(I²DB)

Causal Machine Learning for Real-world Evidence Generation

Linying Zhang, PhD

Assistant Professor of Biostatistics and Medicine

Institute for Informatics, Data Science and Biostatistics (I2DB)

Department of Medicine

Washington University School of Medicine in St. Louis

CSE 4109 Introduction to AI for Health

Nov. 20, 2025

Washington University School of Medicine in St.Louis

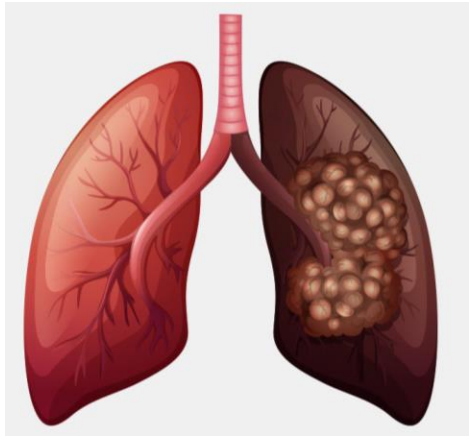
What makes a model or finding “causal”?

Does carrying lighters cause lung cancer?



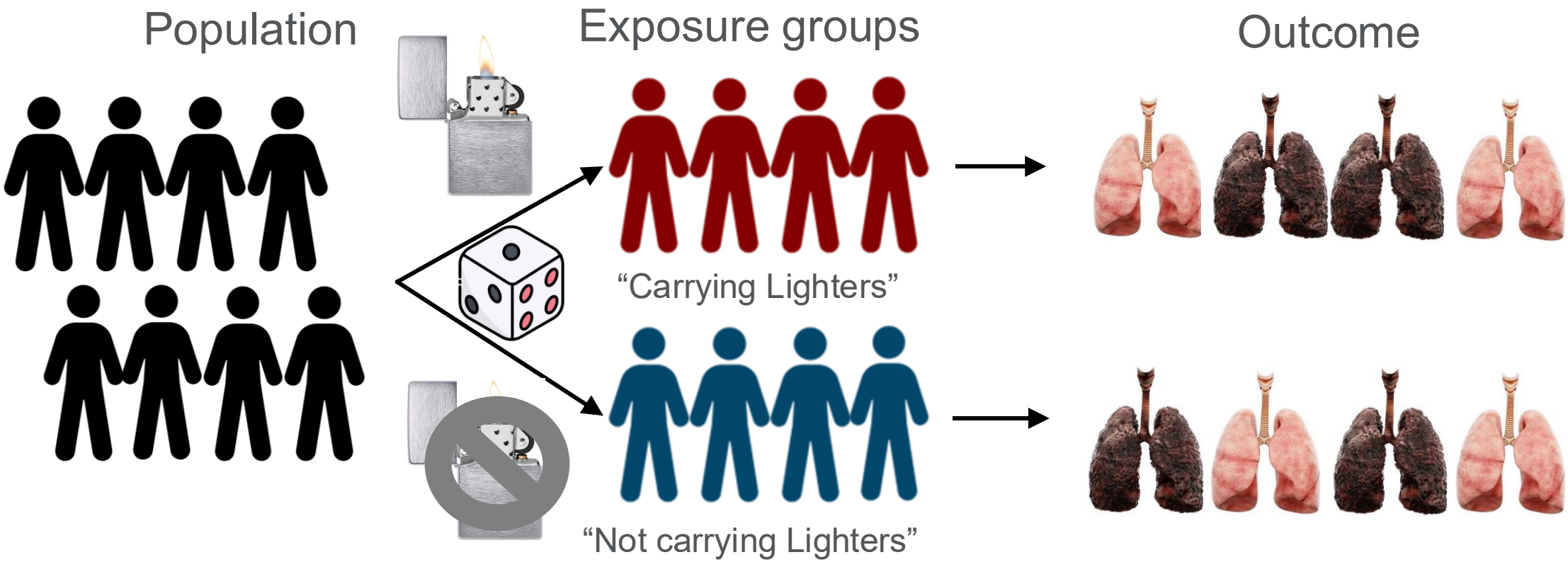
Exposure





Outcome

Randomized controlled trial ("gold standard")

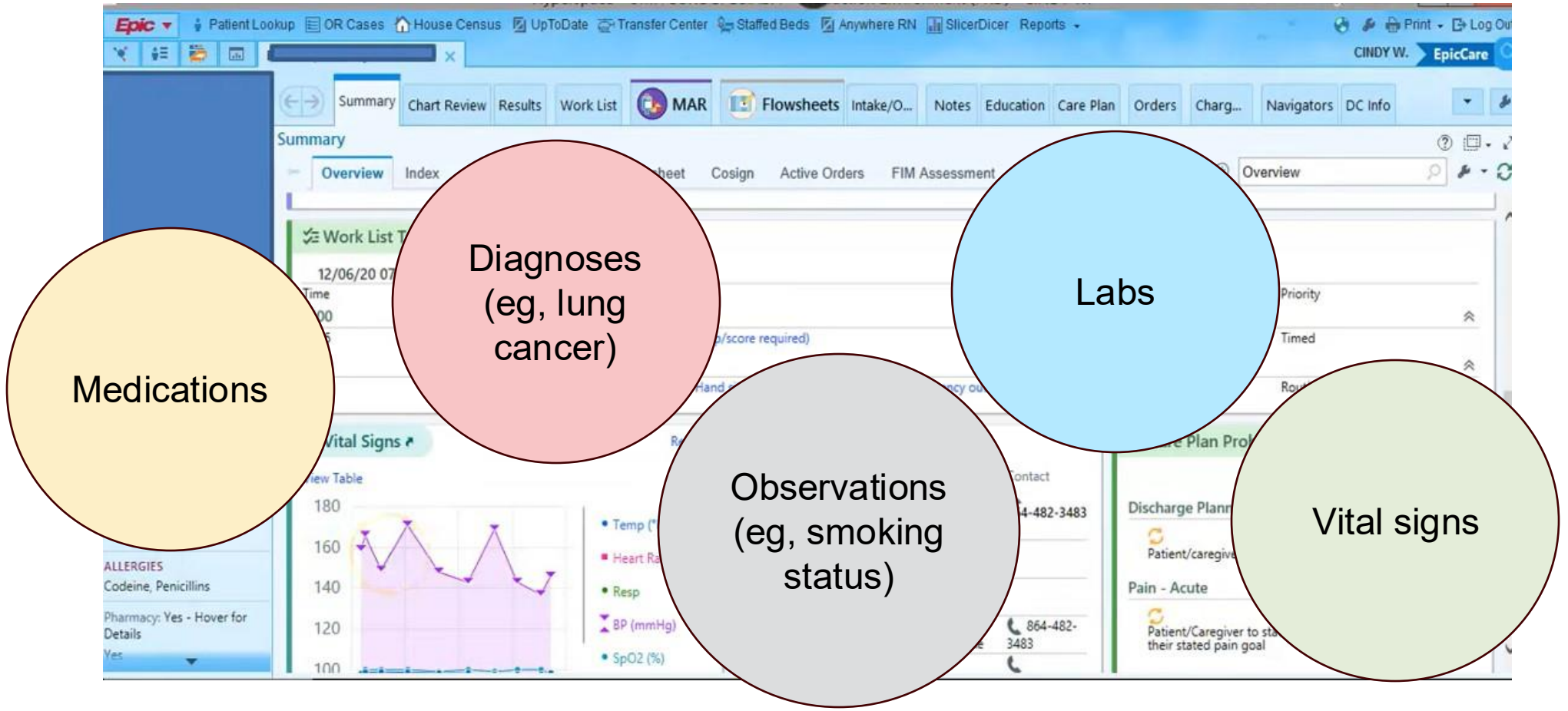


But, no one would fund such a randomized trial ☹

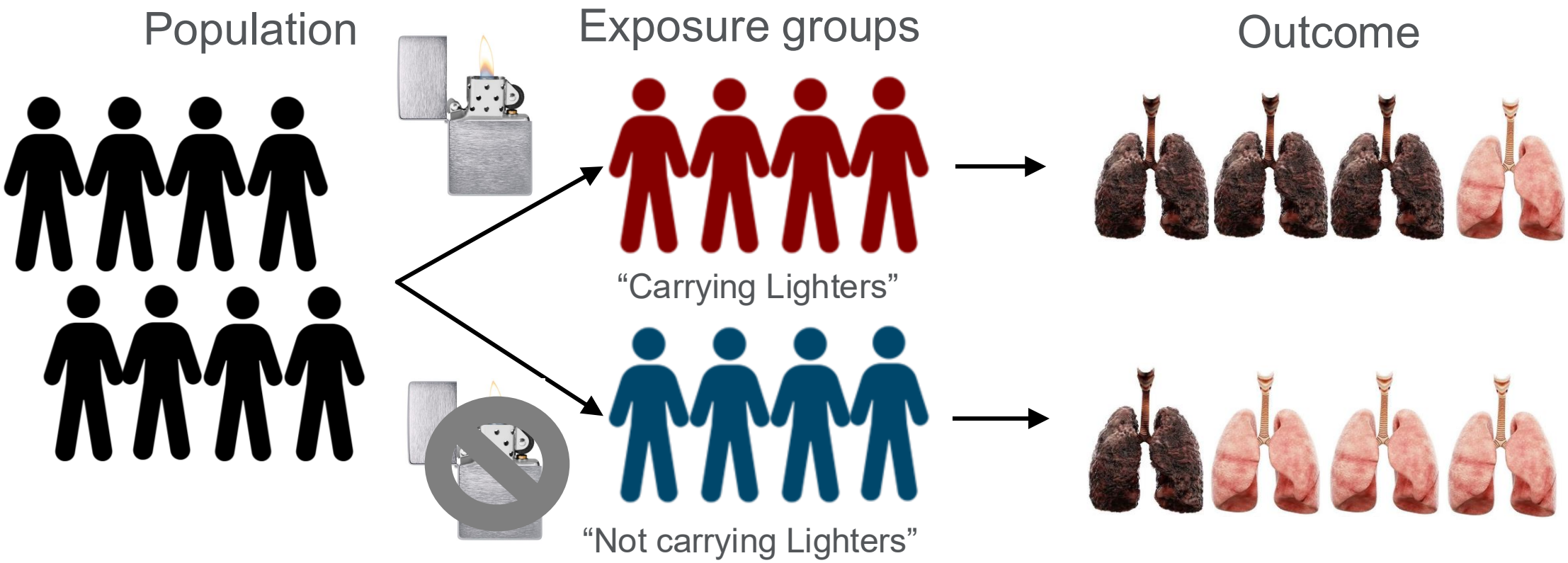
Can we answer the same causal question without a randomized trial?

- **Real-world data,** also known as **observational data,** refer to data generated in non-experimental settings, such as electronic health records (EHRs)

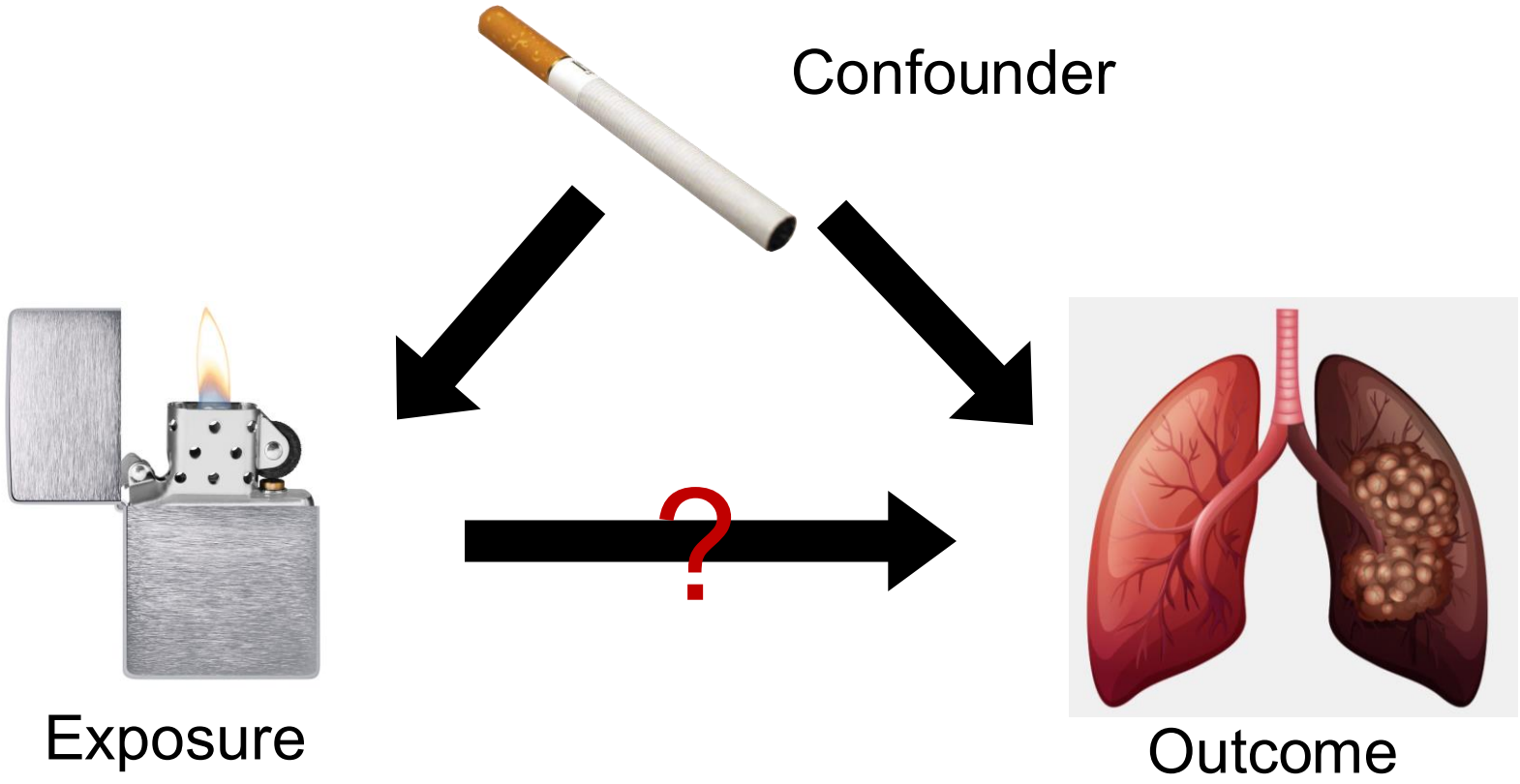
Electronic Health Record (EHR)



Can you get causal answer by comparing the lighters vs no lighters groups from an EHR database?



Confounding bias in observational data



Confounders can exist in any data modalities



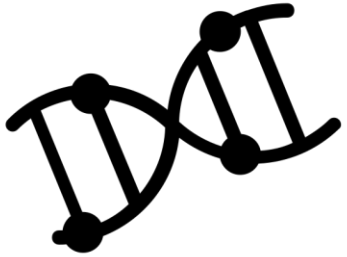
Electronic Health Records (EHRs)



Medical Images



Clinical Notes



Genomics



Demographics Diagnoses Drugs Procedures Labs



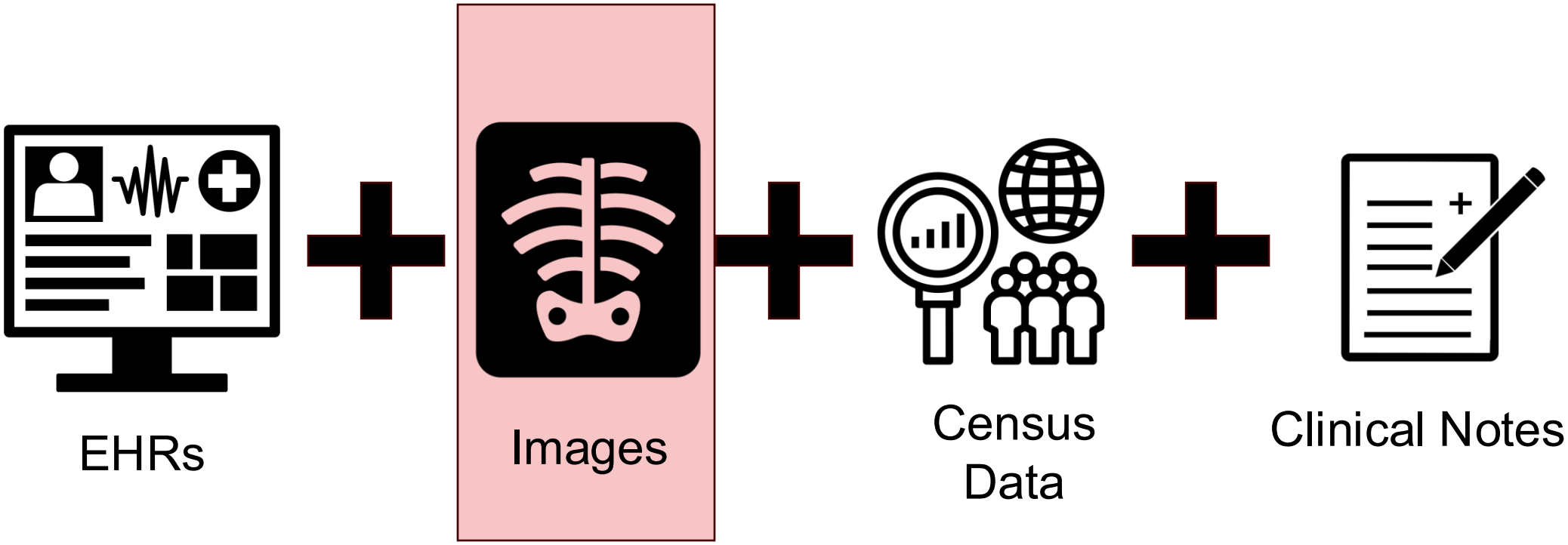
Wearables



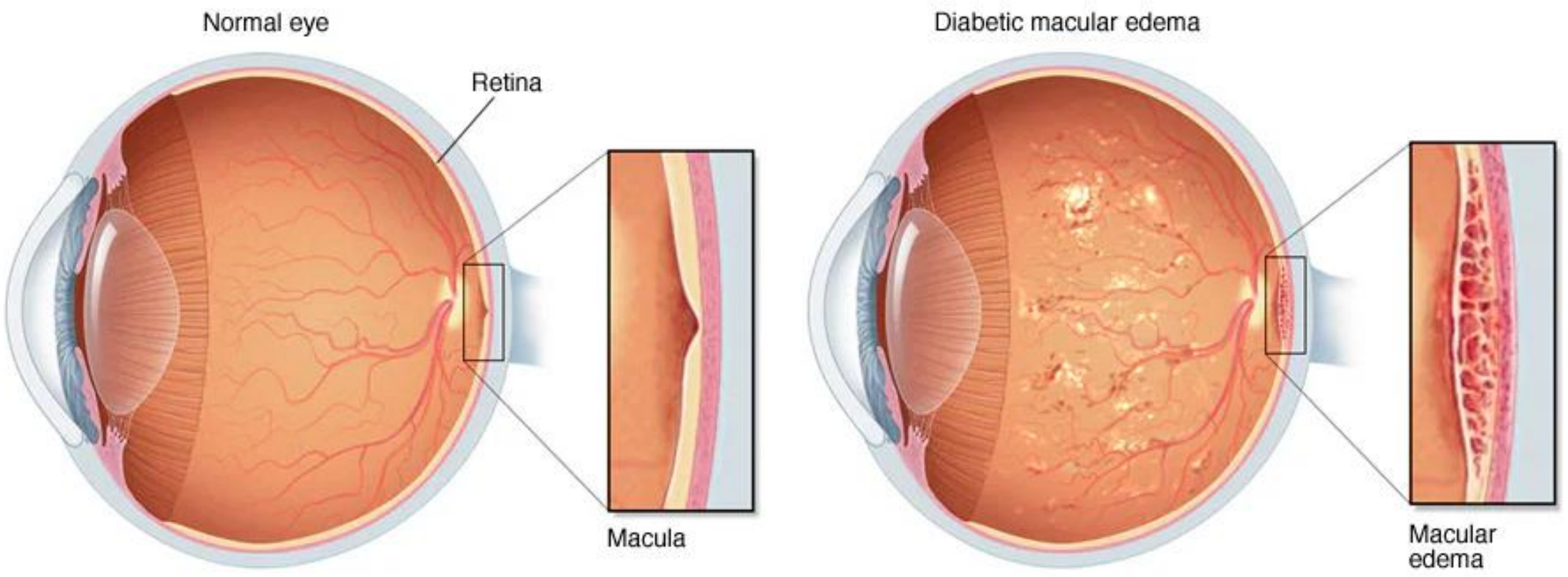
Surveys

(I²DB)

Multi-modal Causal Learning for Health Care



Diabetic Macular Edema (DME)



© MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH. ALL RIGHTS RESERVED.

Optical Coherence Tomography (OCT)



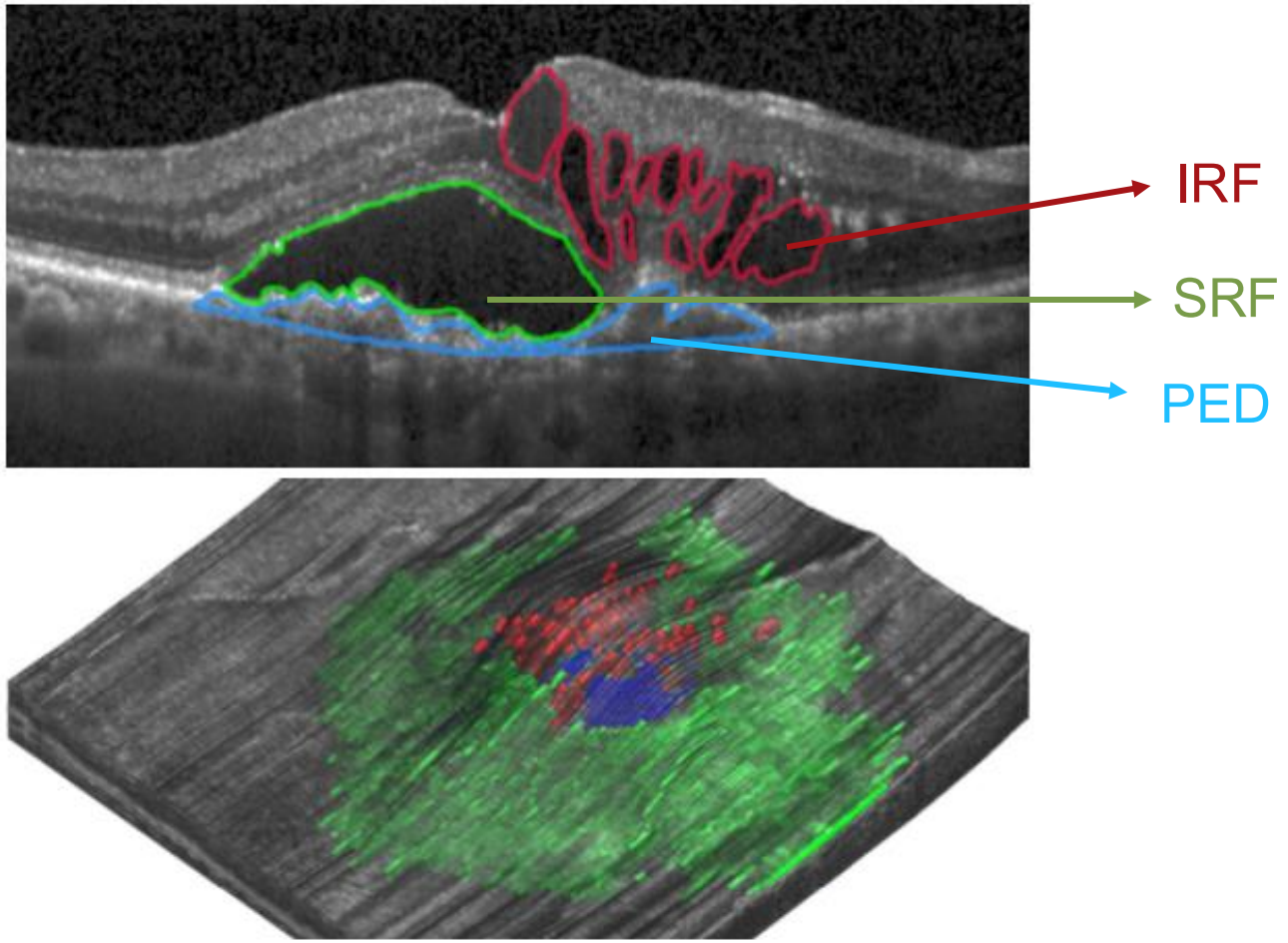
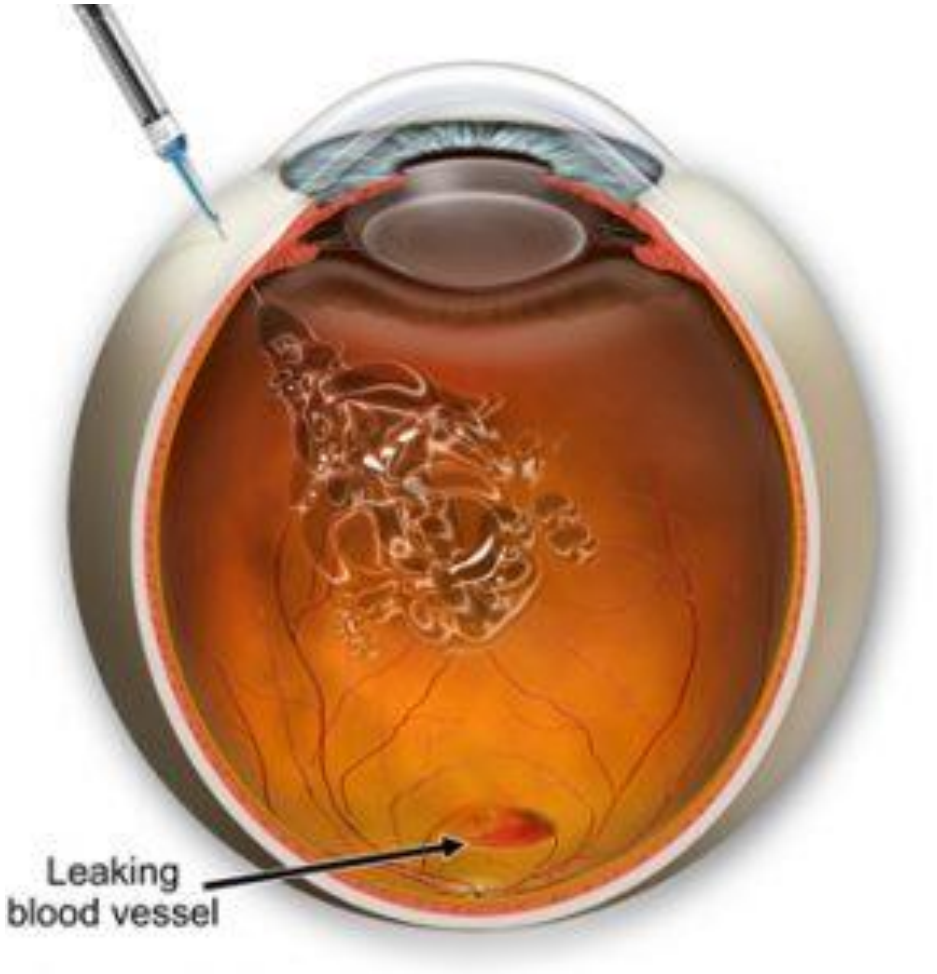


Fig. 2. The three fluid types on a 2D B-scan (above) and as a 3D volume rendering (below): IRF (red), SRF (green) and PED (blue).

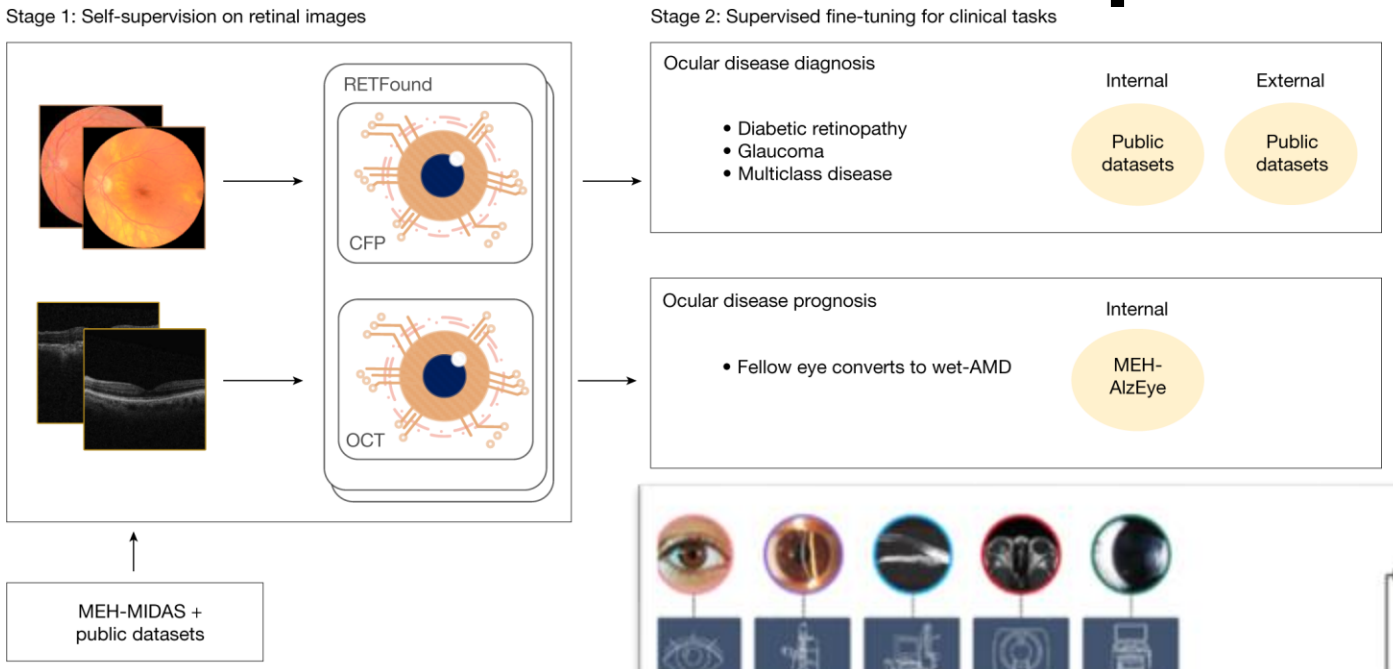
Intravitreal Anti-VEGF Injections



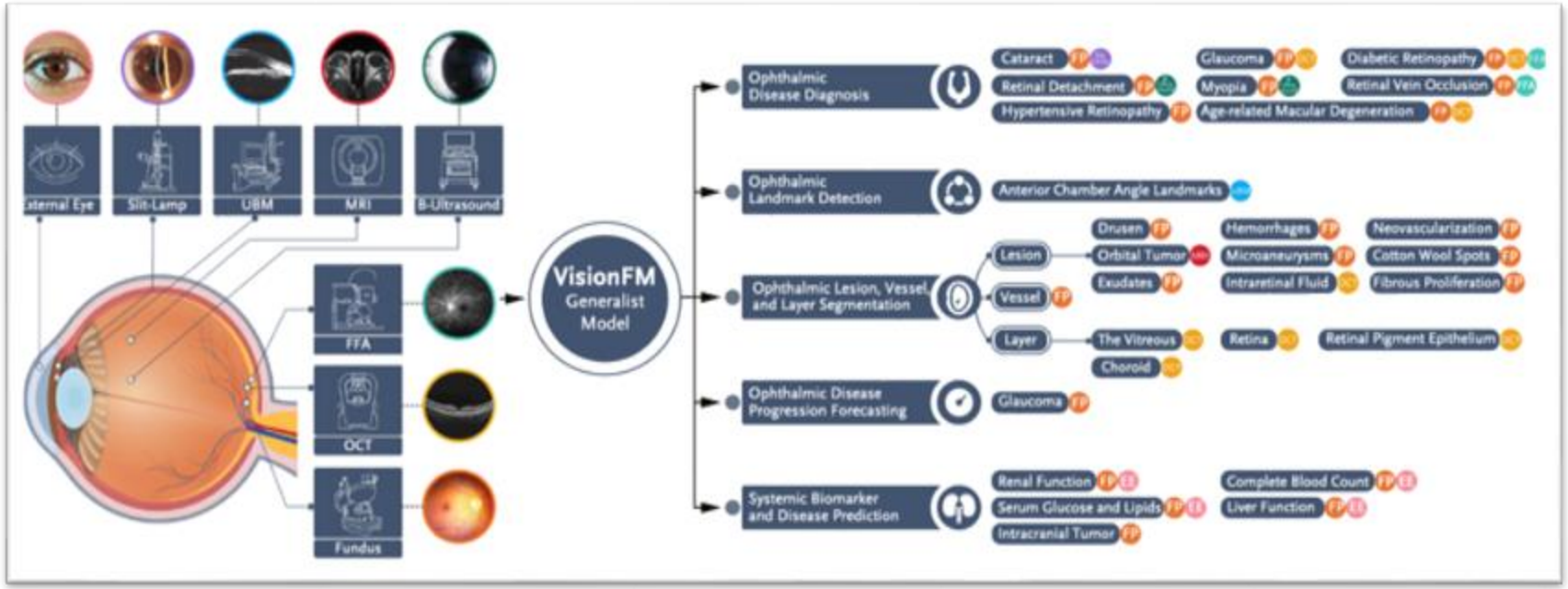
- There are 3 variations of anti-VEGF injections
  - Aflibercept
  - Bevacizumab
  - Ranibizumab

- **Question**: Are the three anti-VEGFs equally effective in reducing vision loss?

Foundation Models in Ophthalmology



RETFound (*Nature* 2023)



VisionFM (*NEJM AI* 2024)

Retinal fluid segmentation model

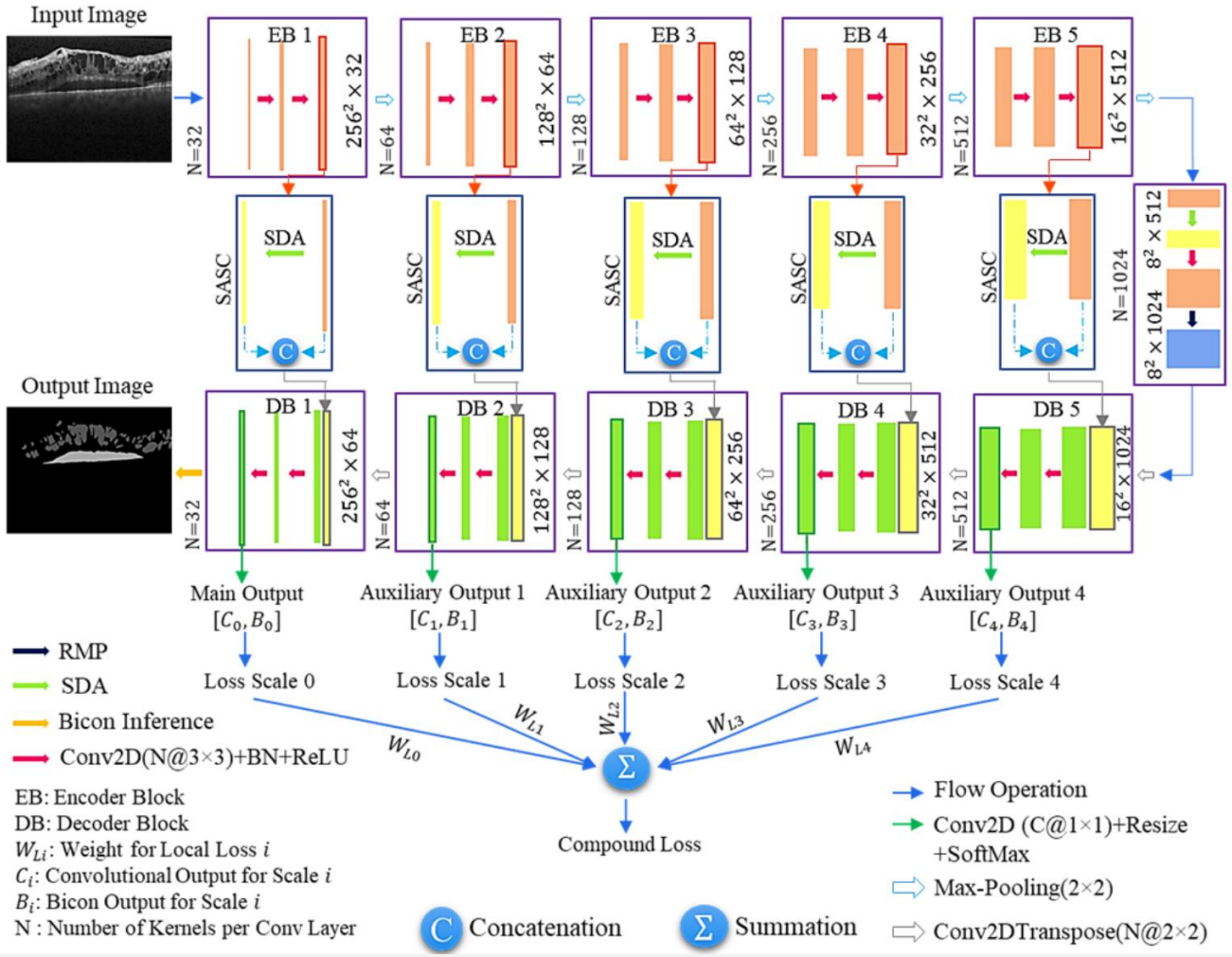


Figure: RetiFluidNet (Rasti et al 2022): U-Net with self-adaptive dual-attention.

(I²DB)

Comparing Foundation and Task-specific Models for Retinal Fluid Segmentation

Dataset:
RETOUCH OCT [1]

- Spectralis: 24 volumes of 49 B-scans, 512×496 pixels
- Cirrus: 24 volumes of 128 B-scans, 512×1024 pixels

Evaluation Metrics:

- Accuracy
- Balanced accuracy
- Dice similarity coefficient
- Intersection-over-Union (IoU)
- Intraclass Correlation Coefficient



Xiaoyu (Klay) Sun

Washington University School of Medicine in St. Louis

Results

RetiFluidNet (RNF)

	Accuracy	Balanced Accuracy	Dice	IoU
Background	99.49±0.17	91.17±1.44	99.74±0.09	99.48±0.17
IRF	99.75±0.09	94.04±0.29	80.67±3.14	76.3±3.37
SRF	99.87±0.08	96.84±1.32	89.83±2.75	87.26±3.99
PED	99.85±0.04	95.4±1.34	86.17±3.49	83.95±3.68

VisionFM (VFM)

	Accuracy	Balanced Accuracy	Dice	IoU
Background	99.29±0.17	87.47±0.07	99.63±0.09	99.27±0.18
IRF	96.56±2.96	88.95±2.22	80.53±1.83	76.77±1.92
SRF	95.25±3.08	90.52±4.84	86.95±6.27	84.37±7.26
PED	95.47±5.14	89.16±5.66	83.62±6.18	81.52±6.18

Table 1: Performance metrics of RetiFluidNet and VisionFM models on OCT fluid segmentation.

	IRF	SRF	PED
Cirrus	0.74 (0.07)	0.74 (0.16)	0.78 (0.11)
Spectralis	0.74 (0.09)	0.66 (0.29)	0.80 (0.15)

Table 2: Mean (Standard Deviation) Inter-Observer Dice Scores from RETOUCH

Results

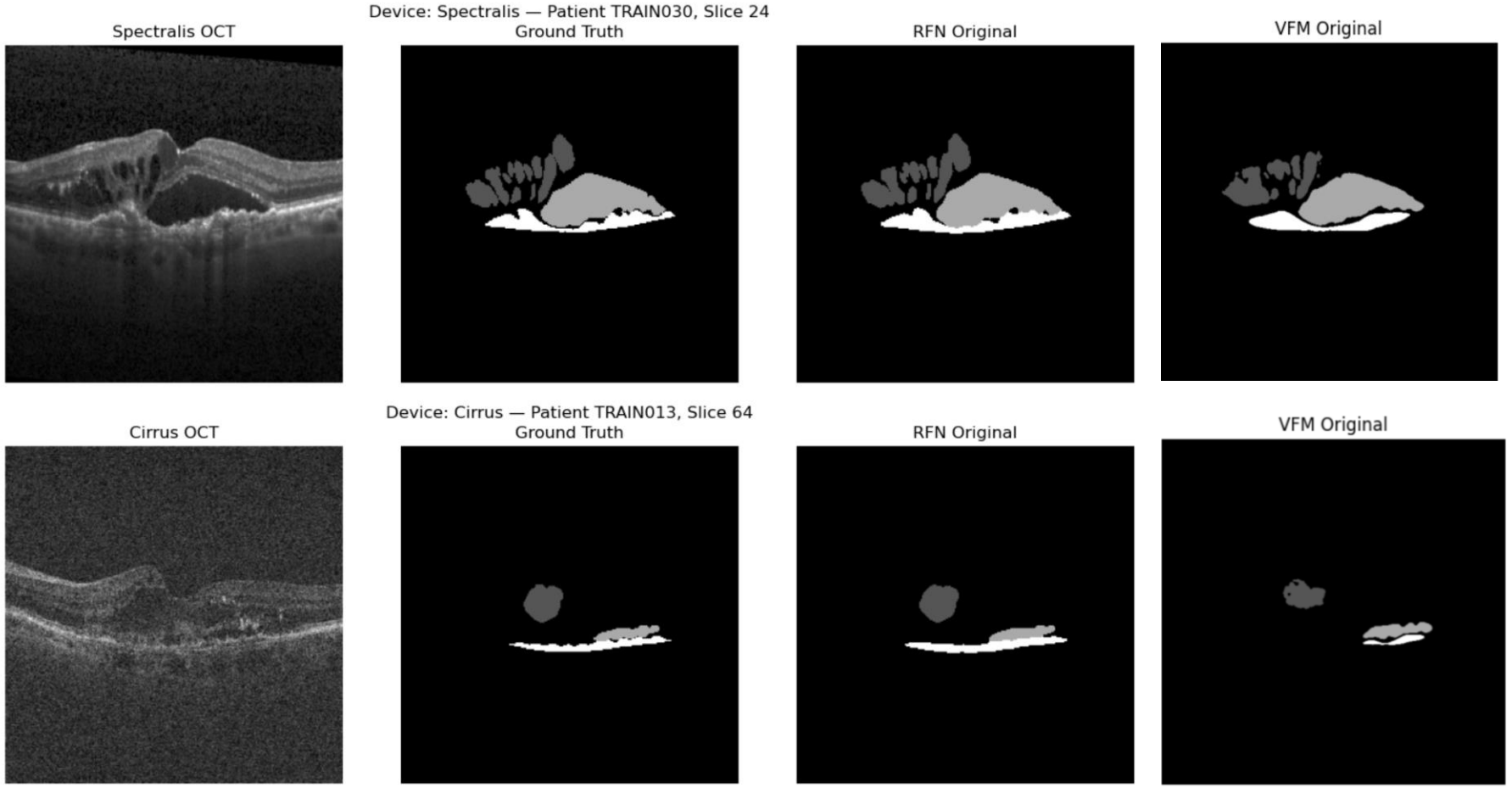


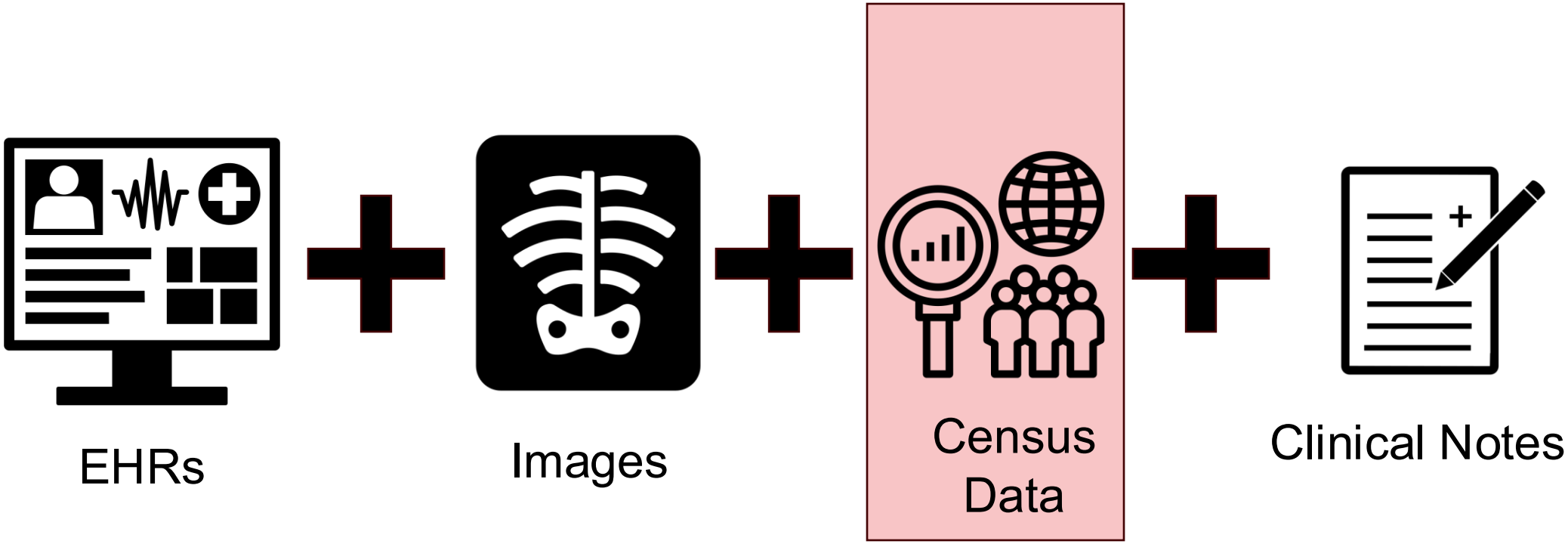
Fig 7: Example OCT segmentation results. Left: Original OCT scans from Spectralis (top) and Cirrus (bottom) devices with ground truth annotations. Middle: Segmentations predicted by RetiFluidNet. Right: Segmentations predicted by VisionFM.

Conclusions

- **Task-specific model (RetiFluidNet)** performed better than foundation model (VisionFM) in retinal fluid segmentation, though the difference is not statistically significant.
- Both models achieved higher agreement with the ground truth than agreement achieved between two human expert annotators.
- **Future Work:** Extend the use of segmentation model and foundation model to incorporate images into multimodal causal learning.

(I²DB)

Multi-modal Causal Learning for Health Care



Social determinants of health (SDoH) and health equity

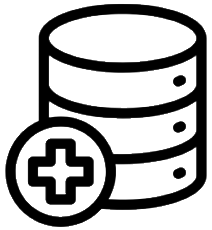
Incorporating SDoH into causal inference is crucial as they can significantly influence the treatment choices and personal health.

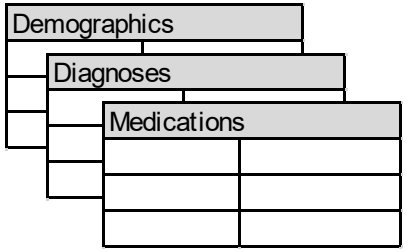


Integration of SDoH and EHR: Workflow



Electronic Health Records (EHRs)





Location Data

location_id	address_1	address_2	city	state	zip	county	FIPS
1	5837 Delmar Blvd	Apt 2	St Louis	MO	63112	St Louis	29510105300
2	1321 Trendley Ave		East St Louis	IL	62201	St Claire	17163504501

Geocoding

ArcGIS

Spatial Join

TIGER Shapefile

United States Census Bureau

Spatial Join

Geospatial Data

Area Deprivation Index

Neighborhood Atlas®

AMERICAN COMMUNITY SURVEY
U.S. CENSUS BUREAU

Social Determinants of Health

PLACES
LOCAL DATA FOR BETTER HEALTH
CDC

: Data Source

: Data Analysis

Temporal Analysis



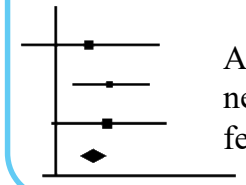
Assess temporal variation of patient population

Geographic Visualization



Visualize patient distribution and neighborhood information

Statistical Modeling



Assess impact of neighborhood features



Luo et al. *[Under review]*

Temporal Trends of Opioid Use Disorder

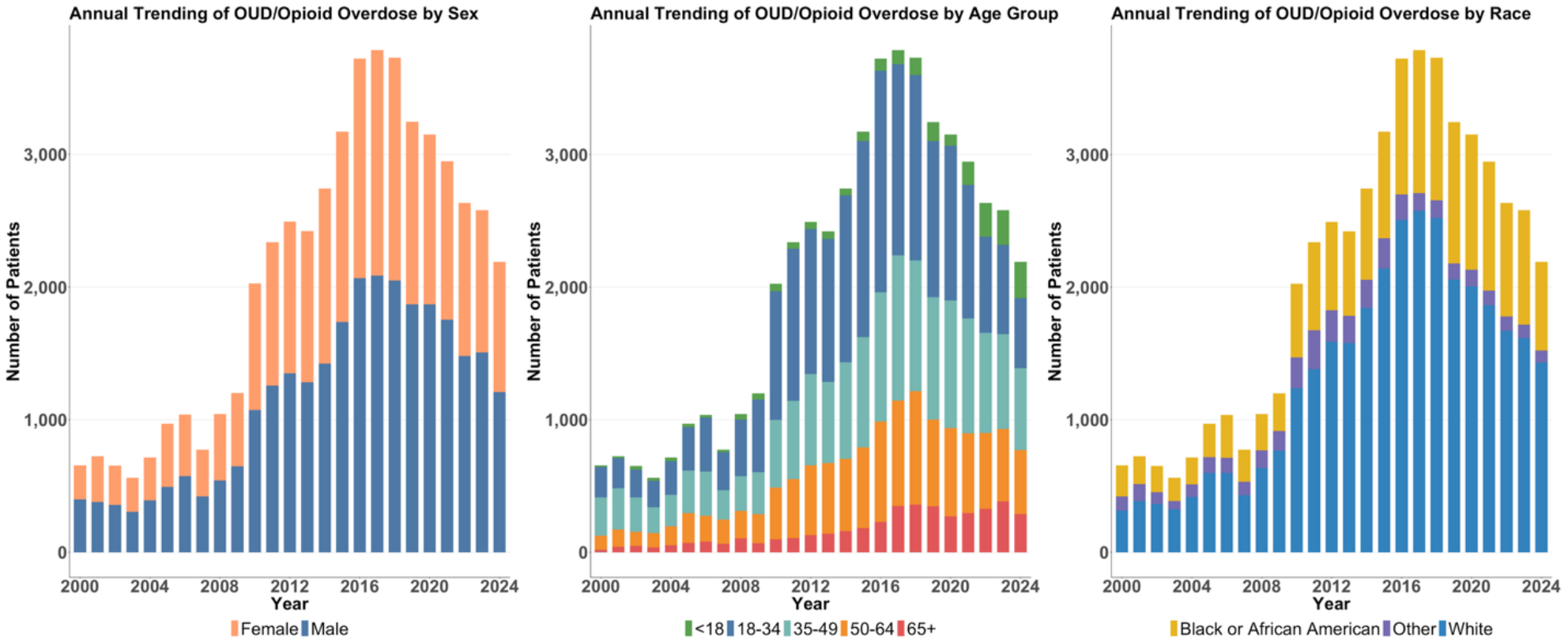
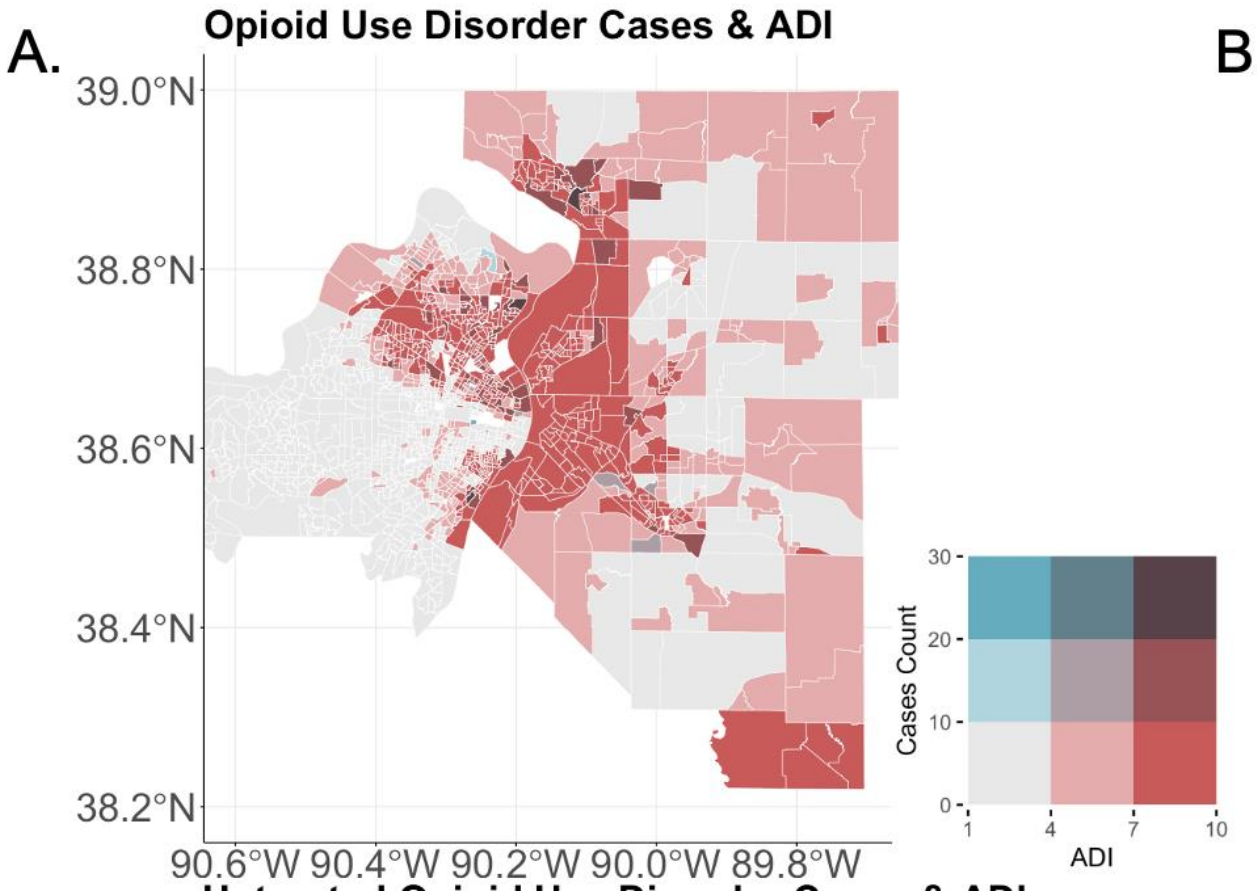
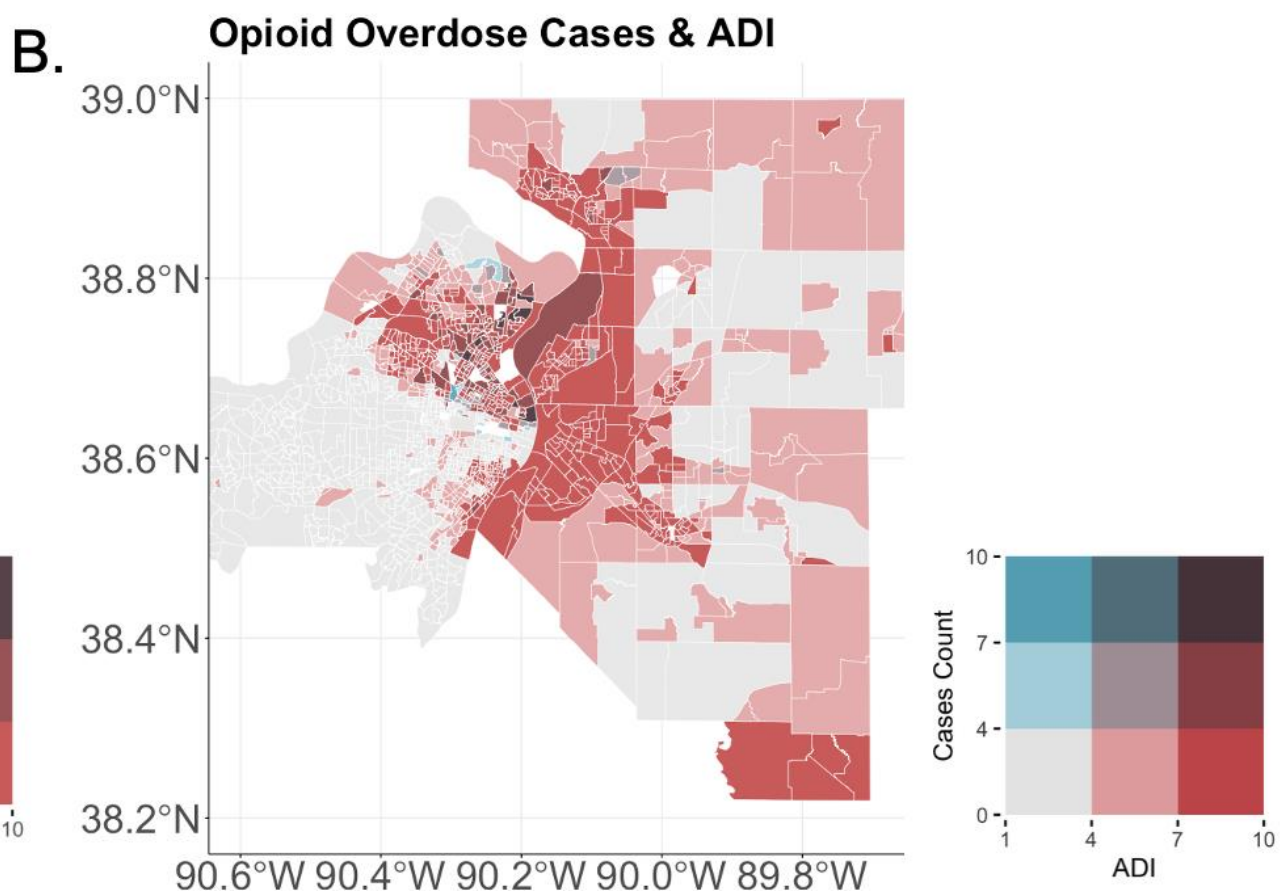


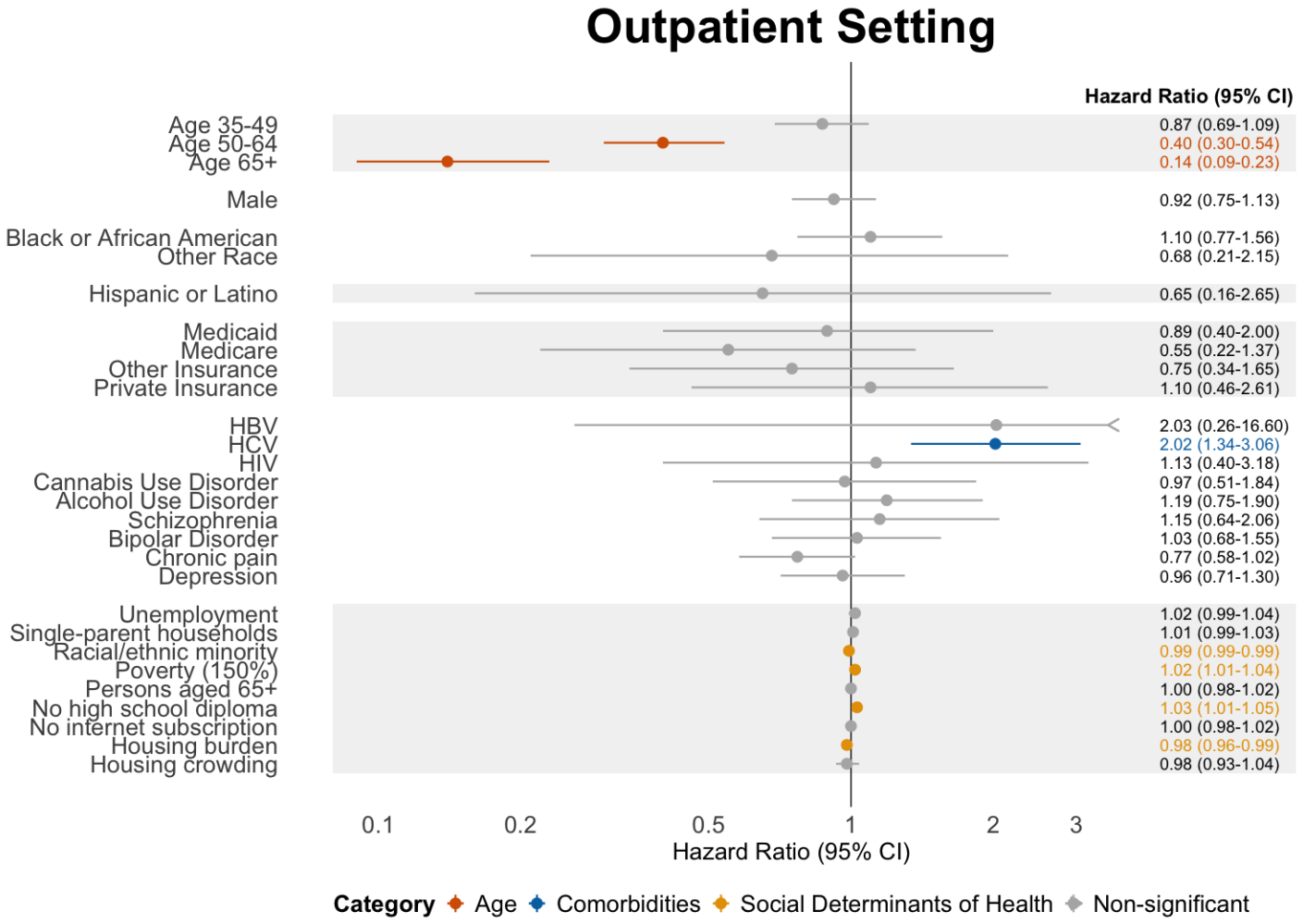
Figure 2. Temporal trends in opioid use disorder (OUD) and opioid overdoes among patients in BJC HealthCare from 2000 to 2024, stratified by sex, age group, race and ethnicity.

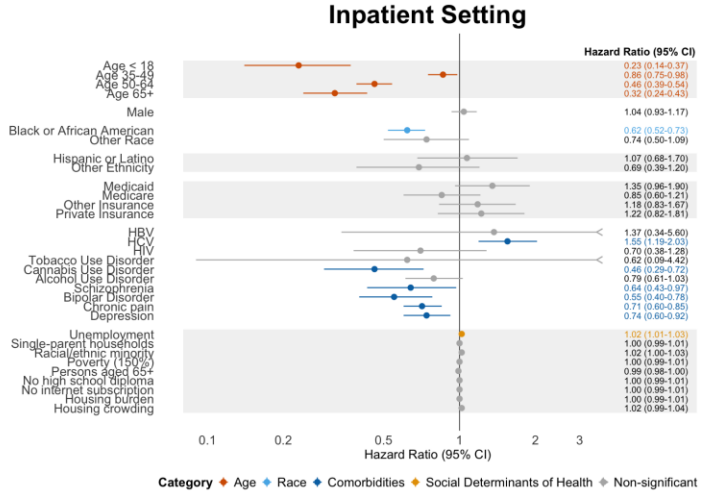
Area Deprivation Index: OUD and Overdose

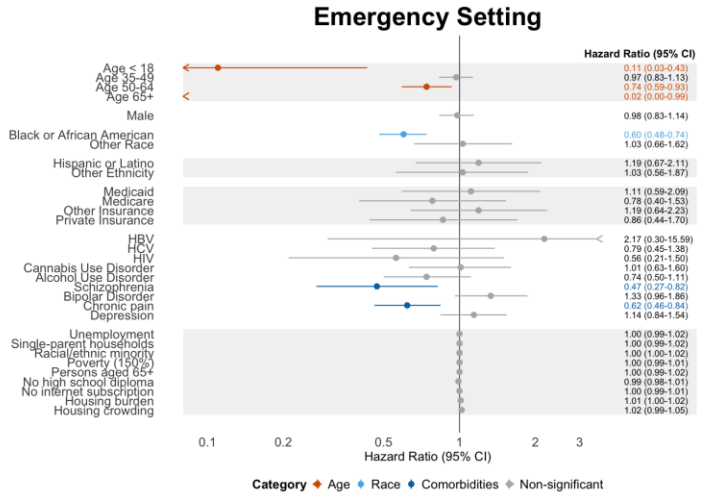




Social determinants most strongly hindered outpatient treatment initiation.

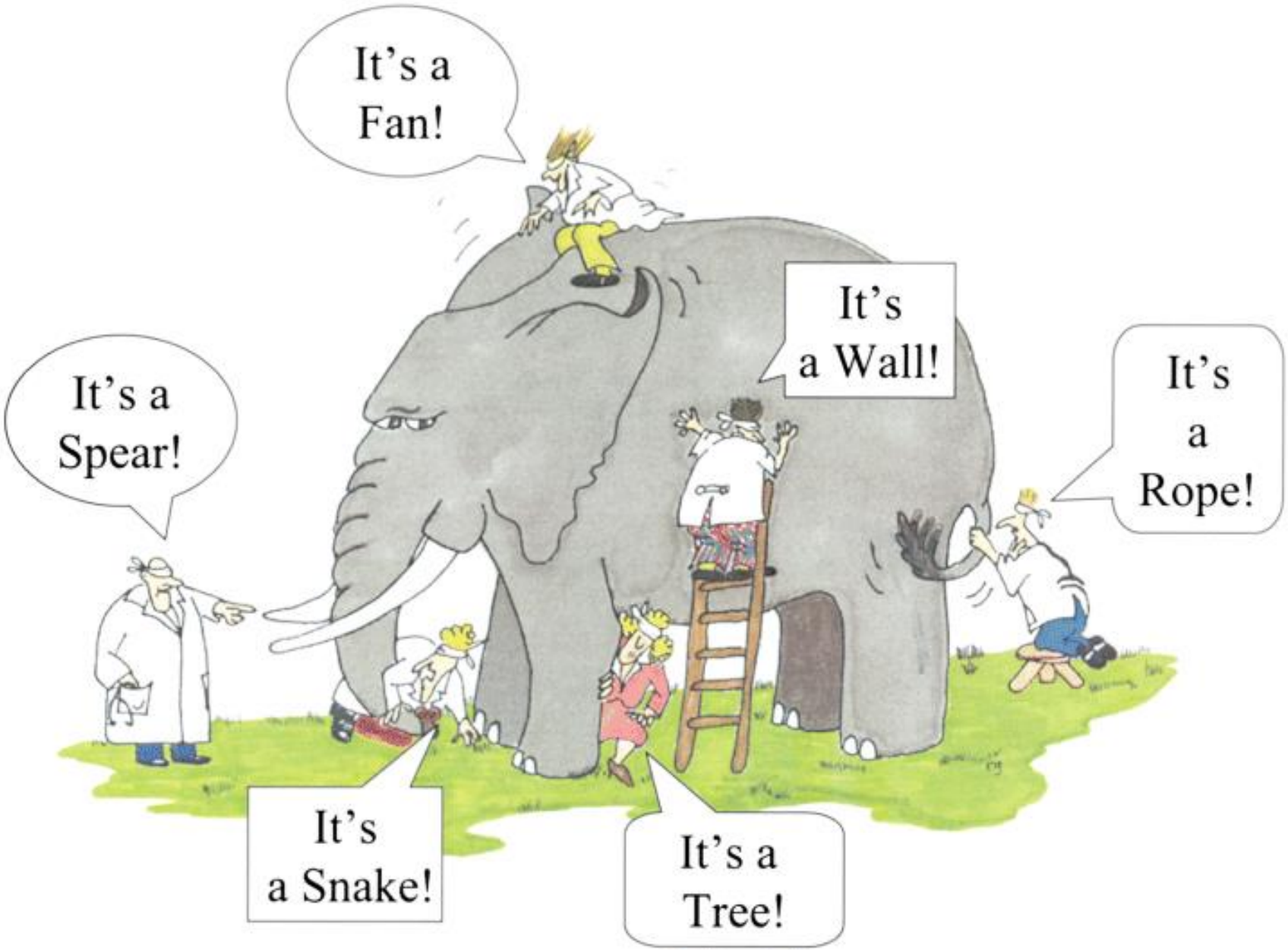






The goal is to integrate causal modeling with machine learning to generate reliable real-world evidence and to build an equitable healthcare system.

(I²DB)



(I²DB)

Join us!

linyingz@wustl.edu

causailab.github.io

