USING EPIC DATA FOR CLINICAL RESEARCH

Anuradha Paranjape, MD MPH

Professor, Medicine

Site PI- PCORnet

TEMPLE HEALTH

The EHR has lots and lots of data!

- Data in EPIC dates back to
  - November 2010 for ambulatory visits
  - August 2016 for hospitalizations

The EHR has lots and lots of data!

- Potential pool for studies is huge:
  - Observational studies
    - 692K patients who have at least one encounter in the past 5 years
  - Trials
    - 289K patients with at least one encounter within the last one year

Note: These are PCORnet definitions

The EHR has lots and lots of data!

- ~1.2 million patients with an EHR record
- 8.1 million ambulatory visits
- 586K ED visits
- ~250K inpatient stays

What kinds of data are available?

- Data are of two general types
  - Structured data:
    - These come from entries made into discrete fields
  - Unstructured data:
    - These are free text notes and reports
    - Scanned in documents

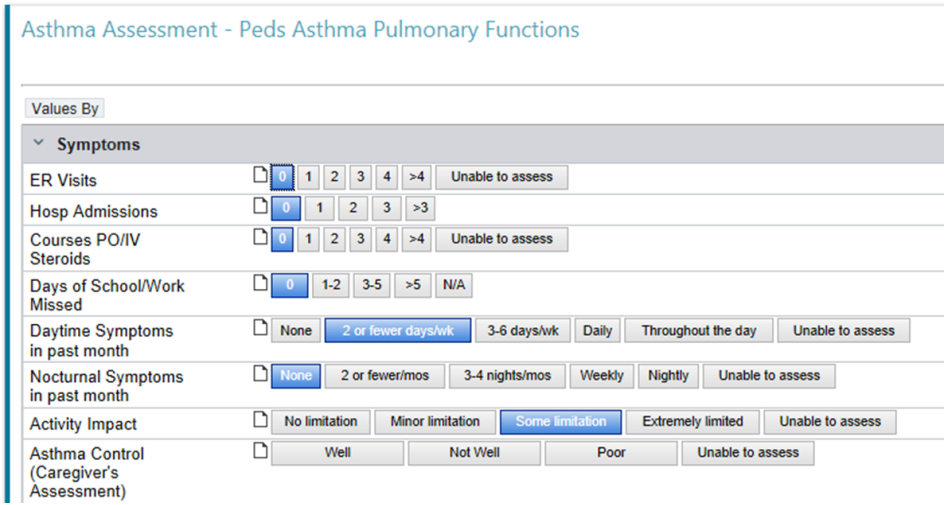
Structured vs unstructured

Structured	Unstructured
• Lab results • Flowsheets • SmartForms • SmartLists • (Mostly anything from a button or dropdown list) • Anything stored is a “discrete data field” ie. “in a box”	• Free text in notes * • Scanned in documents • * we currently have computer “reading” notes for billing, so this is expanding/progressing

Types of data available in epic

Type of data	Comments
Demographics (age, sex, ethnicity…)	Some fields may be missing or inaccurate (preferred language, race)
PMH/PSH/SH/FH	Dependent on user entry *and maintenance (e.g. smoking hx)*
Notes, encounters	Now includes outside info, but not 100% (Epic vs non Epic systems)
Labs, imaging, procedures	Now includes outside info, but not 100% (Epic vs non Epic systems)
Time stamps of all data	Allows calculation of time intervals
Metadata	Insight into user behaviors
"Paper" records scanned in	Internal and external sources

Structured data example



If you need discrete data for your clinical or research work, Epic team can create tools to capture it. (comment from our CMIO Dr. Fleece)

TEMPLE HEALTH

How data is stored and accessed:

- Chronicles
  - The real-time "transactional" database where you do all your work in Epic
- Clarity
  - The storage database that updates daily and serves as a permanent repository.
- **Caboodle:**

  Data Warehouse
  - Mega-database fed by all systems with useful data – clinical, financial, operational, etc.
  - *Coming soon*

Where's the data you want?

- Chronicles (real time) – "My Reports" aka Reporting Workbench in Epic. Self-serve.
  - Examples:
    - Currently admitted patients with Dx <X>
    - Patients discharged in the past 24 hours
    - Patients seen by me in past 30 days with diabetes
    - Current ER census
  - Queries limited to around past 6 months – otherwise, takes too long
- Clarity (long term storage)
  - Any combination of ANY discrete data stored there
  - Custom reports, long lookbacks
  - Won't show you today's data…
  - Must be created by IT report writers for you

What kind of research can I do using EPIC data?

Retrospective Research

- Descriptive epidemiology and patterns of care
  - What are the demographic characteristic of our population with <you name it disease>, and how are they initially managed?
- Comparative Effectiveness
  - How do people treated for <you name it disease> with <therapy A> differ from people receiving <therapy B> and how do their outcomes differ?
- Predictive analytics
  - What characteristics of people with Diabetes are associated with higher rates of re-admission?
    - eDERRI is a multi institutional study led by our own Dan Rubin, MD

Investigations with Epic

Quality Measurement

- Process measures
  - How many patients with Diabetes received appropriate preventative screening or meds (ophtho/ foot exams, ACE-I)
- Intermediate measures
  - How many patients with Diabetes have A1c < 7.0, BP< 130/90 and LDL < 100
- Outcome measures
  - How many patients with Diabetes developed end stage renal disease, AMI, foot amputation
  - We can stratify the above by provider or practice…

Investigations with Epic

Prospective Interventions

I'd like to survey/intervene upon people with the following characteristics

- smokers with COPD who have had an ED visit in the past year
- Patients on 4 BP meds who have BPs > 160/90
- Patients who missed a colonoscopy appointment
- Providers with > 50% panel with BP > 150/90
- Patients hospitalized with CHF who are readmitted within 30 days
- Patients with Lupus with stage 3 renal disease
- Patients with Diabetes whose GFR dropped by 50% in the past 2 years
- Patients on NSAIDs with GI bleed
- Patients started on coumadin yesterday

The Good News

- Epic has a great deal of data on thousands of patients spanning 6 years
- Lots of discrete data on demographics, meds, labs, diagnoses, and changes in these variables over time
- Possible to look up unstructured data (notes, path reports, radiology impressions) on a focused set of patients who meet other clinical criteria
- We can span ambulatory data with coarse information on admissions 2012-2016 and detailed inpatient data since Aug 2016

The challenges

- Defining the “computable phenotype”
  - Who has your disease of interest?
  - Who really has an incident case of your disease of interest?
  - What should count as a drug exposure?
- Regulatory issues in working with data
  - Who is authorized to receive and work with data, and how do patient identifiers matter
  - How can these data be transmitted?
  - Where can these data be stored and used for analysis
- There is a difference between “raw” data and an analytical data set

A word about computable phenotype

- Accuracy of data depends on accurate entry
  - Eg. DM
    - What the MA enters at triage may not be the right DM code
- Defining cohorts requires clinical input
  - DM can be defined as anyone with an A1c value vs anyone with A1c > 7.
    - Former was valid when we did not do screening A1cs

Research vs Quality Improvement

	Human Subjects Research	Quality Improvement
Purpose	**designed to develop or contribute to generalizable knowledge**	**designed to implement knowledge**, assess a process or program as judged by established/accepted standards
Starting Point	knowledge-seeking is independent of routine care and intended to answer a question or test a hypothesis	knowledge-seeking is integral to ongoing management system for delivering health care
Design	follows a rigid protocol that remains unchanged throughout the research	adaptive, iterative design
Benefits	might or might not benefit current subjects; intended to benefit future patients	directly benefits a process, system or program; might or might not benefit patients
Risks	may put subjects at risk	does not increase risk to patients, with exception of possible patients' privacy or confidentialty of data
Participant Obligation	no obligation of individuals to participate	responsibility to participate as component of care
Endpoint	answer a research question	improve a program, process or system
Analysis	statistically prove or disprove hypothesis	compare program, process or system to established standards
Adoption of Results	little urgency to disseminate results quickly	results rapidly adopted into local care delivery
Publication/Presentation	investigator obliged to share results	QI practitioners encouraged to share systematic reporting of insights

https://irb.research.chop.edu/quality-improvement-vs-research

QI Versus Research

What difference does it make?

- Research
  - Need to develop a formal protocol outlining exactly what you intend to do that gets approved by the IRB before you can begin work
  - Need to have a “HIPAA Authorization” for prospective studies when patients are actively enrolled or a an IRB-approved waiver of HIPAA Authorization to use retrospective data that includes identifiers
- Quality Improvement
  - Does not require IRB approval, and use of identifiers is permitted under HIPAA based on its more “operational” imperative.
  - For resident QI – need to track the conduct of the project and justify its status as QI as opposed to research

QI Versus Research

What difference does it make?

- BOTH QI and Research--- No difference when it comes to **DATA SECURITY**
  - Need to be mindful of “minimal necessary use” – justify what you need
  - Need to keep identifiable data on TUHS-connected machines
    - NO portable computers unless they are TUHS-supplied with whole disk encryption
    - Make use of the ”U” Drive – This is your TUHS “My Documents” folder that is accessible from anywhere you log in to a TUHS machine as yourself, or via Citrix
  - Be Mindful about mechanisms of transferring data
    - DO NOT USE external hard disks or thumb drives unless they are encrypted in a manner approved by the CISO

Accessing the Data

- Many personnel have access to Epic, but…
  - Must use Epic consistent with scope of practice
  - Clinicians and medical students looking through current and past patient charts related to clinical care is OK
  - Clinicians and medical students looking through patient charts with a research notion in mind is different – *You cannot just look around*

Accessing Data with Epic

- If the medical students are abstracting charts to pull out labs, medications, vital signs, smoking status, visit activity, *they are wasting their time and your money*!
- All of these data, and other information that is recorded discretely in Epic are available as a batch data dump if you have a list of MRNs of patients for whom you need these data
- If you need abstraction of content buried within notes, or scanned items or EKGs, that still requires Epic access for manual review

Medical Students involved in Research

- 1st and 2nd year medical students do not have automatic access to Epic, but access can be requested for research projects
  - Needs supervision from faculty and specific justification for access
  - No Citrix access*. Goal is to promote supervision by having them work in or around clinical department offices rather than from home
  - Needs to be listed as study personnel on research protocol
    - If exploratory work being done preparatory to research, work need special justification
  - Needs to complete eLearnings for Epic
  - Needs HIPAA and CITI training
  - Needs "EARF" form and special online form completed to capture essential information

*exception for students at St Lukes and for COVID

Medical Students involved in Research

- 3rd and 4th year medical students already have automatic access to Epic, so the path to engage them in research seems more straightforward… BUT
  - Needs supervision from faculty and specific justification for access
  - Goal is to promote supervision by having them work in or around clinical department offices rather than from home – even if they have citrix access
  - Needs to be listed as study personnel on research protocol
    - If exploratory work being done preparatory to research, work need special justification
  - Needs HIPAA and CITI training
  - **More difficult to track centrally, but above requirements are just as important as for 1st and 2nd year students!**

Research Staff Access to Epic

- Recognize that job roles of many, but not all, research staff require access to Epic, and that research staff are likely to be on multiple protocols over time
  - Needs supervision from faculty and specific justification for access
  - Most work should be done on site (Citrix access not settled)
  - Needs to be listed as study personnel on research protocol
    - If exploratory work being done preparatory to research, work needs special justification
  - Needs to complete eLearnings for Epic
  - Needs HIPAA and CITI training
    - HIPAA training different for LKSOM vs TUH employees
  - Needs “EARF” form and special online form completed to capture essential information

Write Access to Epic for Research

- Rarely granted
  - Can provide write access to research staff with appropriate justification
    - Only example where write access is granted is for research RNs administering a research intervention as part of regular treatment
    - Research medications such as those used to rx COVID are listed as “treatment vs placebo” in the electronic record and actual administration record is on paper

Show me the data!

- Are you “just” doing retrospective reviews?
  - No IRB approval required for aggregate data pulls – basically counts of patients meeting criteria that you specify
    - easy to obtain, but not useful for work beyond preliminary explorations
    - Not likely to be enough for a quick abstract submission
  - No IRB approval required for use of truly de-identified individual-level patient data – meaning no names, MRNS, zip codes, dates.
  - However, in many cases, you will still need IRB approval since inevitably, research team will need to look at a representative sample of charts to understand the context of the data a little better
  - Even if your analytic data set does NOT contain identifiers, you still need IRB approval if you began with source data that has identifiers. You will likely be keeping a crosswalk between “fake” and real identifiers so you can explore idiosyncrasies that are unmasked through the analysis
  - It is good practice to obtain a letter from the IRB; for deidentified data the IRB will deem it “not human subjects research”.

Show me the data!

- Do you plan to recruit patients for a research protocol?
  - ALWAYS requires IRB approval
  - Most straightforward when patients opt-in in response to posted fliers or if the investigator reaches out to his own patients when they are being seen for a clinical appointment with that physician
  - Increasing regulatory challenges as recruitment extends to other practices, or if research staff need to recruit via mail, or phone
  - Targeted recruitment will involve developing “computable phenotype” in which a database query is constricted incorporating the various clinical criteria. Patients are stratified by providers they have seen, and the providers, in turn, review these patient lists to authorize recruitment.

Show me the data!

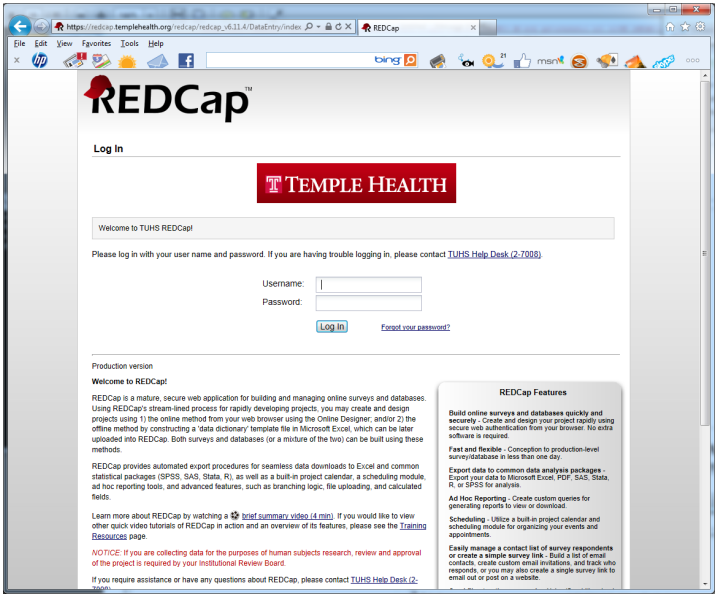
- Sometimes you will have a list of patients that you know you need data
- Sometimes you have a sense of clinical criteria of patients that you are interested in exploring in more depth
- In either case, we can provide structured output of all encounters, diagnoses, labs, medications, etc on the cohort of patients you provide, or that we define through a “computable phenotype”
- **Having the data ≠ being able to work with the data**

Storing Data

Need to move away from Excel for primary storage

- Familiar and convenient, but…
  - Exists as loose files on someone's PC
    - Can only be accessed in one place at one time
    - Security and data loss risks
  - Version control is difficult if two people are working on it
  - No access controls for reading or writing data
  - Can be too easy to use in which individual cells are "overstuffed" with data, or data of the wrong type
  - No audit trail of access or updates to data

Better Solution



REDCap

- A free, web-based data capture tool developed and continually enhanced by a team at Vanderbilt with a broad, mostly academic user community
- Access from anywhere
- ***You already have access with your TUHS windows credentials***
- Online training videos to get you started
- Data stored in a secure TUHS server that is backed up twice daily
- You create custom data collection forms to capture data as text, radio buttons, checkboxes – including use of “branching logic”
- Has data type enforcement to ensure numbers are numbers, dates are dates, phone numbers have all the expected digits

REDCap

- Access a library of standard data collection instruments to incorporate into your research study
- Renders forms meant for viewing and/or completion by authorized research staff
- Renders forms as surveys meant for study subjects to complete
  - Can send links to surveys in a manner where you can keep track of who has responded
  - Can schedule automated survey reminders
- Audit trails of all access and data updates
- Can export data into SAS/SPSS formats (and Excel)
  - Can flag fields as identifiers that will “hash” values or perform date shifting to avoid unwanted exposure of identifiers with export.

Accessing Data with Epic

- Data requests handled by John Turella via links below, unless they are for the COVID database created in the spring
- REDCap –
  - http://redcap.templehealth.org/redcap
- Epic-for-research access form
  - https://is.gd/Epic_Access_Link
- Data access request
  - https://is.gd/EHRDataRequestForm

Data Self Service

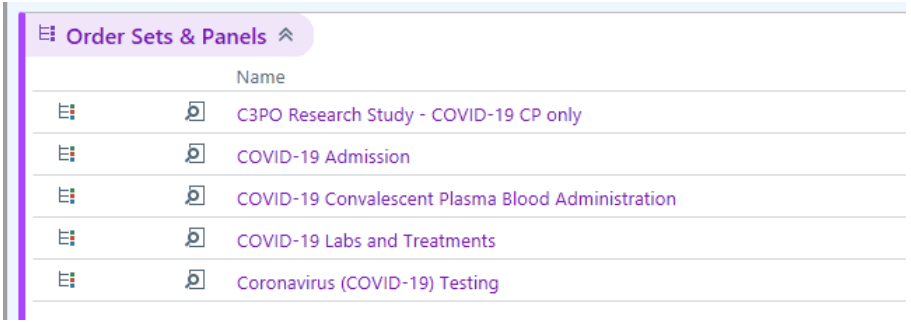
- Epic has a “reporting workbench” capability that allows you to query for patients with certain characteristics.
- Coming soon is COSMOS!
  - Access to multiple institutions data using the EPIC slicer dicer tool.
  - Not live yet
  - Potential for large studies using deidentified data

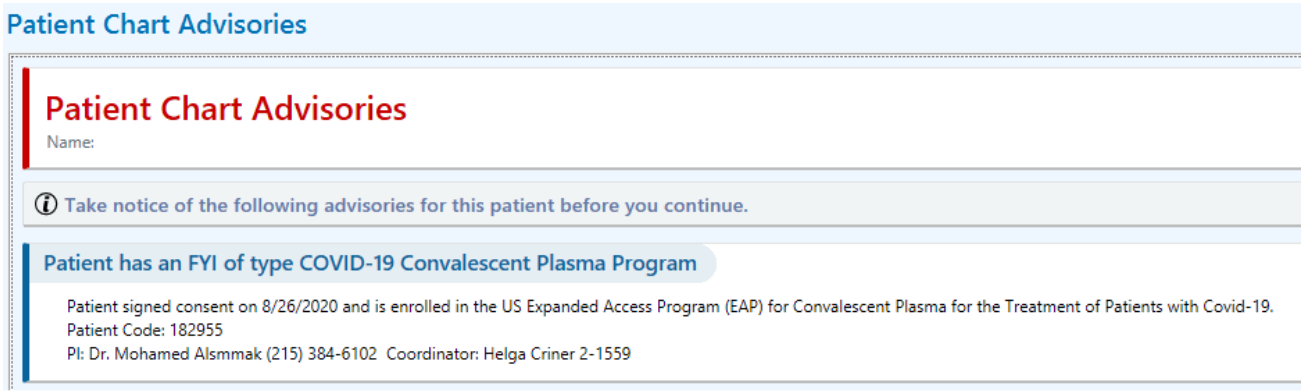
What else can EPIC do for my research?
Decision support tools

- Patient identification and recruitment
  - Alerts or Pt lists of potential study subjects
  - MyTempleHealth messaging
- **Links to study protocols and consent forms**
- **Study protocol order sets**
- **Provider awareness**
  - "This patient is participating in study X. Please notify Y if patient is admitted."

TEMPLE HEALTH

Decision support tools





TEMPLE HEALTH

Questions?

Conclusions

- We have access to a lot of data
- Having data and putting it to meaningful use are two separate issues
- Need to be mindful of the regulatory issues related to access to Epic, and data generated from Epic.