# An Overview of the Possibilities for Research using the Boston Medical Center Clinical Data Warehouse

Linda Rosen, MSEE
Clinical Data Warehouse Research Manager

# What is a Data Warehouse?

A data warehouse is a repository of historical data organized for reporting and analysis. It facilitates data access by having data from many sources in one place, linked together, and easily searchable.

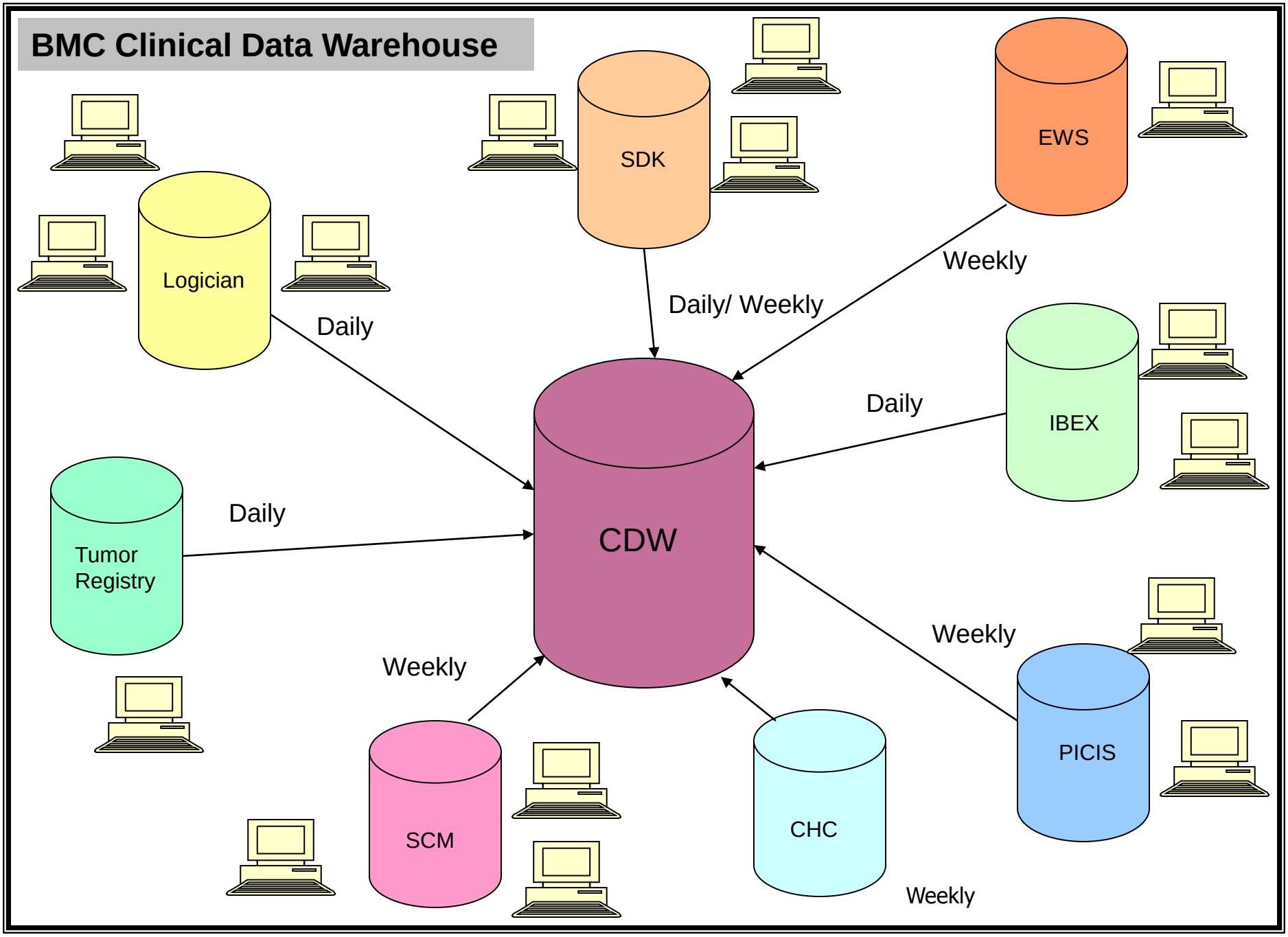
# What is the Clinical Data Warehouse?

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 CDW currently has over 10 years of clinical data.

# The CDW is

- A database containing data from multiple sources
- Historical data extracted from BMC's various clinical software packages
- A database containing data related to each other with unique identifiers
- A database that is only as good as the data entered
- NOT real time data


BMC Clinical Data Warehouse
SDK
EWS
Logician
Daily
Daily/ Weekly
Weekly
IBEX
Daily
CDW
Tumor Registry
Daily
Weekly
PICIS
Weekly
SCM
CHC
Weekly

# Data currently available in the CDW

- SDK –Registration, Visit Dx Info (ICD-9)
- Logician – Outpatient Data
- SCM – Inpatient Data, Lab Data
- IBEX – ED Data
- PICIS – Surgery Data
- Tumor Registry
- EWS – Scheduling
- Anesthesia Manager

# SDK – registration
## starting 10/2003

- Patient demographics
- Patient and provider contact data
- ICD-9 codes fields for each visit
- Op codes for visits
- Admission and discharge date/time
- Service Area/Location
- Visit Type (Inpatient, Outpatient, ED, Observation, Recurring, Group)
- Hospital Charges
- Insurance information
- PCP at time of visit

# Logician
## beginning in 2000

- Patient contact information
- Appointment information (past and future)
- Information entered into templates (observation data)
- Problem List information (active and inactive)
- Referrals and Tests (Orders)
- Medication list (active and inactive)
- Documents from pathology and radiology
- Allergy Info
- Physician Notes
- Directives
- Flags

# SCM – Sunrise Clinical Mgr
## beginning in 2000

- Patient and visit information
- Lab values
- Orders
- Allergies
- Medications
- Locations
- Documents

# IBEX - ED

## beginning 12/2004

- Admission and Discharge dates
- Length of Stay
- ED Location
- Arrival Method
- Complaint
- Diagnosis
- Severity

# PICIS – OR
## beginning 1/2004

- Patient demographics
- Booking information
- Case record data
- Surgical procedure data

# Tumor Registry

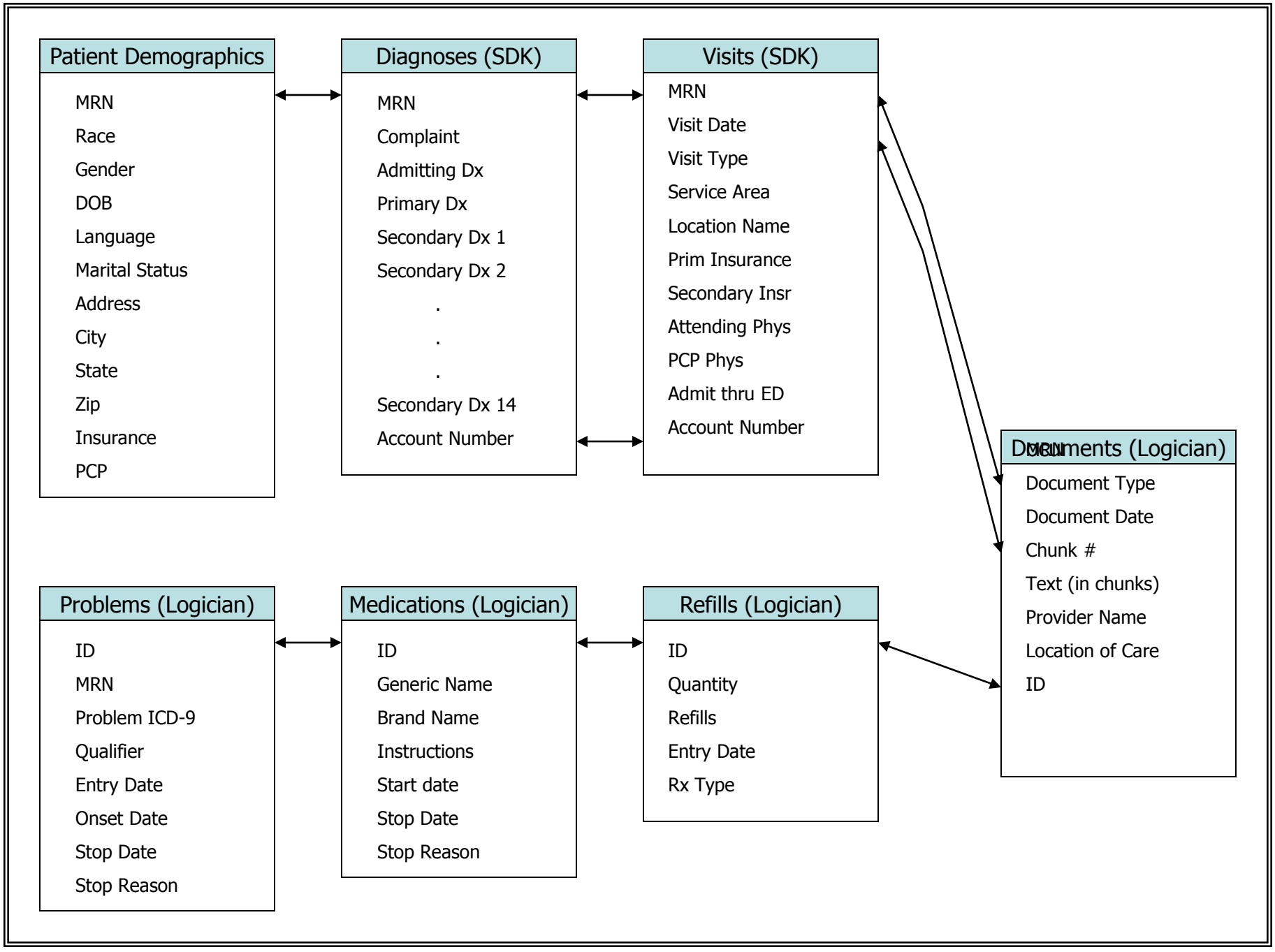
- Patient info
- Tumor site
- Tumor stage
- Date of diagnosis

# EWS - appointments

- Patient Demographics
- Future/Past Appointment Time
- Future/Past Appointment Location
- Status of Appointment
- Provider for Appointment

# CHC Data

- 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"*


Patient Demographics
MRN
Race
Gender
DOB
Language
Marital Status
Address
City
State
Zip
Insurance
PCP
Diagnoses (SDK)
MRN
Complaint
Admitting Dx
Primary Dx
Secondary Dx 1
Secondary Dx 2
.
.
.
Secondary Dx 14
Account Number
Visits (SDK)
MRN
Visit Date
Visit Type
Service Area
Location Name
Prim Insurance
Secondary Insr
Attending Phys
PCP Phys
Admit thru ED
Account Number
Documents (Logician)
MRN
Document Type
Document Date
Chunk #
Text (in chunks)
Provider Name
Location of Care
ID
Problems (Logician)
ID
MRN
Problem ICD-9
Qualifier
Entry Date
Onset Date
Stop Date
Stop Reason
Medications (Logician)
ID
Generic Name
Brand Name
Instructions
Start date
Stop Date
Stop Reason
Refills (Logician)
ID
Quantity
Refills
Entry Date
Rx Type

# What kind of reports are available?

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

# Examples

- Demographic information for a list of MRNs
- Past and future appointments for Pedi patients with referrals to a specialty clinic
- Counts of patients with a diagnosis of ICD-250 or ICD-648
- Data collected from a Logician template for all patients in a study
- Labs, Medications, Problems, Visit Dx for a cohort
- Length of stay information
- Co-morbidities for a cohort
- Identify patients with elevated BP, LDL, TSH
- List of patients with type 2 Diabetes taking Metformin

# *Example Case I – simple counts*

- How many patients were admitted to the ICU in FY 2006 with ARDS (ICD-9 codes 518.5 and 518.82)?
- What was the average length of stay in CY2008 thru CY2011?
- How many of these patients were admitted more that one time in the year following their first admission?

# *Example Case II – de-identified/anonymous data*

- For patients referred by a pediatrician, what is the number of missed appointments at the subspecialty clinic prior to the first kept appointment?
- Number of Appts kept/no show/cancelled
- Birads 4/5 and number of days to follow-up
- Number of ED visits M-F, 8-5 for those with Primary Care provider identified vs. not identified

# *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
  - Echo Conclusion
  - EKG Conclusion

# *Example Case IV – 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 appointments in the next month, have not had a colonoscopy, fecal occult blood test, or flexible sigmoidoscopy and have no family history of colon cancer.

- List of patients with TSH values between 5 and 20; FT4 test with a value < 0.89; exclude FT4i test with a value < 1.0

# Study Report Wednesday, March 12, 2008

| | | | | Logician Information | | SCM Information | | | Emergency Department Info | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Study ID | Last Name | First Name | MRN | Appt With Dr | Appt Date | Current Location | Date | Visit Type | Ed Date | Ed Dep | 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 | | | | |

# One Time Report

- Find all patients with pap smears between 1/1/02 and 6/30/05
- Find subsequent pap smears, colposcopies, LEEP procedures
- Provide patient demographics: age, race
- Provide HIV information – VL, CD4 count before index pap
- Provide Path reports for procedures
- CT- Pulmonary Angiograms with Pulmonary Embolism found

# One Time Report – 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 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

# Data Quality

- Consider these examples:
  - Smoking history
  - Pregnancy at a particular time
  - Medications active/inactive
  - Problem active/inactive

# 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?

# Regulatory Issues

- Same as for accessing other databases, individual paper records or electronic files
- Researchers covered by BUMC
- HIPAA form for counts
- IRB Approval for anything other than counts
- Data Use Agreement

# 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.

# What type of Data do you need?

## Data Counts:

- Preparatory to Research Form only, no IRB needed
  - Submit Prep to Research form to Linda

## Anonymous Data:

- IRB exempt application and relevant HIPAA forms

## Identifiable Data:

- IRB Expedited/Full Board Application and relevant HIPAA forms

# 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

# 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/
  - Contact Linda at LiRosen@bu.edu

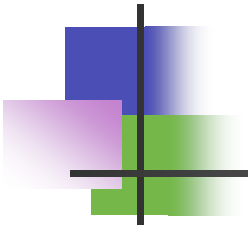
# Accessing Data in the Clinical Data Warehouse

- Assess your data needs
- Plan for data access in your grant application budget
- Plan a meeting or phone call with Linda
- Submit the appropriate IRB forms

# Costs

- $60/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.



Questions?