Introducing the Carolina Data Warehouse for Health

Emily Pfaff

Administrative Director, Informatics & Data Science

NC TraCS Institute

Agenda

- What is the CDW-H?
- Whose data is in the CDW-H?
- How do investigators use the CDW-H?
- What is the process for getting data from the CDW-H?
- A note on data used for participant recruitment
- *(If time allows)* Self-service tools: i2b2 and EMERSE

First, a little about NC TraCS

- TraCS is UNC's CTSA (***C***linical & ***T***ranslational ***S***cience ***A***ward), one of more than 50 CTSAs across the US.
- Located on the second floor of Brinkhous Bullitt, right above the Beach Café.
- Co-PIs: Dr. Tim Carey and Dr. John Buse
- Supports *translational* – "bench to bedside" – science at UNC: regulatory, pilot awards, team science, community engagement, informatics, to name a few

What TraCS IDSci Does

- Serves as a technical and regulatory liaison between the Health Care System and the university for *research* uses of CDW-H data.
- Supports approved secondary uses of those data in informatics and clinical research projects.
- Develops and supports tools for researchers to explore clinical data
- Provides consistent, compliant, low-cost data management infrastructure for clinical researchers to use in their studies.*

*This is REDCap, which will not be covered in this talk.

What is the CDW-H?

- "Warehouse" of electronic health record data collected in UNCHCS, live as of 2009
- Data on ~2.2+ Million unique patients, 1.5 Million active patients, expanding with UNCHCS
  - If it's in Epic (or was in WebCIS), it's in the CDW-H.
- Data collection dates back to:
  - July 2004: Hospital Billed Data
  - July 2008: Physician Billed Data
  - April 2014: Epic Systems Data

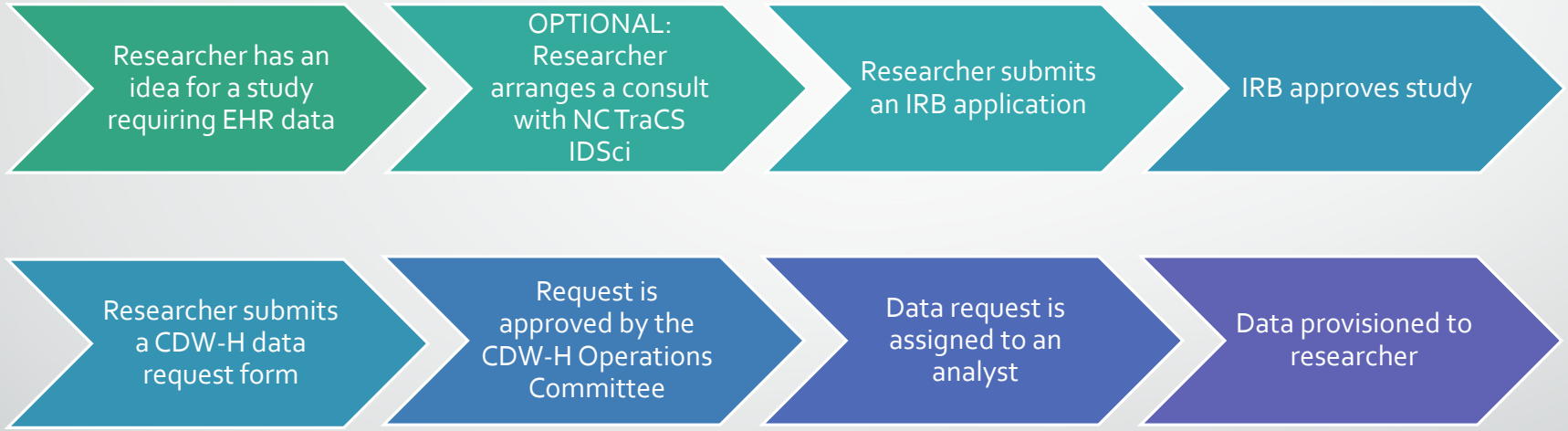
Whose data is in the CDW-H?

- Every patient at UNCHCS, without exception. (Yep, me too.)
- All patients agree that their data can be used for IRB-approved research when they sign UNCHCS's Notice of Privacy Practices.
  - Some may be unaware that they have agreed to this, if they did not read what they were signing.
- There is no way to opt-out of being in the CDW-H, and no way to *blanket* opt-out of having one's data potentially used for IRB-approved research. That includes both retrospective data analysis and studies that directly interact with patients.
- Of course, any patient can opt out of participating in or receiving contact from any one study. It is that study's responsibility to abide by the contacted patient's preferences.

How do investigators use the CDW-H?

- Two kinds of requests: (1) recruitment support and (2) retrospective data analysis.
- "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?"
- "I need a retrospective dataset on patients diagnosed with kidney stones, covering their diagnoses, procedures, medications and lab results for five years preceding their diagnosis."

CDW-H Process



CDW-H Process: The "Ops" Committee

- Meets once a month; chaired by Dr. Don Spencer (UNCHCS CMIO)
- Includes representation from TraCS, UNC SOM, UNCHCS, and IRB
- Main roles:
  1. Compare data requests with IRB protocols and ensure alignment
  2. Determine whether the data requested are appropriate for the study (e.g., minimum necessary)
  3. Ensure data are being stored and handled appropriately, particularly when data are leaving UNC
- The most straightforward requests (according to our "fast track checklist") can receive expedited review. Otherwise, requests are reviewed by the full committee.
- Particularly controversial or precedent-setting requests can be escalated to the CDW-H Oversight Committee.

CDW-H Process: Data Analysts

- Approved requests are assigned to one of our ten Research Informatics Specialists. ("Analysts")
- TraCS analysts serve as Honest Brokers, extracting CDW-H data on behalf of investigators.
- No investigator has direct access to the CDW-H for research purposes.
- Analysts are highly attuned to regulations around the use of data, and are well-versed in the differences between deidentified, HIPAA-limited, and fully identified datasets.

CDW-H Process: Data Provisioning

- Most often, data are provisioned to a secure network drive housed in UNC SOM, accessible by the study team via onyen/password authentication.
- Other arrangements (a secure server in Lineberger, for example) are possible with approval by the Ops Committee.
- The Secure Research Workspace (SRW) and the Longleaf High-Performance Computing Environment are other sanctioned options.
- Local storage of data is *never* sanctioned.

CDW-H Process: Data Provisioning

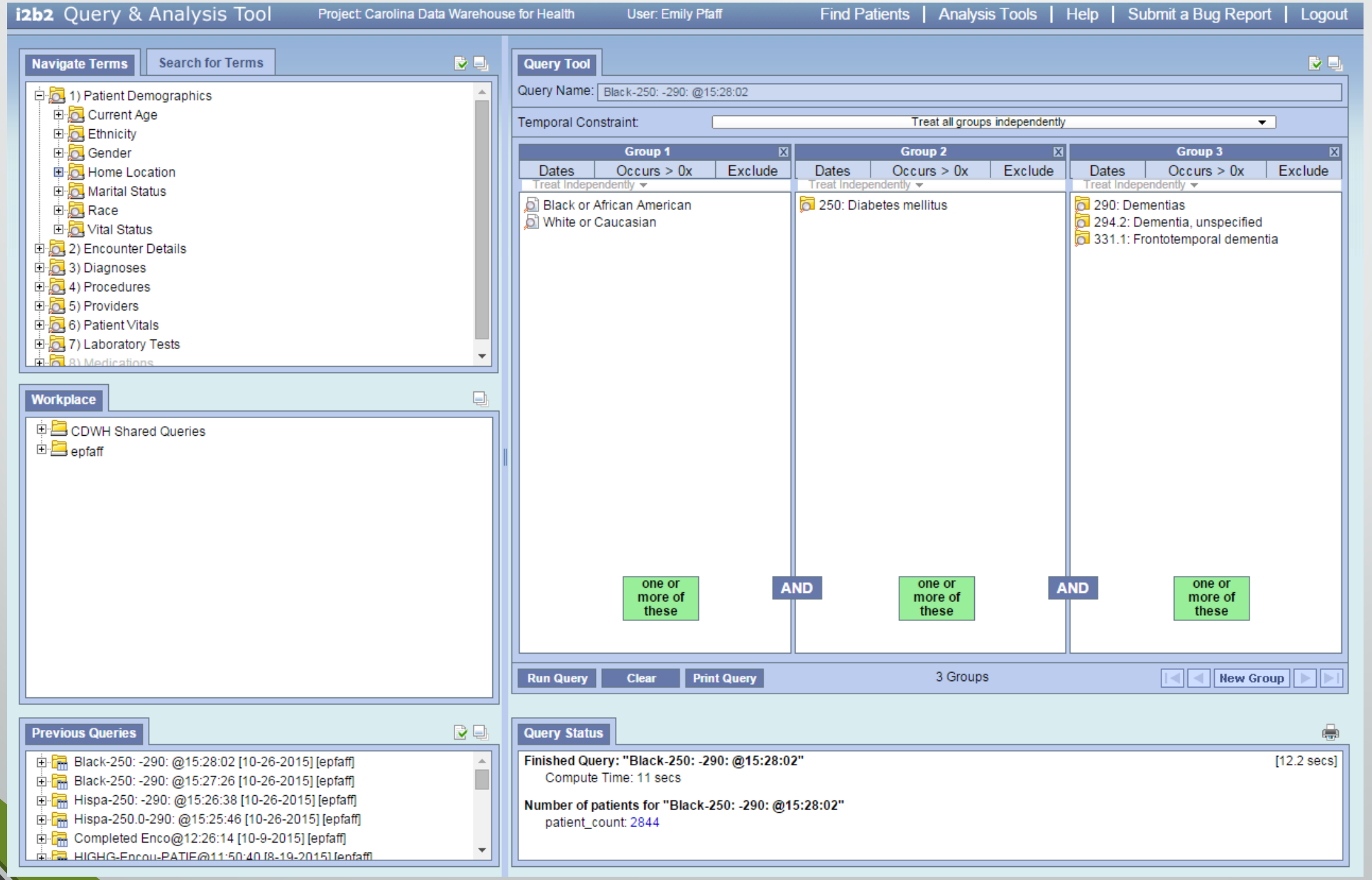
- If HIPAA-limited or fully identified data for non-consented patients are leaving UNC, the Ops Committee will always do a full board review.
- During this review, the Ops Committee ensures the appropriate Data Use Agreement has been executed with the receiving institution.
- NC TraCS has DUA templates specifically written to cover sharing of CDW-H data. We have versions for deidentified, HIPAA-limited, and fully identified datasets.
- Once approved, either TraCS or the investigator may share the data in a secure manner, depending on the needs of the study.

Data Used for Participant Recruitment

- Participant recruitment support requests are more likely to be controversial, because patients are being contacted based on CDW-H data.
- The UNC IRB and TraCS have composed recommended language for study teams to use in their recruitment materials when patients are identified through the CDW-H. [https://tracs.unc.edu/docs/cdwh/Recruitment_letter_language_CDW.pdf]
- If a study team wishes to use email to recruit, special caution is necessary regarding the language used.
- If a study team wishes to use Epic MyChart to recruit, UNCHCS must also sign off on the proposed language (in addition to the IRB).

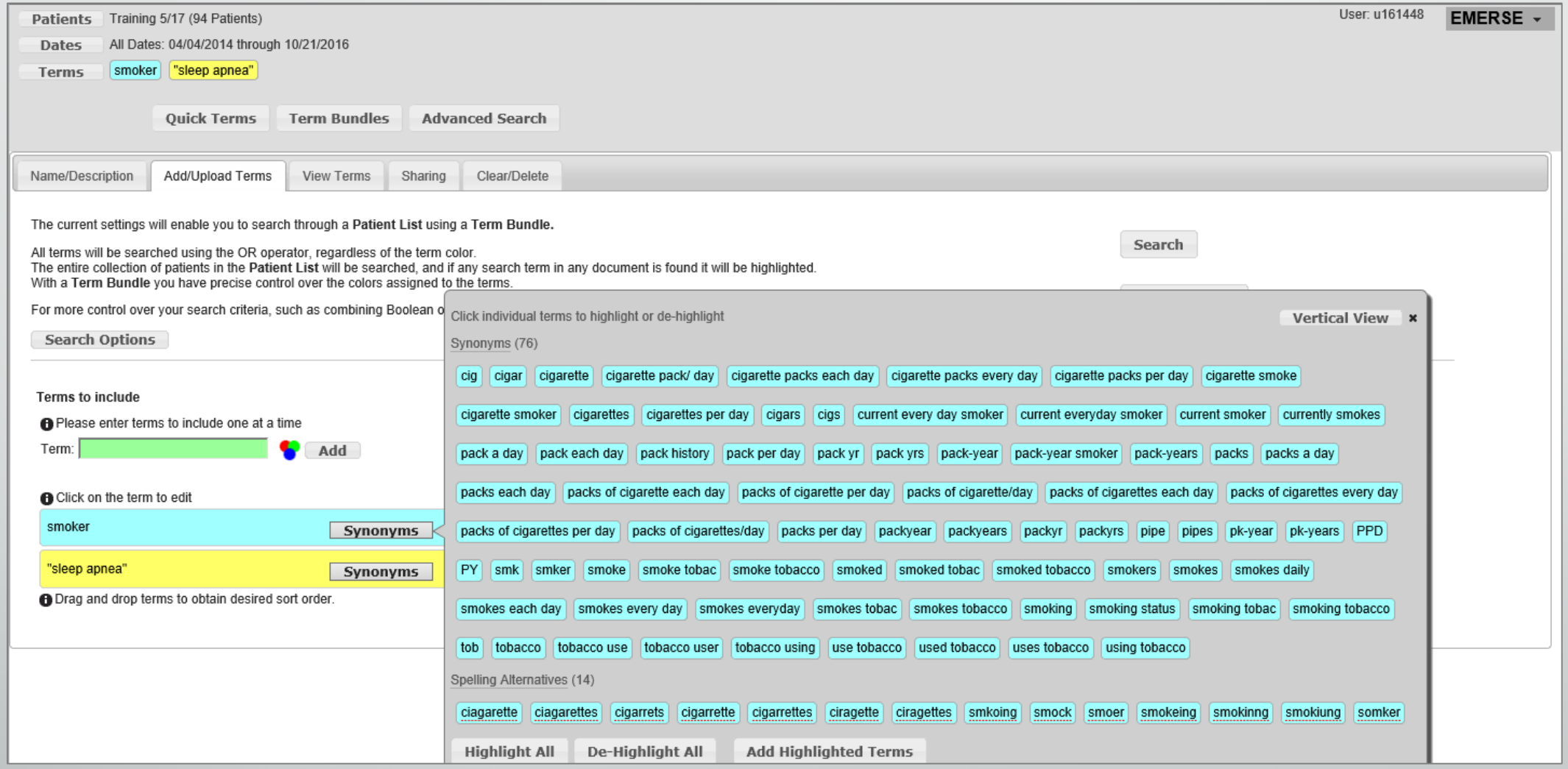
What is i2b2?

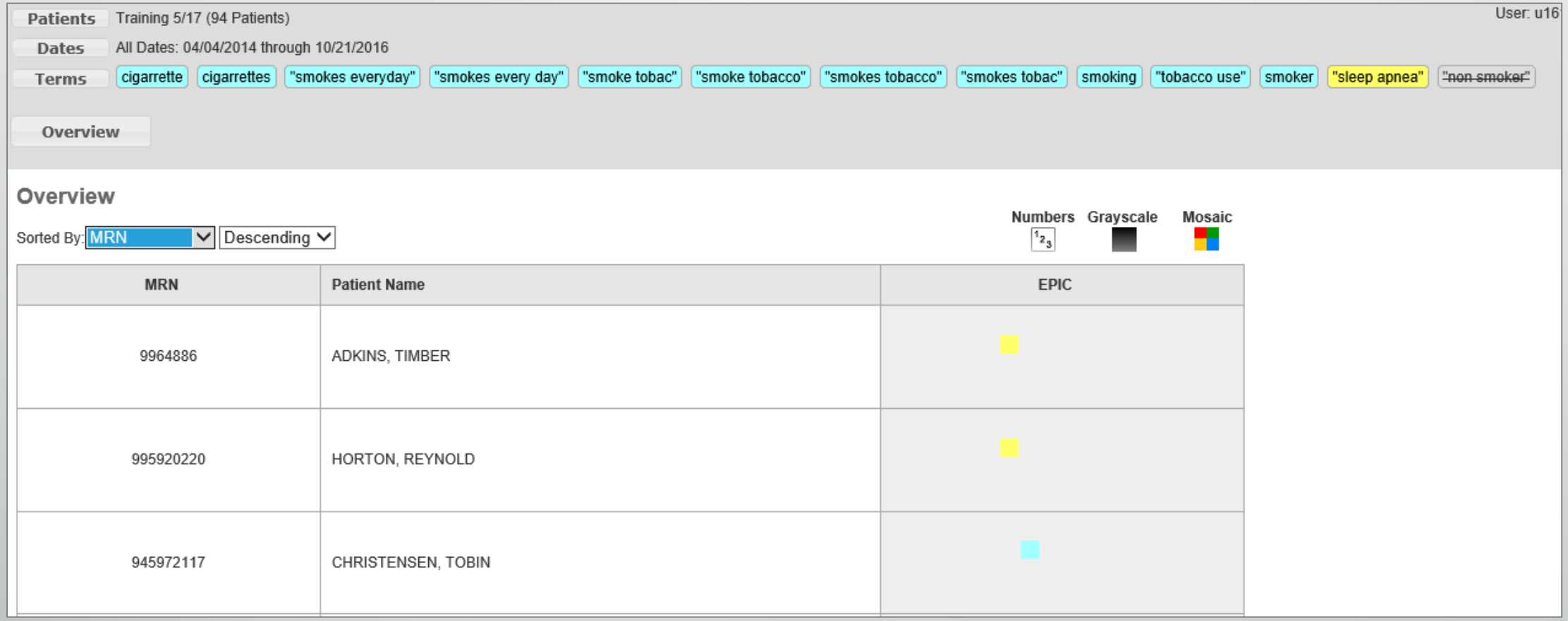
- Self-service querying tool that links with the CDW-H.
- Explore and query *structured* clinical data that has been deidentified and aggregated.
- Does not require IRB to use: output is aggregate count of patients meeting criteria.
- Allows UNC investigators to get hands-on with data in a safe, free way.
- Accessible to anyone with a UNC onyen who takes our online training.


i2b2 Query & Analysis Tool
Project: Carolina Data Warehouse for Health
User: Emily Pfaff
Find Patients | Analysis Tools | Help | Submit a Bug Report | Logout
Navigate Terms
Search for Terms
1) Patient Demographics
Current Age
Ethnicity
Gender
Home Location
Marital Status
Race
Vital Status
2) Encounter Details
3) Diagnoses
4) Procedures
5) Providers
6) Patient Vitals
7) Laboratory Tests
Workplace
CDWH Shared Queries
epfaff
Previous Queries
Black-250: -290: @15:28:02 [10-26-2015] [epfaff]
Black-250: -290: @15:27:26 [10-26-2015] [epfaff]
Hispa-250: -290: @15:26:38 [10-26-2015] [epfaff]
Hispa-250.0-290: @15:25:46 [10-26-2015] [epfaff]
Completed Enco@12:26:14 [10-9-2015] [epfaff]
Query Tool
Query Name: Black-250: -290: @15:28:02
Temporal Constraint: Treat all groups independently
Group 1
Dates
Occurs > 0x
Exclude
Treat Independently
Black or African American
White or Caucasian
one or more of these
AND
Group 2
Dates
Occurs > 0x
Exclude
Treat Independently
250: Diabetes mellitus
one or more of these
AND
Group 3
Dates
Occurs > 0x
Exclude
Treat Independently
290: Dementias
294.2: Dementia, unspecified
331.1: Frontotemporal dementia
one or more of these
Run Query
Clear
Print Query
3 Groups
New Group
Query Status
Finished Query: "Black-250: -290: @15:28:02"
[12.2 secs]
Compute Time: 11 secs
Number of patients for "Black-250: -290: @15:28:02"
patient_count: 2844

What is EMERSE?

- **E**lectronic **Me**dical **R**ecord **S**earch **E**ngine
- EMERSE allows investigators to search clinical notes using key words; can support research, quality improvement, and clinical operations
- Originally developed by University of Michigan in 2005; launched at UNC-Chapel Hill in October 2017
- Contains fully identified notes from UNCHCS since the *first* Epic Go Live on April 4, 2014
- Allows for data exploration within the tool—does not allow for any data extraction.
- Research users must attest under an active IRB with an attached HIPAA waiver before they are allowed to use the tool. All user activity is audited, and training is required.


Patients
Training 5/17 (94 Patients)
Dates
All Dates: 04/04/2014 through 10/21/2016
Terms
smoker
"sleep apnea"
User: u161448
EMERSE
Quick Terms
Term Bundles
Advanced Search
Name/Description
Add/Upload Terms
View Terms
Sharing
Clear/Delete
The current settings will enable you to search through a Patient List using a Term Bundle.
All terms will be searched using the OR operator, regardless of the term color.
The entire collection of patients in the Patient List will be searched, and if any search term in any document is found it will be highlighted.
With a Term Bundle you have precise control over the colors assigned to the terms.
For more control over your search criteria, such as combining Boolean o
Search
Search Options
Terms to include
Please enter terms to include one at a time
Term:
Add
Click on the term to edit
smoker
Synonyms
"sleep apnea"
Synonyms
Drag and drop terms to obtain desired sort order.
Click individual terms to highlight or de-highlight
Vertical View
Synonyms (76)
cig
cigar
cigarette
cigarette pack/ day
cigarette packs each day
cigarette packs every day
cigarette packs per day
cigarette smoke
cigarette smoker
cigarettes
cigarettes per day
cigars
cigs
current every day smoker
current everyday smoker
current smoker
currently smokes
pack a day
pack each day
pack history
pack per day
pack yr
pack yrs
pack-year
pack-year smoker
pack-years
packs
packs a day
packs each day
packs of cigarette each day
packs of cigarette per day
packs of cigarette/day
packs of cigarettes each day
packs of cigarettes every day
packs of cigarettes per day
packs of cigarettes/day
packs per day
packyear
packyears
packyr
packyrs
pipe
pipes
pk-year
pk-years
PPD
PY
smk
smker
smoke
smoke tobac
smoke tobacco
smoked
smoked tobac
smoked tobacco
smokers
smokes
smokes daily
smokes each day
smokes every day
smokes everyday
smokes tobac
smokes tobacco
smoking
smoking status
smoking tobac
smoking tobacco
tob
tobacco
tobacco use
tobacco user
tobacco using
use tobacco
used tobacco
uses tobacco
using tobacco
Spelling Alternatives (14)
ciagarette
ciagarettes
cigarrets
cigarrette
cigarrettes
ciragette
ciragettes
smkoing
smock
smoer
smokeing
smokinng
smokiung
somker
Highlight All
De-Highlight All
Add Highlighted Terms


Patients
Training 5/17 (94 Patients)
User: u16
Dates
All Dates: 04/04/2014 through 10/21/2016
Terms
cigarrette
cigarrettes
"smokes everyday"
"smokes every day"
"smoke tobac"
"smoke tobacco"
"smokes tobacco"
"smokes tobac"
smoking
"tobacco use"
smoker
"sleep apnea"
"non smoker"
Overview
Overview
Numbers
Grayscale
Mosaic
Sorted By: MRN
Descending
MRN
Patient Name
EPIC
9964886
ADKINS, TIMBER
995920220
HORTON, REYNOLD
945972117
CHRISTENSEN, TOBIN

Thank you!