


Using Epic Data for Research


OCTRI Biomedical Informatics Program

Robert Schuff, MS
Oct 20, 2014



Agenda

- Research Uses of Epic
 - Cognos / Clinical Data Marts
 - Research Data Warehouse
 - Cohort Discovery




Research Uses of Epic

- Patient counts to determine study feasibility
- Patient lists for study recruitment
- Large-scale retrospective “data-only” studies

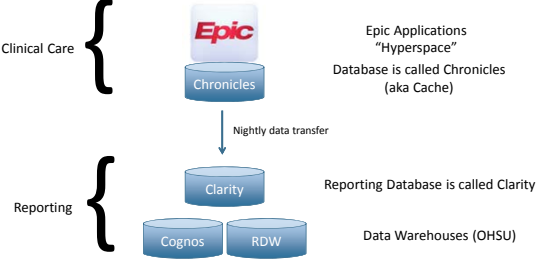
Clarity or Data Warehouse

- Collecting structured data for research
- Clinical trial alerts for recruitment
- Monitoring patient/participant events

Epic / Hyperspace



Epic Data Lingo




Clinical Care { Epic Chronicles Database is called Chronicles (aka Cache)

Reporting { Clarity Reporting Database is called Clarity



Cognos RDW Data Warehouses (OHSU)

Nightly data transfer

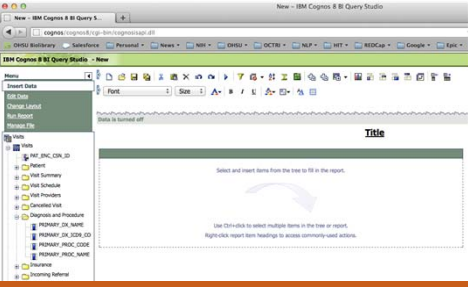



Cognos (Clinical Data Marts)

- Available through Cognos
- Web-based report writer similar to MS Access
- Several data structures available
- Training and access available through Clinical Reporting Group
- Primarily designed for clinical operations reporting

Cognos – (Clinical Data Marts)

Research Data Warehouse (RDW)



What is the RDW?

- IRB-approved repository of clinical and research data created and maintained by OCTRI *specifically for research purposes*
- Contains Epic data as primary data source
 - Also includes Pathology, Cancer Registry
 - Integrates eIRB and opt-out data to facilitate compliance
- Organized to allow aggregation across millions of patient records quickly



Types of Research Activities Supported by the RDW

- Preparatory to Research
 - Feasibility (counts and aggregates)
 - Cohort identification
- Conduct of Research Protocol
 - Patient lists for recruitment
 - Data extracts chart review, safety monitoring, registries etc.
 - Large dataset extraction (e.g. epidemiology, notes for NLP projects)



How much data are in the RDW?

- Epic Ambulatory since 2005
- Epic Inpatient and ED since May 2008
- >2.8 million patients
- >600,000 patients with at least 1 “visit”
- >10.3 million “visits”
- Epic data is complex!
 - Clarity contains over 8,700 data tables and over 75,000 data elements



What types of data are available?

- Demographic information
 - gender, race, ethnicity, preferred language, employment status
- Provider, insurance coverage
- Birth history
 - date of birth, gestation, birth weight
- Vitals
 - blood pressure, weight, height, pain ratings
- Contact and census information
 - Addresses, phone, census information
- Lab orders and results
- Diagnoses, problem lists, medical history (ICD-9, ICD-10)
- Hospital encounter information
- Medication lists and orders
- Ambulatory encounter information
 - clinic date, provider, department/clinic, PCP, chief complaint, cancel reason
- Procedures
- Surgeries
- Insurance coverage for patient encounter
 - benefit plan, insurance class, co-pay



Accessing the RDW


- Data can be released to investigators in three ways
 - Counts - no IRB required.
 - De-identified - IRB determination of “non-human subjects research”
 - Fully identifiable - requires IRB approval



Accessing the RDW

Self-service with Cohort Discovery


- Data can be released to investigators in three ways
 - Counts - no IRB required.
 - De-identified - IRB determination of “non-human subjects research”
 - Fully identifiable - requires IRB approval



Accessing the RDW


Through OCTRI Analysts

- Data can be released to investigators in three ways
 - Counts - no IRB required.
 - De-identified - IRB determination of “non-human subjects research”
 - Fully identifiable - requires IRB approval




Cohort Discovery

Self-service RDW queries for patient counts

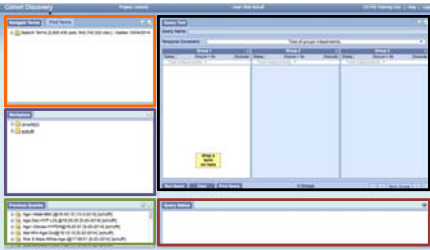


Cohort Discovery - Purpose


Provide researchers with a self-service web-based tool to determine study feasibility or discover patient cohorts during the “preparatory to research” phase by obtaining counts of patients based upon specified inclusion and/or exclusion criteria



Cohort Discovery - Screen




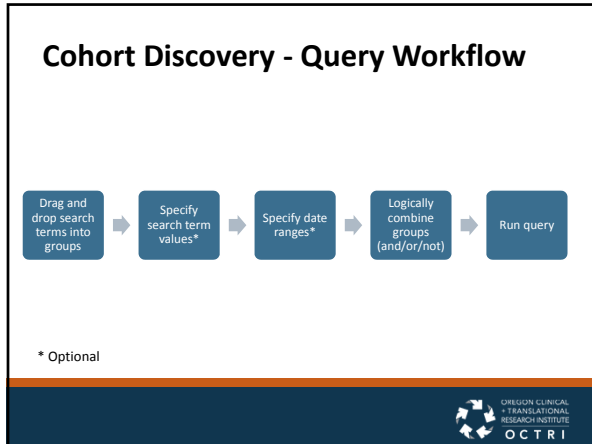
- Navigate/Find Terms** used to find and select concepts to include in query
- Query Tool** area where queries are built and specified
- Query Status** query results displayed
- Workplace** used to save queries to your private or a shared area for collaborative development
- Previous Queries**



Cohort Discovery - Data Organization

- Data are organized into a collection of “facts” (observations) about patients
- Each **observation** belongs to a **patient** and has a value, a date, and a link to a **concept** describing the observation





- ### Protecting Patient Privacy
- All data are de-identified before being loaded
 - Surrogate patient identifiers
 - All dates associated with a given patient are shifted by a random amount at least +/- 15 days
 - Patients older than 89 are all recorded as 90 yo
 - True patient counts are obfuscated
 - Returned count is perturbed by up to +/- 3 each time the query is invoked
 - If the true count is less than 10 the returned result is "Less than 10"
 - Accounts locked if activity is suspicious
-

Cohort Discovery Demo

Accessing RDW and Cohort Discovery

<p style="text-align: center;">Research Data Warehouse</p> <p><i>IRB-approved research repository of Epic data for researchers</i></p> <ul style="list-style-type: none"> fully identified datasets for recruitment or current participant data de-identified datasets for retrospective/epidemiological studies 	<p style="text-align: center;">Cohort Discovery</p> <p><i>Search the RDW for a set of patients meeting certain inclusion or exclusion criteria</i></p> <p style="text-align: center;">Cohort Discovery Tutorials Offered Monthly</p> <p style="text-align: center;">Next scheduled Thursday, October 30, 2014, from 10:00 am to 12:00 pm</p>
---	--

Email octri@ohsu.edu or octrihelp@ohsu.edu for details

Thank You!

For more information please email me!

schuffr@ohsu.edu