The Boston Medical Center Clinical Data Warehouse
…lots of answers
… what's the question?

Linda Rosen, MSEE
Clinical Data Warehouse Research Manager

What is a Database?

A collection of information organized to provide efficient retrieval.

What is the Clinical Data Warehouse?

In 2005, Boston Medical Center embarked on a major project to collect data spread throughout its many electronic systems into a consolidated, organized and accessible database for analysis, reporting and research purposes.

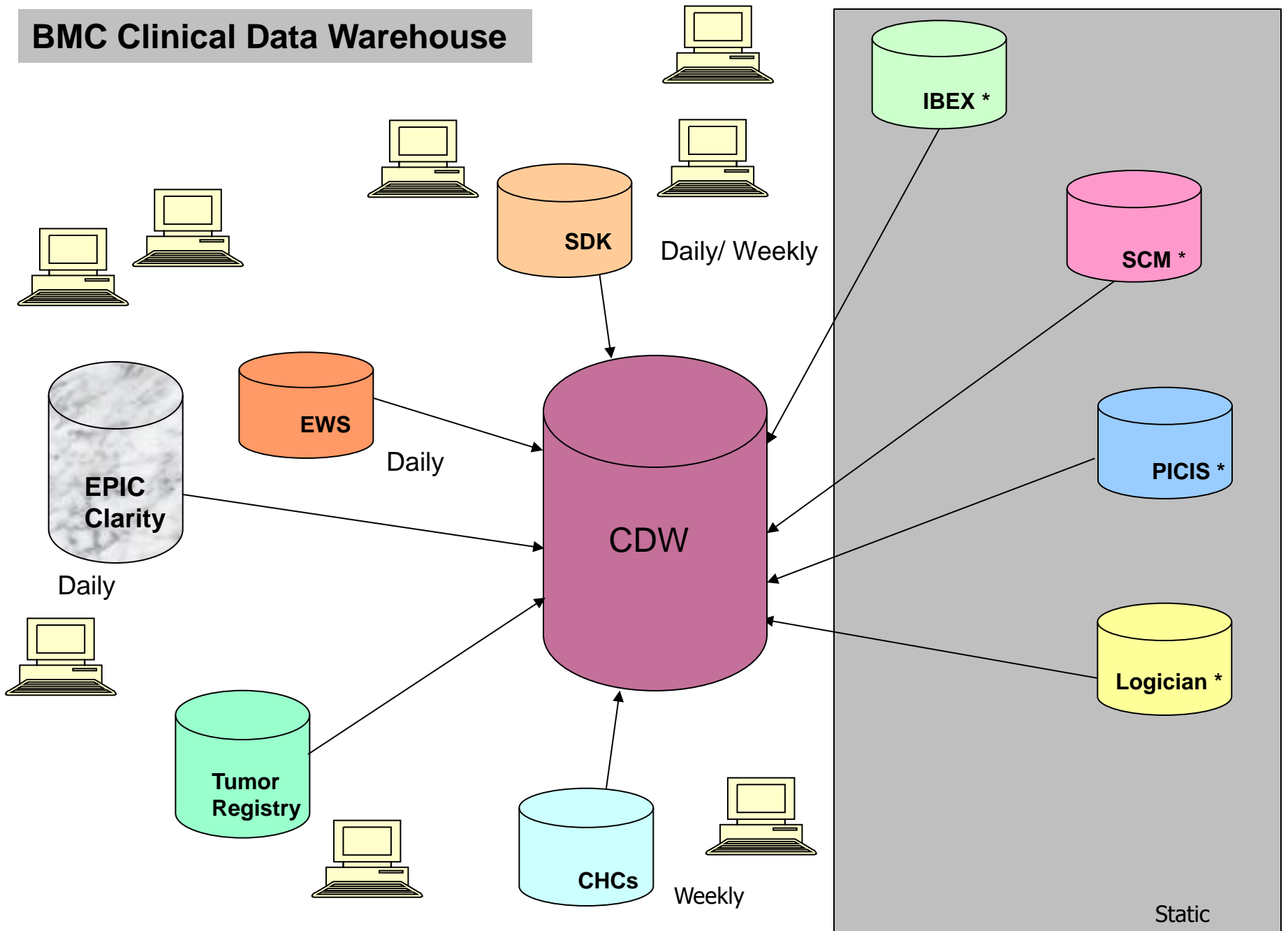The Clinical Data Warehouse currently has about 12-15 years of clinical data -- depending on the application.

The CDW is a resource for many different purposes — Quality Control, Improvement in Care, Regulatory Reporting, Daily Business Reporting, State Reporting… and for research.

The CDW is

- A database containing data from multiple sources.
- Many years of historical data extracted from BMC's various clinical software packages – some now static.
- A database containing data related to each other with unique identifiers.
- A database that is only as good as the data entered.
- A significant resource for researchers.
- NOT real time data.

BMC Clinical Data Warehouse

SDK Daily/ Weekly

EWS Daily

EPIC Clarity Daily

Tumor Registry

CHCs Weekly

CDW

IBEX *

SCM *

PICIS *

Logician *

Static

How to use the CDW for Research?

- What's available?

- How do I get data?

- Why can't I search myself?

- When do I need IRB approval?

Data currently available in the CDW

- SDK – Registration, billing, visit dx Info (ICD-9 codes)
- *Logician – Outpatient, Labs, Vitals, Visits, Problems, Meds
- *SCM – Inpatient Data, Labs, Documents, Meds
- *IBEX – Emergency Department, Meds,
- *PICIS – Surgery
- Tumor Registry
- EWS – Appointments
- Anesthesia Manager
- EPIC
- EPIC Notes – accessible, but not in the CDW
- Some Community Health Center Data
- * static data

Person
- PID
- EXTERNALID
- SEX
- GENDER
- DATEOFBIRTH
- ADDR1
- ADDR2
- CITY
- STATE
- ZIP
- PCP_ID
- STATUS
- LANGUAGE

Encounter
- PAT_ID
- PAT_MRN_ID
- ACCOUNT_NUMBER
- PAT_NAME
- PAT_ENC_CSN_ID
- CONTACT_DATE
- INPATIENT_DATA_ID
- ADMIT_DTM
- DISCHARGE_DTM
- ED_DEPARTURE_DTM
- TRIAGE_STATUS
- HOSP_ADMSN_TYPE
- APPT_STATUS
- ENCOUNTER_TYPE
- REGISTRATION_STAT
- DISCH_DISPO
- HEIGHT
- WEIGHT
- SMOKING _STATUS

Orders
- PAT_MRN_ID
- ACCOUNT_NUMBER
- PAT_ID
- ORDER_PROC_ID
- CHILD_ORD_PROC_ID
- ORDERING_DATE
- ORDER_TIME
- PROC_CODE
- ORDER_CLASS
- COSIGNED_USER_ID
- PROC_ID
- STANDING_PRDER
- STANDING_OCCURS
- STAND_INTERVAL
- INSTANTIATED_TIME

Medications
- PAT_ID
- MRN
- ACCOUNT_NUMBER
- MAR_DESCRIPTION
- ORDER_ID
- DISPLAY_NAME
- START_DATE
- END_DATE
- ORDER_START_TIME
- ORDER_END_TIME
- TAKEN_TIME
- SIG
- INFUSION_RATE
- TASK_STATUS
- DOSE_UNIT
- MEDICATION_DOSE
- MEDICATION_FREQ

ClientVisits
- IDCODE
- GUID
- CLIENTGUID
- ADMITDTM
- DISCHARGDTM
- TYPECODE
- VISITIDCODE
- SERVICEGUID

Observations
- PID
- HDID
- OBSDATE
- OBSVALUE
- SDID
- CHANGE_DATE
- OBS_TYPE
- DESCRIPTION
- RANGE
- STATE
- USRID
- STATE

Patient
- PAT_MRN_ID
- PAT_UD
- PAT_NAME
- BIRTH_DATE
- SEX
- PAT_FIRST_NAME
- PAT_LAST_NAME
- PAT_MIDDLE_NAME
- ADD_LINE_1
- ADD_LINE_2
- CITY
- STATE
- ZIP
- HOME_PHONE
- WORK_PHONE
- LANGUAGE
- ETHNICITY
- MARITAL_STATUS
- RACE

Appointments
- ID
- DTM
- STATUS
- SCHED_PROV
- DEPARTMENT
- LOCATION
- CANCEL DT
- RESECHED DT
- COMMENTS
- WHEN SCHED
- DURATION
- REASON

Visits SDK
- MRN
- ACCOUNT_NUMBER
- ADMIT_DTM
- DISCH_DTM
- PCP_CODE
- SERVICE_AREA
- LOCATION_CODE
- COMPLAINT
- CHARGES
- DX1...DX100
- BALANCE
- DISCH_STATUS

ClientVisitLocation
- GUID
- CLIENTVISITGUID
- CLIENTGUID
- TRANSFERREQUESTDTM
- LOCATIONGUID
- CREATEDWHEN

Results
- ORDER_PROC_ID
- PROC_ID
- LINE
- PAT_ID
- SERV_AREA_ID
- COMPONENT_NAME
- RESULT_TEXT
- RESULT_NUM
- SPECIMN_TAKEN_TIME
- RESULT_FLAG_TXT
- RESULT_DATE
- COMPONENT_ID
- ORDER_TIME

Problems
- PROBLEM_LIST_ID
- PAT_ID
- DX_ID
- ICD9_CODE
- NOTE_DATE
- PROBLEM_STATUS
- CLASS_OF_PROBLEM
- PROBLEM_CMT
- PRIORITY
- RESOLVED_DATE
- CHRONIC_YN
- PRINCIPAL_PL_YN
- HOSPITAL_PL_YN
- IS_PRESENT_ON_ADM

Problems
- PID
- CODE
- DESCRIPTION
- QUALIFIER
- START_DATE
- STOP_DATE
- STOP_REASON
- CHANGE_DATE
- ANNOTATE
- USED_ID
- ENTRY_DTM
- SIGNED_DTM
- STATUS

IBEX visits
- MEDREC
- COMPLAINT
- ID
- ACUITY
- ADMIT_DATE_TIME
- EXIT DATE
- DISCH STATUS
- LNAME
- FNAME
- DOB
- GENDER
- ACCTNUM

Documents
- SDID
- PID
- SUMMARY
- DOC_TYPE
- USR_ID
- LOCATION_OF_CARE
- CHANGE_DATE
- PUBUSER
- PUBTIME
- JOBTITLE
- CLINICALDATE

Surgery
- CASE_ID
- SCHED_SURGERY_DT
- CASE_SCHED_STATUS
- CANCEL_REASON
- CANCEL_DATE
- SURGERY_DATE
- ACCOUNT_NUMBER
- PAT_ID
- PAT_MRN
- RESULT_FLAG_TXT
- RESULT_DATE
- COMPONENT_ID
- PNL_1_PRIM_PROC_NM_WID

Observations
- ITEMNAME
- CLIENTGUID
- CLIENTVISITGUID
- VALUE
- GUID
- ORDER_GUID

Patient SDK
- MRN
- GENDER
- DOB
- ADDR
- CITY
- STATE
- ZIP
- RACE1
- RACE2
- RACE3
- ETHNICITY1
- ETHNICITY2
- PATIENT_NAME

FAQs

- Where do I start?

FAQs

- Where do I start?

    Start with a research question (or two).

FAQs

- Where do I start?

    Start with a research question (or two).

    Get cohort size to see if pursuing an IRB protocol is worthwhile.

FAQs

- Where do I start?

  Start with a research question (or two).

  Get cohort size to see if pursuing an IRB protocol is worthwhile.

- What kind of data →IRB or Not?

  Counts

  De-identified Data

  Identified Data

  Do you need subjective data?

  Is your research dependent on highly textual data?

Regulatory Issues

- Permission is the same as for accessing other databases, individual paper records or electronic files
- IRB approval for anything other than counts
- Researcher must be part of the covered entity
- Researchers must complete Human Subjects Research training

What type of Data do you need?

Data Counts:

- Fill out online data request.

Anonymous Data:

- IRB exempt application and relevant HIPAA forms

Identifiable Data:

- IRB Expedited/Full Board Application and relevant HIPAA forms

Applying to Use the data in the Clinical Data Warehouse

- Am I doing human subjects research?

- Do I need to have data that identifies patients?

- Will I need to return to find additional information after a data set is created?

- Does my data request restrict a count to a very small number?

FAQs

- Where do I start?
- What do you need from me to process my data request?

FAQs

- Where do I start?
- What do you need from me to process my data request?
- What can I do to make the process easier?

FAQs

- Where do I start?
- What do you need from me to process my data request?
- What can I do to make the process easier?
- What does it mean to query a database?

How do you query data?

- Contrary to popular belief, there is no magic button.
- Computers don't spit out info. Software is needed.
- Data is often text information even if it looks like a number.
- Each single piece of data requested may require more than one query, but some can be retrieved in bulk queries.
- How many ways can info be represented?
- Data often needs to be cleaned or transformed.
- Access to the Clinical Data Warehouse is restricted.
- SQL is the database software structured query language.
- Each researcher's set queries is unique.

FAQs

- Where do I start?
- What do you need from me to process my data request?
- What can I do to make the process easier?
- How will I get the data?

FAQs

- Where do I start?

- What do you need from me to process my data request?

- What can I do to make the process easier?

- How will I get the data?
    - Password protected Excel file(s) sent via secure email or via Box.com

FAQs

- Where do I start?
- What do you need from me to process my data request?
- What can I do to make the process easier?
- How will I get the data?
- How should I send data (to merge) to you?

FAQs

- Where do I start?
- What do you need from me to process my data request?
- What can I do to make the process easier?
- How will I get the data?
- How should I send data (to merge) to you?
  - Use secure email when sending any protected health info or private info

FAQs

- Where do I start?
- What do you need from me to process my data request?
- What can I do to make the process easier?
- How will I get the data?
- How should I send data (to merge) to you?
- Can I send a list of patients for whom I need lab data, demographics, comorbidities, address info?

FAQs

- Where do I start?
- What do you need from me to process my data request?
- What can I do to make the process easier?
- How will I get the data?
- How should I send data (to merge) to you?
- Can I send a list of patients for whom I need lab data, demographics, comorbidities, address info?
  - Use secure email when sending the list

FAQs

- Where do I start?
- What do you need from me to process my data request?
- What can I do to make the process easier?
- How will I get the data?
- How should I send data (to merge) to you?
- Can I send a list of patients for whom I need lab data, demographics, comorbidities, address info?
- I'm not an Excel wizard, can you help me through issues?

FAQs

- Where do I start?

- What do you need from me to process my data request?

- What can I do to make the process easier?

- How will I get the data?

- How should I send data (to merge) to you?

- Can I send a list of patients for whom I need lab data, demographics, comorbidities, address info?

- I'm not an Excel wizard, can you help me through issues?
  - Save a copy of the original data before you start reorganizing.
  - ALWAYS save files on secure BMC/BU network drives.
  - KEEP the file password protected.
  - KEEP the key, if there is one, in a separate file

FAQs

- Where do I start?
- What do you need from me to process my data request?
- What can I do to make the process easier?
- How will I get the data?
- How should I send data (to merge) to you?
- Can I send a list of patients for whom I need lab data, demographics, comorbidities, address info?
- I'm not an Excel wizard, can you help me through issues?
- To capture medications, do you want the trade name or generic name?

FAQs

- Where do I start?
- What do you need from me to process my data request?
- What can I do to make the process easier?
- How will I get the data?
- How should I send data (to merge) to you?
- Can I send a list of patients for whom I need lab data, demographics, comorbidities, address info?
- I'm not an Excel wizard, can you help me through issues?
- To capture medications, do you want the trade name or generic name?
- ICD-9 or ICD-10?

FAQs

- Where do I start?
- What do you need from me to process my data request?
- What can I do to make the process easier?
- How will I get the data?
- How should I send data (to merge) to you?
- Can I send a list of patients for whom I need lab data, demographics, comorbidities, address info?
- I'm not an Excel wizard, can you help me through issues?
- To capture medications, do you want the trade name or generic name?
- ICD-9 or ICD-10?
- What is the turn around time for getting data?

FAQs

- Where do I start?
- What do you need from me to process my data request?
- What can I do to make the process easier?
- How will I get the data?
- How should I send data (to merge) to you?
- Can I send a list of patients for whom I need lab data, demographics, comorbidities, address info?
- I'm not an Excel wizard, can you help me through issues?
- To capture medications, do you want the trade name or generic name?
- ICD-9 or ICD-10?
- What is the turn around time for getting data?
- What is the cost, who gets charged?

Costs

- $70/hour
  - Users are charged when the service takes 1 hour or more. Researchers are encouraged to include these costs in grant proposal budgets as either a service, consultation, or as percent full time equivalent (FTE) for the data warehouse manager, as appropriate.

  - Investigators wanting to access data from the data warehouse for research purposes may ask for an exemption to being charged by providing a brief written justification which will be reviewed by the Office of Clinical Research. These will be considered for trainees (students, residents, fellows) and unfunded (faculty) research. Other justifications may be considered.

FAQs

- Where do I start?
- What do you need from me to process my data request?
- What can I do to make the process easier?
- How will I get the data?
- How should I send data (to merge) to you?
- Can I send a list of patients for whom I need lab data, demographics, comorbidities, address info?
- I'm not an Excel wizard, can you help me through issues?
- To capture medications, do you want the trade name or generic name?
- ICD-9 or ICD-10?
- What is the turn around time for getting data?
- What is the cost?
- Are researchers limited to a certain number of variables?

What kind of reports are available?

- Data counts
- Recurring reports
- Online reports
- One time data sets
- Cross-referenced data (multiple sources)
- Data for study recruitment

Examples: Data Counts

- Number of patients with a diagnosis of ICD-250 or ICD-648 AND on hypertension meds
- Number of patients who on mechanical ventilation in 2010-2015
- Number of patients, over 65, with advanced directives.
- Number of ICU visits for those between 18 and 65.
- Number of patients with newly diagnosed Breast Cancer yearly.

Examples – Recurring report

- For a list of Primary Care Physicians provided, find English speaking patients between 50 and 75 years-old (PCP contact info, patient contact info, insurance info, patient demographics) who have a PC appointment in the next week, have not had a colonoscopy, fecal occult blood test, or flexible sigmoidoscopy and have no family history of colon cancer.

- List of kids (5-16) with upcoming appointments in Ophthalmology who were previous diagnosed with visual issues.

- Upcoming appointments for patients enrolled in study who have been lost to follow-up.

Example – online reports

- Mostly for recruitment
- Software written and packaged to collect requested data
- Software is scheduled to run automatically to produce the report
- Report is posted to the Business Objects portal

File | Edit | View | Insert | Format | Tools | Data | Window | Help

Type a question for help

F19

Study Report   Wednesday, March 12, 2008

Study ID	Last Name	First Name	MRN	Logician Information		SCM Information			Emergency Department Info			
				Appt With Dr	Appt Date	Current Location	Date	Visit Type	Ed Date	Ed Dept	Admit Room	Exit C
241	Leonard	John	1492206			H Burc	3/5/08 3:37 PM	Clinic				
241	Leonard	John	1492206			H Dental Clinic	3/5/08 1:40 PM	Clinic				
266	Jones	Susam	3482616			H Burc	3/7/08 1:00 PM	Clinic				
279	Michaels	Sally	8492769			M Gastroenterology BUMG	3/10/08 12:30 PM	Pre Clinic				
300	Thomas	Stephen	6632553			H P/C Clinic Team #4	3/5/08 11:30 AM	Pre Clinic				
309	Jetson	George	2912235	JWORCESTER130	3/19/08 10:19 AM							
309	Jetson	George	2912235			H P/C Clinic Team #3	3/10/08 10:20 AM	Pre Clinic				
319	Jones	James	6442849	SMORRISSEY129	3/13/08 11:19 AM							
319	Jones	James	6442849			H4W04-1	3/5/08 3:03 PM	Inpatient				
319	Jones	James	6442849			H P/C Clinic Team #2	3/13/08 11:20 AM	Pre Clinic				
319	Jones	James	6442849						3/5/2008	NA	H4W04 1	FLR
325	Youle	David	2252130			E Pulmonary	3/11/08 1:30 PM	Pre Clinic				
325	Youle	David	2252130			E Pulmonary	3/11/08 2:00 PM	Pre Clinic				
325	Youle	David	2252130			E Pulmonary Lab	3/11/08 1:00 PM	Pre Clinic				
325	Youle	David	2252130			H P/C Clinic Team #4	3/7/08 10:00 AM	Pre Clinic				
325	Youle	David	2252130			H P/C Clinic Team #4	3/11/08 9:00 AM	Pre Clinic				
326	Smith	Barbara	0842921			H Burc	3/5/08 3:44 PM	Clinic				
341	Whitman	Robert	4512946						3/8/2008	UC		DC
342	Lipman	Jon	6192276			E Physical Therapy	3/6/08 3:30 PM	Clinic				
342	Lipman	Jon	6192276			E Physical Therapy	3/6/08 3:30 PM	Pre Clinic				
342	Lipman	Jon	6192276			E Physical Therapy	3/11/08 3:30 PM	Pre Clinic				
342	Lipman	Jon	6192276			E Physical Therapy	3/13/08 3:30 PM	Pre Clinic				
342	Lipman	Jon	6192276			H ID/GI	3/6/08 2:17 PM	Clinic				
358	Franks	Matthew	1162316			E Adult Psychiatry/Ology	3/12/08 10:30 AM	Pre Clinic				
367	McDonald	Hope	4112694	BVINER268	3/18/08 3:00 PM							
367	McDonald	Hope	4112694			H P/C Clinic Team #4	3/18/08 3:00 PM	Pre Clinic				
369	Shaw	Kelly	5692916			E Sleep Studies	3/9/08 8:00 PM	Pre Clinic				
373	White	Jennifer	2833447	Georgia Montouris	3/19/08 11:45 AM							
373	White	Jennifer	2833447			E EEG/EMG	3/5/08 1:00 PM	Pre Clinic				

AKL / CAM / DAD / ESW / JKW / LAW / SMH / Unassigned /

Online Report

Example – one time data set

- Patients with a Primary Care Visit between each November 1 and April 1st for last 5 years
- Flu shot or No Flu shot
- Find ED, Inpatient, Observation hospital visits in the 10 months subsequent to the date of the shot.
  - Patient demographics: age, gender, race, ethnicity, insurance status
  - Diagnoses for the visits
  - Length/Cost of hospital stay
- Clinic Visits, Specialty clinic visits
- Lab data

One Time Data set – a misnomer

- Complicated reports may take many iterations to get the complete data set.
- The data provided are only as good as the request received.
  - Provide ICD-9/10 codes when diagnoses are involved a request.
  - Identify the source of the data if there is a known preference.
  - Indicate if the request is a one-time data set or if there will be future request for the same data with different dates.
  - Specify if there are multiple ways to denote a data item of interest.
    - e.g., CIN I/CIN II/CIN III could also be listed as Mild/Moderate/Severe Dysplasia
- Often it is not until the researcher reviews a set of data that the request (and subsequent new report) can be refined to meet the study's requirements – understand that the process is iterative.
- It is the researcher's responsibility to understand the data.
  - Ask questions!

Example: De-identified/anonymous data

- Patients with Crohn's disease:
  - Demographics, meds, surgeries, co-morbidities, selected labs

- Patients with prostate cancer
  - Date of dx, demographics, Surgery, PSA pre/post treatment, Cancer stage info, treatment (radiation, chemo), pathology

- Women with a dx of infertility (by ICD-9 code)
  - BMI, smoking status, FSH, Estradiol test, hysterosalpingogram info, myomectomy

- Patients with septic arthritis
  - Age at dx, gender, site of infection, microbiology info, LOS, comorbidities, blood cultures

Example Case III – identifiable data

- For the provided set of medical record numbers and hospital admission dates:
    - Value and date of the most recent WBC
    - C-Reactive Protein Values
    - Viral Load/CD4 Count
    - Path Report findings
    - Op Reports: Pre/Post Diagnosis, Title, Surgeon
    - Echo Conclusion
    - EKG Conclusion

Data Quality

- Consider these examples:
  - Smoking history
  - Pregnancy at a particular time
  - P's and G's
  - Medications active/inactive
  - Problems active/inactive
  - BMI, Height and Weight
  - Patients with seizures
  - CIN 1, 2, 3, I, II, III, mild, severe dysplasia

CDW and the IRB

- All CDW requests *except for some of those asking for simple counts* require submission of a proposal for review to the IRB.
- If the project is human subjects research, then it must be approved by the IRB.

Clinical Research Step-by-Step Summary

- Determine your study hypothesis and the patient population inclusion and exclusion criteria.
- Ask if there are enough patients that meet the study criteria to know if the IRB process is worthwhile (if it is an issue).
- Complete Human Subjects Training – You are required to show documentation of human subjects protection training if you have contact with human subjects or their IDENTIFIABLE data while doing research at BUMC.
- Determine your data needs; meet/speak with Linda Rosen to discuss.
- Plan for data access in your grant application budget.
- Schedule a meeting with Mary-Tara Roth to get clarity on any regulatory issue, clinical research training, or issue related to general services provided by the CRRO – Clinical Research Resources Office.
- Submit your IRB protocol through INSPIR II.
- Send an online data request once you have approval from the IRB.

How do researchers access the CDW?

- Through the Clinical Data Warehouse Research Manager
  - http://www.bumc.bu.edu/ocr/clinical-research-clinical-warehouse-data-access/clinical-data-warehouse-request.form
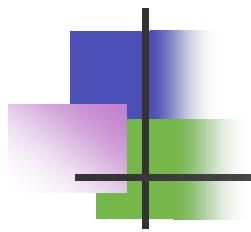 - Contact Linda at LiRosen@bu.edu

IRB forms

- De-identified or partially de-identified or will be used for Prep to Research activities
  - HIPAA Prep-to-Research form
  - HIPAA De-Identified Data form
  - HIPAA Limited Data Set form ***Note: in order to obtain a Limited Data Set the investigator must sign a Data Use Agreement.*
  - HIPAA Decedent Research form
- Informed consent/HIPAA Authorization
  - Clinical Research Form
- Waiver of informed consent/HIPAA Waiver of Authorization
  - HIPAA Waiver of Authorization form

Useful Links

- **CTSI:** Clinical and Translational Science Institute
  - http://ctsi.bu.edu/index.php/resources/tools/
- **RedCap:** Research Electronic Data Capture
  - http://www.redcap.org, http://ctsi.bu.edu/index.php/redcap-users-group/
- **BUMC Clinical Data Warehouse:** *lirosen@bu.edu*
  - http://www.bumc.bu.edu/ocr/clinical-research-clinical-warehouse-data-access/
- **Data Request Form**
  - http://www.bumc.bu.edu/ocr/clinical-research-clinical-warehouse-data-access/clinical-data-warehouse-request-form/
- **ICD-9 codes:** http://icd9cm.chrisendres.com/index.php?action=contents
- **ICD-10 codes:** http://apps.who.int/classifications/icd10/browse/2015/en
- **IRB:** http://www.bumc.bu.edu/irb/
- **INSPIR II:** http://www.bumc.bu.edu/irb/inspir-ii/
- **Profiles** (research networking): http://ctsi.bu.edu/index.php/resources/profiles/

Questions?

CHC Data -

South Boston, South End, Harvard St., Dorchester House, Codman Square, Mattapan, Roslindale, Boston Health Care for the Homeless, Whittier, Uphams Corner, Roxbury

- ## Boston Health Net
  - ### Associate Director, Judy Henderson
    - Judy.Henderson@bmc.org

  - ### To use the BHN health centers for research purposes:
    - Complete a Project Summary Form
    - Present the study at the BHN Research Subcommittee meeting (monthly)

    - Executive Director's signature is required
  - See the Clinical Research Times Article *The Boston HealthNet: "A Strategy to Recruit Boston's Diverse Populations"*