EHR for psychiatric discovery

Roy Perlis, MD MSc
MGH Center for Quantitative Health
rperlis@mgh.harvard.edu

JAMA+ AI

CENTER FOR EXPERIMENTAL DRUGS & DIAGNOSTICS

MASSACHUSETTS GENERAL HOSPITAL

HARVARD MEDICAL SCHOOL

Disclosure

Dr. Perlis has received payment for service on scientific advisory boards of Genomind, Circular Genomics, Alkermes, and Atella

He has received payment (and a really cool fleece) for service as Editor in Chief of JAMA+ AI, and as AI Editor at JAMA Network Open

Electronic health records are not designed for your research

Epic was (and is) a tool to ensure that hospitals can bill for services … everything else came later

EHR data is not clinical trial data: it is artifacts of routine care, collected by people whose primary goal is to take care of patients while avoiding medicolegal risk and getting paid for their time.

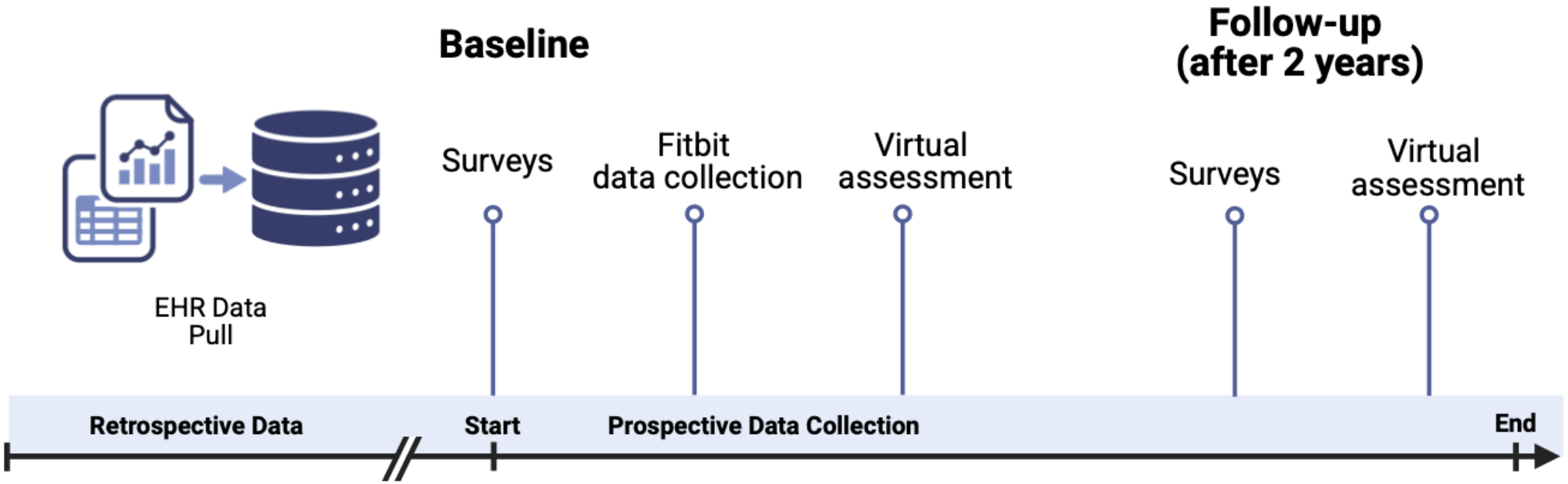
- Diagnoses are not necessarily correct (misclassification)
- Missing data is not random (the open system problem)
- Treatment assignment is not random (confounding by indication)
- Much of the signal in EHR reflects patterns of service or other confounding.

Mantra for EHR research

"You can't always get what you want…
but if you try sometime you'll find
you get what you need."

- noted scholar of EHR research Mick Jagger (1969)

SPARC-XP



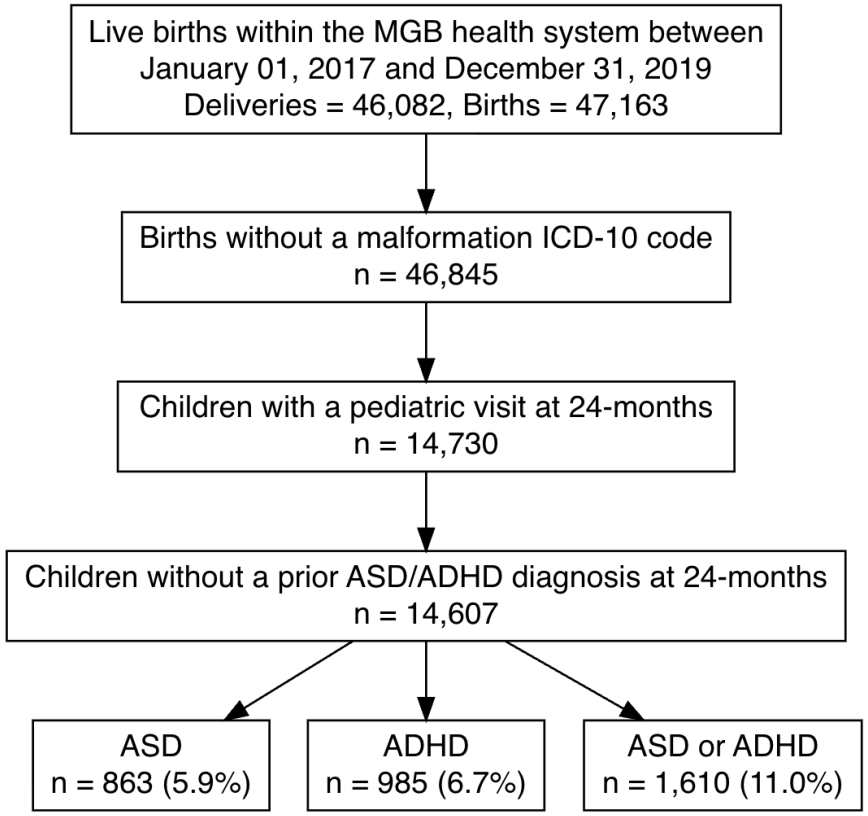
EHR Data
- Maternal infections, diagnoses, and medications
- Infant neurodevelopmental outcomes

Questionnaires
- Psychosocial History
- SWAN Scale
- SRS-2
- Vineland-3

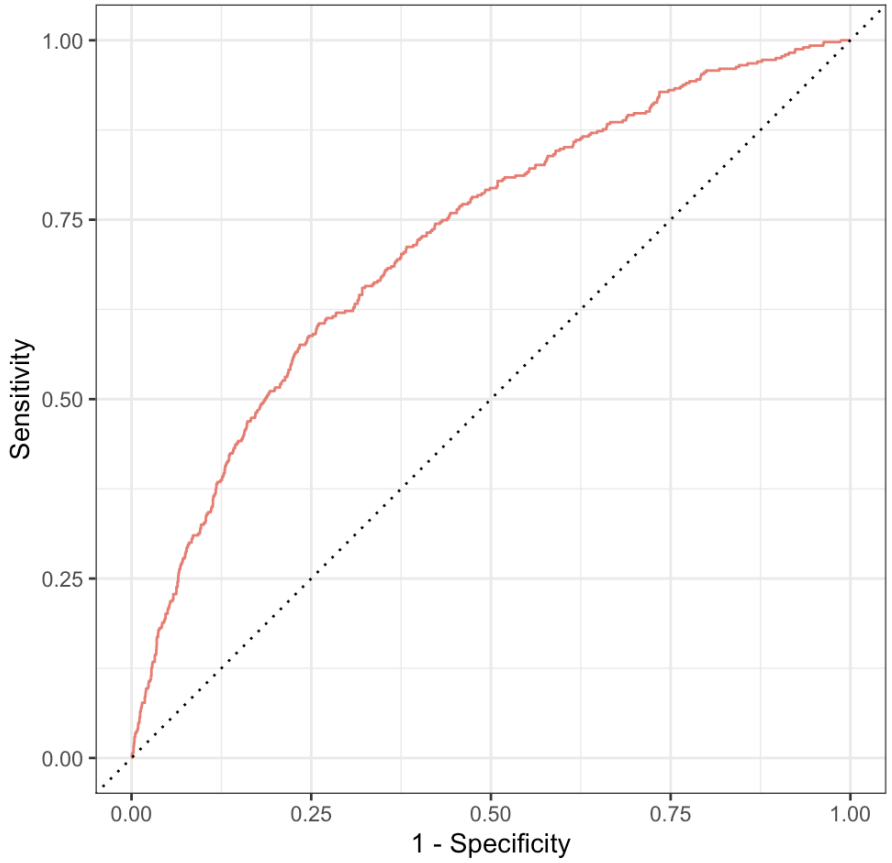
Cognitive Assessments
- BRSA-4
- KBIT-2
- Expressive Vocabulary Test

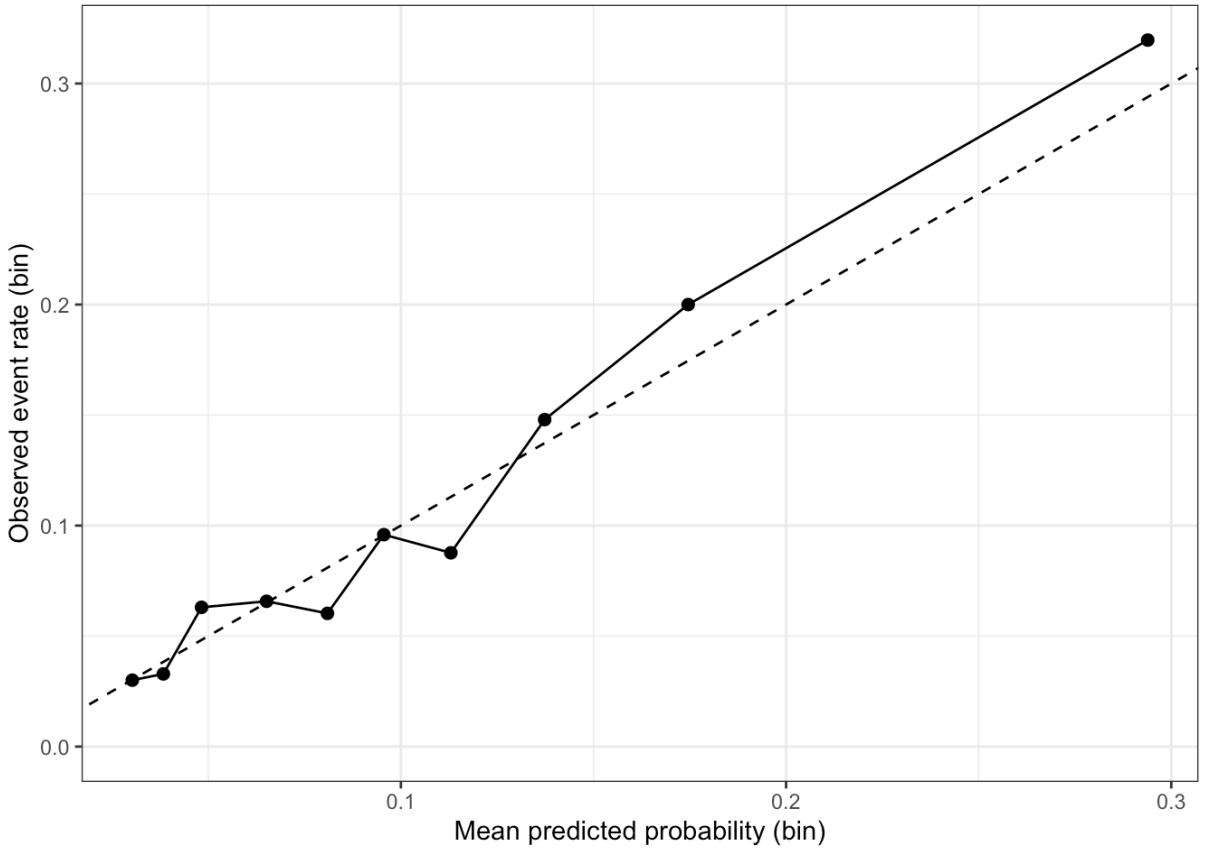
Fitbit Data
- Activity levels
- Sleep Patterns
- HR and SpO2 levels

SPARC-XP CONSORT diagram (training/testing set)



Baseline model performance (test set)



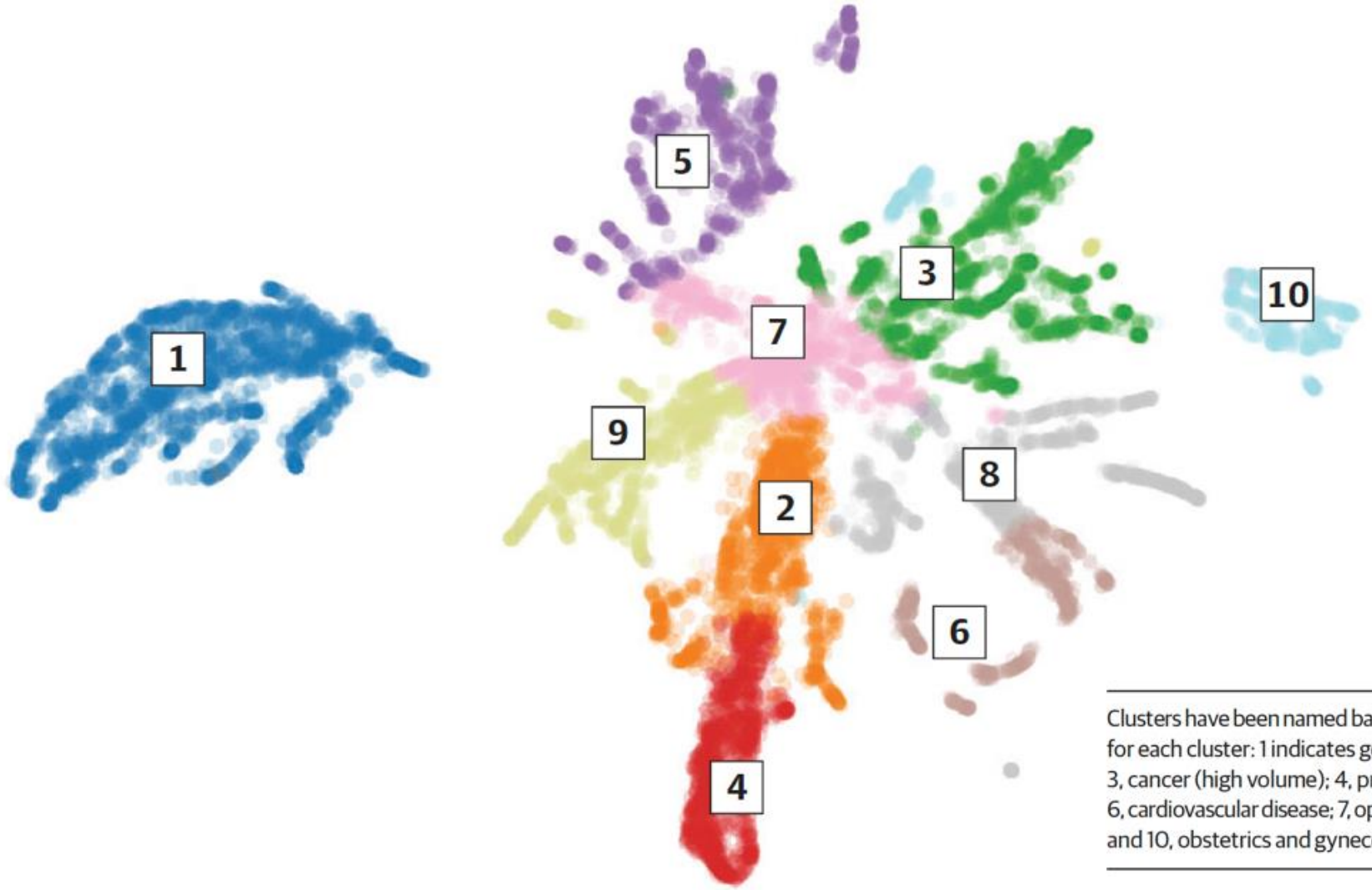


AUC 0.72, NPV 0.92, PPV 0.29

Beware system-specific signal

Example: when is antidepressant response not just antidepressant response?

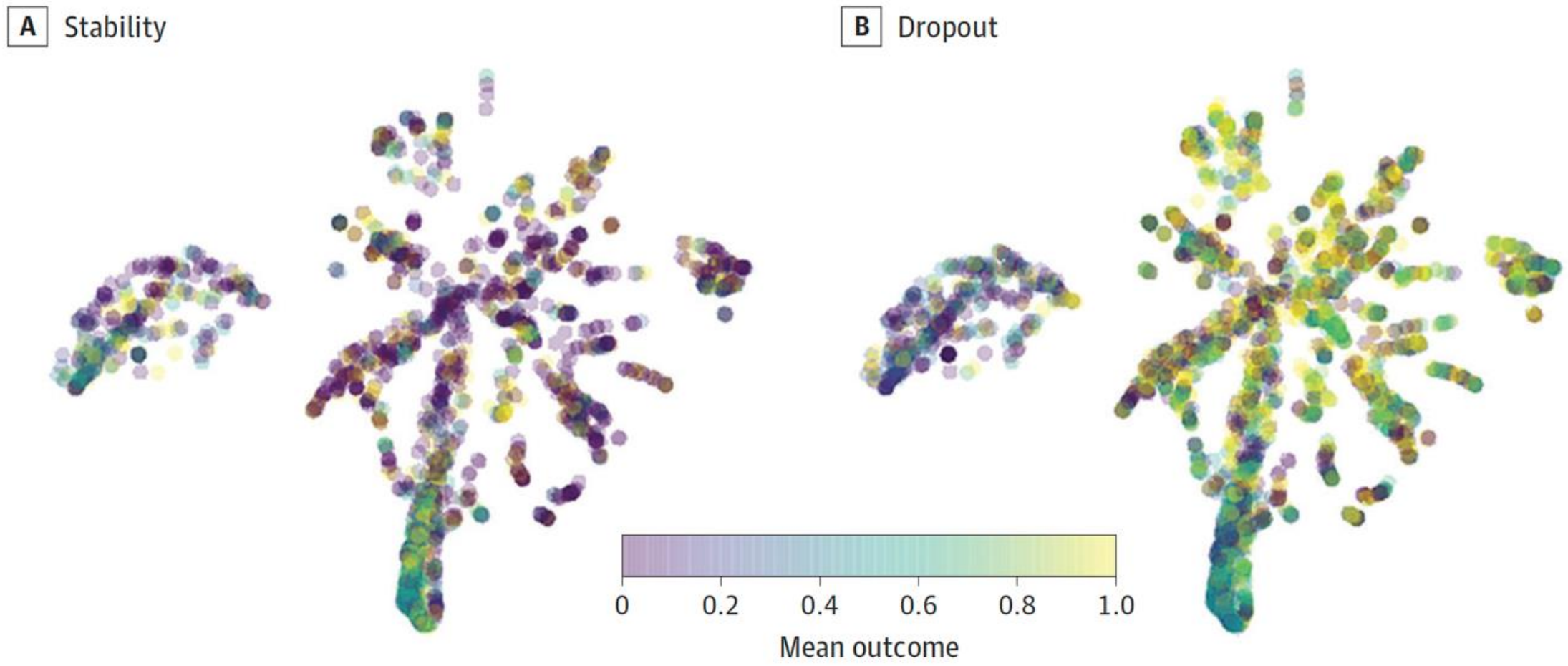
A map of antidepressant prescribers at Mass General - Brigham



Clusters have been named based on our review of the predominant diagnostic codes for each cluster: 1 indicates general psychiatry; 2, primary care (low volume); 3, cancer (high volume); 4, primary care (high volume); 5, musculoskeletal pain; 6, cardiovascular disease; 7, ophthalmology; 8, kidney disease; 9, cancer (low volume); and 10, obstetrics and gynecology.

Rathnam JAMA Psych 2024

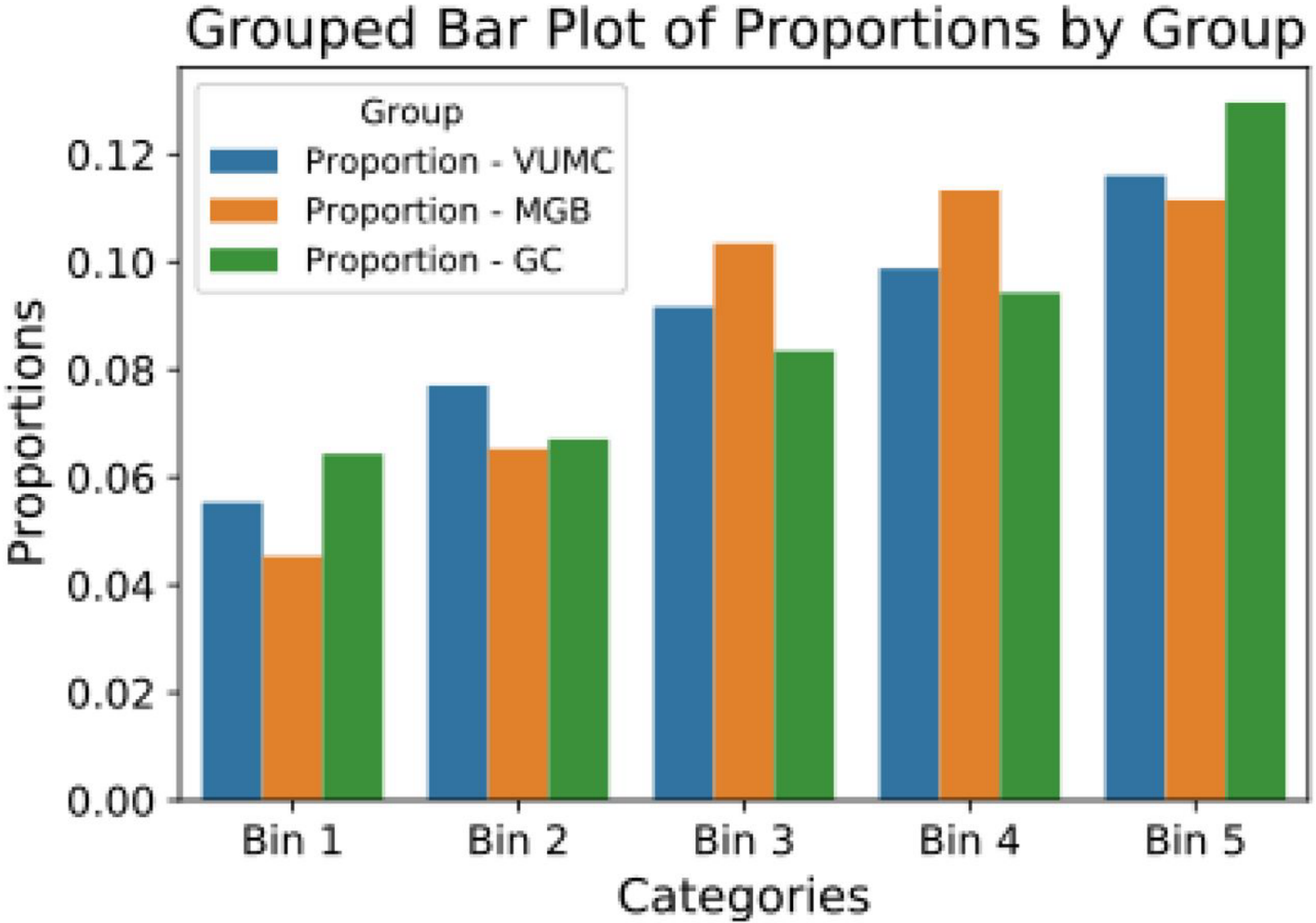
Outcome map for antidepressant prescribers* at Mass General-Brigham



* I'm the little purple dot in the cluster on the far left

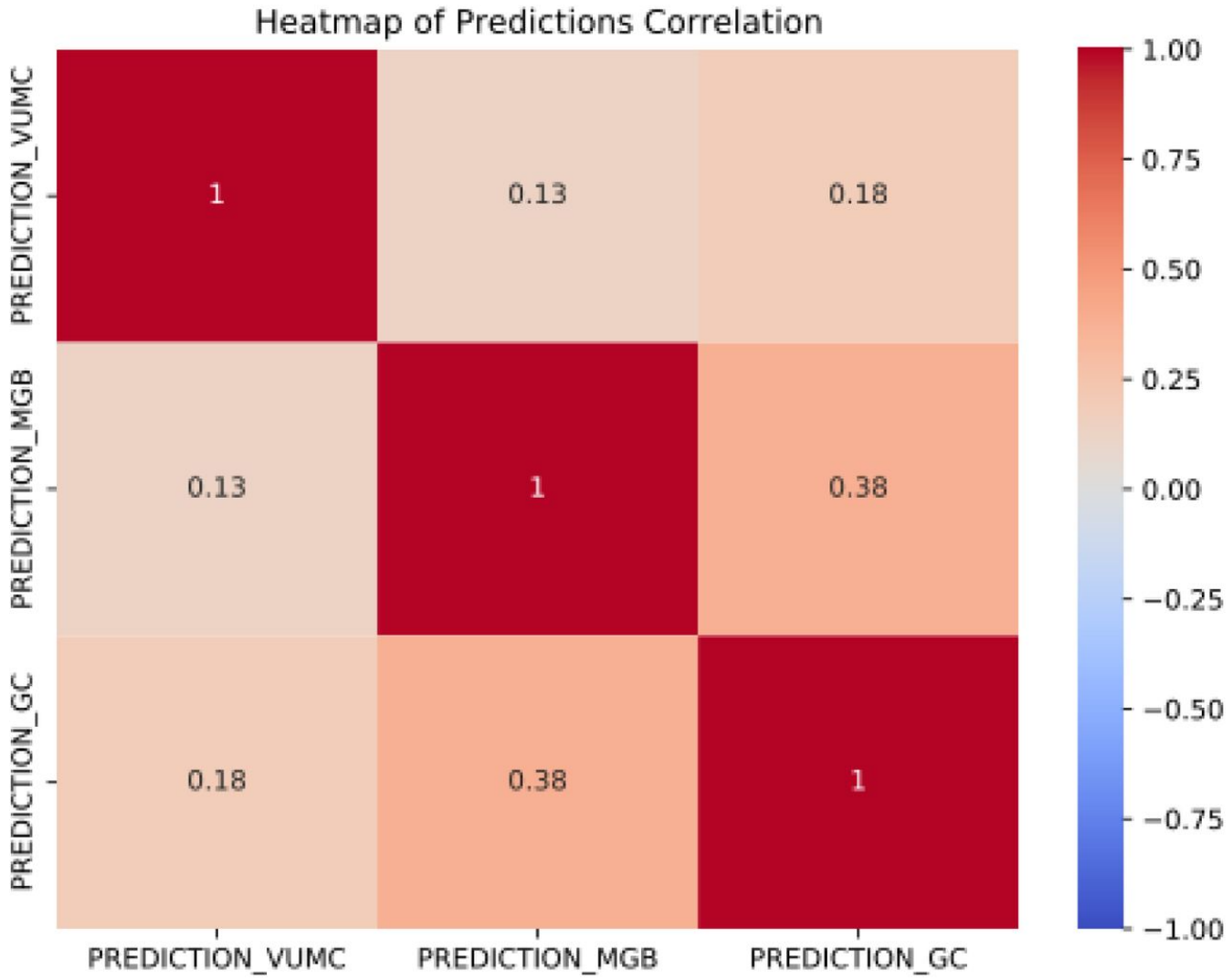
Rathnam JAMA Psych 2024

And models do show signal across systems, but…



Walsh in revision

There's not much signal!



Walsh in revision

How do we keep EHR data from breaking our hearts? 3 hypotheses*

1. The signal is there, we just need better prediction tools
   (Unlikely in most clinical contexts but hope springs eternal)

2. The signal is there, we just need better feature engineering
   LLMs for concept identification
   Dimensional measures

3. The signal is weak, we need orthogonal data sets

* Note similarities to genomics in determination to boil the ocean!

Dimensional phenotyping

Most coded EHR data reflects categories – presence/absence of diagnosis

Most EHR NLP effort has attempted to more reliably identify categories

But what about *dimensions* – e.g., RDoC; personality; depression severity?

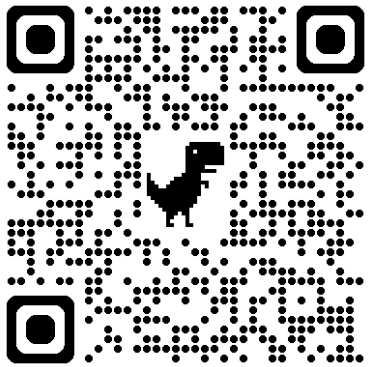
McCoy AJP 2015; McCoy Biol Psych 2018; McCoy Biol Psych 2024

Dimensional phenotyping

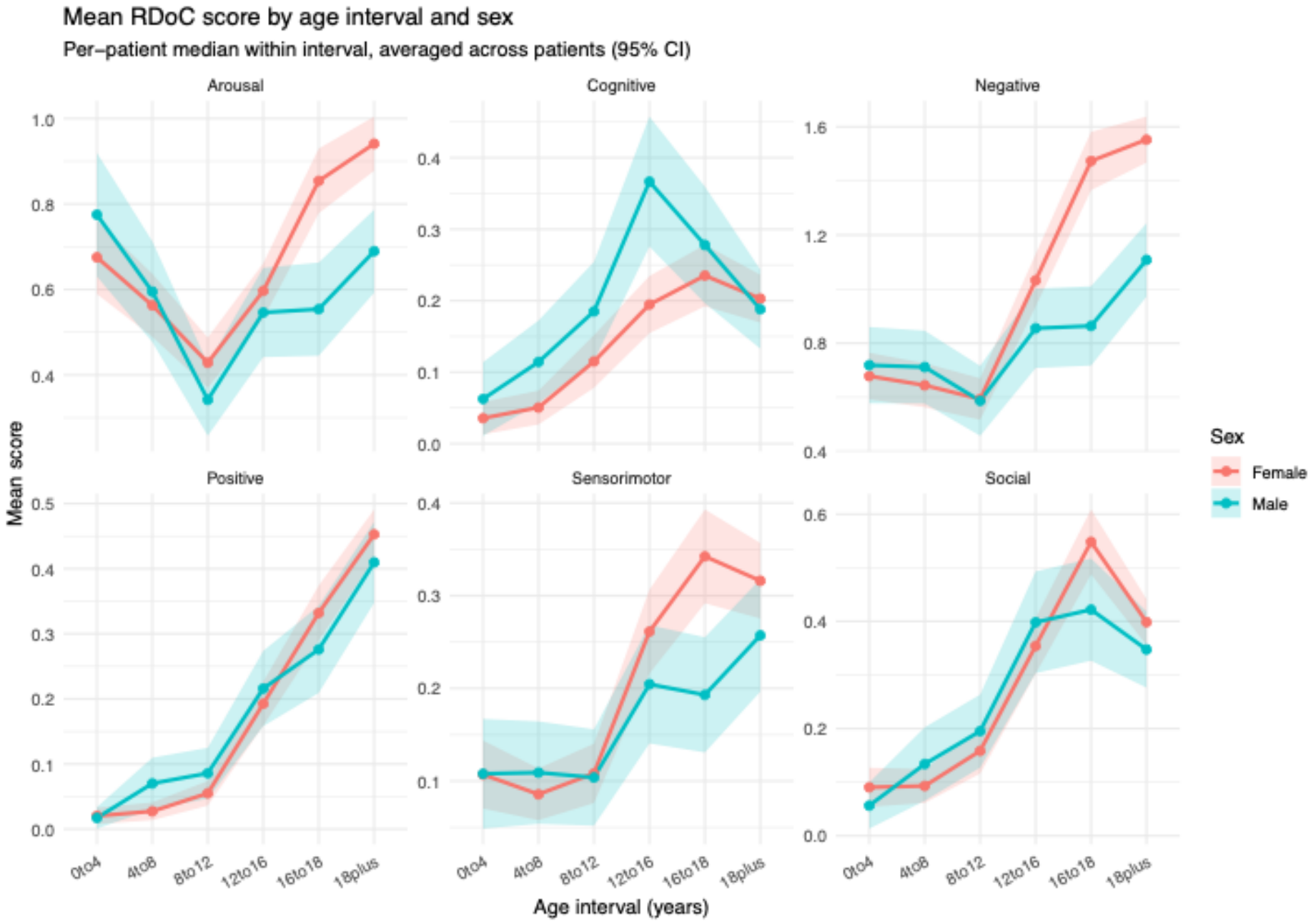
Try simple unigram and bigram counting, almost instantaneous!

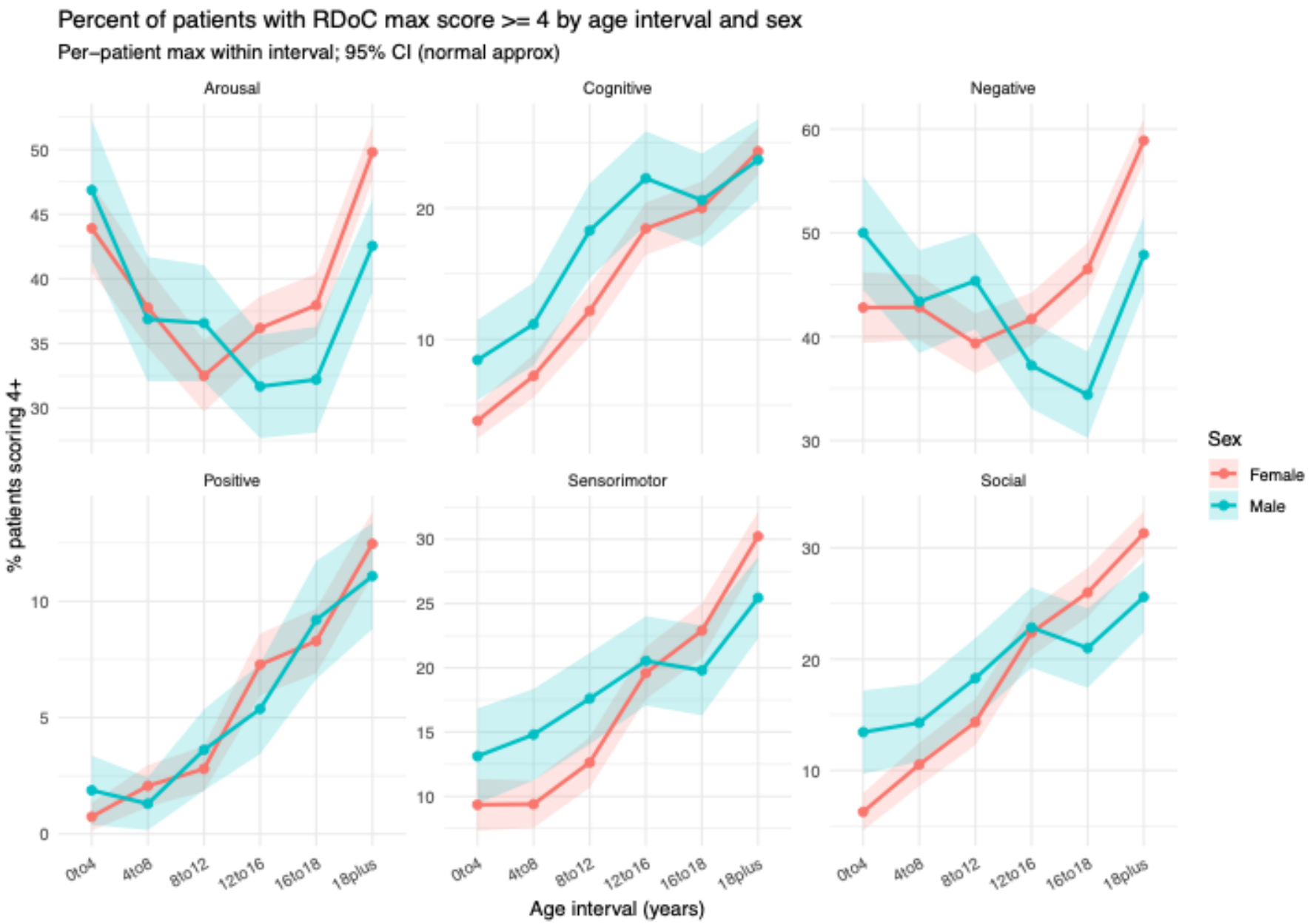
Does not require frontier LLM … local models (qwen or gemma family, e.g.) acquit themselves quite well.

Example: JASPer-MH prospective cohort (~160k notes among ~4500 individuals age 18-23)



McCoy BMJ Mental Health 2025; Perlis unpublished


Mean RDoC score by age interval and sex
Per−patient median within interval, averaged across patients (95% CI)
Arousal
Cognitive
Negative
Positive
Sensorimotor
Social
Mean score
Age interval (years)
0to4
4to8
8to12
12to16
16to18
18plus
Sex
Female
Male


Percent of patients with RDoC max score >= 4 by age interval and sex
Per–patient max within interval; 95% CI (normal approx)
Arousal
Cognitive
Negative
Positive
Sensorimotor
Social
% patients scoring 4+
Age interval (years)
0to4
4to8
8to12
12to16
16to18
18plus
Sex
Female
Male

Filling in the gaps in EHR

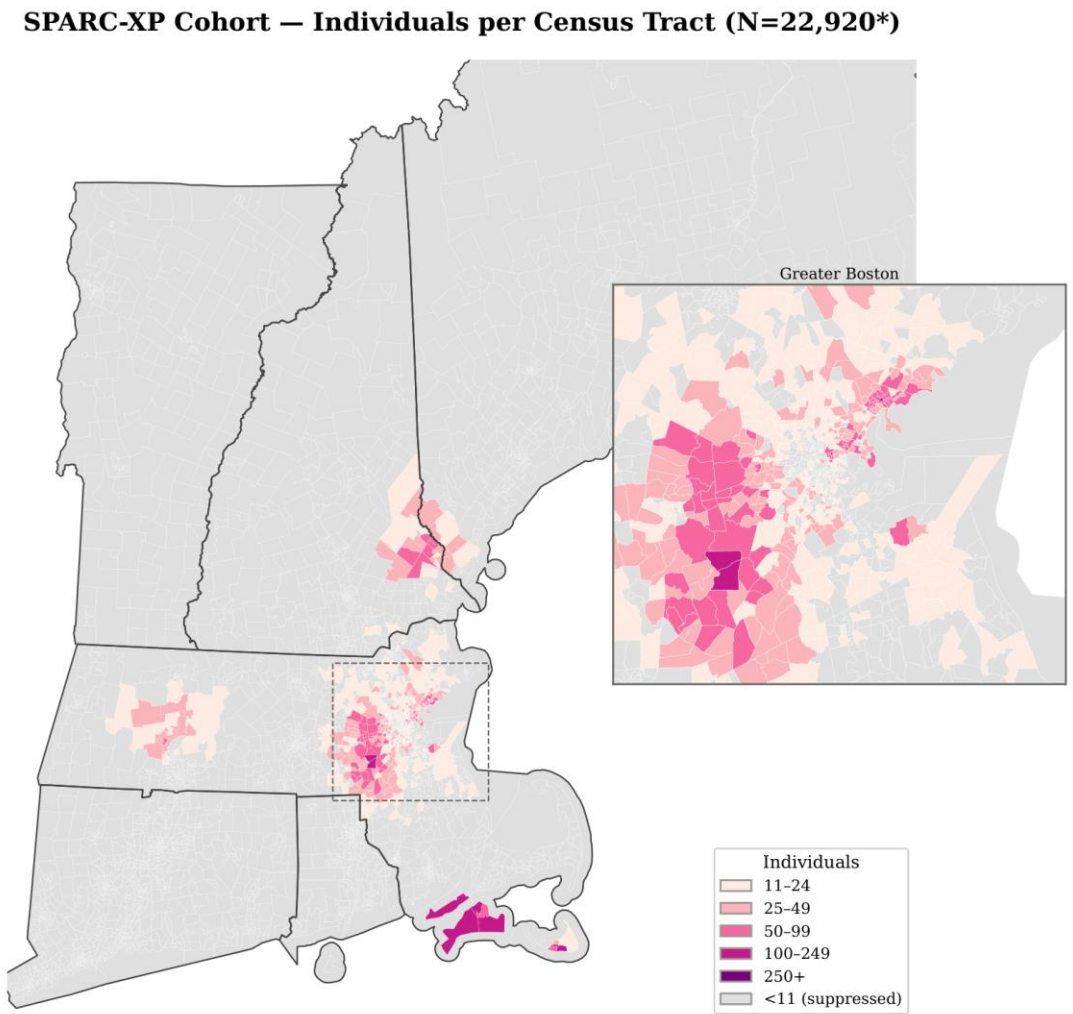
Example: Geocoding with linking to pregnancy exposures
SPARC-XP cohort (~22k births)

Challenges

- Local geocoding to protect PII
- Movement
- High-resolution exposures
- Appropriate analytic approaches

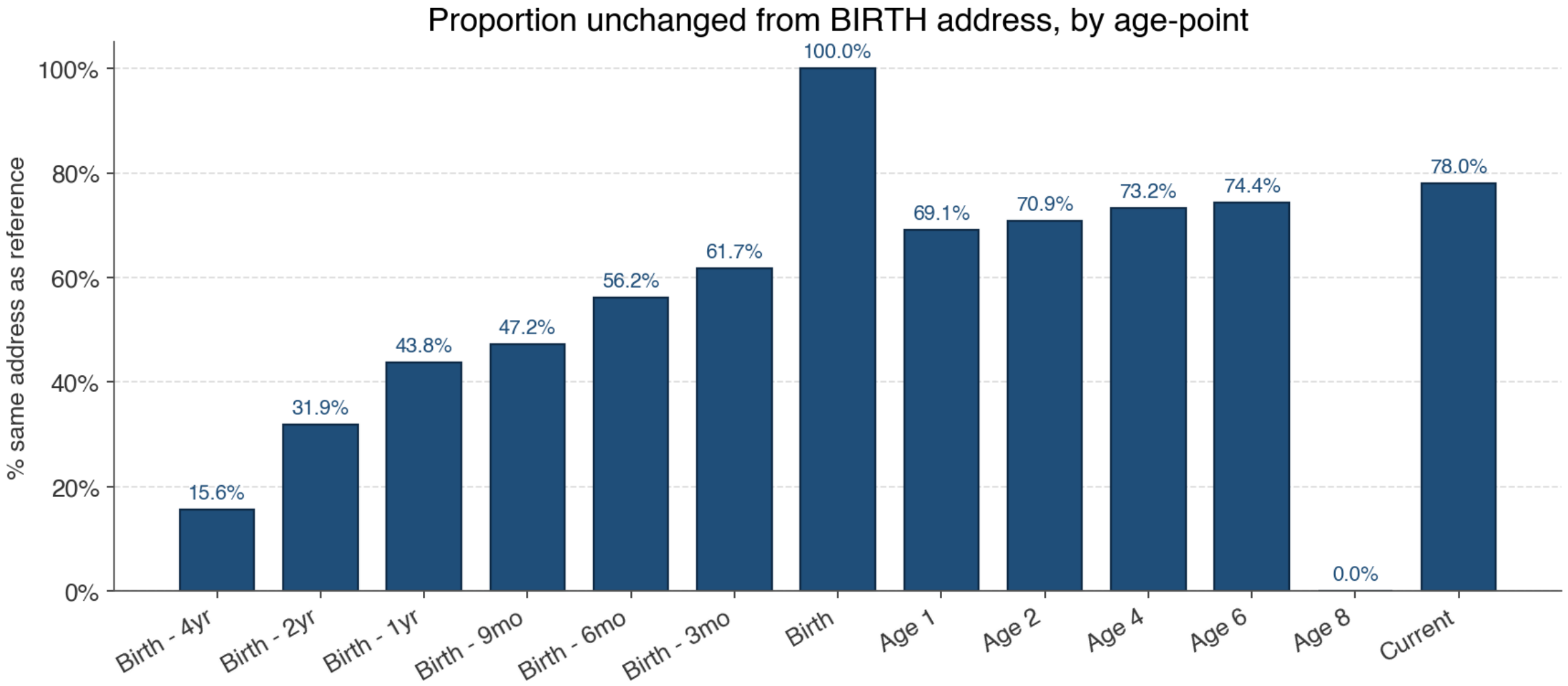
Map C: Cohort Count by Census Tract

Absolute counts | New England states | Tracts with <11 suppressed



*22,798 individuals shown in displayed states. Tracts with fewer than 11 individuals are suppressed to preserve privacy.

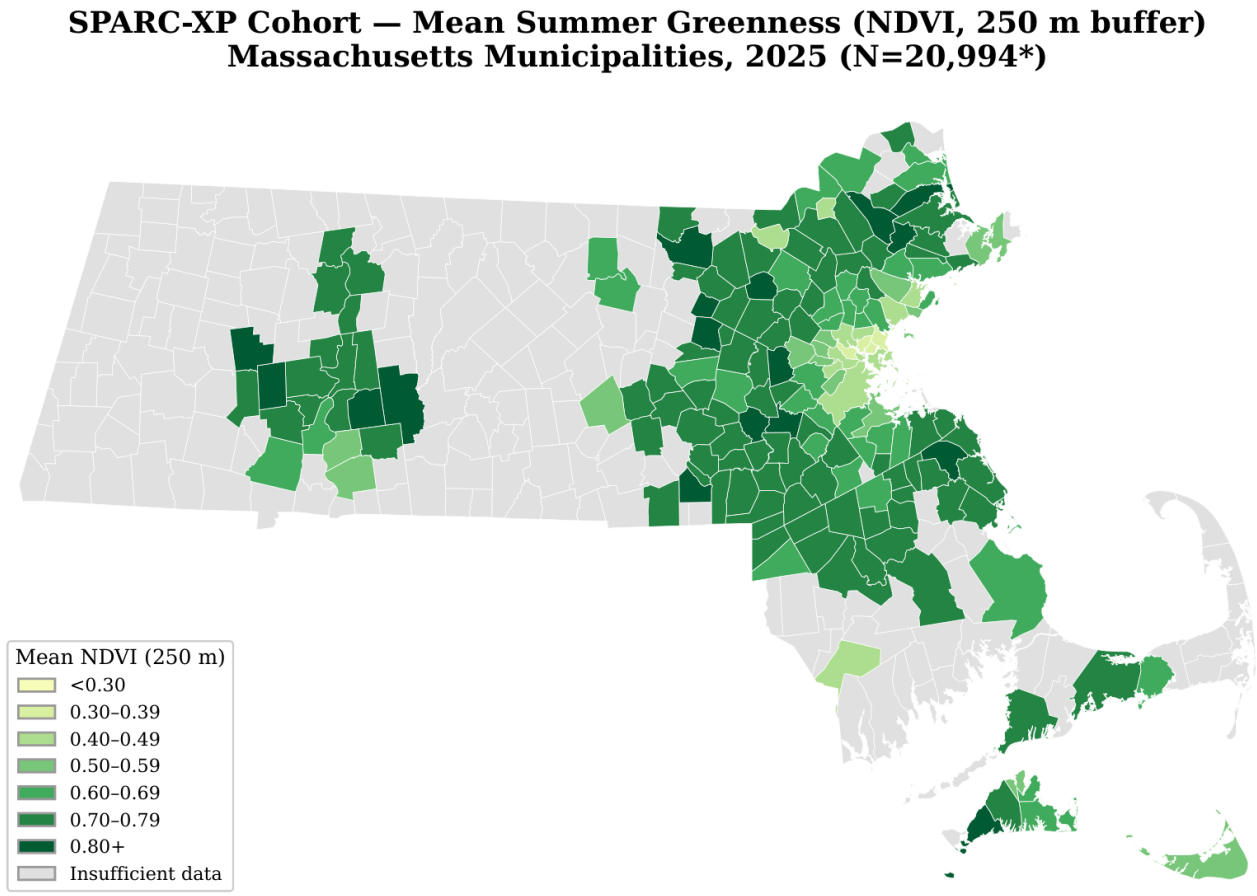
% unchanged from BIRTH address, by age-point



Denominator: persons with non-missing address at both BIRTH and the listed age-point.

Map D: Mean Summer Greenness (NDVI, 250 m buffer)

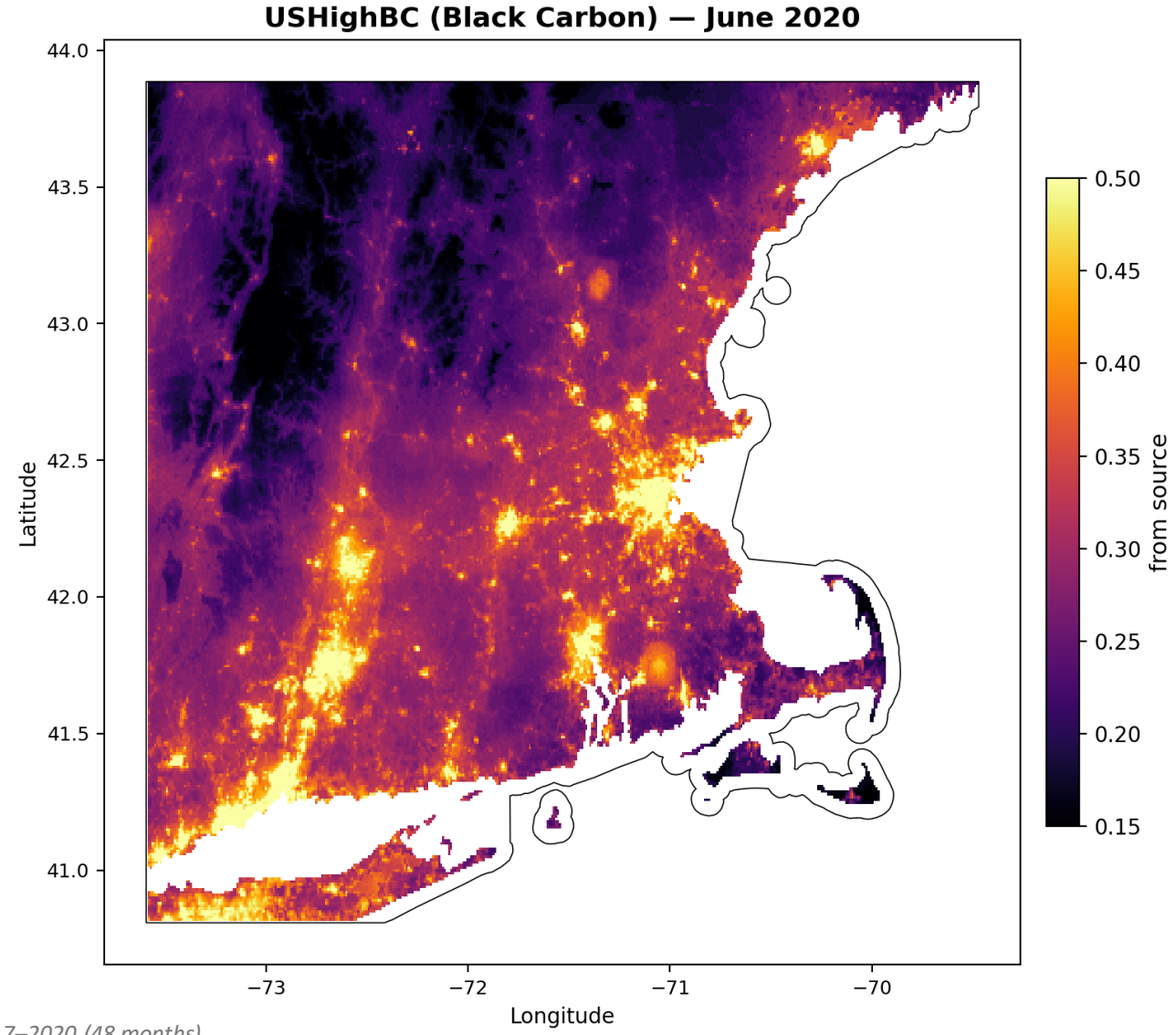
Massachusetts municipalities | NASA HLS 30 m satellite imagery | Summer 2025



*20,994 addresses with valid summer 2025 NDVI. Municipalities with fewer than 11 individuals are suppressed. Source: NASA Harmonized Landsat Sentinel-2 (30 m).

USHighBC (Black Carbon)

US 1 km monthly black carbon | USHAP ensemble



Thank you!

- NIMH
- NICHD, NHGRI, NSF
- Dozoretz Family
- Barnett Family

rperlis@mgh.harvard.edu
roy.perlis@jamanetwork.org

JAMA+ AI

CEDD CENTER FOR EXPERIMENTAL DRUGS & DIAGNOSTICS

MGH 1811 MASSACHUSETTS GENERAL HOSPITAL

HARVARD MEDICAL SCHOOL