Innovations in Research Approaches: Data Aggregation and Linkage

NASEM Prepregnancy BMI and Gestational Weight Gain Workshop

September 5, 2025

Stephanie A. Leonard, PhD, MS
Assistant Professor
Dunlevie Center for Maternal-Fetal Medicine
Stanford University School of Medicine

Stanford University

Disclosures

- No conflicts of interest

Outline

- Promises & pitfalls of studying prepregnancy BMI and gestational weight gain using big data sources:
  - Vital statistics data
  - Medical claims data
  - Electronic health record (EHR) data
- Conclusions & research gaps

Vital statistics data

- Data are collected on Certificates of Live Birth and Fetal Death
- Typically, forms are completed during the birth hospitalization then reported to state health departments and the U.S. National Vital Statistics System of the CDC
- The CDC revised forms in 2003 to include fields for height and for weight prepregnancy and at birth
  - Fully implemented in the U.S. in 2016

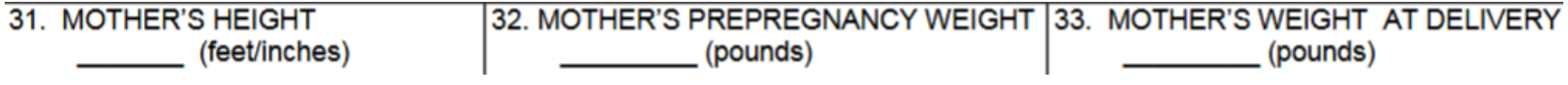
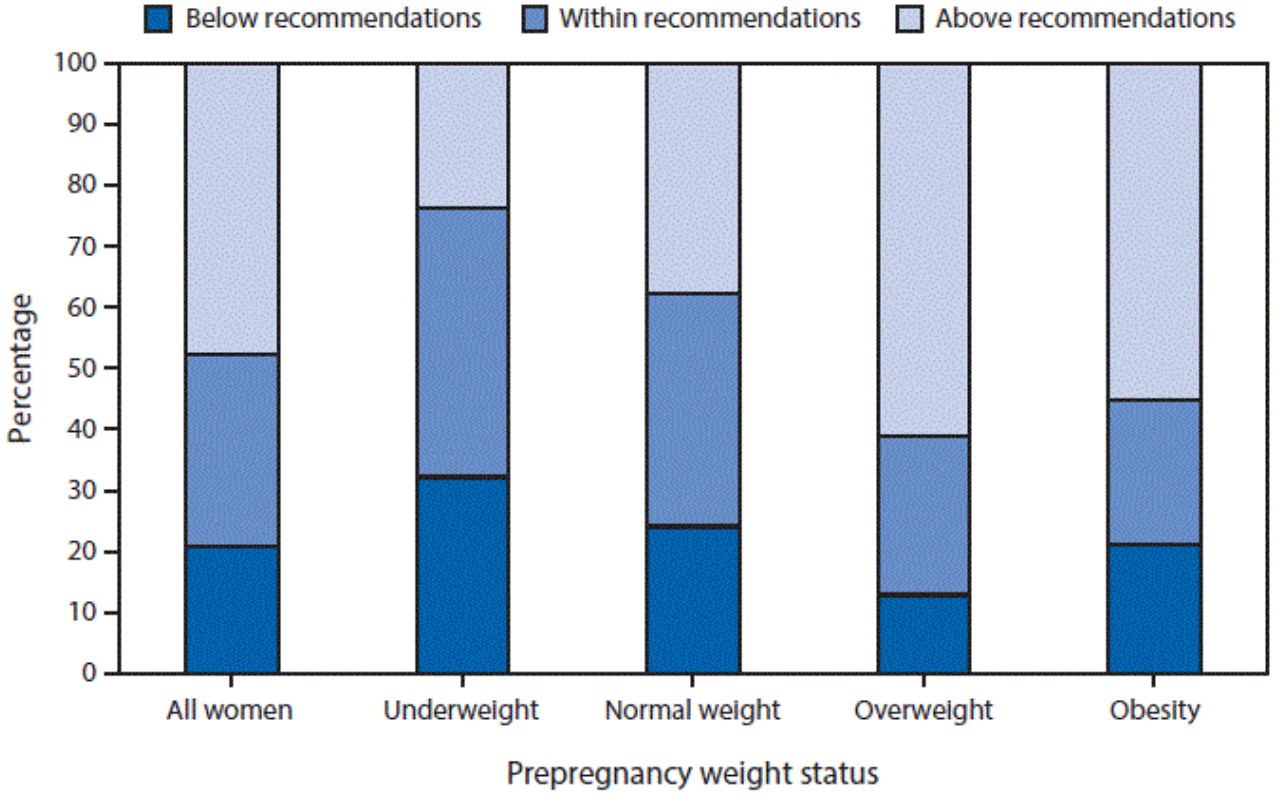


Image from: https://www.cdc.gov/nchs/data/dvs/birth11-03final-acc.pdf

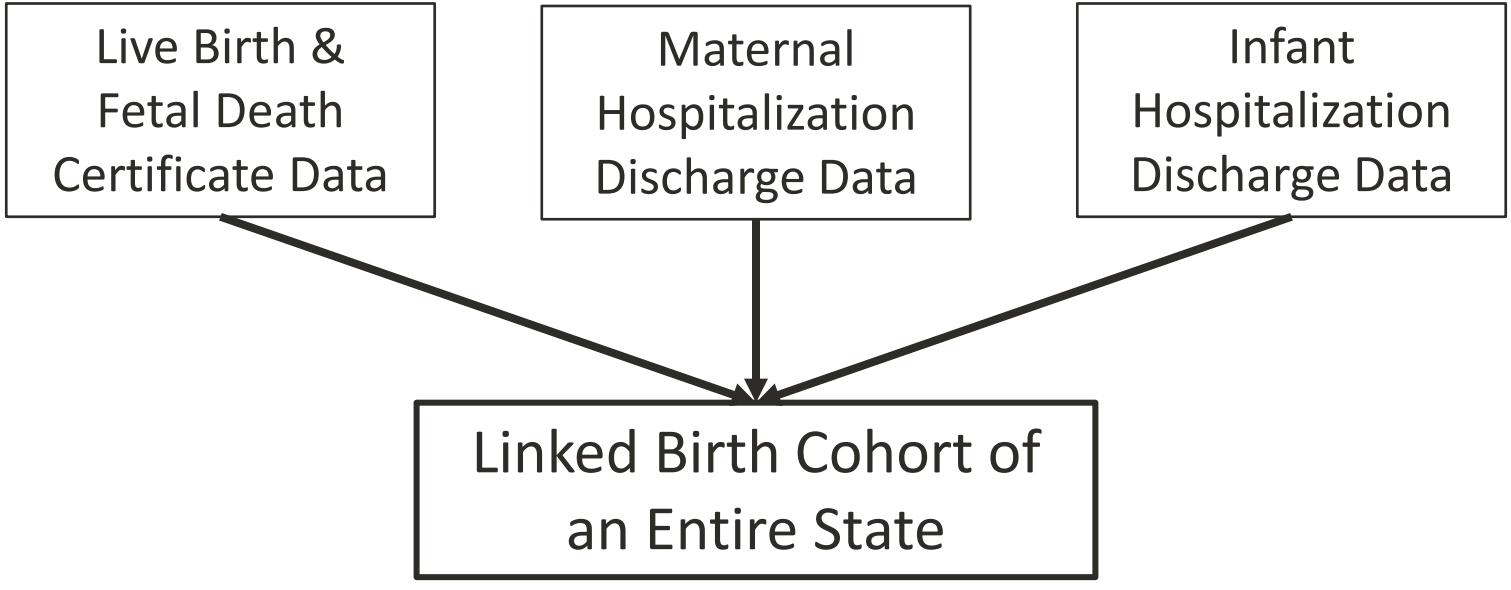
Promises of vital statistics data

Large sample size and high generalizability



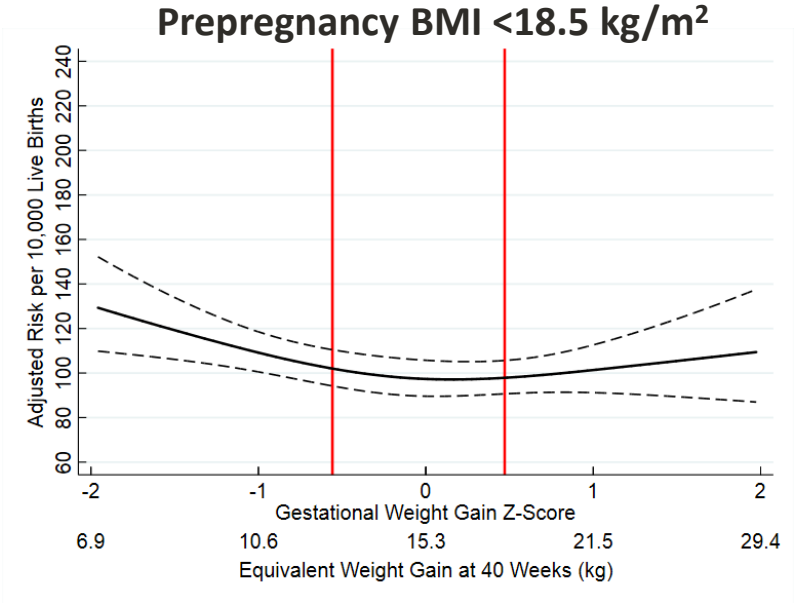
Publicly accessed at: https://www.cdc.gov/mmwr/volumes/65/wr/mm6540a10.htm. *QuickStats: Gestational Weight Gain Among Women with Full-Term, Singleton Births, Compared with Recommendations — 48 States and the District of Columbia, 2015. MMWR Morb Mortal Wkly Rep 2016;65:1121. DOI: http://dx.doi.org/10.15585/mmwr.mm6540a10.*

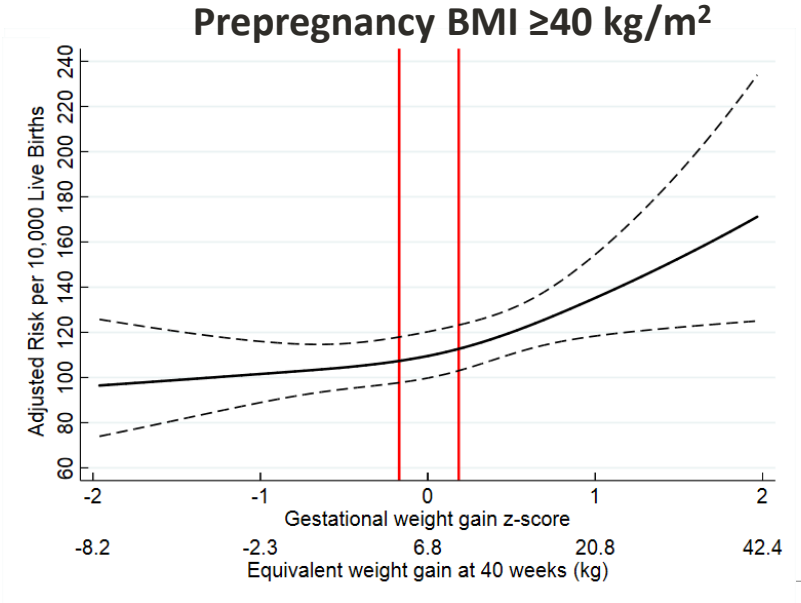
Ability to link to other data sources



Strength in size, generalizability, and linkage

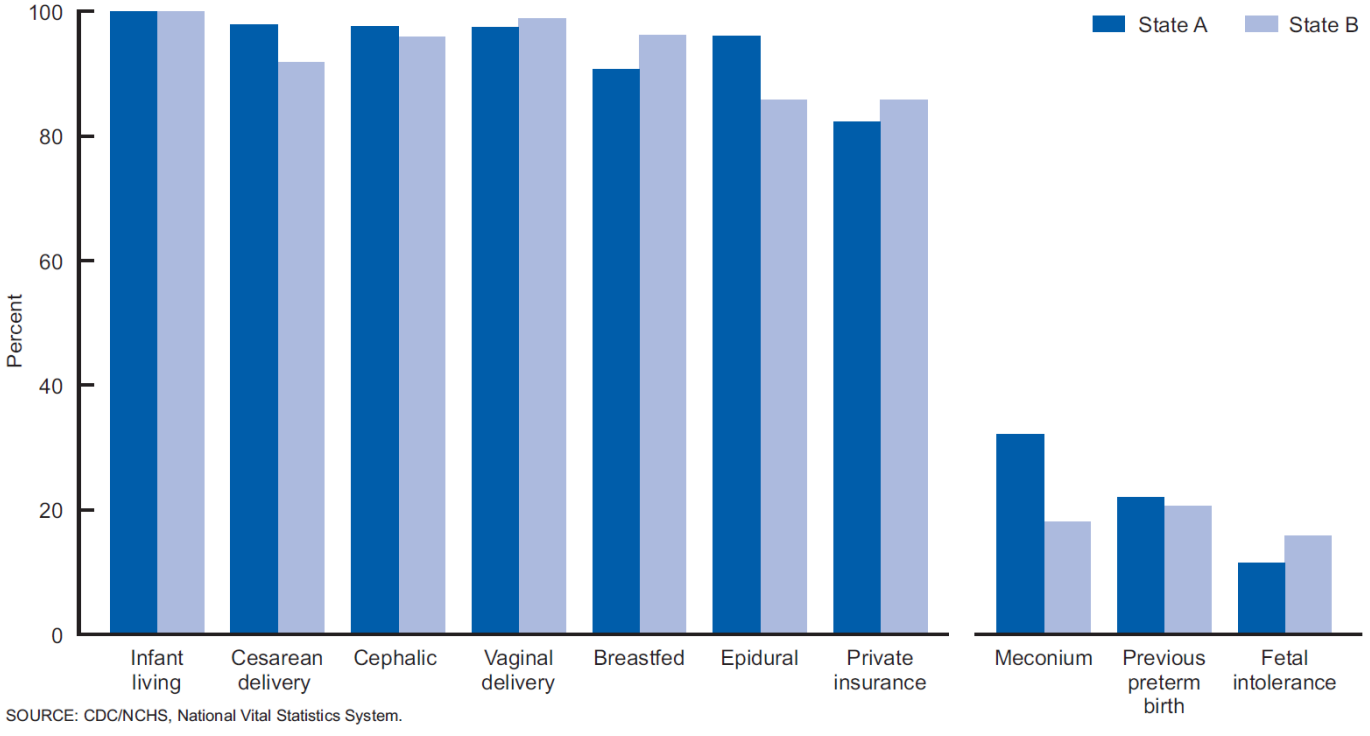
Example of assessing the risk of severe maternal morbidity across gestational weight gain values and stratified by prepregnancy BMI group in 2.5 million people





Reported in Leonard SA, Abrams B, Main EK, Lyell DJ, Carmichael SL. *Am J Clin Nutr* 2020;111:845-853.

Multiple useful data fields have high validity



Publicly accessed at: https://www.cdc.gov/nchs/data/nvsr/nvsr62/nvsr62_02.pdf. Martin JA, Wilson EC, Osterman MJK, et al. *Nat Vital Stat Rep* 2013;62(2)

Pitfalls of vital statistics data

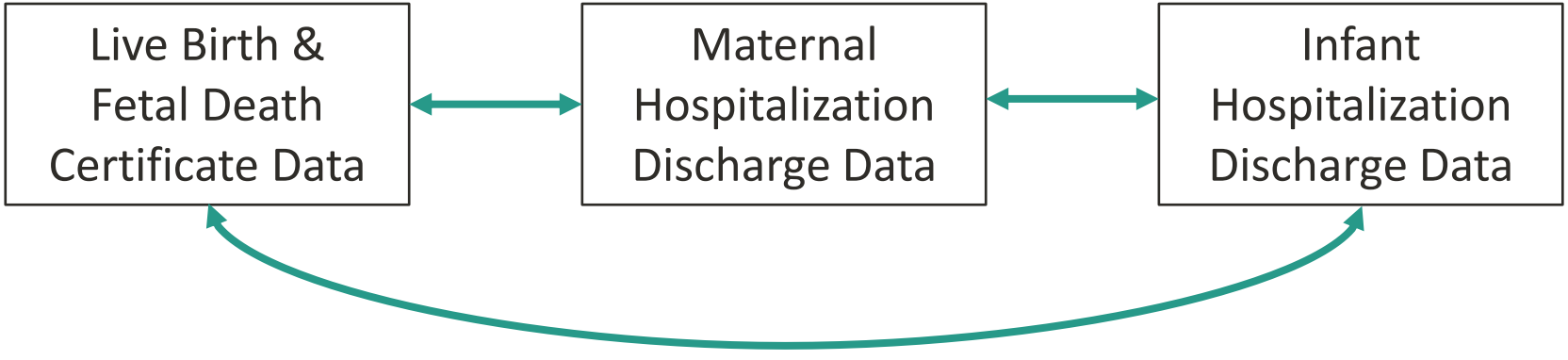
Validity of weight measurements

- Prepregnancy BMI and gestational weight gain reported in vital statistics data have been found to be overall consistent with medical records[1,2]
- However, validity found to vary by patient characteristics[1,2]
  - Separate studies have reported lower validity for Black patients than White patients
- Systematic review found people tend to underreport prepregnancy and delivery weight and overreport gestational weight gain → moderate misclassification of BMI/GWG categories that largely does not bias associations[3]

[1]Bodnar LM, Abrams B, Bertolet M, et al. *Paediatr Perinat Epidemiol* 2014;28:203-212
[2]Deputy NP, Sharma AJ, Bombard JM, et al. *Epidemiology* 2019;30(1):154-159
[3]Headen I, Cohen AK, Mujahid M, Abrams B. *Obesity Rev* 2017;18(3):350-369

Challenges in linking to other datasets

- Valid clinical information in vital statistics is limited, but linkage to other datasets is typically probabilistic and often prohibitive



Medical claims data

Medical claims data

- Claims data include patient encounter information, including diagnoses and procedures
- Claims data may also include information on:
  - Billed and paid amounts
  - Demographics
  - Enrollment in an insurance program
  - Medications
  - Laboratory values
- Commonly used nationwide claims datasets include Medicaid research files, Merative™ MarketScan®, Optum Clinformatics, Veterans Affairs, National Inpatient Sample, others

Promises of claims data

- Large, generalizable patient groups
- Data often longitudinal
- Detailed information on diagnoses, procedures, and often medications
- Multiple types of pregnancy outcomes (not limited to live births)
- Can be linked to other datasets: vital statistics, EHR data

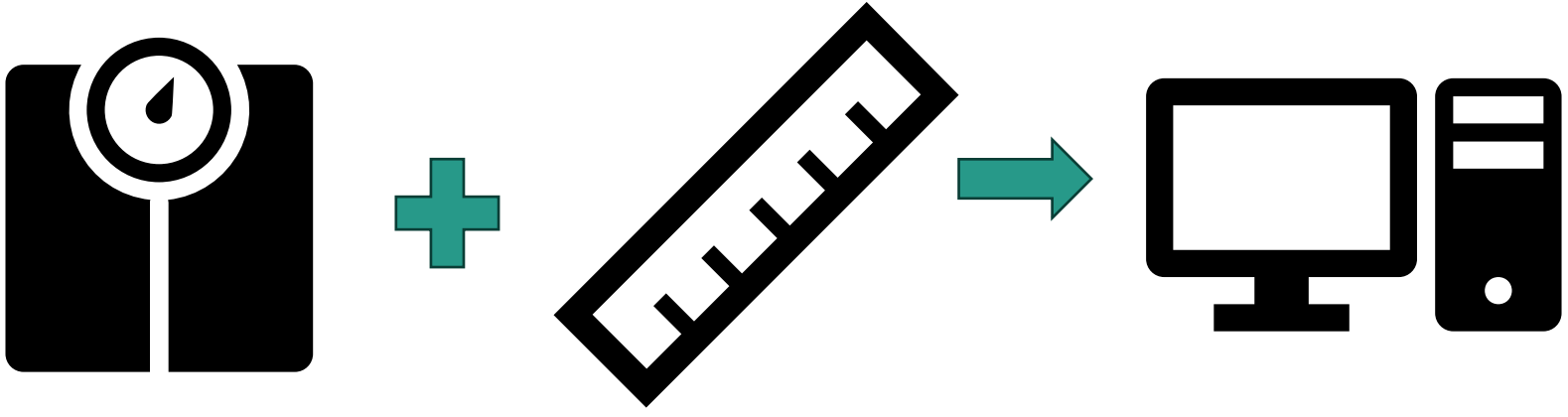
Pitfalls of claims data

- Limited information on weight, height, and demographic details
  - Often reliant on ICD-10-CM codes for BMI group
- Barriers to linking maternal-child pairs
- Lag time from data collection to research
- Often costly or difficult to obtain access
- Data management, storage, and analysis can be challenging

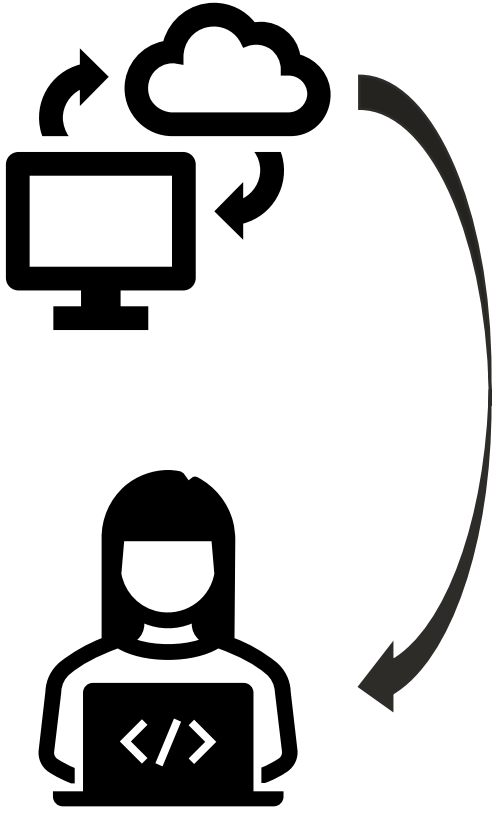
Electronic health record (EHR) data

Promises of EHR data

Weight & height measured by healthcare professionals & recorded in EHR system



Data aggregated and shared for research purposes



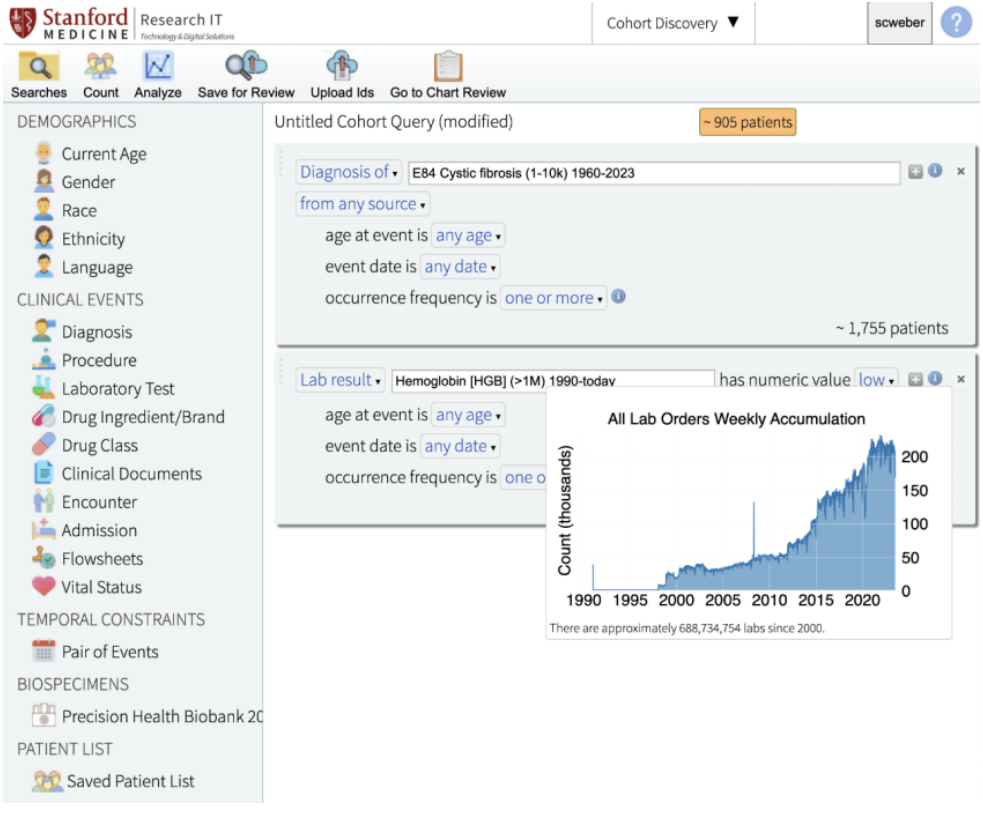
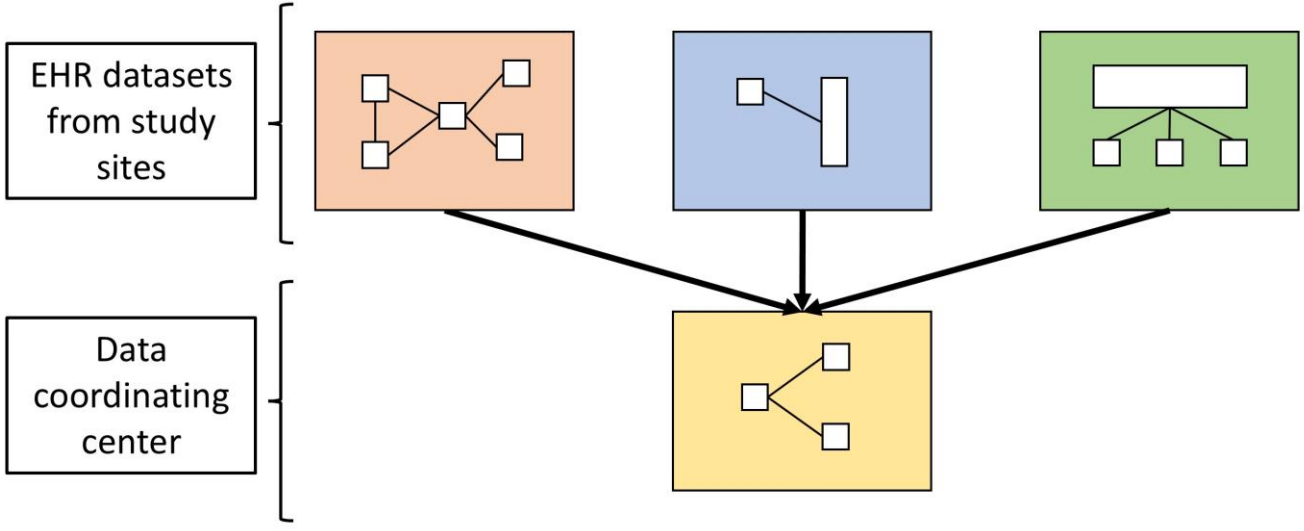
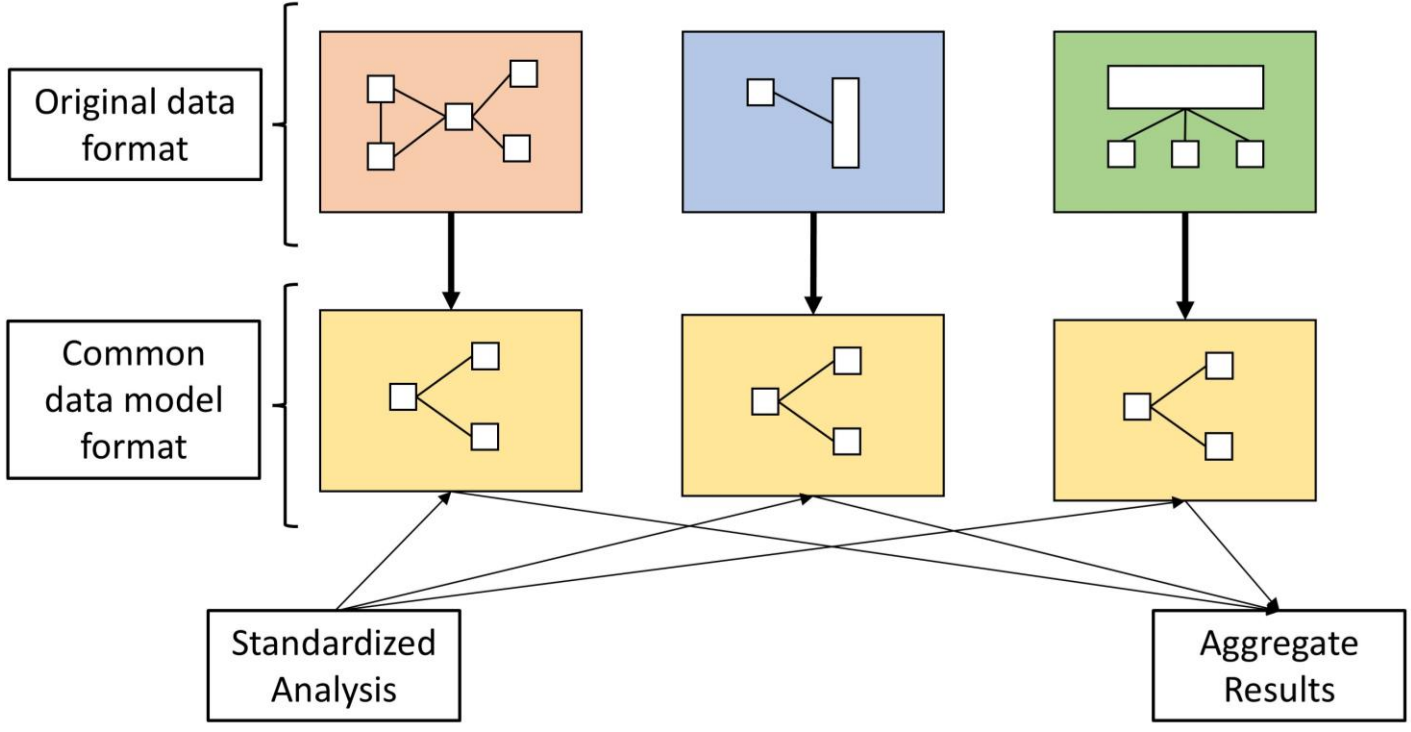


Image from: https://med.stanford.edu/starr-tools/cohort-discovery.html

Multi-site EHR cohorts can have higher generalizability and sample size

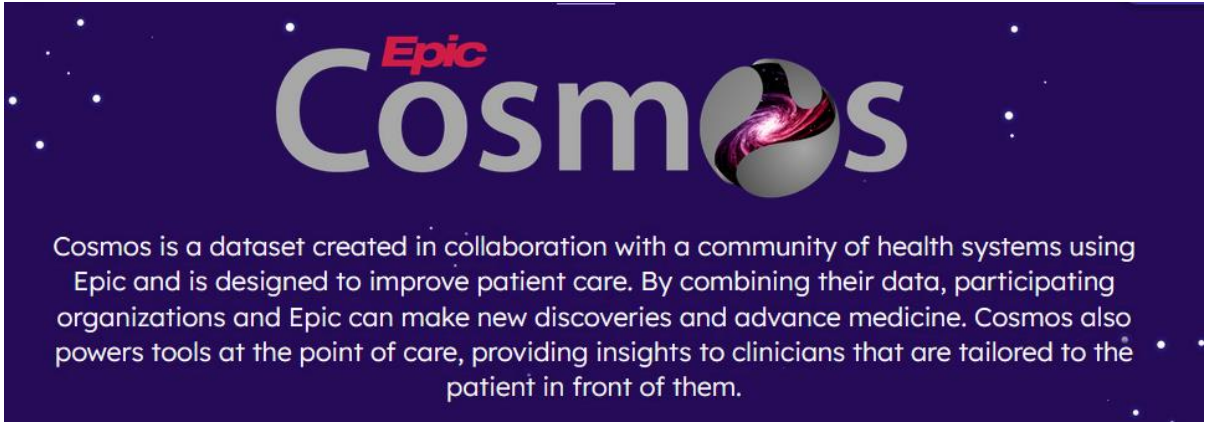


Distributed data network studies leverage common data models for reproducibility & rigor: OHDSI



OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Centralized EHR datasets have massive sample sizes: Epic Cosmos



Images from: https://cosmos.epic.com/

Pitfalls of EHR data

Multiple potential sources of bias

- Missingness of weight & height data can be high and differentially higher in healthier people[1,2]
- Challenging to identify and correct erroneous values[3]
- Weight is not universally measured at start and end of pregnancy
  - Weight at first prenatal care visit is often used as proxy measure of prepregnancy weight, and weight at last prenatal care visit as proxy measure of delivery weight
  - Depending on visit timing, this could result in biasing prepregnancy BMI up and gestational weight gain down

1. Rea S, et al. *AMIA Jt Summits Transl Sci Proc* 2013;214-218
2. Baer HJ, et al. *JAMA Internal Med* 2013;173(17):1648-1652
3. Guide A, et al. *J Biomed Informatics* 2024; 104660

Conclusions and research gaps

Conclusions

- Big data sources have advanced in recent years and have important strengths of large size and generalizability
- Researchers and policy makers should consider the strengths and weaknesses of big data sources for prepregnancy BMI and gestational weight gain

Research gaps

- Updated validation studies for trajectory of weight from before to the end of pregnancy in different big data sources
- Methods to improve the internal validity of research using maternal weight, height, and BMI values in big data sources

Thank you

stephanie.leonard@stanford.edu
https://med.stanford.edu/leonard