Health Analytic Group

Research Data Management

Objectives

- Specify several different appropriate data sources
- Describe and appreciate limitations of data sources
  - Timing
  - Workflow process
  - Completeness
  - Accuracy
- Understand the difference between manual chart abstraction and programmatic field extraction

Objectives

- Identify Protected Health Information and determine if necessary for project
- Explain how to submit requests
- Understand impact on turn-around times of requests
- Understand impact of changing scope of request

Identifying Data Sources

- **EPIC EMR** (electronic medical record)
  - Most recognized
  - PreEPIC, multiple EMRs and even paper.
  - View care timeline across facilities
  - Resulting large data set:

    -pull target research populations (ie. labs, bmi, dx)

    -pull data to analyze historical trends and predictive models (ie. to reduce admission rates)

    -comparison models (ie. age, bmi)

Identifying Data Sources

➢NATIONAL REGISTRIES

- ie. National Trauma Registry (current request)
- ie. National Death Index (older request)
- Unusable to Usable form; raw data sometimes Gigs
- Integrate with Epic records
- Fee?

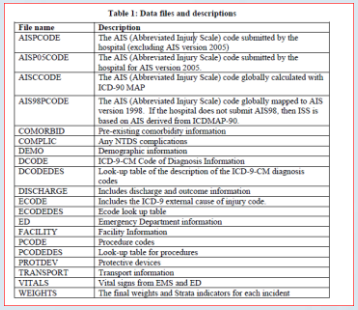


Table 1: Data files and descriptions

File name	Description
AISPCODE	The AIS (Abbreviated Injury Scale) code submitted by the hospital (excluding AIS version 2005)
AISP05CODE	The AIS (Abbreviated Injury Scale) code submitted by the hospital for AIS version 2005.
AISCCODE	The AIS (Abbreviated Injury Scale) code globally calculated with ICD-90 MAP
AIS98PCODE	The AIS (Abbreviated Injury Scale) code globally mapped to AIS version 1998. If the hospital does not submit AIS98, then ISS is based on AIS derived from ICDMAP-90.
COMORBID	Pre-existing comorbidity information
COMPLIC	Any NTDS complications
DEMO	Demographic information
DCODE	ICD-9-CM Code of Diagnosis Information
DCODEDES	Look-up table of the description of the ICD-9-CM diagnosis codes
DISCHARGE	Includes discharge and outcome information
ECODE	Includes the ICD-9 external cause of injury code
ECODEDES	Ecode look up table
ED	Emergency Department information
FACILITY	Facility Information
PCODE	Procedure codes
PCODEDES	Look-up table for procedures
PROTDEV	Protective devices
TRANSPORT	Transport information
VITALS	Vital signs from EMS and ED
WEIGHTS	The final weights and Strata indicators for each incident

298599 LOC47 999999925898 X-X--X--------XX-34-X---- 925995 -5489 X45 422859522 9895J589 25Q5269 22445 3F592 65I552 95X49 F591 I5009 J519 C3259

Identifying Data Sources

➢ DEPARTMENT REGISTRIES (cont.)

- Local Registries

  -RMH Trauma; mandated?;upload to NTR

- Diabetes/Cardio related EPIC registries

  -membership inclusion criteria, datapoints of significance, scheduled

Identifying Data Sources

➢OTHER EXTERNAL DATABASES EXIST

- ie. Virginia All Payer Claims Database (APCD)

  -under the authority of the Virginia Department of Health; analyze your Carilion dataset alongside state data

Sources and Status of APCD Data Collection

Payers	Currently Collected	Planned Collection
Commercial Payers	Yes	--
Third Party Administrators/Self-Funded	Yes	--
Medicaid	Yes	--
Medicare	--	Yes
TriCare	--	Yes

Types of Data Collected	Currently Collected	Planned Collection
Medical Claims	Yes	--
Eligibility	Yes	--
Pharmacy	Yes	--

Typically Included APCD Information

-Encrypted social security
-Patient demographics(date of birth, gender, residence, relationship to subscriber)
-Type of product (HMO, POS, Indemnity, etc.)
-Type of contract (single person, family, etc.)
-Diagnosis codes (including E-codes)
-Procedure codes (ICD, CPT, HCPC, CDT)
-NDC code / generic indicator / other Rx
-Revenue codes
-Service dates
-Service provider (name
-tax id, payer id, specialty code, city, state zip code)
-Prescribing physician
-Plan charges & -payments
-Member liabilities (co-pay, coinsurance, deductible)
-Date paid
-Type of bill
-Facility type

Identifying Data Sources

- **Enterprise Data Warehouse (EDW)**
  - Can be populated by external entities
  - Claims information
  - Example (medicare patient claims)

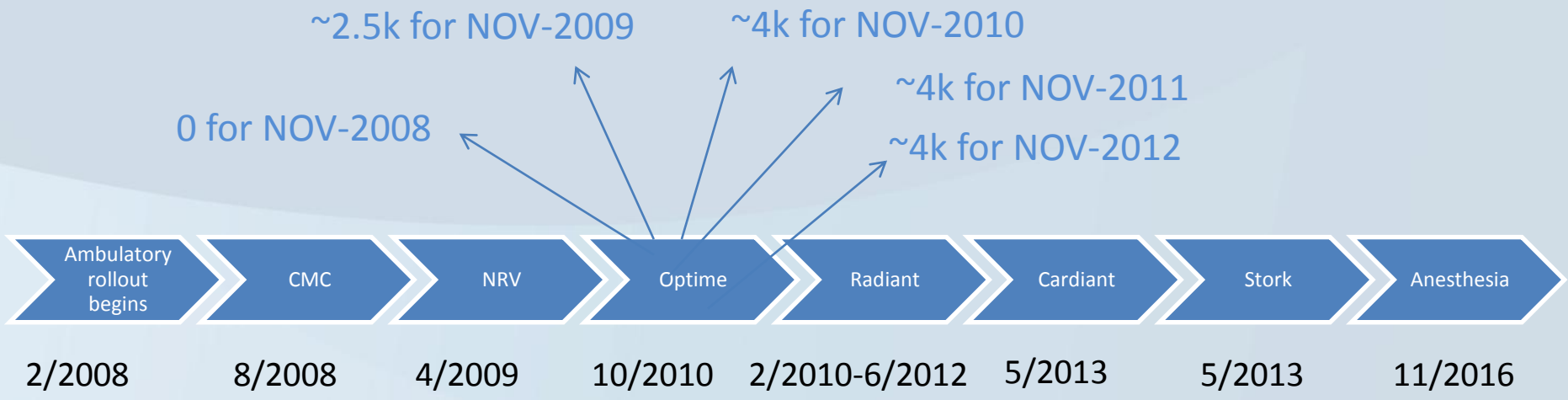
Identifying Data Sources

- **SURVEYS**
  - Important, data not documented elsewhere
  - Tie survey data to Epic patient or provider related data
  - Examples (surgery resident procedures, patient experiences)

Data Source Limitations

- Timing
- Workflow/Process
- Completeness
- Accuracy

Data Source Limitations: Timing



If you attempt to pull populations prior to these go live dates, you will have incomplete data without significant manual chart review, through paper charts. The data may not even be available.

Data Source Limitations:
Workflow/Process

- Who records the data?
  - Department or external
- What is the process that is followed?
  - Great way to gain perspective on your data
- Are there inconsistencies in the process?
  - If so, many reasons (i.e.'s)
- Have workflows changed over time?
  - If so, many reasons (i.e.'s)

Data Source Limitations:
Workflow/Process

Example:

- Code Blue documentation changed over time. In addition, the original workflow was inconsistently followed and there weren't standard processes for nursing.
- Solution: Limit the time period needed to a consistently documented time frame.

Data Source Limitations: Completeness

- How complete is the documentation on the population of interest?
- Are there significant gaps/holes that would bias the results?
- How critical are the data points that may be incomplete

Data/Source Limitations: Completeness

Example:

- "For injury research, external cause-of-injury codes (E- codes), which are used to classify injury incidents by intent (e.g., unintentional, homicide/assault, suicide/self-harm, or undetermined) and mechanism (e.g., motor vehicle, fall, struck by/against, firearm, or poisoning), are incomplete in the majority of state inpatient and ED databases."
  http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3744773

Other examples:

- Ethnicity/race, employer, history of disease

Data Source Limitations: Accuracy (datapoints are present, but are they reliable?)

- Is the data discrete or free text?
- Is it collected the same way each time?
- Are the collectors consistent?
- Changes during time period that would impact

Data/Source Limitations: Accuracy

Example:

- NEDOC
  - Score is a measure of overcrowding in the ED.
  - Formula calls for #ED beds
  - January 2016 is when hospital bed increased from 504 to 532

Example:

- Ethnicity/Race
  - Visual observation or actual?

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3744773/

Manual Chart Abstraction / Programmatic Field Extraction

- Describe manual vs. programmatic extraction
- Which method is faster? Which method is used?
  - How many records result from the pull criteria
  - Data points
    - # of data points
    - # EPIC screens and # Database tables
    - type of datapoints
      i. clear,discrete field?
      ii. most recent data point; historical flags (we need to look at every table where this may be recorded)
      iii. calculated?
    - masking PHI
- Data collection tool during request submission helps us to ascertain how to achieve best turnaround times.

Protected Health Information

- Name
- All Dates/Times
- Address information including zip code
- Date of birth (and age if >89)
- Telephone/Fax numbers
- Email
- SSN
- MRN, account numbers, insurance numbers

Protected Health Information

- Certificate/license numbers
- Vehicle, device numbers
- URLs/IP addresses
- Biometric identifiers
- Full face photographic images
- Unique numbers or codes including initials
- Rare conditions

Protected Health Information

- If you don't need it, don't collect it
- Forms available through Research and Development:
  - https://www.carilionclinic.org/research-development/forms
  - PHI form (if you are interested in accessing PHI for your research project)
  - research Login Account

Submitting Requests

- https://mytsc.carilionclinic.org
- General and Research/QAQI submission forms
- Codes (if known)
  - Procedure codes (ICD/CPT)
  - Diagnosis codes
  - Lab codes
  - Provide detailed names
- Inclusions/Exclusions Criteria
- Deadline and Deadline driver
- Feasibility counts request?

Turn Around Times on Requests

- Triage
- The more complex, the longer it takes
- The more focused and specific the request, the less time it takes

Impact of Changing the Request

- Some changes small
- Others are complete rewrites and will impact the timeframe of delivery
  - May involve many more tables
  - May be much more complex

Questions

?