (0.198-1.961)
Middle class (58,021-94,000)	1.000 (reference)	1.000 (reference)
Upper-middle class (94,001-153,000)	1.343 (1.094-1.627)	2.362 (0.764-5.787)
Upper class (>153,000)	1.121 (0.769-1.542)	4.553 (0.166-11.686)
Education		
low education level	1.000 (reference)	1.000 (reference)
high education level	1.233 (1.024-1.471)	1.908 (0.751-4.315)
Insurance status		
low insurance level	1.000 (reference)	1.000 (reference)
high insurance level	1.291 (1.056-1.575)	1.696 (0.724-3.508)
PM 2.5 status		
low PM2.5 level	1.000 (reference)	1.000 (reference)
high PM2.5 level	0.818 (0.684-0.981)	0.743 (0.333-1.541)

Discussion

- Contribution

- We explore the relation between SDoHs and the diagnosis and treatment of IPF.
- We extract demographic-specific averages from zip code-level data using the American Community Survey (via the U.S. Census Bureau API) as proxies for individual SDoH characteristics. Comparing with composite measures such as the Area Deprivation Index (ADI), our approach allows for a more **fine-grained** investigation of the associations between individual SDoH features and clinical outcomes.
- We build two machine learning models—logistic regression and XGBoost classification—to examine the impact of SDoH variables on clinical outcomes.

Discussion

- Limitation

- limited Cohort Size: only 1,029 patients in the analysis of antifibrotic treatment and 679 patients in the analysis of time-to-diagnosis.
- Single-Registry Data: The analysis is based solely on data from UTHealth OMOP CDM, a single registry. This could introduce bias.

Thanks!

The code is released at: https://github.com/ruilialice/IPF_SDOH/tree/main?tab=readme-ov-file