United States Environmental Protection Agency

EPA Tools and Resources Webinar: Use of Electronic Health Records to Address Pressing Environmental Health Questions

Dr. Cavin Ward-Caviness
US EPA Office of Research and Development

April 20, 2022

EPA United States Environmental Protection Agency

Outline for Presentation

Electronic health record (EHR) Considerations

Environmental Health Research Using EHRs

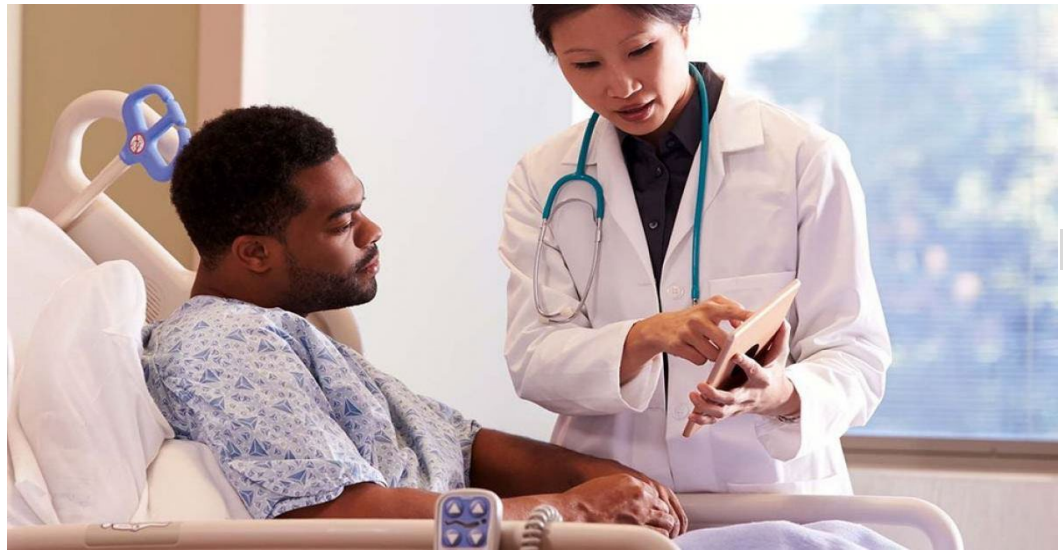
Next steps in EHRs and Environmental Health

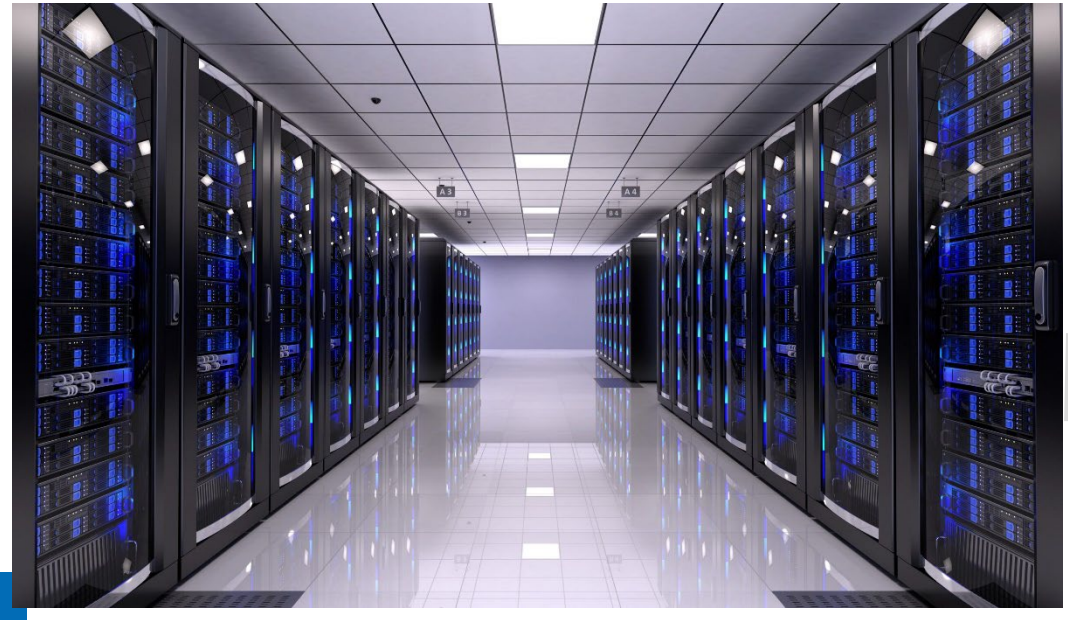
EPA United States Environmental Protection Agency

What Key Questions can EHRs Help Us Answer?

- Who are the most vulnerable?
- What are the health risks of the most vulnerable as compared to the general population?
- What clinical markers presage a major event?
- Can progression of disease be modified?
- How can we track environmental risks at community and individual levels?

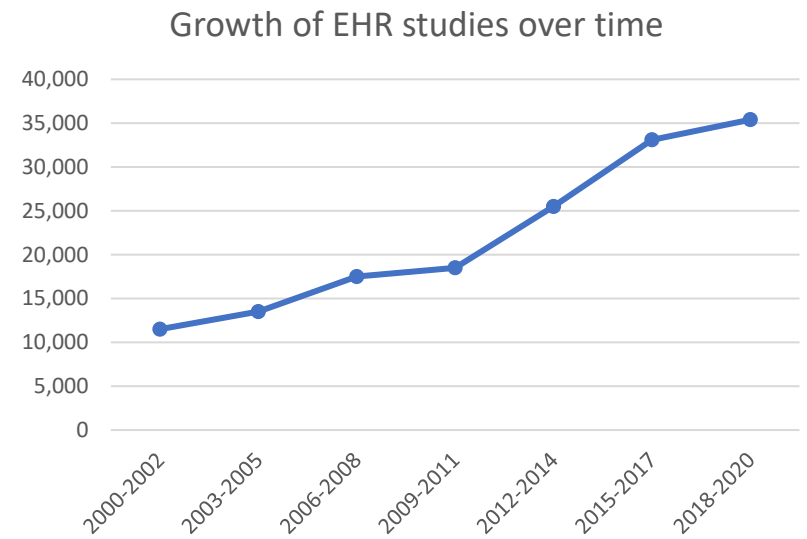
Rise of EHRs in Health Studies





Patient ID	Visit Date	ICD-10 Code	Clinical Lab	Vitals
A	March 1 2000	I10; K45.0	3.4	89
B	Feb 4 2020	J01.3	5.6	103
C	August 19 2014	I11	19.2	79
D	August 19 2014	I10	1.3	98

Patient ID	Age	Sex	Address
A	23	M	123 Main St
B	78	F	45 East Ave
C	67	F	9 Broad Dr
D	34	U	4 West St



EPA United States Environmental Protection Agency

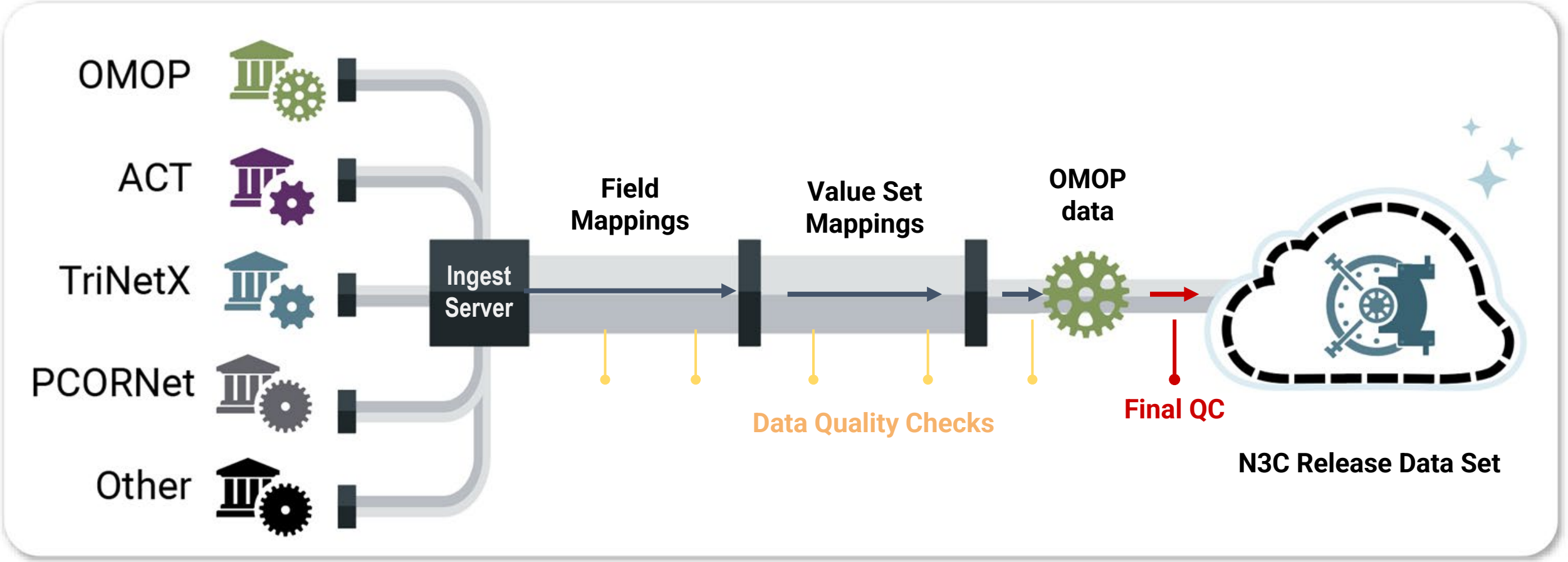
Utility of EHRs for Environmental Health

- Individual Level Data
- Local -> Regional -> National Harmonization
- Longitudinal Data
- Deep Phenoytping
- Address History
- Continual Collection

Key Considerations for EHRs: Harmonization

- How to insure that data are the same across EHRs?
- How to construct Computable Phenotypes?
- Demographics (is race always encoded the same?)
- What does an address represent?
- When are column entries the same?
- Is the structure of your data similar, e.g. Common Data Models (CDM)
- How to bring in national/international standards, e.g. Observational Medical Outcomes Partnership (OMOP)

Example: N3C Data Ingestion & Harmonization Pipeline



N3C = National COVID Cohort Collaborative (www.covid.cd2h.org)

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

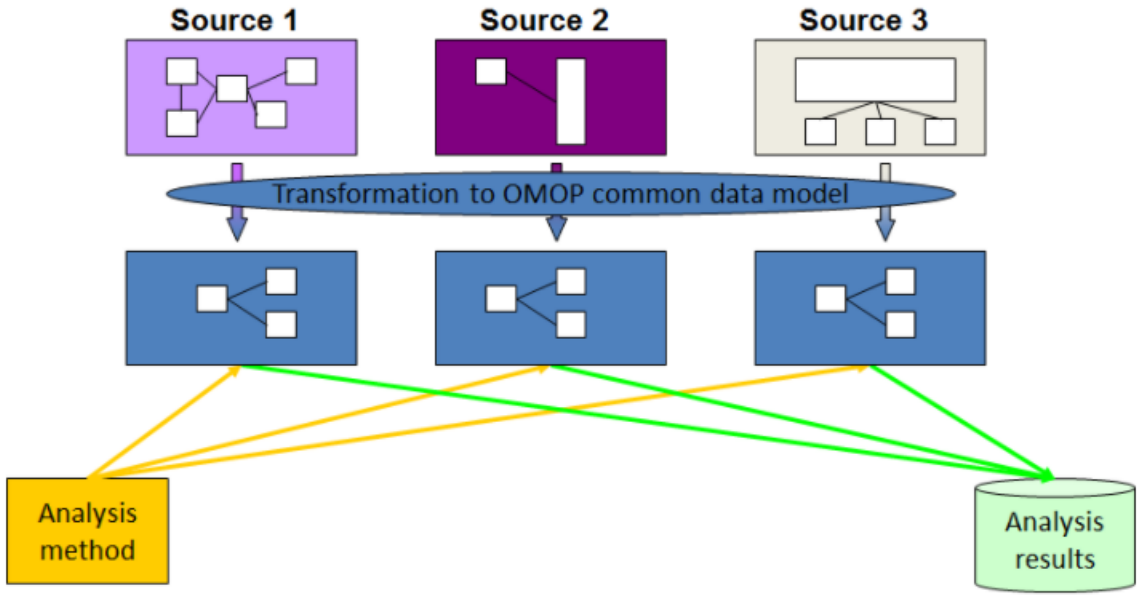
Who We Are | OHDSI Updates & News | Standards | Software Tools | OHDSI Studies | Book of OHDSI | Resources | New To OHDSI?

EHDEN Academy | This Week In OHDSI/Community Calls | Events/Collaborations | Workgroups | Forms For Workgroups, MS Teams

Our Journey – Where The OHDSI Community Has Been, And Where We Are Going | OHDSI2022 Symposium | Follow OHDSI/Newsletters

Home › Data Standardization › OMOP Common Data Model

OMOP Common Data Model



What is the OMOP Common Data Model (CDM)?

An international organization that seeks to "improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care" https://www.ohdsi.org/who-we-are/mission-vision-values/

Developers & maintainers of the OMOP Common Data Model which is a way of representing EHRs that is compatible with many existing systems and allows for sharing of queries/code and a firm understanding of concepts across data sources

EPA United States Environmental Protection Agency

Common Data Models and the Environment

Common Data Models are the means by which hospitals communicate

Most common, e.g. OMOP, are built around the person whereas many exposures are built around place and time

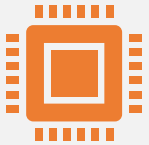
Efficiently moving between the two “at scale” (billions to trillions of records) requires novel thinking and new structures for ingesting and exporting data

Key Considerations for EHRs: Timeliness

- Environmental exposures vary based on space and time
- Not all EHR records align with available exposures, and EHRs may change over time
- Capturing specific events, e.g. the coronavirus pandemic or past exposures, may also impose timeliness constraints on EHRs

EPA United States Environmental Protection Agency

Key Considerations for EHRs: Computable Phenotypes



Computable phenotypes refers to a set algorithm which can use standardized EHRs to determine an condition/disease state

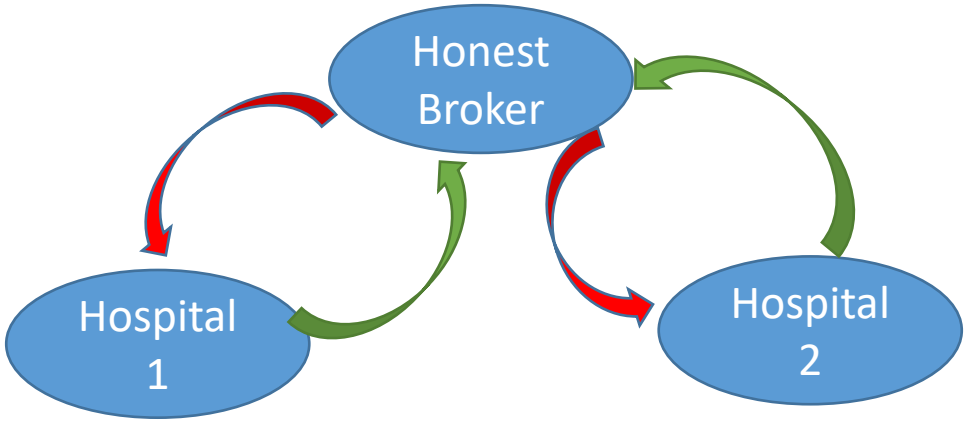
This requires alignment of EHR concepts, units, and agreement on a clear definition of the condition/outcome considered

Is essential to avoid time consuming hand review for every patient and allow for efficient analyses

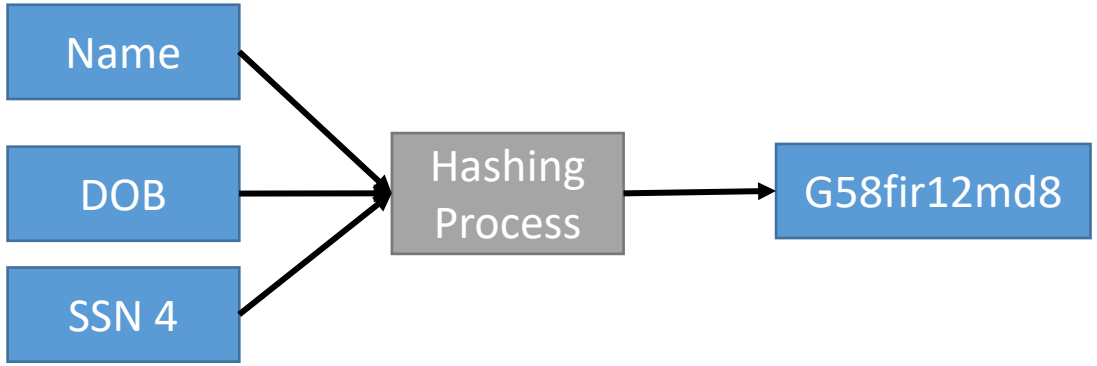
Preserving Privacy

Institutional Review Boards are a key component in protecting privacy

Honest Broker
A third party that receives identifiers and allows for matching across datasets without direct sharing of identifiers. May receive identified or de-identified data depending on the agreement



Patient Preserving Privacy Linkage
A means of creating "hashes" which allow for identifying patients without directly sharing identifiers like name or SSN. An honest broker may hold or even create the hashes



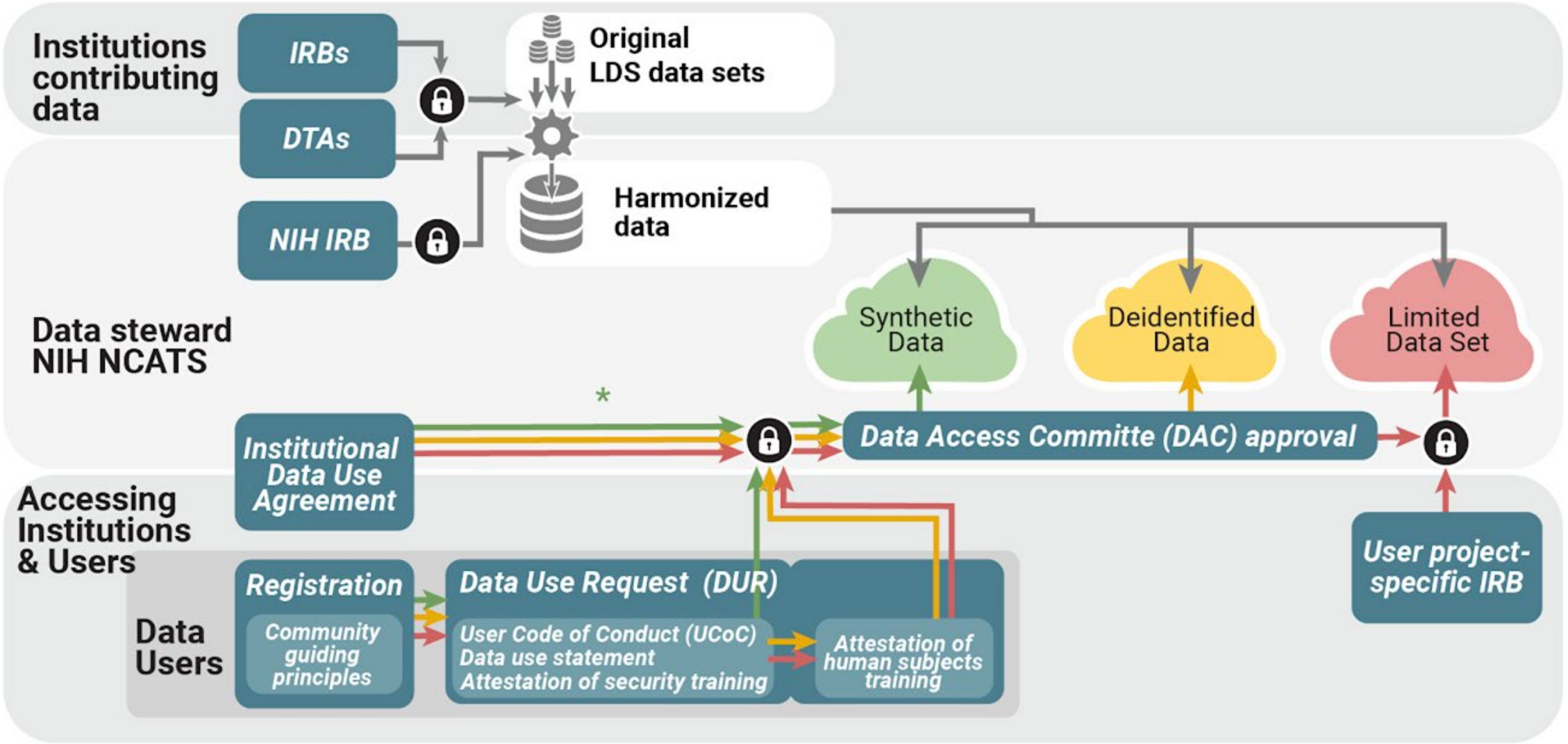
EPA United States Environmental Protection Agency

Infrastructure & Access Challenges

- Shared or local computing?
- How to manage resource allocation?
- How to share code & data and manage provenance?
- Tiered access or one level for all?
- How to manage & approve IRBs / project scope?

National COVID Cohort Collaborative

Bringing it all together: Example from N3C



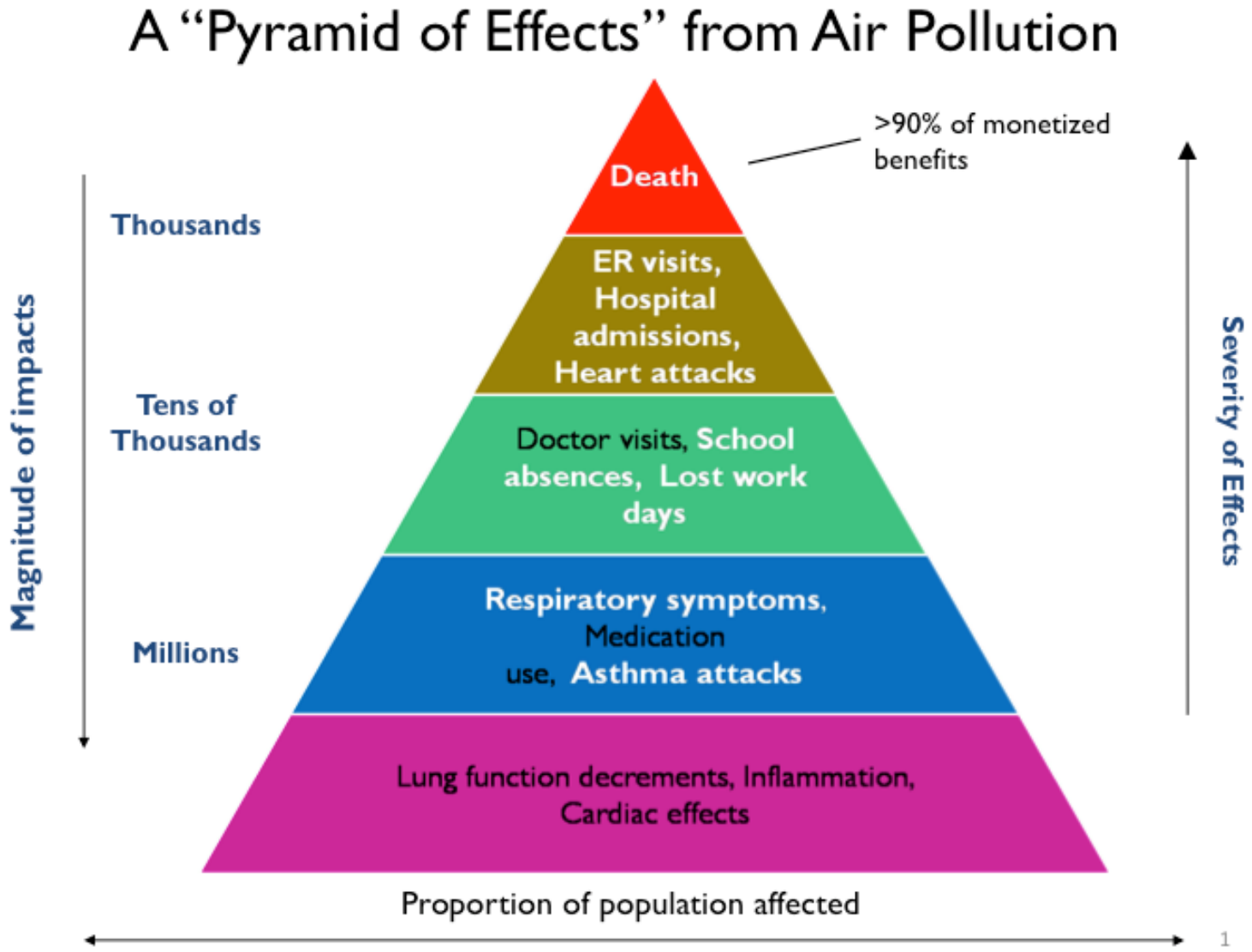
Research Using EHRs: Air Pollution, Readmissions, and Procedures

Problem

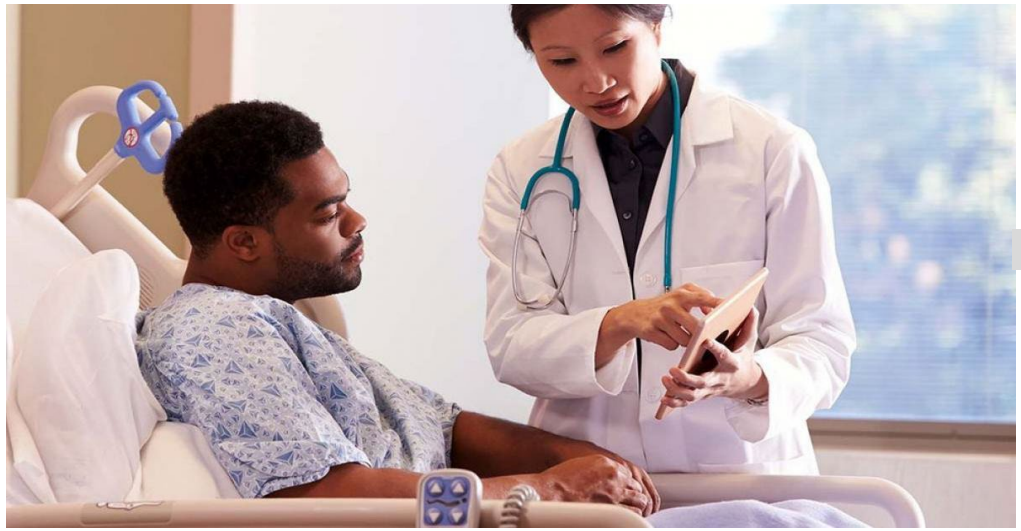
Environmental Health risks are greatest for the most vulnerable yet they are the least studied

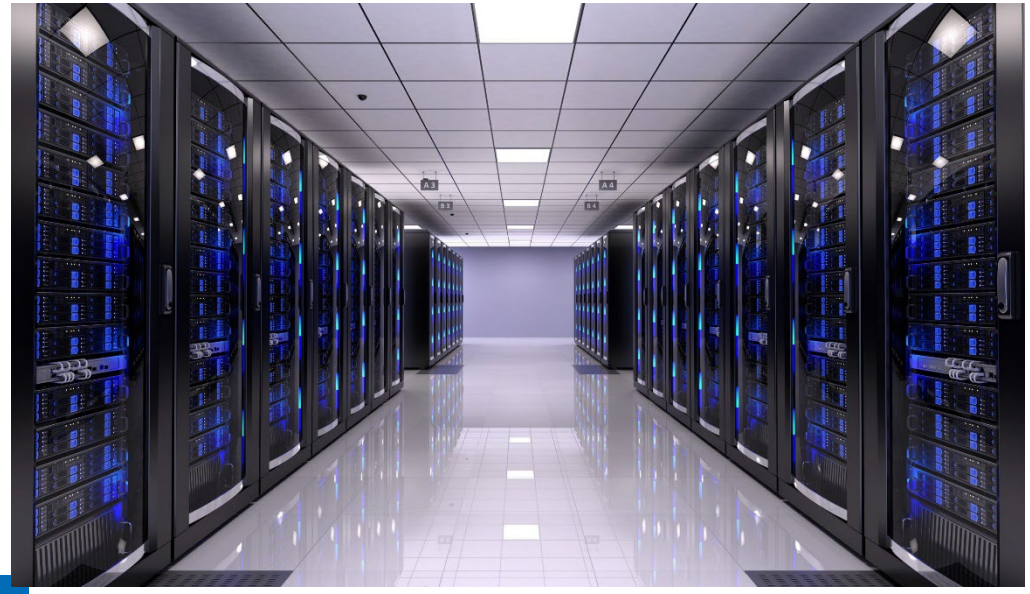
Problem

Large, longitudinal cohorts with individual data on vulnerable populations are hard, expensive, and time consuming to collect



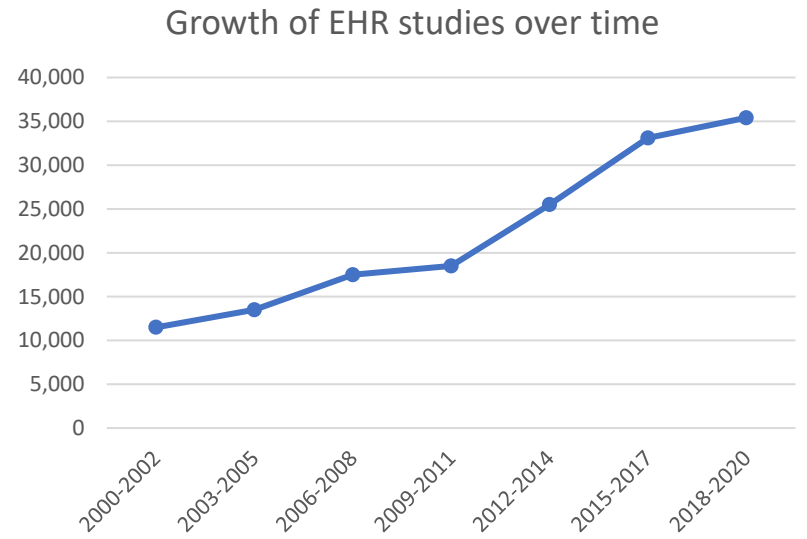
Solution: Electronic Health Records





Patient ID	Visit Date	ICD-10 Code	Clinical Lab	Vitals
A	March 1 2000	I10; K45.0	3.4	89
B	Feb 4 2020	J01.3	5.6	103
C	August 19 2014	I11	19.2	79
D	August 19 2014	I10	1.3	98

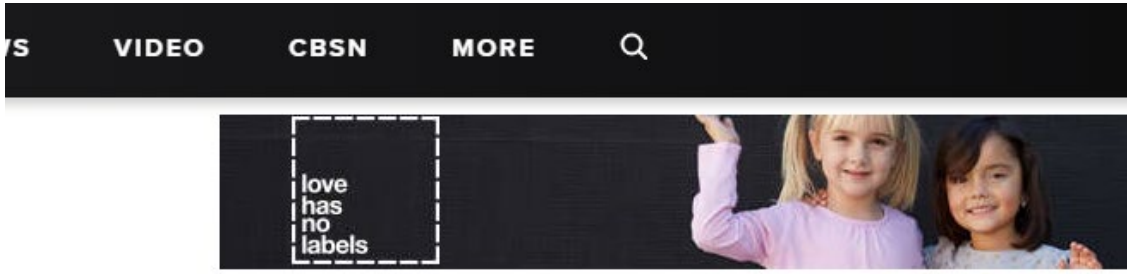
Patient ID	Age	Sex	Address
A	23	M	123 Main St
B	78	F	45 East Ave
C	67	F	9 Broad Dr
D	34	U	4 West St



Heart Failure

- 550,000 new cases per year
- Projected prevalence of 3% (1/33 individuals) in US by 2030
- Most common diagnosis in individuals 65+
- Represents 1 out of 8 deaths
- 5-year mortality rate ~50%
- Costs $30.7 Billion annually
- Known risk factors unchanging

https://www.cbsnews.com/news/heart-failure-rates-on-the-rise/



By ROBERT PREIDT / HEALTHDAY / *January 27, 2017, 1:49 PM*

U.S. heart failure rates on the rise

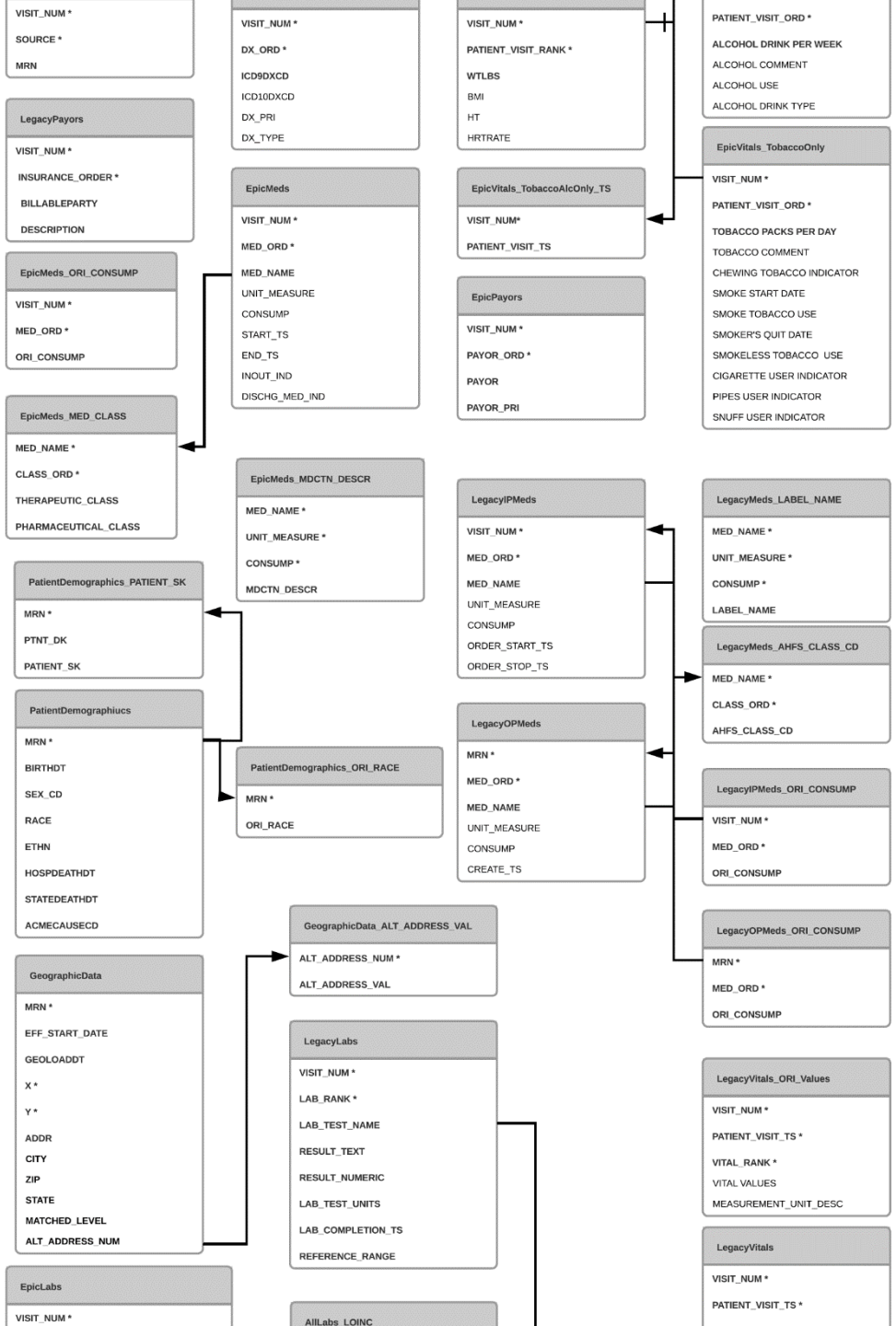
Share / Tweet / Reddit / Flipboard / Email

https://www.cdc.gov/dhdsp/data_statistics/fact_sheets/fs_heart_failure.htm
https://www.emoryhealthcare.org/heart-vascular/wellness/heart-failure-statistics.html
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5463661/pdf/zjch-7-1264696.pdf

http://circ.ahajournals.org/content/early/2017/01/25/CIR.0000000000000485

EPA Clinical and Archived records Resource for Environmental Studies (EPA CARES)

- A solution to the difficulty of rapidly constructing the needed longitudinal, well phenotyped cohorts that can answer pressing environmental health questions
- Merges EHRs from UNC Healthcare System (and soon others) with environmental exposure data to create a powerful environmental health resource
- Currently contains patient populations for heart failure (~40K), myocardial infarction survivors (~30K), a random sample of children (~30K) and adults (~20K)
- More than a dozen exposures incorporated including air pollution, radiation, contaminated sites, social determinants of health, …



EPA CARES is housed in a custom SQL database (~50 tables & > 10M observations) behind a secure medical firewall. Allows for secure, rapid data merging & extraction

EPA CARES Research Example: $PM_{2.5}$ & Hospital Visits & Readmissions

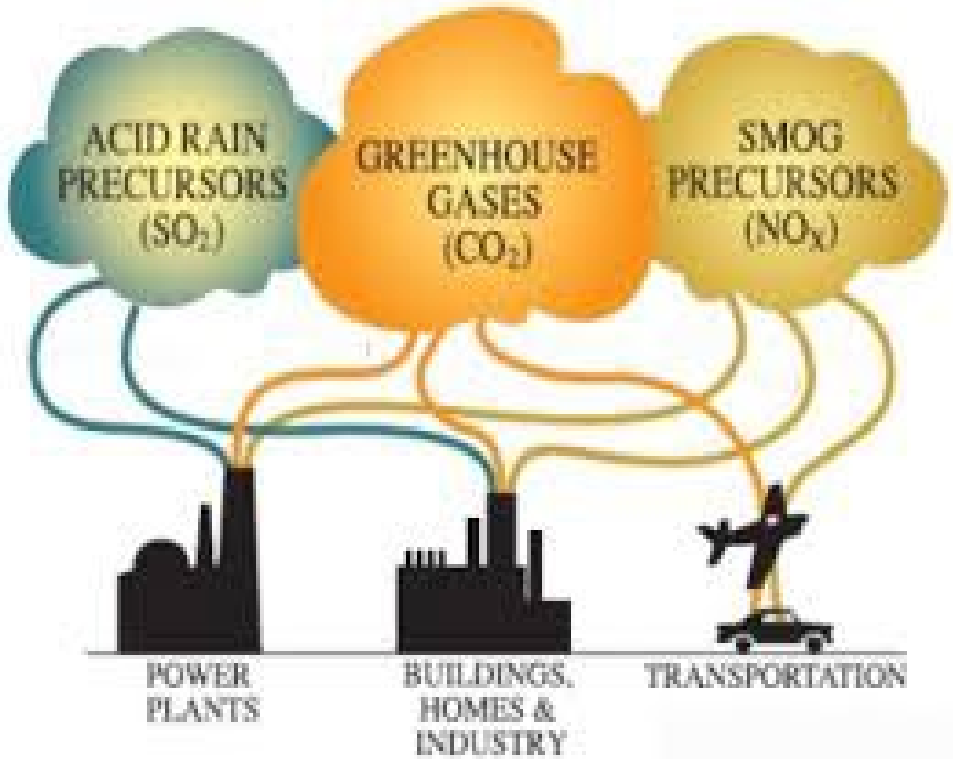
- Most current studies of hospitalization focus on short-term exposures using time-series studies. Very few studies of long-term exposures with individual patient data

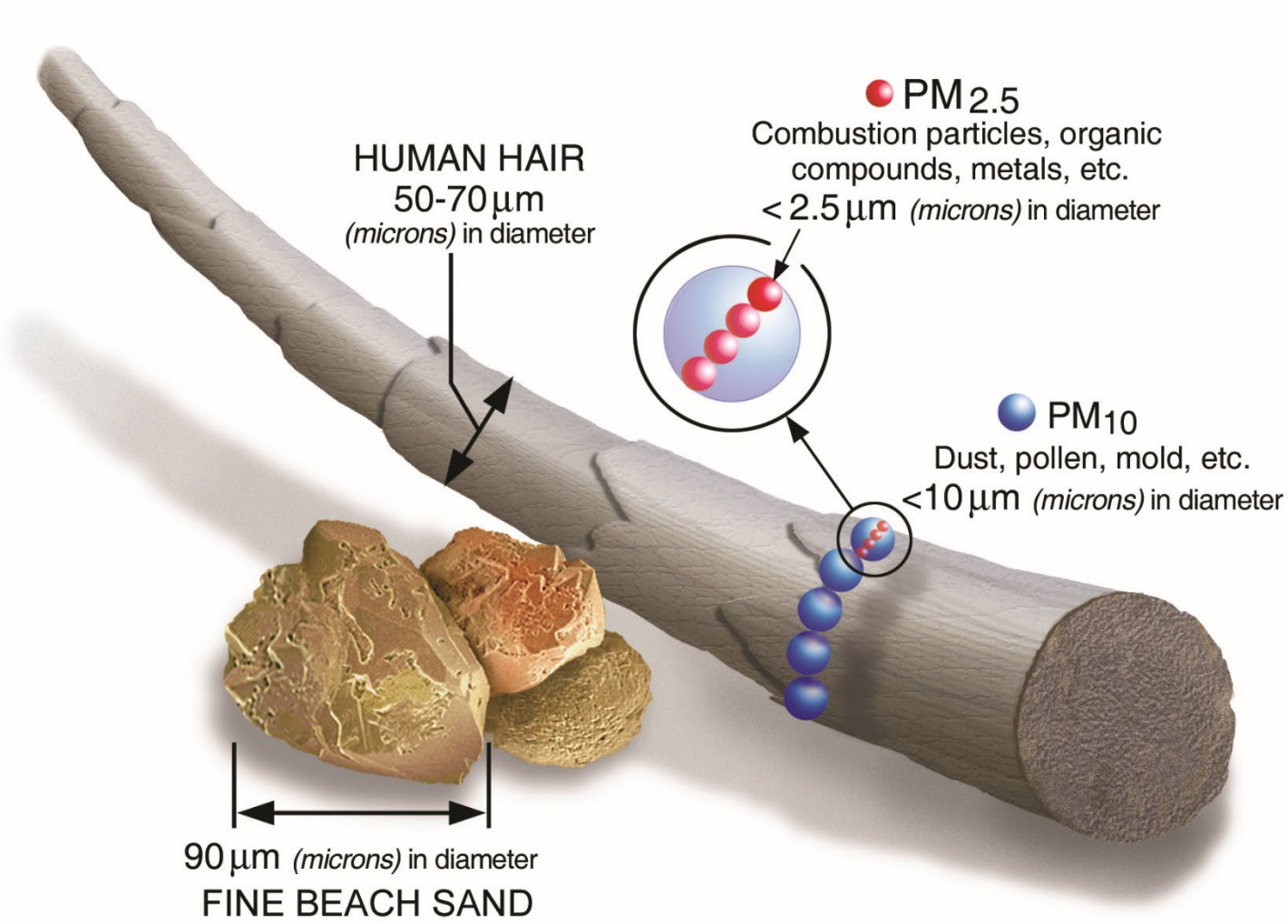
- Examined 442,244 visits from 20,920 Heart Failure patients (2004 – 2016) to understand if elevated $PM_{2.5}$ at the time of diagnosis is related to increased hospital visits

- Used Poisson regression with an offset of the log-follow up days while adjusting for age, race, sex, distance to monitor, urbanicity, year of HF diagnosis, neighborhood level socioeconomic status, and county-level access to healthcare, crimes, and access to healthy foods

- Used Inverse Probability Weighting to add causal interpretation and adjust for competing risk of death

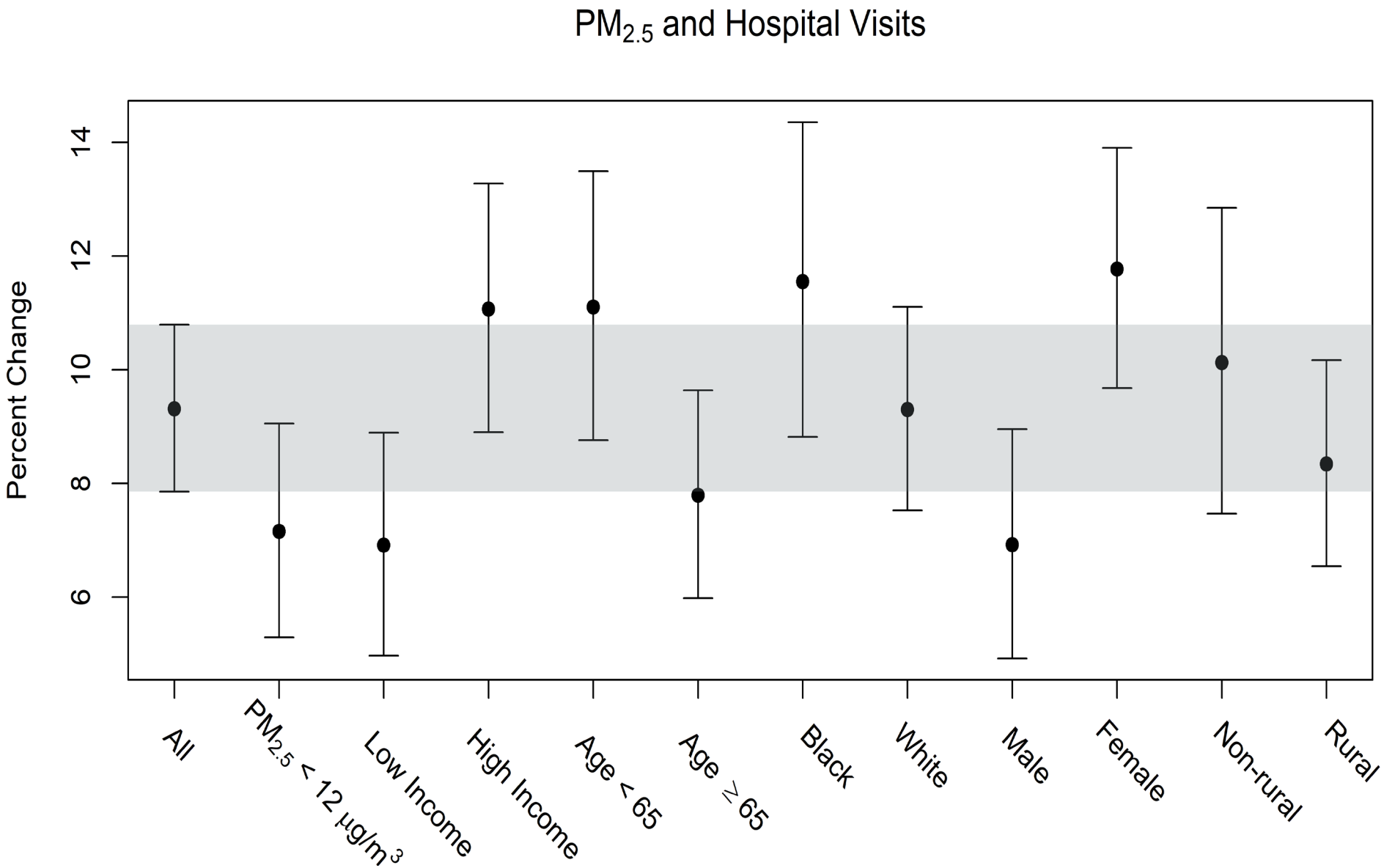
Various forms of Air Pollution



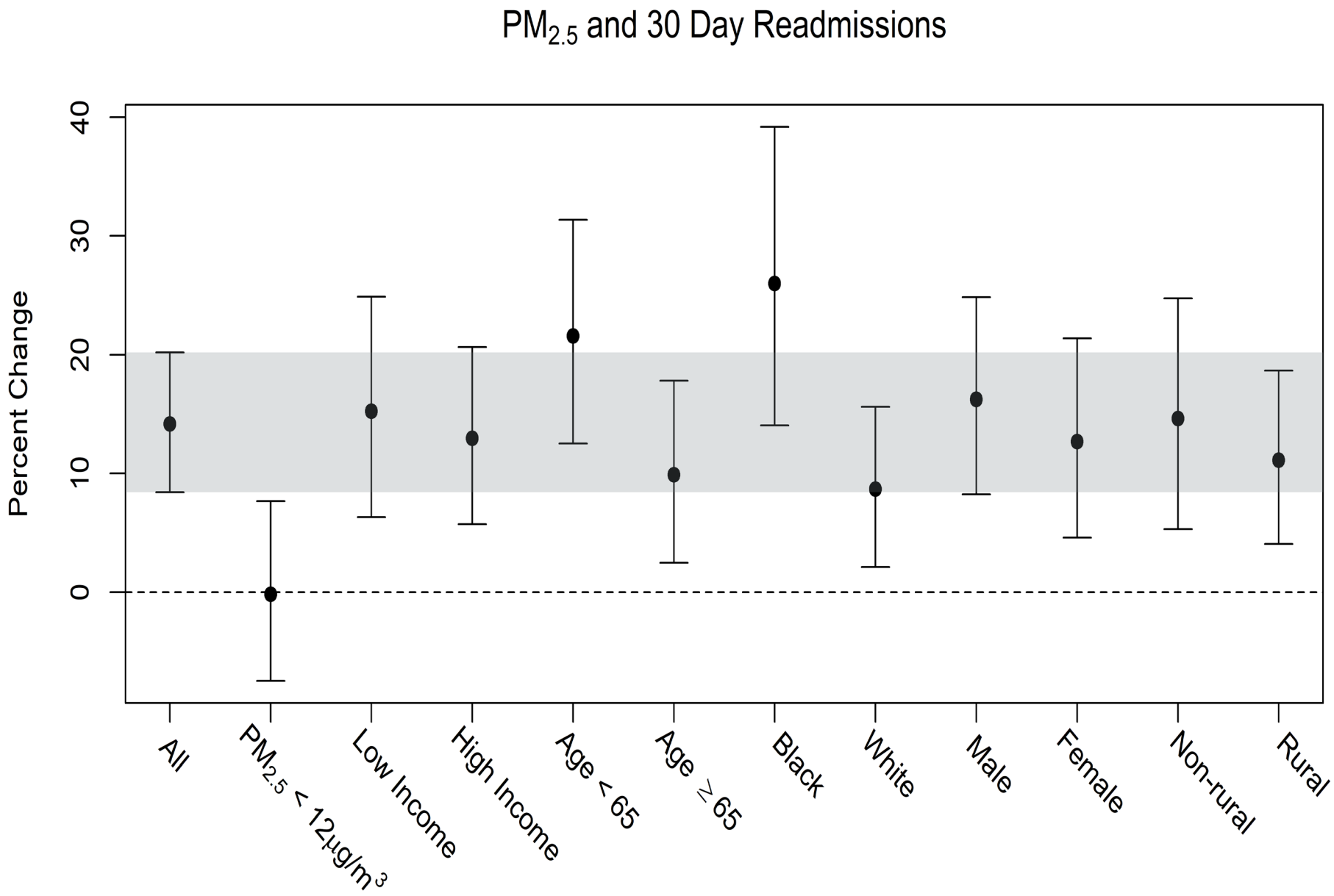


EPA United States Environmental Protection Agency

	Study Cohort (N = 20920)		Inpatient Admission Cohort (N = 12474)	
	Mean	SD	Mean	SD
Age (y)	68.8	14.8	69.0	15.1
Total Visits (N)	14.2	14.3	16.4	15.5
Inpatient Admissons (N)	1.34	1.77	2.29	1.79
30 day readmissions (N)	0.49	1.10	0.49	1.10
Follow up time (y)	2.79	3.05	3.02	3.21
$PM_{2.5}$ (ug/m3)	9.87	1.75	10.2	1.84
	N	%	N	%
Females	10998	52.6	6657	53.4
Males	9922	47.4	5817	46.6
Race - White	13875	66.3	8345	66.9
Race - Black	5564	26.6	3423	27.4



$PM_{2.5}$ health risks are highly varied even among the most vulnerable, an important observation for the field



$PM_{2.5}$ health risks are consistently elevated among Black study participants, highlighting the environmental risks faced by this community

Ongoing EPA CARES Projects

- What is the impact of contaminated sites (brownfields) on health and how can we estimate the impact of their remediation?
- How does air pollution impact the most vulnerable individuals?
- How does smoke exposure impact those with CVD differently than the general population?
- How do different sources of air pollution (traffic, biomass burning) differentially impact health outcomes?
- What are the interactions between social factors (e.g. poverty) and environmental exposures?
- What role did the environment have in exacerbating the coronavirus pandemic for people with existing CVD?
- Many more…

EPA United States Environmental Protection Agency

Successes & Lessons Learned

- One of the largest individual data studies of heart failure patients, air pollution, and hospital visits + readmissions
- Spent first months setting up database to house data – rarely done for epidemiology but KEY!
- Reusable, pre-cleaned data facilitates studies
- Health risks differ in some expected (higher risks among Black HF patients) and unexpected (higher risks for younger diagnosed patients) ways

CARES Current Impact

- Key insights into the environmental health risks of heart failure and myocardial infarction (MI) patients
- New connections between environmental exposures and hospital utilization (visits, admissions, procedures)
- Ongoing studies to quantify benefits of reducing contamination experienced by most vulnerable
- Generalizable lessons for next wave of environmental health & EHR intersections

Next Steps

1

Continued analysis of local EHRs & tapping into unique national resources (N3C)

2

Expanding Common Data Models to Incorporate Environmental Data

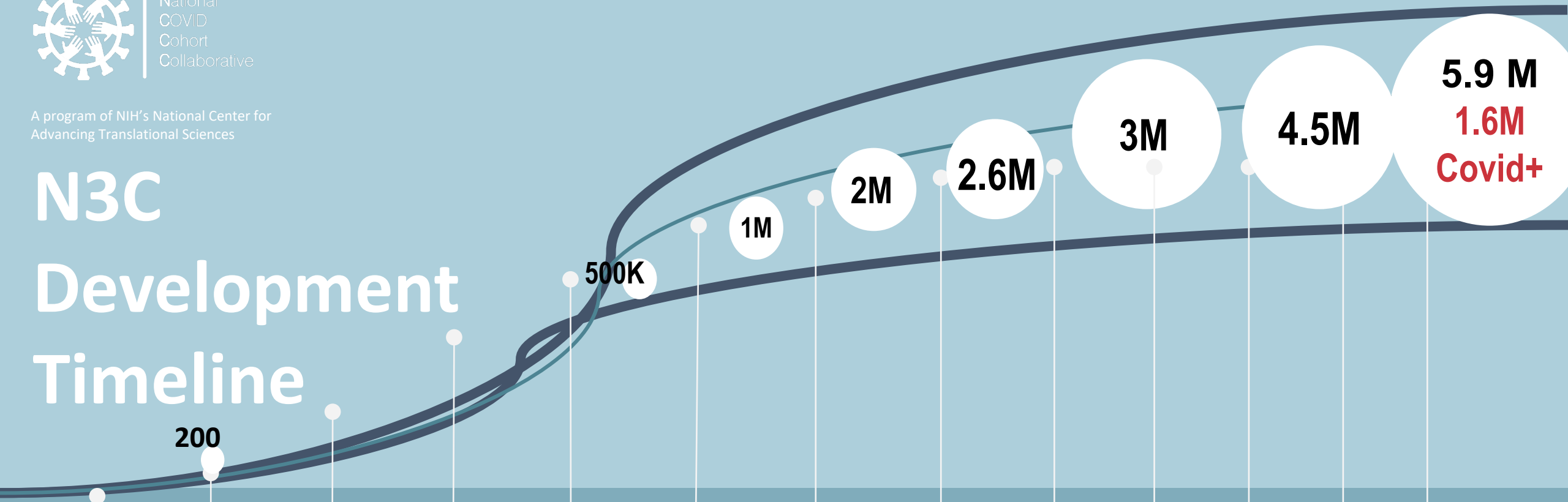
3

Establishing best practices that can be replicated

National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

N3C Development Timeline



Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
Kick off	1st DTA	17 DTAs	43 DTAs 4 DUAs	48 DTAs 29 DUAs	59 DTAs 88 DUAs 30 DURs 12 Sites Avail.	69 DTAs 116 DUAs 67 DURs 25 Sites Avail.	73 DTAs 129 DUAs 103 DURs 34 Sites Avail.	75 DTAs 139 DUAs 140 DURs 36 Sites Avail.	4 papers submitted	80 DTAs 160 DUAs 210 DURs 42 Sites Avail.	86 DTAs 186 DUAs 248 DURs 47 Sites Avail.	87 DTAs 199 DUAs 248 DURs 47 Sites Avail.	88 DTAs 199 DUAs 248 DURs 56 Sites Available
DTA Done	Palantir	DUA Complete	COC Complete	DAC Charter	NIH IRB	Synthetic Avail	201 Objects Va	1st IDeA CTR Ext Data Set Knowledge Store Cohort Paper	15 papers in Publication committee review	N3C Public Dashboard	24 Domain teams		2385 Registered Users
IRB JHU	Phenotype 1.0	CDMs Mapped	1st Manuscript	Initial Training	Support Desk		PPRL Contract	516 Invest		External datasets	40+ pre/ publications		199 Research Projects
G-Suite			Domain Teams		Harm/QA Comp.								

17May21

EPA United States Environmental Protection Agency

Planned N3C Projects

- What are the environmental health risks faced by those who had a coronavirus infection or developed COVID19?
- Does air pollution or other environmental exposures lead to a greater risk of long COVID?
- Did environmental exposures exacerbate the pandemic – particularly for marginalized & historically discriminated communities?
- Would machine learning algorithms for clinical risks be enhanced by incorporating environmental exposures?

Best Practices

As we move forward it is important to build upon some of the best practices we have learned and generalize/formalize them

- How to construct environmental subsections of EHR databases?
- Procedures for constructing “analysis ready” datasets?
- Methodologies for analyzing environmental data with EHRs
- Reporting that makes analyses useable by clinicians and public health officials

Conclusions

- Electronic health records are an increasingly utilized tool in environmental health studies
- Though filled with challenges, EHR resources are emerging that facilitate novel and important studies
- Increasing harmonization (via CDMs) and incorporation of environmental data into CDMs will drive a new wave of environmental health studies
- Performing studies in a way that can be utilized by patients, public health officials, and communities will be key

Contact

Dr. Cavin Ward-Caviness
Computational Biologist (Principal Investigator)
Public Health and Integrated Toxicology Division
US Environmental Protection Agency
ward-caviness.cavin@epa.gov

Acknowledgements

- Matthew Buranosky
- Wayne Cascio
- Bob Devlin
- David Diaz-Sanchez
- Anne Weaver
- Joshua Moyer
- Emily Pfaff
- Marshall Clark
- All members of the Ward-Caviness lab, the IT staff supporting these resources, and the many people who make these efforts possible