



Accessing Data for Clinical Researchers: The Boston Medical Center Clinical Data Warehouse



Linda Rosen, MSEE

Clinical Data Warehouse Research Manager

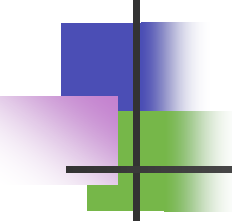
Richard Saitz, MD, MPH

Associate Director, Office of Clinical Research



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 systems into a consolidated, organized and accessible database for analysis, reporting and research purposes.



The CDW is...

- A database containing data from multiple sources
- Data extracted from the databases of BMC's Clinical software packages
- A database containing data related to each other with some unique identifier
- A database that is only as good as the data entered



Data currently available in the CDW

- SDK – Registration and Visit Data
- Logician – Outpatient Data
- SCM – Inpatient Data
- IBEX – ER Data
- PICIS – Surgery Data
- Tumor Registry



PICIS - OR

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



SDK - registration

- Patient demographics
- Patient and provider contact data
- Diagnosis (ICD-9 codes)
- Procedure (CPT codes) information related to a visit
- Admission and discharge date/time
- Visit location
- Insurance information



Logician – Outpatient

- Patient contact information
- Appointment information (past and future)
- Information entered into templates (known as observation data)
- Problem list information
- Referrals and tests (Orders)
- Documents from pathology and radiology
- Allergies and medications
- Physician Notes
- Flags



SCM – Sunrise Clinical Manager

- Patient and visit information
- Lab values
- Orders
- Allergies



IBEX - ED

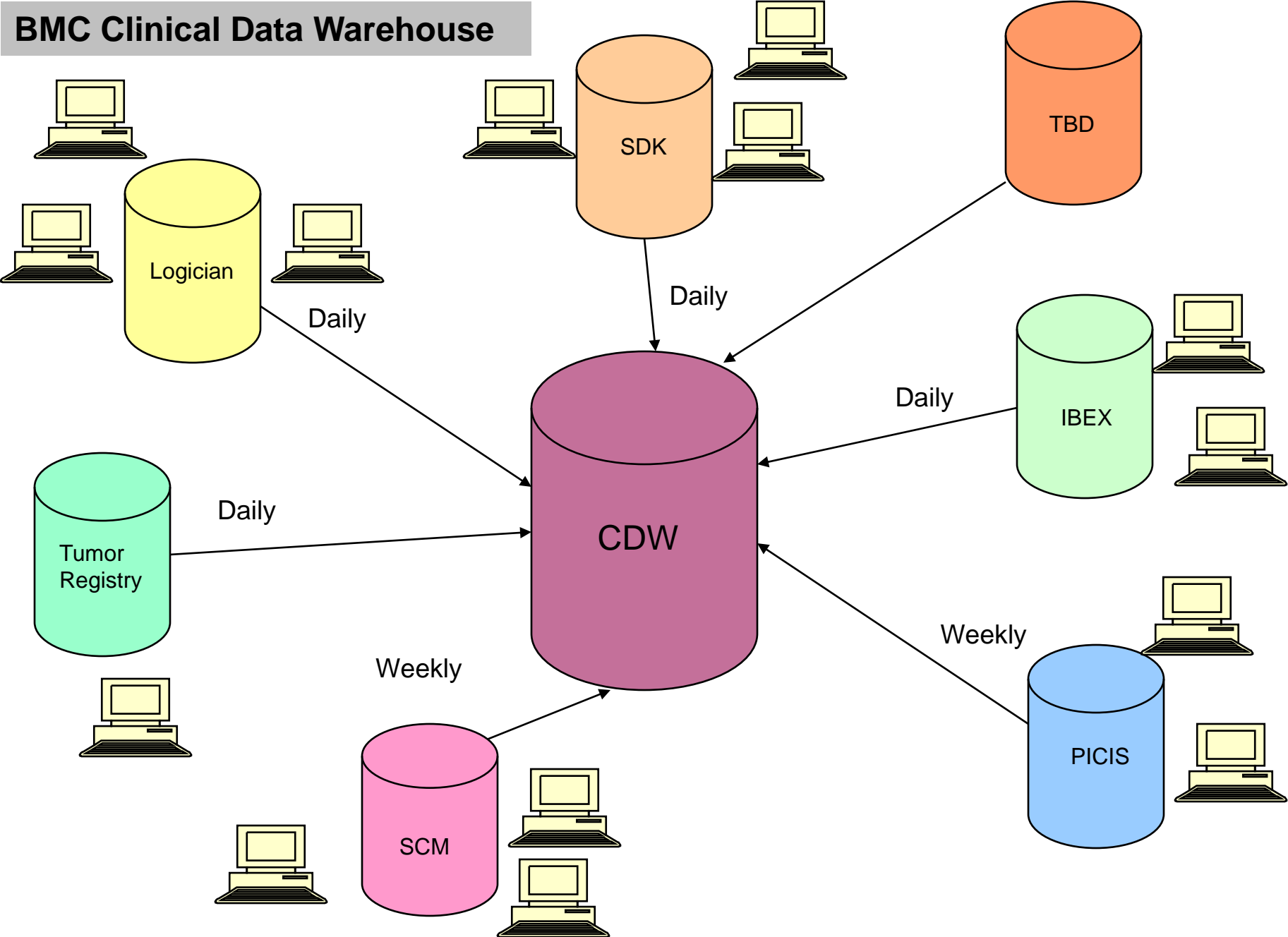
- Admission and Discharge dates
- Length of Stay
- ED Location



Tumor Registry

- Patient info
- Tumor site
- Date of diagnosis

BMC Clinical Data Warehouse





What kind of reports are available?

- Recurring reports
- One time data sets
- Online reports
- Data counts
- Cross-referenced data
- Data from a particular set of Logician templates, updated weekly



Examples

- Demographic information for a list of MRNs
- Past and future appointments for Pediatric patients with referrals to a specialty clinic
- Counts of patients with a diagnosis of ICD-202.10 or ICD-202.20
- Data collected from a Logician template for all patients in a study



More Examples

- Length of stay information for set of patients
- Contact information for adolescents who have asthma
- Count/Identify patients who are G6PD deficient
- 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?
- What proportion of these admissions ended in death?

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?
- What is the number of rescheduled appointments prior to the first kept appointment?



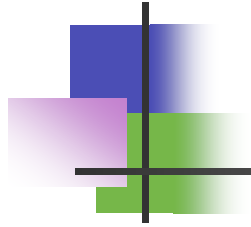
Example Case III – Identifiable Data

- For the provided set of medical record numbers and hospital admission dates:
 - what was the value and date of the most recent WBC
 - C-Reactive Protein
 - Pre-albumin before the admission date?



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.



Case Study

How has the Clinical Data Warehouse helped improve subject identification in our study?

Julie Glick, MPH
GI Clinical Coordinator II
Section of Gastroenterology
Department of Medicine

Study Overview

Principle

Investigator: Dr. Paul C. Schroy, III, MD, MPH

Design: Randomized clinical trial

Setting: BMC primary care units

Participants: Average-risk, English-speaking,
50 to 75 years of age, previously
unscreened, except for Fecal
Occult Blood Test

Changes in Identification Strategies

Method to Identify Subjects	Dates	# Chart Audits	Hours Spent By 3 R.A.s	Subjects Enrolled
RAs Chart Audit	Mar '05- Sept '05	1,500	1,440 hrs	50
RAs Chart Audit (with MS Access db)	Sept '05- March '06	3,250	1,440 hrs	170
Logician Reports (through Clinical Data Warehouse)	Oct '06 - April '07	3,700	120 hrs	170

Changes in Recruitment Strategies

- It is important to note that the number of subjects enrolled also increased due to changes in recruitment strategies.
- PCP “clicks” in Logician + Letters from PCPs
→ Direct calls by RA to patients

Cost Analysis

Method to Identify Subjects	Cost for 6 month period
RA's Chart Audit and Rely on PCP for Recruitment	\$ 1,300 (Clicks) + <u>\$25,700</u> (Letters) \$27,000
RA's Chart Audit (with MS Access) with Direct Calls	\$23,300
Logician Reports (from CDW) with Direct Calls	\$11,800 – <u>\$ 3,000</u> (Initial startup cost) \$ 8,800 (going forward)

Tweaking the Reports

- Since July '06, we have worked with Linda and the CDW to identify patients that are appropriate and eligible for our study.
- The majority of our work was done in the first few months, but it has been an ongoing process.
- Medical records contain detailed information and providers document patient information in different places in the Logician charts. Therefore, we continue to provide feedback to Linda to enable her to improve our reports.

Conclusion

- During the months that Linda has been working with us, we have saved time and reduced our costs significantly.
- We now have more time to spend on recruiting and meeting with subjects, as well as working on other studies.
- We are currently writing a paper on the importance of utilizing medical record technology to improve patient identification and recruitment.

Recurring Report Example

Microsoft Excel - Gastro_MAY_2007.xls

File Edit View Insert Format Tools Data Window Help Type a question for help

J4 fx

	A	B	C	D	E	F	G	H	I
1	DR_NAME	APPTDATE	LASTNAME	FIRSTNAME	ELIGIBLE	MRN	DATEOFBIRTH	PHONE	WORKPHONE
2	BATTAGLIA	05/01/07 01:00 PM	SMITH	JANET	YES	8058782	9/27/1932	(555)447-8969	
3	BATTAGLIA	05/01/07 03:00 PM	METZGER	BARBARA	YES	9976900	4/19/1937	(555)343-8496	
4	DAVID	05/03/07 09:19 AM	JOY	KAREN	YES	9699808	7/24/1936	(555)364-8780	(555)863-8910
5	DAVID	05/10/07 08:40 AM	TAYLOR	MELISSA	YES	8008586	4/8/1948	(555)834-7989	
	DAVID	05/16/07 01:00 PM	JONES	TIMOTHY	YES	8880869	1/9/1934	(555)838-3744	

Ready NUM

N41

fx

AHEAD Study Report															
					Logician Information		SCM Information			Emergency Department Information					
Ra Assigned	ID	Lname	Fname	MRN	Appt With Dr	Appt Date	Current Location	Date	Visit Type	Ed Date	Ed Dept	Admit Room	Exit Code	Exit Date	Admit D
CAM	28	Doe	John	3333333	SJALISIT77	4/6/07 9:00 AM									
CAM	28	Doe	John	3333333			M Otolaryngology	4/6/07 9:00 AM	Pre Clinic						
CAM	100	Jones	Mark	4444444			E Adult Psychiatry/Olog	4/5/07 6:00 PM	Pre Clinic						
CAM	171	Smith	John	5555555	ARSEN41	4/10/07 2:30 PM									
CAM	171	Smith	John	5555555			H P/C Clinic Team #1	4/10/07 2:30 PM	Pre Clinic						
CAM	200	Wells	Sarah	6666666						4/1/2007	ACUT		A/WOL	4/1/2007	
CAM	208	Walsh	Jane	7777777	HPHAM313	4/11/07 1:30 PM									
ESW	10	Johnson	Joe	8888888	KVIOLA268	4/11/07 3:00 PM									
ESW	37	Wilson	Mary	9999999			E Foot Care Specialty	4/10/07 1:00 PM	Pre Clinic						
ESW	54	Black	Sam	2222222						3/28/2007	UC		DC	3/28/2007	
ESW	60	Long	Hector	2121212			E Adult Psychiatry/Olog	4/4/07 11:00 AM	Pre Clinic						
ESW	76	Williams	David	3131313						4/1/2007	PEDI		DC	4/1/2007	
ESW	79	Ramirez	Joseph	4141414						4/2/2007	ACUT		DC	4/2/2007	
ESW	164	French	Jason	5151515	LSTEAD123	4/3/07 2:30 PM									
ESW	164	French	Jason	5151515			H P/C Clinic Team #2	4/3/07 2:30 PM	Pre Clinic						
ESW	164	French	Jason	5151515			H P/C Clinic Team #4	4/4/07 10:00 AM	Pre Clinic						
ESW	174	Morris	Clement	6161616						3/31/2007	NA		DC	4/1/2007	
ESW	174	Morris	Clement	6161616						4/2/2007	UC		DC	4/2/2007	
LAW	192	O'Connor	Barbara	7171717	CHIMAR602	4/6/07 7:30 AM									
LAW	192	O'Connor	Barbara	7171717			H P/C Clinic Team #4	4/4/07 10:30 AM	Pre Clinic						
LAW	211	Shoemaker	Morris	8181818	AKURBAN61	4/5/07 10:45 AM									
LAW	211	Shoemaker	Morris	8181818	DGUNTHER4000	4/6/07 9:15 AM									
LAW	211	Shoemaker	Morris	8181818			H Dermatology Clinic	4/5/07 10:45 AM	Pre Clinic						
LAW	211	Shoemaker	Morris	8181818			H Family Medicine	4/6/07 9:15 AM	Pre Clinic						
RHS	165	Yee	Pamela	3131331			H P/C Clinic Team #2	4/4/07 2:30 PM	Pre Clinic						
RHS	169	Roberts	Jim	3232323	KVIOLA268	4/5/07 9:00 AM									
RHS	169	Roberts	Jim	3232323			H P/C Clinic Team #4	4/5/07 9:00 AM	Pre Clinic						
RHS	182	Troy	Helen	4242424	DVEGA254	4/5/07 3:19 PM									
RHS	182	Troy	Helen	4242424			H OB/GYN Team B	4/5/07 3:20 PM	Pre Clinic						
RHS	185	Green	Joan	5252525	RPRABAKARAN12	4/10/07 3:10 PM									
RHS	185	Green	Joan	5252525			H P/C Clinic Team #2	4/10/07 3:10 PM	Pre Clinic						
RHS	202	Green	Gretta	6262626	ICHAKRABATY123	4/11/07 4:30 PM									



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
- HIPPA form for counts
- IRB Approval for anything other than counts



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.



Data Counts:

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



Anonymous Data:

- IRB exempt application and relevant HIPAA forms



Identifiable Data:

- IRB Expedited/Full Board Application and relevant HIPAA forms



Preparatory to Research



Principal Investigator:

Email address:

Phone number:

Research Staff needing access to protected health information:

As approved by IRB in INSPIR Section A

[Investigator to change if different from INSPIR Section A]

Study Title or Study Idea:

Number of records needed: ☐ ≥ 50 ☐ < 50

The Privacy Rule (45 CFR 164.512) allows the use or disclosure of protected health information required in order to prepare a research application or proposal, provided that certain criteria are met. Please read the following statements. If you agree, please sign below. Also complete item #4.

- 1.The use or disclosure requested will be limited to the preparation of a research protocol or for similar purposes preparatory to research.
- 2.No protected health information will be removed from the covered entity by the researcher in the course of the review.
- 3.The requested information constitutes the minimum necessary data to accomplish the goals of the research.
- 4.Please attach a list of the selection criteria for records required (e.g.; all asthmatics seen in the Asthma Clinic), the dates of the records required (e.g.; clinic visits from July 1, 1998 through December 31 2000), and data fields required for the research.

By submitting this form with an INSPIR application, the PI attests to the following:

I declare that the requested information constitutes the minimum necessary data to accomplish the goals of the research.

I agree that the protected health information will not be re-used or disclosed to any other person or entity, except as required by law, for the authorized oversight of the research study, or for other research for which the use or disclosure of protected health information would be permitted by the Privacy Regulation (45 CFR 164.512)

DATA AND/OR RECORDS NEEDED FOR RESEARCH PROTOCOL

1. Selection Criteria (e.g.; asthmatics seen is Asthma Clinic)
2. Dates of required records: from ____/____/____ through ____/____/____
3. Data fields required (list fields required from an electronic data base)
4. Anticipated sources of information (check all that apply)
 - ☐ Paper medical records
 - ☐ Electronic files
 - ☐ Other _____



How do Researchers Access the CDW?

- Through the Clinical Data Warehouse Research Manager
 - <http://www.bumc.bu.edu/ocr/dataaccess>
 - Contact Linda Rosen at LiRosen@bu.edu



A simple query:

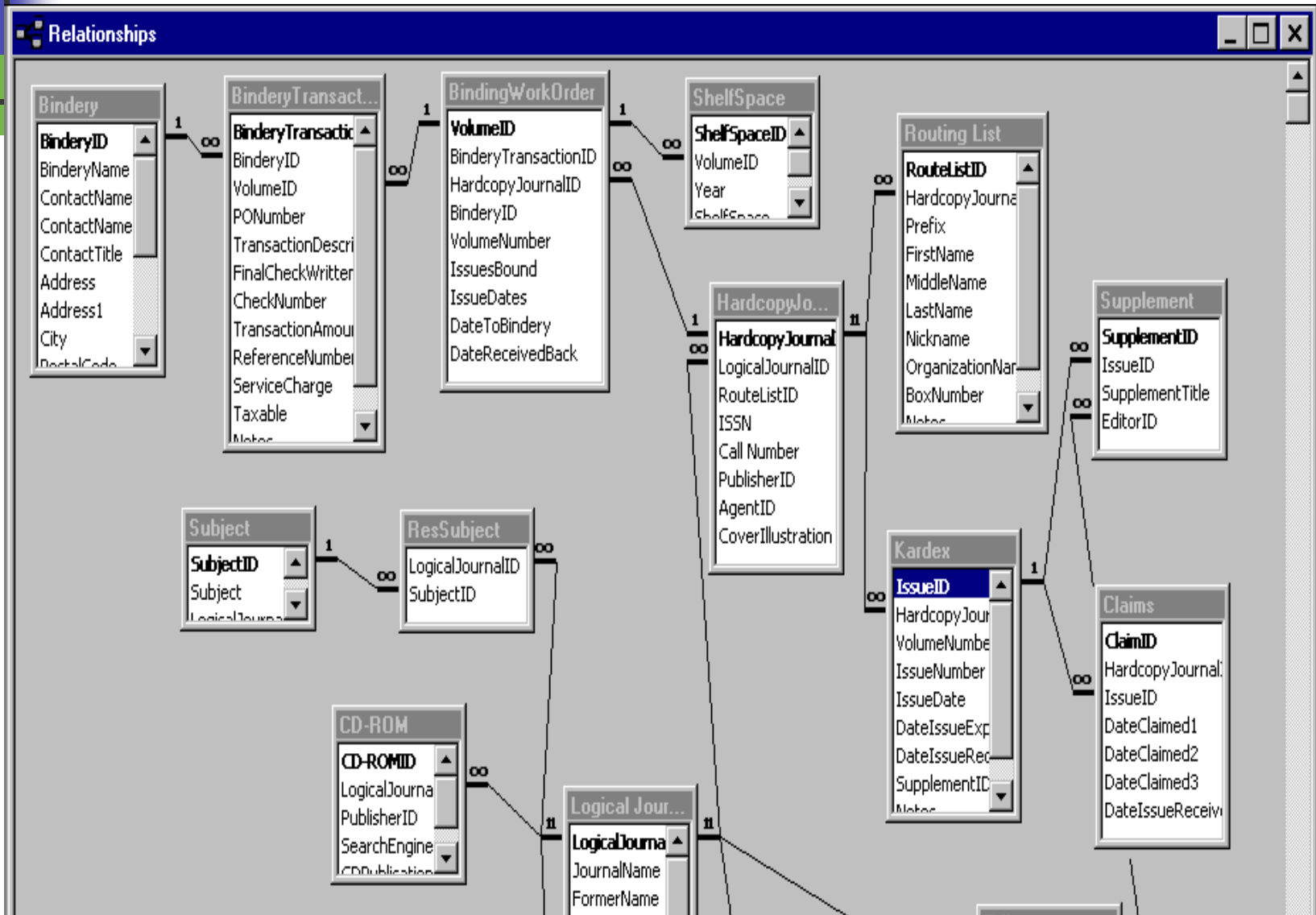
```
SELECT count(distinct(mrn)) FROM clinical_intel.sdk_visit sdk_visit
WHERE to_date(sdk_visit.visit_date, 'mm/dd/rr')
      BETWEEN TO_DATE('01/01/05', 'mm/dd/rr')
              AND TO_DATE('12/31/05', 'mm/dd/rr')
AND (sdk_visit.visit_type = 'E'
     OR sdk_visit.admit_source = 'EMERGENCY ROOM');
```



Database Relationship Diagram

Example

Database Relationship Diagram



Navigating Between Tables Using Keys

AUTHOR

au_id (PK)	au_lname	au_fname	address	city	state
172-32-1176	White	Johnson	10932 Bigge Rd.	Menlo Park	CA
213-46-8915	Green	Marjorie	309 63rd St. #411	Oakland	CA
238-95-7766	Carson	Cheryl	589 Darwin Ln.	Berkeley	CA
267-41-2394	O'Leary	Michael	22 Cleveland Av. #14	San Jose	CA
274-80-9391	Straight	Dean	5420 College Av.	Oakland	CA
341-22-1782	Smith	Meander	10 Mississippi Dr.	Lawrence	KS
409-56-7008	Bennet	Abraham	6223 Bateman St.	Berkeley	CA
427-17-2319	Dull	Ann	3410 Blonde St.	Palo Alto	CA
472-27-2349	Gringlesky	Burt	PO Box 792	Covelo	CA
486-29-1786	Locksley	Charlene	18 Broadway Av.	San Francisco	CA

AUTHOR_TITL F

au_id (PK)	title_id (FK)
172-32-1176	PS3333
213-46-8915	BU1032
213-46-8915	BU2075
238-95-7766	PC1035
267-41-2394	BU1111
267-41-2394	TC7777
274-80-9391	BU7832
409-56-7008	BU1032
427-17-2319	PC8888
472-27-2349	TC7777

PK = Primary Key Column
FK = Foreign Key Column

TITLE

title_id (PK)	title	type	price	pub_id (FK)
BU1032	The Busy Executive's Database Guide	business	19.99	1389
BU1111	Cooking with Computers	business	1.95	1389
BU2075	You Can Combat Computer Stress!	business	2.99	736
BU7832	Straight Talk About Computers	business	19.99	1389
MC2222	Silicon Valley Gastronomic Treats	mod_cook	19.99	877
MC3021	The Gourmet Microwave	mod_cook	2.99	877
MC3026	The Psychology of Computer Cooking	UNDECIDED		877
PC1035	But Is It User-Friendly?	popular_comp	22.95	1389
PC8888	Secrets of Silicon Valley	popular_comp	20	1389
PC9999	Net Etiquette	popular_comp		1389
PS2091	Is Anger the Enemy?	psychology	10.95	736

PUBLISHER

pub_id (PK)	pub_name	city
736	New Moon Books	Boston
877	Binnet & Hardley	Washington
1389	Algodata Infosystems	Berkeley
1622	Five Lakes Publishing	Chicago
1756	Ramona Publishers	Dallas
9901	GGG&C	München
9952	Scootney Books	New York
9999	Lucerne Publishing	Paris



Accessing Data in the Clinical Data Warehouse

- Assess your data needs
- Submit the appropriate IRB form(s)
- Plan for data access in your grant application budget
- Plan a meeting with Linda Rosen



Questions?