BOSTON MEDICAL CENTER

Clinical Data Warehouse for Research

Using Clinical Data for Research: Real-World Case Studies in Data Literacy

Melissa Hofman, MSIS – Research Informatics Director, CDW for Research
Heather Hsu, MD MPH – Scientific Director, CDW for Research
June 8, 2022

CDW for Research – Two-Part Series

BOSTON MEDICAL CENTER
Clinical Data Warehouse for Research

Real-World Case Studies in Data Literacy

June 8, 2022

Learning Objectives:

1. Describe key concepts in data literacy and the data lifecycle that are important to understand when using electronic health records data for research.

2. Understand some basic "truisms" of using electronic health records data for research.

3. Consider *what* you are measuring and *why* you are measuring it when planning your next research study.

Watch Part 1: BMC Clinical Data CDW for Research

Recorded May 11, 2022

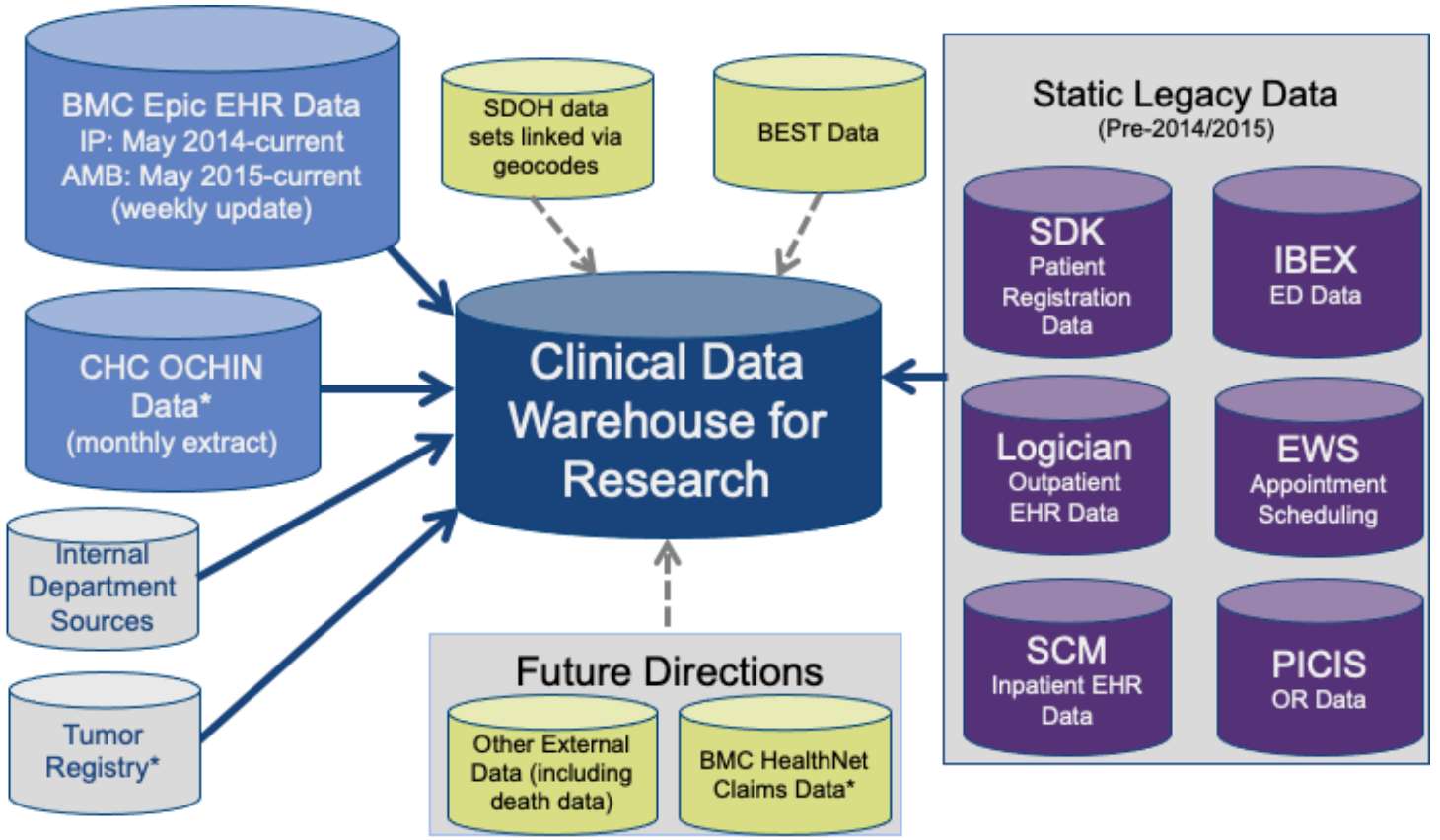
Learn the nuts & bolts of the CDW for Research, including data available in the data warehouse, services offered by the CDW for Research, and how to request data.

tinyurl.com/CDWpart1

The BMC Clinical Data Warehouse for Research

Boston Medical Center — Clinical Data Warehouse for Research

- Centralized resource to **access patient-level and population-level data for research**
  - Electronic health records data (Epic)
  - Historical data from BMC's legacy clinical systems
  - CHC OCHIN EHR data
  - Piloting BMCHP claims data for research purposes
- Provides **simple aggregate counts** to prepare for grants or studies
- Provides **data extracts**
  - Cohort identification, recruitment lists, and data sets for prospective studies
  - Data sets for observational/retrospective studies
  - **Linked data sets** from multiple sources related to each other with **patient-level unique identifiers**



*special permission required to access

*The CDW for Research also **collaborates with Departments and Divisions to increase research infrastructure** and **leverage data** for research purposes.*

BOSTON MEDICAL CENTER

Clinical Data Warehouse for Research

DATA LITERACY: What is it?

Clinical Research Data Literacy

- The ability to understand the data through its lifecycle to allow for the proper **conceptualization, operationalization, extraction, and interpretation and communication** of the data

- **Data literacy** enables you to transform data into information, information into knowledge, and knowledge into wisdom

- Essentially, without a **basic understanding of data and intentionality**, one is susceptible to the misuse of data and misinterpretation

Research Data Lifecycle

BOSTON MEDICAL CENTER
Clinical Data Warehouse for Research

RESEARCH
DATA
LIFECYCLE

Key Concepts

- **We need to move beyond 'Garbage in, Garbage out'**
  - Data may be "good" going in, but the way that the data are stored and/or extracted impact your ability to analyze it
  - You can't "analyze away" bad data collection and documentation practices

- **Measure twice, cut once principle**
  - Answers to questions related to what you are measuring and why can take a data request in many different directions
  - Particularly with observational research, the bulk of the work happens in conceptualizing your concepts, obtaining and cleaning the data and setting up your data set
  - Actually "crunching the numbers" should be trivial when your data are set up well

- **It's not just about the data that you ask for – *it is about what the data lifecycle allows you to measure***
  - Misspecification in the early stages will snowball into bigger problems across the lifecycle
  - For interventional research, thinking about how to improve data capture in Epic/notes ***before*** *a* trial starts greatly increases what you can confidently measure and say with the data during trial analysis

BOSTON MEDICAL CENTER

Clinical Data Warehouse for Research

REAL-WORLD CASE STUDIES: Getting to ‘quality in, quality out’

Gaining an Appreciation for “the Trees”

Conceptualizing & Operationalizing a Research Question

Boston Medical Center
Clinical Data Warehouse for Research

Conceptualization:

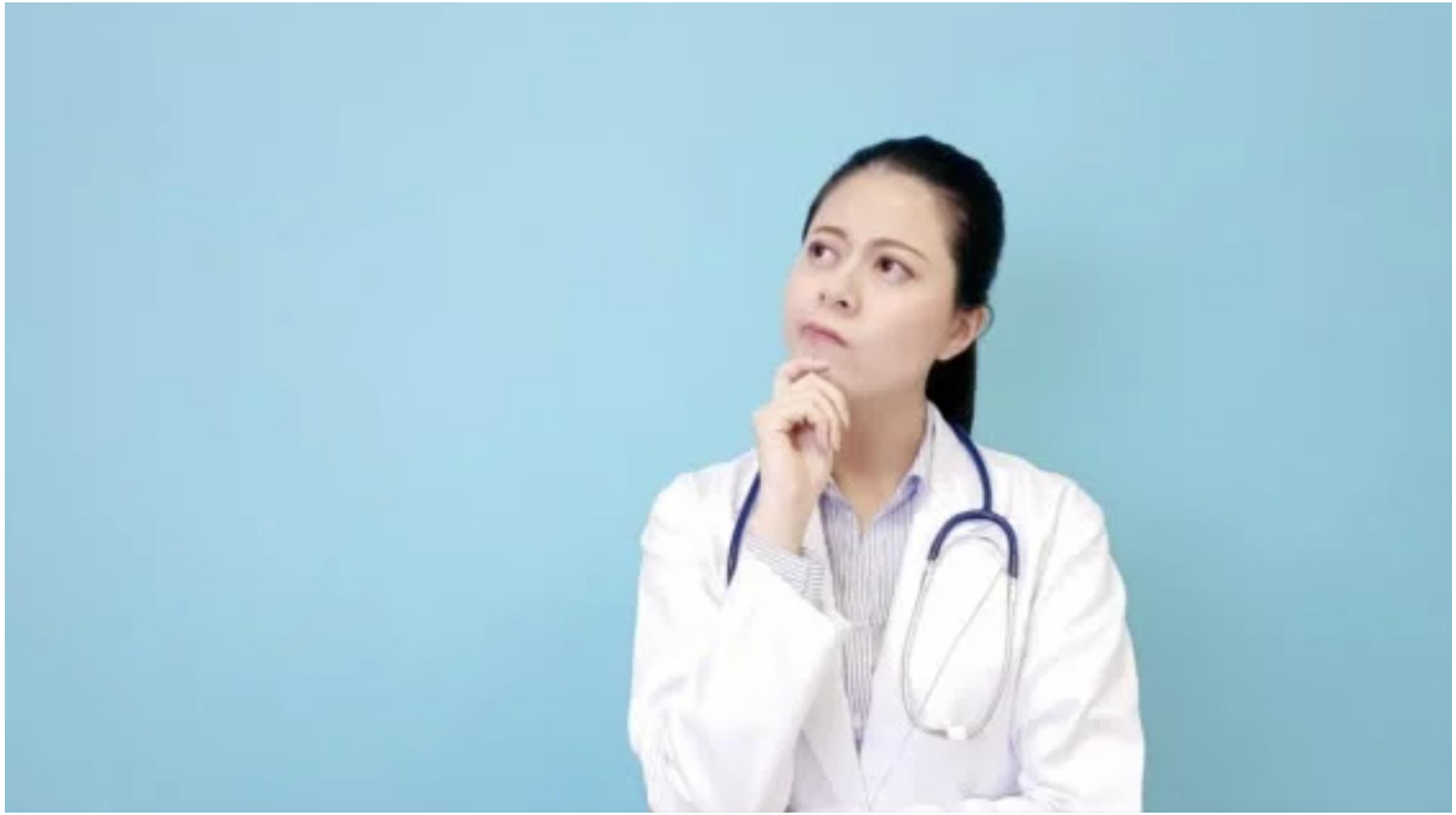
- What are the concepts (i.e., variables & outcomes) that you plan to study?
  - Examples of concepts: age, obesity, long COVID, ectopic pregnancy, social disadvantage, medications for opioid use disorder, readmissions, delay in diagnosis
- What is your research question?
- What study design will you use to answer your research question?

Operationalization:

- How will you define and measure the concepts that you plan to study?
- What data will you use?
  - Options: primary data collection vs. secondary use of data

Case Study: Delayed Presentation for Gallbladder Disease

BOSTON MEDICAL CENTER
Clinical Data Warehouse for Research



Case Study: Delayed Presentation for Gallbladder Disease

Boston Medical Center
Clinical Data Warehouse for Research

Study question: Is a patient's socioeconomic status associated with delayed presentation for gallbladder disease?

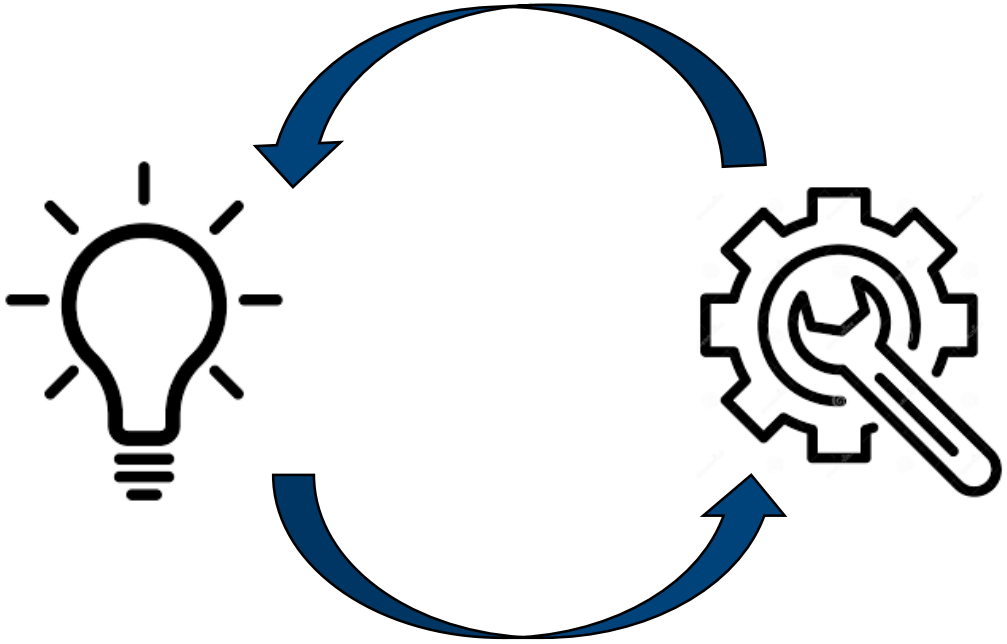
Hypothesis: Patients with lower socioeconomic status will present later in the course of gallbladder disease and will be more likely to require emergency surgery for gallbladder removal

Study design: Observational crossectional study

Case Study: Delayed Presentation for Gallbladder Disease

Boston Medical Center
Clinical Data Warehouse for Research

Is a patient's socioeconomic status associated with delayed presentation for gallbladder disease?



**Define concepts*
&
Refine research question**

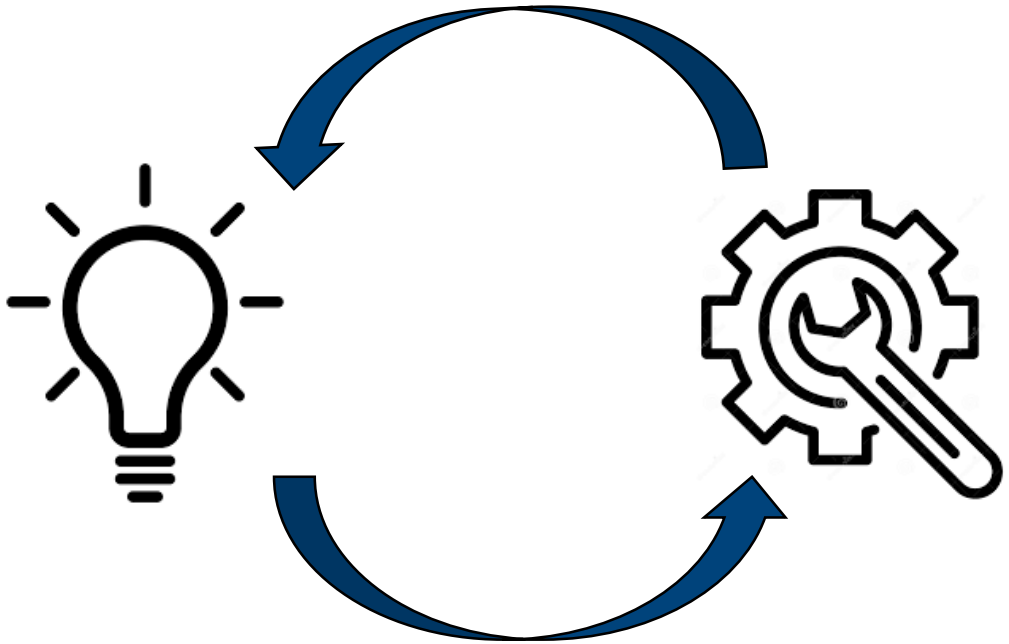
**Reminder: concepts are the variables and outcomes you want to study*

Case Study: Delayed Presentation for Gallbladder Disease

BOSTON MEDICAL CENTER
Clinical Data Warehouse for Research

Refined Question

Is a patient's socioeconomic status associated with acuity at presentation for gallbladder surgery?



Concepts

- "A patient"
- Socioeconomic status
- Acuity
- Gallbladder surgery
- More variables?

Case Study: Delayed Presentation for Gallbladder Disease

Defining the concept of "acuity"

- Need a combination of expertise in clinical workflow and documentation practices (research team) & data expertise (CDW for Research team)
- Iterative approach ("measure twice, cut once")
- Initial approach: classify patients as high/low acuity using the "elective" or "urgent" OR case flags

! The sickest patients with gallbladder disease often have percutaneous cholecystostomy drains placed & then return for elective surgery

Case Study: Delayed Presentation for Gallbladder Disease

Defining the concept of "acuity"

- Need a combination of expertise in clinical workflow and documentation practices (research team) & data expertise (CDW for Research team)
- Iterative approach ("measure twice, cut once")
- Initial approach: classify patients as high/low acuity using the "elective" or "urgent" OR case flags

! The sickest patients with gallbladder disease often have percutaneous cholecystostomy drains placed & then return for elective surgery

→ Modified approach: high acuity defined as either requiring urgent surgery or placement of a drain

Case Study: Delayed Presentation for Gallbladder Disease

What data did the research team get?

- For all OR cases that meet inclusion criteria (based on 8 CPT codes), CDW for Research team provided:
  - Urgent vs elective OR case flag
  - History (Y/N) of percutaneous cholecystostomy tube placement (based on 11 CPT codes) within 1 year of the cholecystectomy procedure date
  - 4 iterations of text strings from the operative note to indicate whether a percutaneous cholecystostomy tube was removed during the cholecystectomy (necessary for patients who had a tube placed elsewhere)

Case Study: Delayed Presentation for Gallbladder Disease

BOSTON MEDICAL CENTER
Clinical Data Warehouse for Research

What do the data look like? → Encounter-level data

Variables used to define the concept of "acuity"

Study ID	Age_yrs	Pre-op diagnosis	Procedure date	OR Flag	Procedure	History_Drain	Drain text string
1	36	Gallstone pancreatitis	1/1/2020	Elective	Laparoscopic cholecystectomy	yes	<text from op note>
1	36	Gallstone pancreatitis	1/1/2020	Elective	ERCP	yes	not found
2	84	Cholecystitis	2/5/2019	Elective	Laparoscopic cholecystectomy	no	not found
3	47	Gallstones	4/7/2017	Elective	Laparoscopic cholecystectomy	no	not found
4	52	Gallstone pancreatitis	2/15/2020	Urgent	Open cholecystectomy	no	<text from op note>

Case Study: Delayed Presentation for Gallbladder Disease

Boston Medical Center
Clinical Data Warehouse for Research

What do the data look like? → Patient-level characteristics

Study ID	Sex	Insurance	Preferred language	SVI
1	M	Medicaid	English	0.8581
2	F	Medicaid	Haitian Creole	0.1143
3	F	Commercial	Spanish	0.9157
4	F	Other gov't	English	0.6512
5	M	Medicare	English	0.3335

Defining the concept of "socioeconomic status" using the CDC's Social vulnerability index (SVI)

- *SVI uses census data by census tract to create a vulnerability score based on 4 themes*
  - *Socioeconomic status/poverty, household composition, race/ethnicity/language, housing/transportation*
- *Higher score = more socially vulnerable*
- *Used in disaster planning & epidemiologic research*

Case Study: Delayed Presentation for Gallbladder Disease

BOSTON MEDICAL CENTER
Clinical Data Warehouse for Research

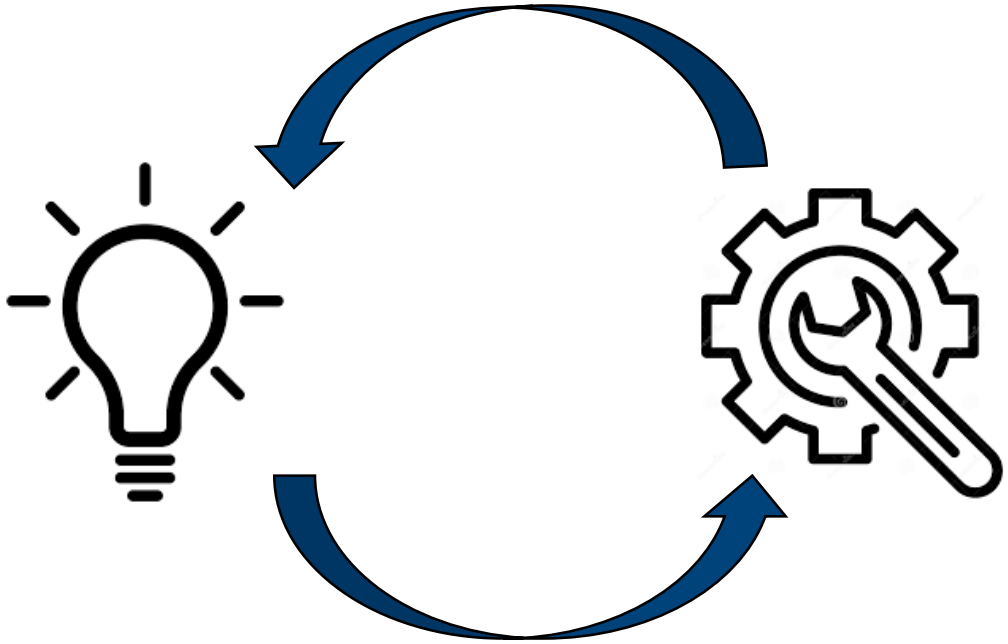
What can YOU make the data look like?

Study ID	Age_yrs	Sex	SVI	Diagnosis	Acuity
1	36	M	0.8581	Gallstone pancreatitis	High
2	84	F	0.1143	Cholecystitis	Low
3	47	F	0.9157	Gallstones	Low
4	52	F	0.6512	Gallstone pancreatitis	High
5	63	M	0.3335	Gallstones	Low

Research team can use data fields provided by the CDW for Research to derive an "acuity" variable

Research team can combine encounter-level data and patient-level characteristics to create an analytic dataset that has one patient per row

Case Study: Delayed Presentation for Gallbladder Disease



Operationalizing a research question that uses EHR data well requires cooperation between the study team & the CDW for Research team

Case Study: Example of Phenotyping Housing Insecurity

BOSTON MEDICAL CENTER
Clinical Data Warehouse for Research

Objective: Find the proportion of BMC patients* experiencing housing insecurity in 2021 for the purpose of ...

THE "WHY"
Why do you want to know this information?

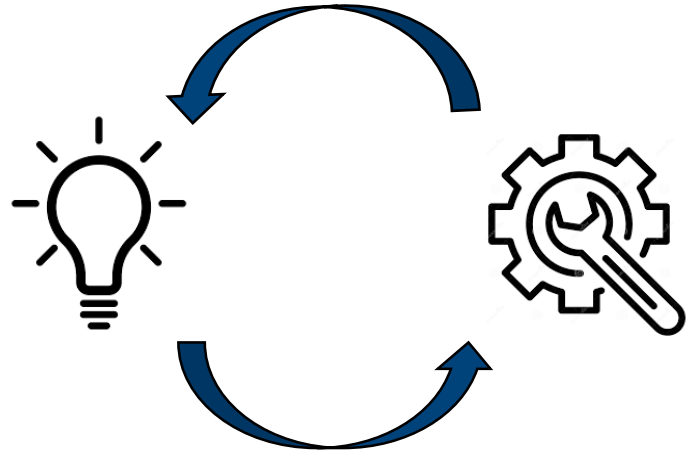
THE "WHAT"
How would you approach identifying housing insecurity in the EHR?

*BMC patient defined as: *Patients with at least one completed office visit, telehealth visit, ED encounter, and/or inpatient encounter at BMC between January 1, 2021 and December 31, 2021 (211,726).*

Case Study: Example of Phenotyping Housing Insecurity

BOSTON MEDICAL CENTER
Clinical Data Warehouse for Research

Objective: Find the proportion of BMC patients* experiencing housing insecurity in 2021



- **Approach #1**: Patients who screened positive on either BMC THRIVE screening housing questions → **N= 3,926 (1.9%)**
- **Approach #2:** Patients with ICD-10 Z59.0 (Problems related to housing & economic circumstances) in the medical record → **N= 5,423 (2.6%)**
- **Approach #3:** Patients identified as experiencing homelessness or housing insecurity using the CDW for Research's housing insecurity algorithm (7 EHR data elements) → **N= 14,962 (7.1%)**

*BMC patient defined as: *Patients with at least one completed office visit, telehealth visit, ED encounter, and/or inpatient encounter at BMC between January 1, 2021 and December 31, 2021 (211,726).*

Understanding what you are trying to measure and why is crucial for identifying the 'best' data to answer your research question.

Case Study: Example of Phenotyping Housing Insecurity

The CDW for Research Housing Insecurity Algorithm is an example of a **computable EHR phenotype**

- **Definition**: a clinical condition, characteristic, or set of clinical features that can be determined solely from data in EHRs (+/- other ancillary data sources) and does not require chart review or interpretation by a clinician
- EHR phenotypes can either be locally tailored or more broadly generalizable to data streams from multiple health systems

- **Why important?**
  - Efficiency
  - Supports reproducible queries of EHR data

Customizing EHR Phenotypes

Boston Medical Center
Clinical Data Warehouse for Research

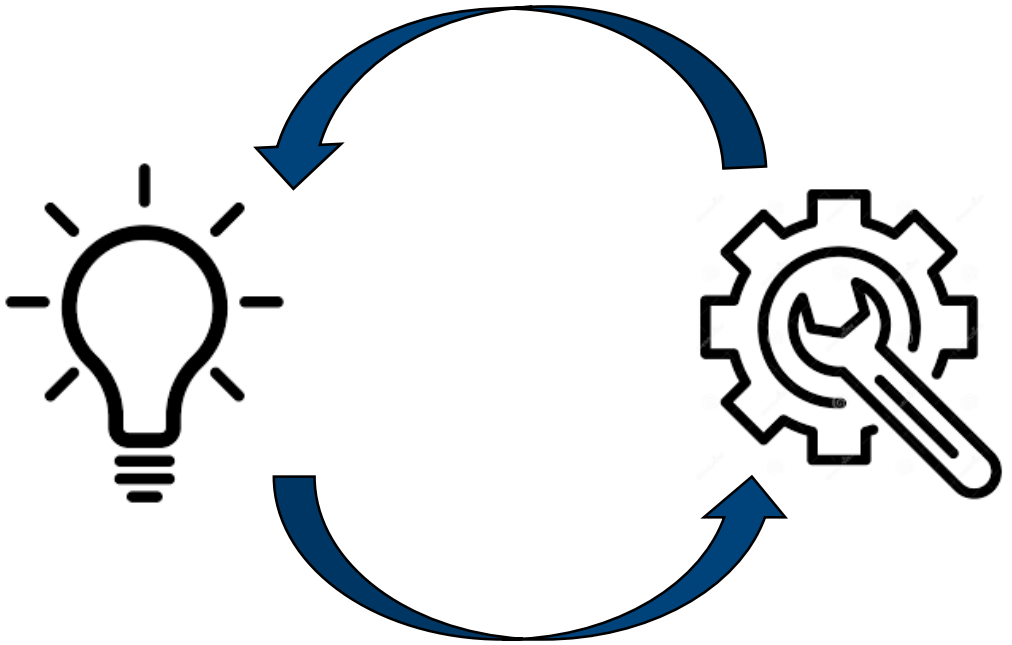
CDW for Research is actively working on development & validation of locally tailored EHR phenotypes

- 4 social determinants of health domains of focus: housing, food security, utility needs, transportation

- Clinical conditions of major research importance:
  - Opioid use disorder
  - Sickle cell disease
  - Diabetes
  - Pregnancy
  - Asthma

- We are always seeking partnerships with researchers who want to collaborate on development and validation of EHR phenotypes that they will use in their research!

What research questions are you considering?



Final Take Home Points

BOSTON MEDICAL CENTER
Clinical Data Warehouse for Research

- ✓Think about data & the data life cycle early in your next research project
- ✓Give careful thought to WHAT you are trying to measure and WHY
- ✓Remember that the CDW for Research team is available for consultation & collaboration

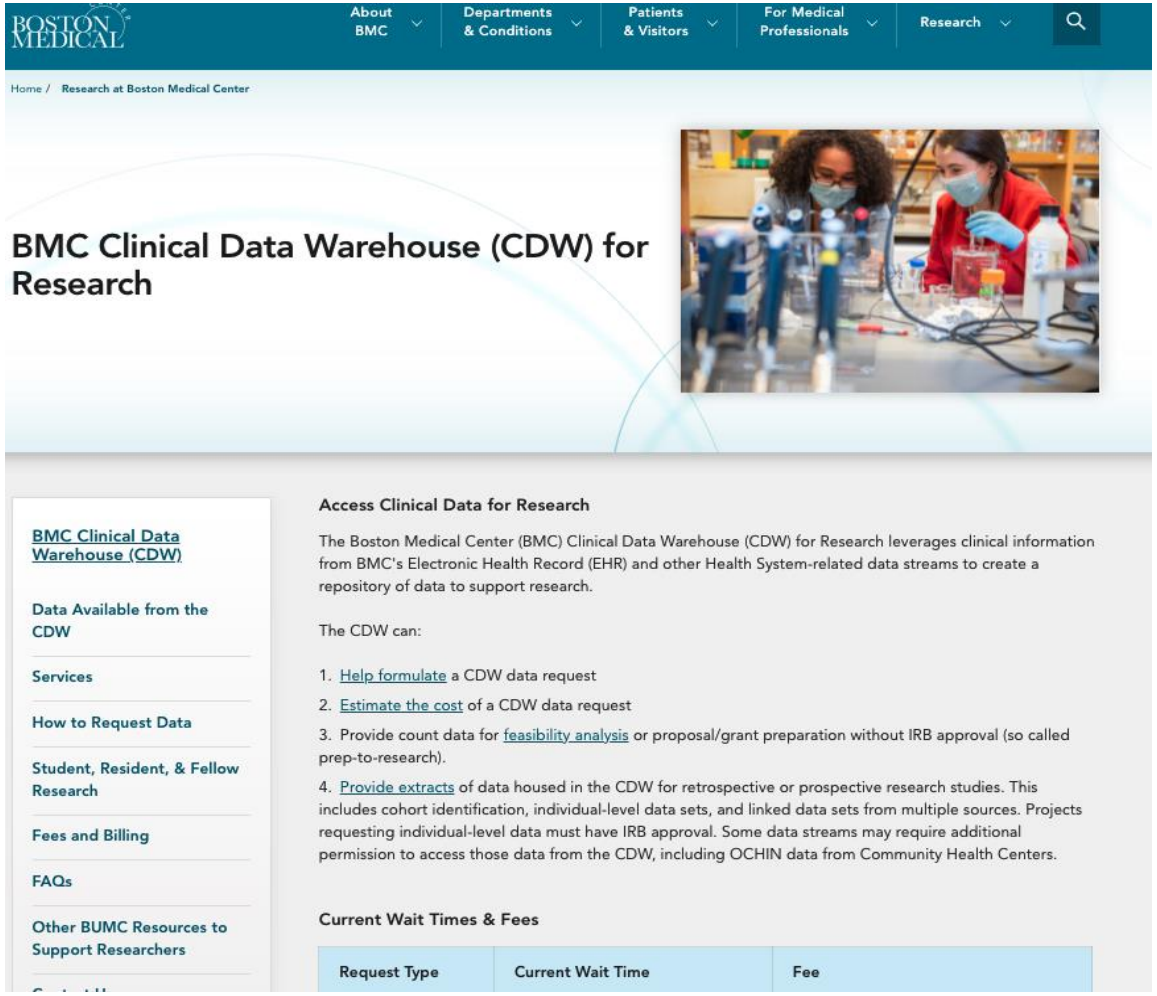
CDW for Research goal:
QUALITY IN, QUALITY OUT

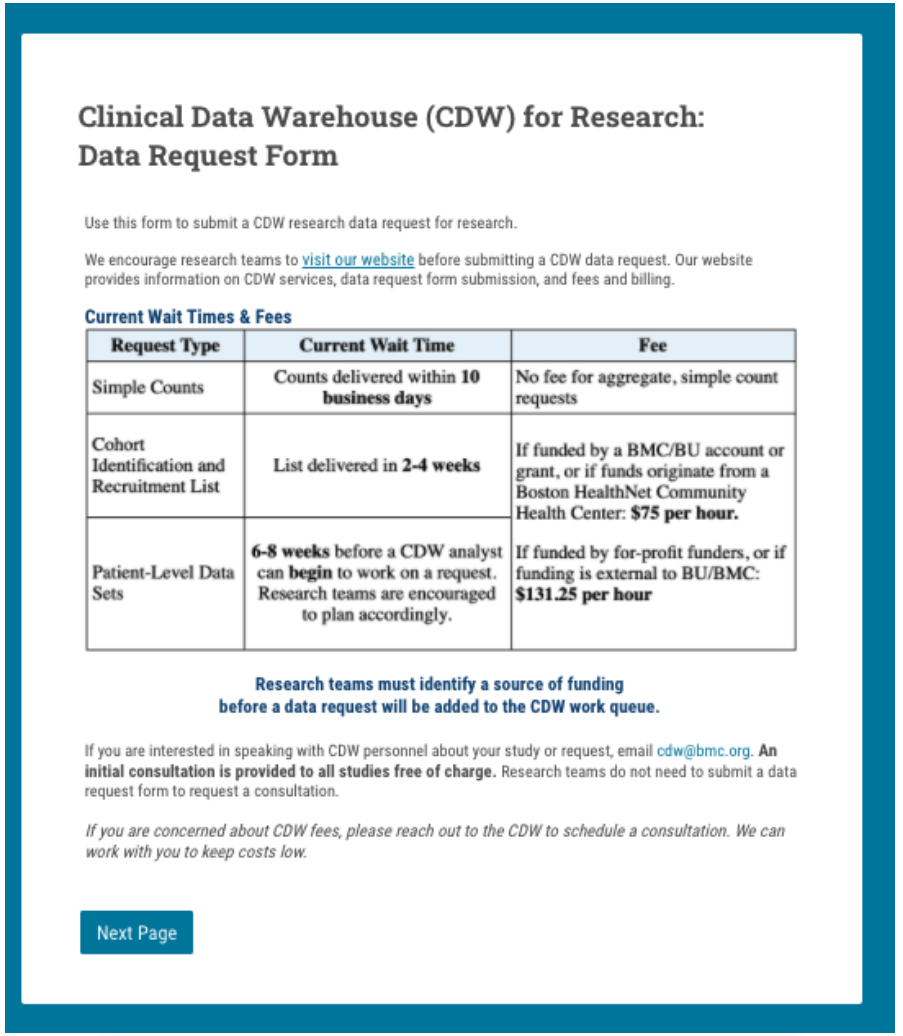
For more information CDW for Research services and fees, or to request data…

Boston Medical Center — Clinical Data Warehouse for Research

⇒ **Visit our website:** https://www.bmc.org/research/clinical-data-warehouse-cdw

⇒ **Email:** cdw@bmc.org





⇒ **Submit a CDW for Research Data Request using our online form**