



SCHOOL OF MEDICINE

North Carolina Translational and Clinical Sciences Institute

Working with the Carolina Data Warehouse for Health (CDWH)

North Carolina Translational and
Clinical Sciences (NC TraCS) Institute

What is the CDWH?

- Aggregate of electronic health record data collected in the UNC Health Care system
- Clinical, research, administrative, billing data represented
- Nearly all of the time: it's in Epic (or was in the pre-Epic EHR systems), it's in the CDWH

What *else* is the CDWH?

- 9 data analysts: set up queries, provision data. Meet with study teams to strategize
- 3 project managers: initial contact for new projects. Meet with study teams to assess project needs, discuss feasibility, costs, timelines, regulatory and legal matters.

Types of data requests

- Investigators come in with different goals for their CDWH data:
 - To recruit patients/identify patients and get recruitment lists
 - To obtain data on patients enrolled in a clinical trial (for example, ongoing reports with data on recent clinic visits and lab levels).
 - Retrospective datasets

Data available in the CDWH

- Patient demographics
- Encounter details
- Diagnoses
- Procedures
- Providers
- Patient vitals
- Lab tests
- Medications
- Orders
- Notes ★
- Charges and Payors ★
- Surgery
- Labor and delivery
- Patient-reported data (including SDOH) ★
- Clinical flowsheets ★
- Custom data elements
- ...and more!

Different needs/different cadences

CDWH data can be provided—

- On a one-time basis
- As automated recurring reports

Some examples of data requests—

- "I need a retrospective dataset on patients diagnosed with kidney stones in the last 5 years, covering their diagnoses, procedures, medications and lab results."
- "I am recruiting for a trial and would like a pool of potential participants to contact. Can you give me names and contact information for female patients who were newly diagnosed with Crohn's disease in the past year?"

More discussion on defining your clinical question— including issues you are facing in your actual current research— to match available data types and query strategies shortly!

Data availability by date

- 7/1/2004 - 4/3/2014: data from legacy EHR system WebCIS. Medical Center data only.
- 2007-2014 pre-Epic inpatient orders
- 2008-2014 pre-Epic physician billing data
- **4/4/2014: Epic goes live at UNC. Health System-wide data.**

More on dates: Epic Go Live dates by location

BLUE RIDGE VALDESE (Day Op and ED ONLY)	07/23/2022	PARDEE	06/18/2016
BLUE RIDGE MORGANTOWN	07/23/2022	REX	06/20/2014
CALDWELL	06/18/2016	ROCKINGHAM - Physician Billing	08/01/2018
CHATHAM	04/04/2014	ROCKINGHAM - Hospital Data	05/15/2021
HILLSBOROUGH *	04/04/2014	SOUTHEASTERN REGIONAL MEDICAL CENTER	07/23/2022
HIGH POINT REGIONAL	05/20/2016-Sept 2018*	UNC	04/04/2014
JOHNSTON	05/20/2016	UNC Physician Network	Varies - Initial practices went live on 04/04/2014 and 06/20/2014. Practices continue to Go Live as they are added to UNC.
LENOIR	05/15/2021	WAYNE HEALTH	09/22/2018
NASH	09/22/2018		



SCHOOL OF MEDICINE

North Carolina Translational and Clinical Sciences Institute

How to create a data question

1. Write down three things:

- The overall **decision(s)** the health data will help support
- The **immediate goal** the data directly helps achieve, based on the
- **Context** of the problem and what else is known about it

1. Example:

- **Decision:** Whether to give a person higher priority for COVID therapeutics based on specific risk factors.
- **Context:** There is a specific need to show whether Type 1 diabetes is associated with worse outcomes in COVID patients
- **Immediate goal:** figure out whether type 1 diabetes impacts COVID severity

2. Write down the question in plain English

- Example: "Are people with Type 1 diabetes at higher risk of worse COVID outcomes?"

3. Use the PICOT framework

- **P**opulation, **I**ntervention, **C**omparison, **O**utcomes, **T**imepoint(s)
- Each of these components will help further specify the data request!

3. Example

- **Population:** people with type 1 diabetes
- **Intervention:** COVID infection
- **Comparison:** people without type 1 diabetes, who had COVID infection as well
- **Outcomes:** hospital admission, mortality
- **Timeframe:** from 12/2019 - onward

4. Define your nouns!

- Your research question probably has several nouns in it at this point. These need to be carefully defined!

4. Example:

- Population: people with ***type 1 diabetes***
- Intervention: ***COVID infection***
- Comparison: ***people without type 1 diabetes***, who had COVID infection as well
- Outcomes: ***hospital admission, mortality***
- Timeframe: from 12/2019 - onward

4. Example: people with type 1 diabetes

- *Population:* ICD-Coded diagnoses and glucose medication: ICD-9 250.x1, ICD-9 250.x3, and ICD-10 E10.xx. AND
- Labs showing an undetectable c-peptide result and positive diabetes autoantibody component result

4. Example: DM1 and COVID, continued:

- *Intervention:* COVID infection defined as positive PCR test for COVID-19
- *Comparison:* equal number of randomly selected people who do not meet the inclusion criteria defined in the population
- *Outcomes:* Hospitalization encounter or death within 21 days of first positive COVID test.
- *Time frame:* Patients with face-to-face encounters from 12/2019 onward and a positive COVID PCR test from that time point as well.

5. Define your nouns again!

- ICD-Coded diagnoses and glucose medication:
ICD-9 250.x1, ICD-9 250.x3, and ICD-10
E10.xx. AND
- Labs showing an ***undetectable c-peptide result*** and ***positive diabetes autoantibody component result***

6. What are your Inclusion/Exclusion criteria?

- Your PICOT framework will help you!

6. Example

- Inclusion criteria: type 1 diabetics with COVID infection, as defined by the "population" and "intervention" of our PICOT formatted question, AND an equal number of randomly selected people without DM1 who also had COVID infection
- Exclusion: type 1 diabetics who never had COVID infection, people without type 1 diabetes not included in the random sample

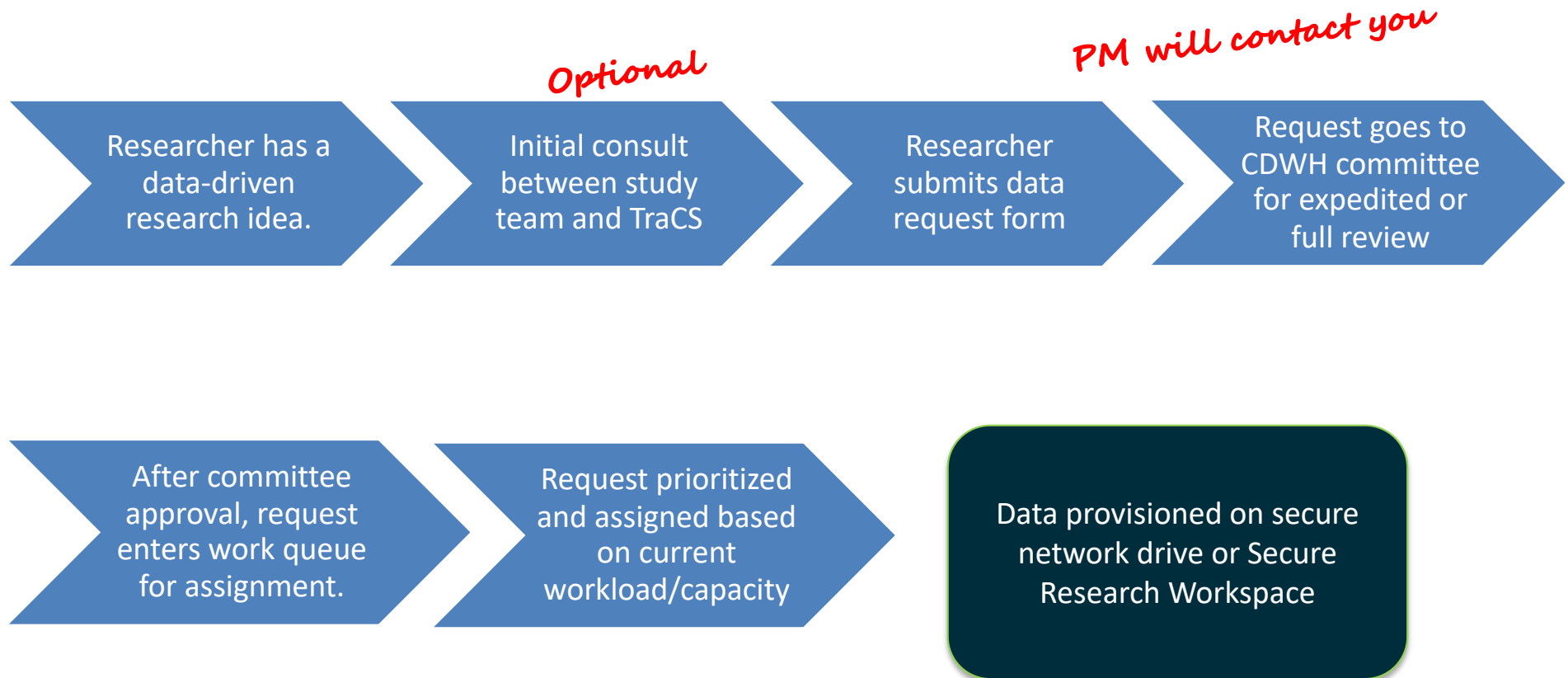


SCHOOL OF MEDICINE

North Carolina Translational and Clinical Sciences Institute

Requesting CDWH Data

Life of a CDWH data request



Your data request

The basics

Basic Request Information

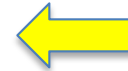
IRB number
* must provide value

Have you already spoken with a TraCS analyst about this request?
* must provide value

Yes
 No

reset

Short title for your project
* must provide value

Brief description of your research question and data needs
* must provide value
1000 characters remaining
Expand

You must have an approved IRB to submit your full data request form

Your data request

The basics

Basic Request Information

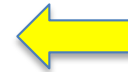
IRB number
* must provide value

Have you already spoken with a TraCS analyst about this request?
* must provide value

Yes
 No

reset

Short title for your project
* must provide value

Brief description of your research question and data needs
* must provide value
1000 characters remaining
Expand

You must have an approved IRB to submit your full data request form

Your data request

Defining your cohort

Description of Data Needs

Please list the **INCLUSION** criteria to define your cohort of interest

* must provide value

1000 characters remaining

Expand

Be as specific as possible!

Please list the **EXCLUSION** criteria to define your cohort of interest

* must provide value

1000 characters remaining

Expand

Be as specific as possible!

Did you perform an initial i2b2 query that you would like to use as the basis for this request?

* must provide value

Yes

No

reset

Your data request

What data do you want to receive?

Please list the specific data elements you would like in your output. *E.g., names, MRN, DOB, age, lab tests, procedures, medications, diagnoses (please provide diagnosis codes and/or CPT codes if applicable)*



*** must provide value**

2000 characters remaining

Expand

Your data request

Date coverage

Start date for data * must provide value	<input type="text"/>		Today	M-D-Y
End date for data	<input type="text"/>		Today	M-D-Y

- Epic data: 4/4/2014-now
- Legacy data: 7/1/2004-4/2/2014

Your data request

Obtaining your data

Data Output Requirements

Requested frequency of data pulls
** must provide value*

One-time
 Recurring

reset

Names and roles of persons requiring access to this data
** must provide value*

300 characters remaining

Expand

Please select the data storage option that you would like to use for your request:
** must provide value*

The TraCS file server
 The Secure Research Workspace environment
 Trust Zone server within the NC TraCS server pool
 Another university or health care system server, with explicit approval of the CDW Operations and/or CDW Oversight Committee

reset

Do you have a strong preference for the end format of your data? (E.g., .csv files, SAS files, etc.)
** must provide value*

Yes
 No

reset

Your data request

Sharing data outside UNC

External Data Sharing

Will this data or derivatives of it be shared with anyone outside of UNC? Yes No

* must provide value reset

If you are sharing patient-level data you will need

- A data use agreement: CDWH staff will provide the right starter document
- Specific approval by the TraCS Operations Committee

Your data request

Funding

Funding Information

Our standard rates are \$105/hour. Requests that will be invoiced externally are billed at \$165/hour. Limited i2b2 and EMERSE based services can be performed at no charge; please enter a consult request for details.

Please select your funding situation

* must provide value

- I have gotten an estimate from an analyst already, and have funding available.
- I need an estimate from an analyst
- I have arranged to pay for a % of an analyst's salary on my grant budget
- I do not have funding available

reset

We're happy to talk to you to come up with a plan that fits your research budget.

Regulatory issues

Regulatory Questions

Which of the following types of patient identifiers do you need in your dataset? (Check all that apply)

* must provide value

- "Direct" identifiers (e.g., name, medical record number, phone number, email address, etc.)
- Dates of service (e.g., visit dates), dates of birth, and/or dates of death
- Geographic locations at the town, city, or ZIP level (or larger)
- Geographic locations smaller than town, city, or ZIP level (like a home address or geocode)
- None of these

Based on your responses, you are requesting an identified dataset. The following questions will help to ensure that your IRB approval is in line with this type of data request.

In section A.9 of your IRB, do the identifiers you have checked off match the identifiers that you are requesting here?

* must provide value

- Yes
- No

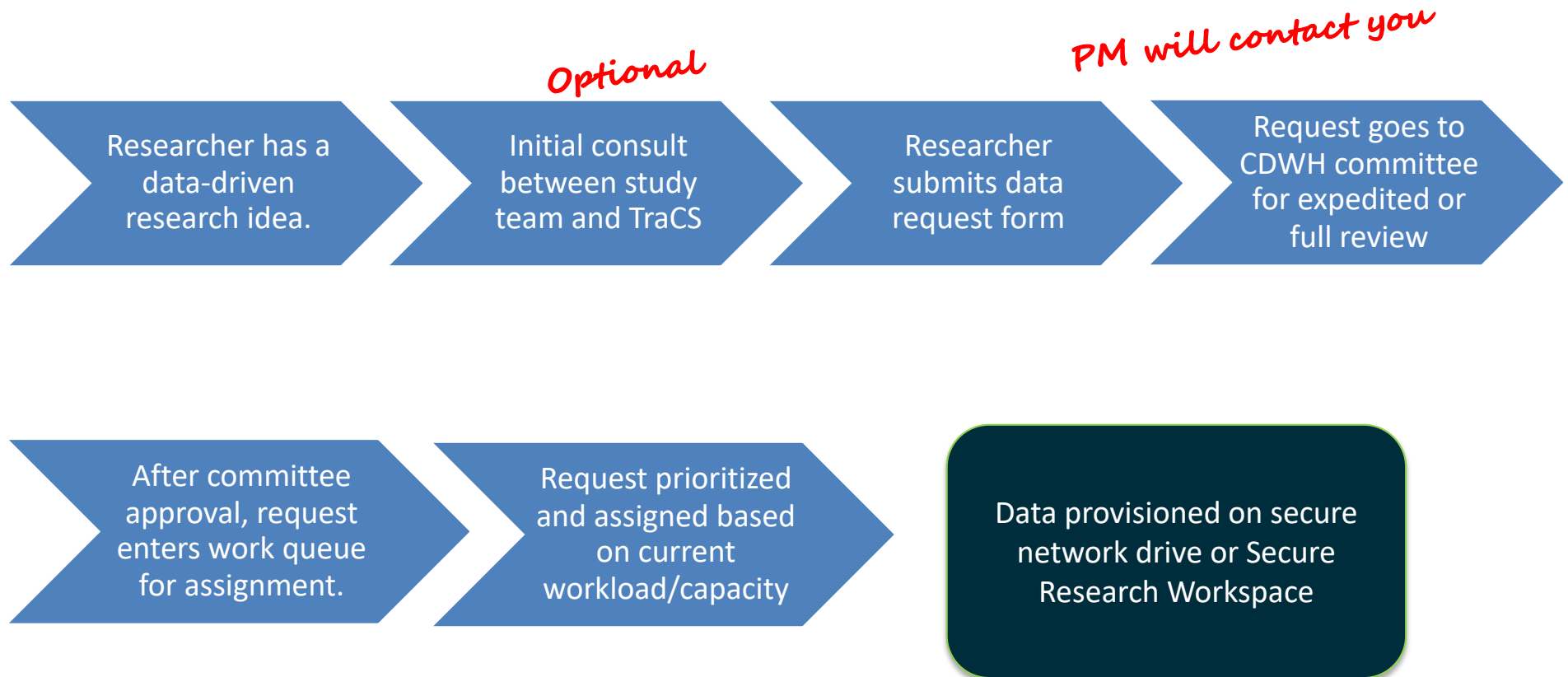
reset

In section C.1 of your IRB, have you selected both "medical records in any format" and "Carolina Data Warehouse for Health" as data sources? Note that if Epic access is needed for chart reviews or other purposes, "Electronic medical records" must also be selected.

- Yes
- No

reset

Life of a CDWH data request





SCHOOL OF MEDICINE

North Carolina Translational and Clinical Sciences Institute

i2b2

i2b2

- “Informatics for Integrating Biology and the Bedside”
- Self-service querying tool that links with the Carolina Data Warehouse for Health
- Explore and query structured clinical data that has been deidentified and aggregated (i.e., counts)
- Includes data from the entire UNCHCS
- Why use i2b2?
 - Identify cohort counts on your own
 - Test different exclusion and inclusion criteria
 - Get the information you need for early stages of research more quickly
 - An analyst can use your initial queries when working on your data request

i2b2 interface

The screenshot displays the i2b2@UNC interface. On the left is a navigation tree with categories: 01) Patient Demographics, 02) Encounter Details, 03) Diagnoses, 04) Procedures, 05) Point of Care, 06) Patient Vitals, 07) Laboratory Tests, 08) Medications, 09) Clinical Observations, 10) Social History, 11) Insurance Payors (Epic), and COVID-19. The main area is the 'Query Tool' for a query named 'Femal-E11: -Hemog@14.01.45'. The query timing is set to 'Non-Temporal Query: Treat all groups independently'. The query is structured into three groups:

Group 1	Group 2	Group 3
Group 1 Dates Occurs > 0x Exclude Treat Independently Female	Group 2 Dates Occurs > 0x Exclude Treat Independently E11: Type 2 diabetes mellitus	Group 3 Dates Occurs > 0x Exclude Treat Independently Hemoglobin A1c/Hemoglobin.total.MFr.Pt.Bld.Qn > 5.7 % [≥01/01/2019]

Each group is connected to the next by an 'AND' operator. Below each group is a green box containing the text 'one or more of these'. At the bottom of the query tool, there are buttons for 'Run Query', 'Clear', and 'New Group', along with the text '3 Groups'.

Data breakdowns in i2b2

Number of patients for "Femal-E11: - Hemog@14:01:45"

patient_count: 49994

Race patient breakdown for "Femal-E11: - Hemog@14:01:45"

American Indian or Alaska Native: 275

Asian: 883

Black or African American: 17611

Hispanic/Latino (Not Accurate - Use Ethnicity): 0

Native Hawaiian or Other Pacific Islander: 53

Other: 3126

Patient Refused: 144

Unknown: 635

White or Caucasian: 27267

Age patient breakdown for "Femal-E11: - Hemog@14:01:45"

0-9: 36

10-17: 218

18-34: 1871

35-44: 3503

45-54: 7203

55-64: 11610

65-74: 13466

75-84: 8594

>= 65: 25553

>= 85: 3493

Greater Than or Equal to 18 Years Old: 49740

Less Than 18 Years Old: 254

Not recorded: 0

Gender patient breakdown for "Femal-E11: - Hemog@14:01:45"

Female: 49994

Male: 0

Unknown: 0

Vital Status patient breakdown for "Femal-E11: - Hemog@14:01:45"

Alive: 47036

Deceased: 2901

Unknown: 57

Using i2b2

- Training is required and offered online by NC TraCS
- i2b2 queries can be done “preparatory to research”; an IRB is not required
- There is no cost to use i2b2@UNC
- Access to i2b2 requires a valid UNC onyen, as well as completion of a 1-hour online training session.
- More information and training registration links are available at <http://tracs.unc.edu/services/biomedical-informatics/i2b2>
- i2b2 itself can be found at i2b2.unc.edu.

Using i2b2 to get patient-level data

- Even though you can only view aggregate data within i2b2, you can also use it to get some basic patient-level data as well.
- Just submit a data request form, list your i2b2 query and user name on the form. For no charge we can send you the following data elements for the patients identified by your i2b2 query: **MRN, living status, name, DOB, current age, sex, race, and ethnicity**. Note that any additional data elements will likely involve data fees, however.
- Why are only these data elements available for free? Because we have a semi-automated process to provide these. Other data requires custom programming.



Helpful resources handout, slides, and recording links

- Download links at bottom of series webpage

go.unc.edu/clinical-data-literacy

CLINICAL DATA
LITERACY SERIES:
ELECTRONIC HEALTH DATA BASICS



End of seminar series feedback survey (very short!)

