Introduction to FSM Research Analytics

Daniel Schneider
Manager of Research Analytics

First Mondays Faculty Development
05/02/2016
Agenda

• What is the EDW?
• What types of research can it help with?
• How to get started!
• Q/A
What is the EDW?
The EDW Provides a Campus Data Platform for Your Approved Projects

Participating Entities

Contribute
Report
Analyze

EDW

Load
Cleanse
Enhance

Campus Data

Analyse Data and Create Reports

Provide Campus Data Platform
What is the EDW?

The EDW is a SQL database that contains data from many of the transactional systems used by the campus.
What is the EDW?

Clinical data from NMG (EPIC)

Billing data from NMG (IDX/Epic)

Clinical data from NMH (Cerner)

Billing data from NMH (PRIMES)

More clinical systems (ancillary)

Research systems from FSM
What does FSM Research Analytics do?

The end goal of the EDW is to provide a more comprehensive dataset and options for interacting with and understanding it.
Intro to the EDW

Common Workflows

Research, Operations or Quality

**Research** IRB approval is required if EDW is used for research purposes, especially when results may be published in a journal.

**Operational** Reports that display metrics used by administrators on a day-to-day basis.

**Quality** Reports must support an NMH DMAIC, be sponsored by a physician department chair at NMG or otherwise officially sponsored.

*Research reports should not be used for operational or quality purposes; the reverse typically holds true, as well.*
Research Uses of the EDW

• Study Feasibility
• Clinical Trial Subject Recruitment
• Retrospective Chart Review
• Prospective, Longitudinal Studies
Study Feasibility

• Aggregate counts of patients meeting defined criteria
• No approval needed for prep-to-research data containing no PHI
• i2b2 is currently available and is being actively developed

Example:

“How many patients we've had over the years with ovarian cancer and a recurrence in the brain...”
Clinical Trial Subject Recruitment

• Real-time scheduling data from NMG
• Primary care physician contact information
• Daily updates to the subject population

Example:

“Patients with the diagnoses of peripheral arterial disease and/or coronary artery disease...”
Retrospective Chart Review

• Data in the EDW reaches back about to 1998...
• ... but becomes robust after 2000
• Discrete fields can be used to efficiently define and capture data elements for a population
• Complex NLP can be leveraged for free-text, though not as reliably as discrete fields

Example:

“The study is evaluating patients at NMG Ophthalmology with the following ICD9 codes between Jan 1, 2004 and Jan 1, 2011...”
Prospective, Longitudinal Studies

- Disease, other study registries can be built to keep track of a population, events and outcomes
- Data can be integrated with custom applications

Example:

“For consented patients provide the following data points for the first available Cardiac Cath Report: 1. Mean RA  2. Mean PA  3. PA Saturation %  4. Mean PCW  5. Mean LV  6. Fick Output...”
Getting Started
Getting Started...

EDW
NORTHERN MEDICAL DATA WAREHOUSE

MY REPORTS

**Avg Hours by Month**
Avg Hours by Month - Average of dev_hrs and variance of dev_hrs for each hrs_entered_dts Month broken down by hrs_entered_dts Year. Color shows details about hrsLogged_by. The data is filtered on re...
Path: /Reports/EDW/EDW Operational Metrics/Analyst Workbook/Avg Hours by Month

**Dollars YTD**
Dollars YTD - Sum of Hours2Cash for each hrsLogged_by. Color shows details about task_id. Details are shown for summary. The data is filtered on request_type, status and hrs_entered_dts. The request...
Path: /Reports/EDW/EDW Operational Metrics/Analyst Workbook/Dollars YTD

**EDW Realtime**
A finger on the pulse of the EDW. Shows report executions, application accesses, and AskEDW queries.
By: Parsons, Ford
Path: /Reports/EDW/Security/EDW Realtime

eNOTISelRBAuthorizedPersonnel
Given a study id [param name irb_number] this query returns the user_id (aka netid), first_name,
Getting Started...

- Submit Data Request using the EDW Portal
- Work with an EDW Analyst to refine requirements
- Service Agreement
  - Data Steward approvals
- Reports retrieved from the EDW Portal

It’s an iterative process!
Requesting an EDW Report

- IRB and NMHC approval is needed
- Discrete data elements are best
- Be able to send screen shots of the data elements you’re looking for
- Free text is difficult to work with
  - NLP can be used but is time consuming (expensive) and rarely as accurate as a discrete field
- Recharge rate of $75/hour
What is a Data Steward?

• The data in the EDW is still owned by the contributing organizations

• Data Stewards are representatives from these organizations and approve data and access requests

More information on the EDW Wiki
http://edw.bioinformatics.northwestern.edu/edwwiki/index.php/Data_Steward_Policy
Future Directions
Future Directions

• Integrated Data Structures (IDS)
  – Incorporate new source data (Cadence, CDH)

• Predictive Analytics/Biostatistics
  – Integrate R into our analytics infrastructure
  – Provide preliminary statistical analysis
  – Utilize advanced data visualizations

• Self-Service Query and Annotation
  – REDCap
  – Abstractor

• Predictive Analytics
  – RAPID

• Advanced Natural Language Processing
  – Siddhartha Jonnalagadda (sid@northwestern.edu)
Quick Reference

NMEDW@northwestern.org
http://edw.northwestern.edu
https://grants.nubic.northwestern.edu/
Questions?