Using the VDW

Terry Field, Gene Hart,
Debra Ritzwoller

Why develop an HMORN data warehouse?

- Each health plan has data about its enrollees and their utilization and diagnoses in various electronic databases.
- These data can make many aspects of epidemiologic and clinical research more efficient.
- However... pulling together the data elements needed from each site on a project by project basis was time-consuming and expensive.

What is the *Virtual* data warehouse?

- A federated database where each site stores their data locally in identical data structures
- Each site creates a series of datasets across selected content areas based on common definitions and concepts.
- Purpose: enhance efficiencies in performing multi-site research
- Each site retains ownership and control of their respective data.

Why use the VDW?

- Data are easy to pull, good file structure.
- Huge number of source databases, combined, cleaned and simplified.

- Clean, tested data
- Ease of use – excellent tool for feasibility inquiries and for acquiring comparable data across other sites.

- Can leverage programmer time – “One stop shopping!”
- Site programmers are experienced and committed
- Macros and algorithms already programmed

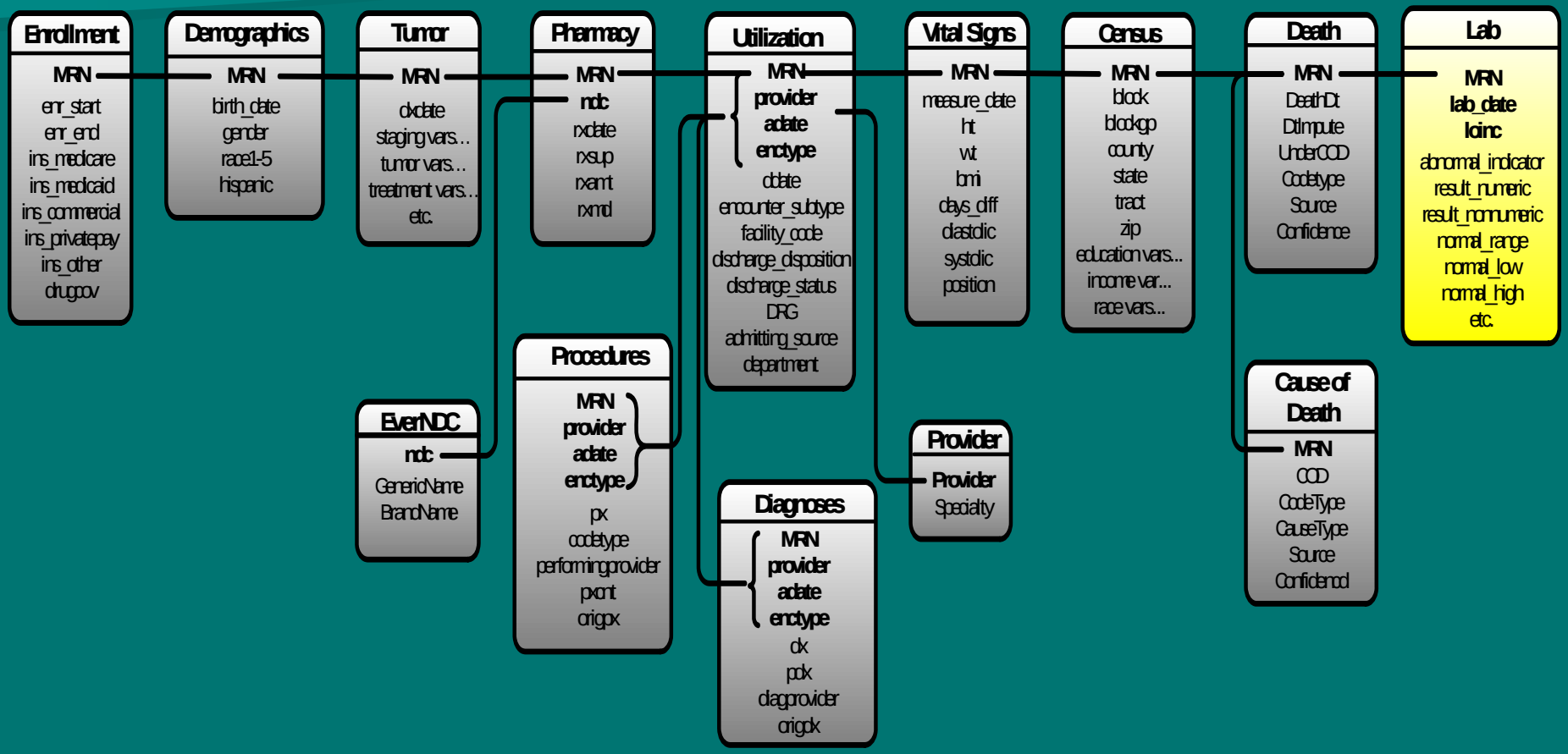
The HMO CRN Cancer Research Network

Sites Participating in the VDW

VDW Content Areas

- Enrollment/Demographics
- Census
- Pharmacy
- Utilization
  - procedure and diagnosis codes
  - inpatient and outpatient events
- Vital Signs
- Laboratory
- Tumor Registry
- More to come . . .

VDW Data Structure



Virtual Data Warehouse

Each site runs local programs to pull data from legacy data systems

Real databases set up identically at all sites

Programs written using VDW data dictionary

Output = datasets from the sites

Research team

Using the VDW: Acquiring Data

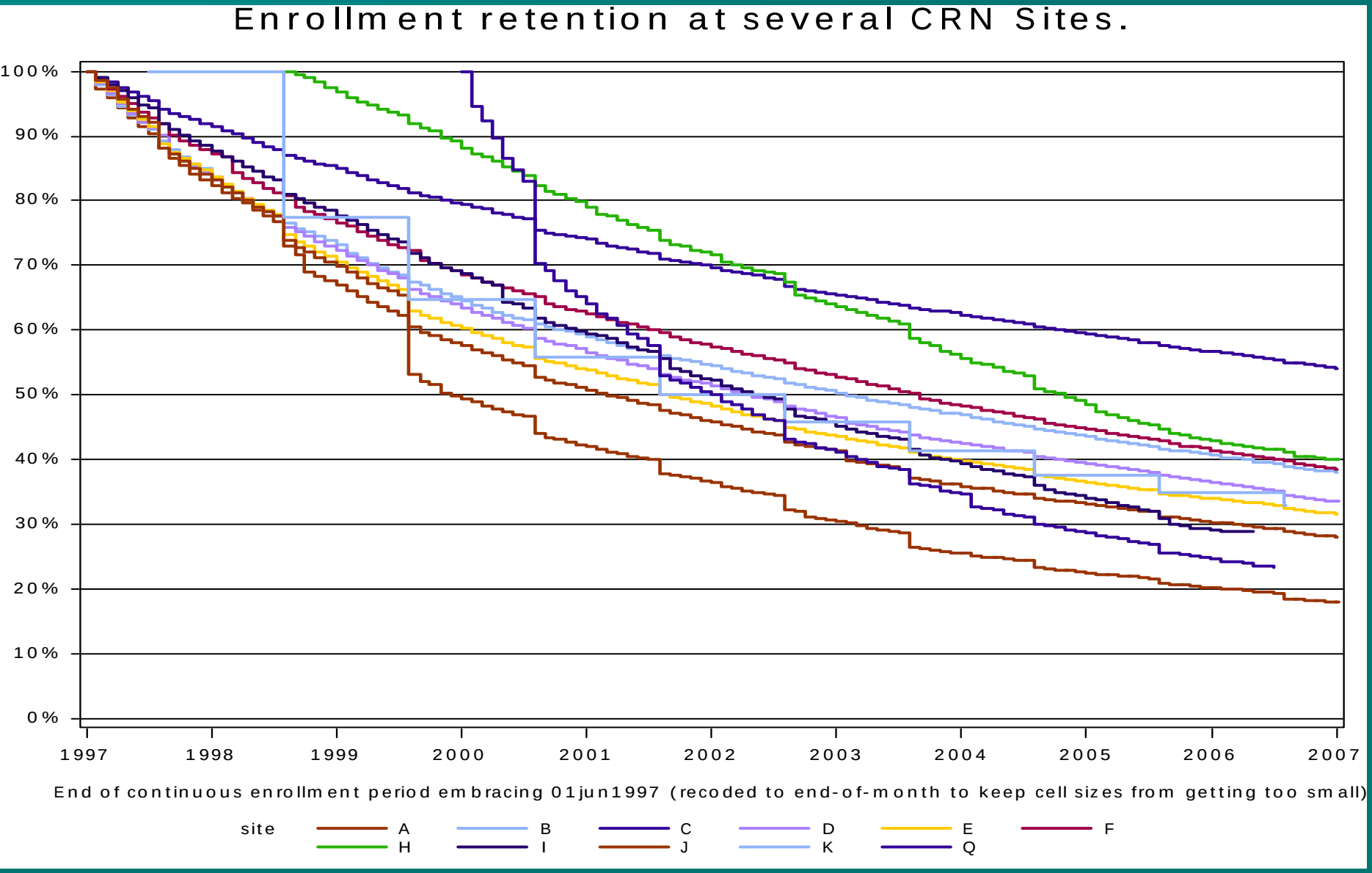
- Project team develops specifications
- Write SAS program to pull data from the VDW at lead site
- Test VDW program at another site
- Distribute program to each participating site
- Run program at each site and return results to lead site
- Analyze and publish

Enrollment

Enrollment
MRN
enr_start
enr_end
ins_medicare
ins_medicaid
ins_commercial
ins_privatepay
ins_other
drugcov

- Macros for:
- 2 Different types of continuous enrollment definitions (disregarding gaps of specified # of months)
- Finding date of first disenrollment after an index date

VDW: Enrollment/Demographics



Demographics

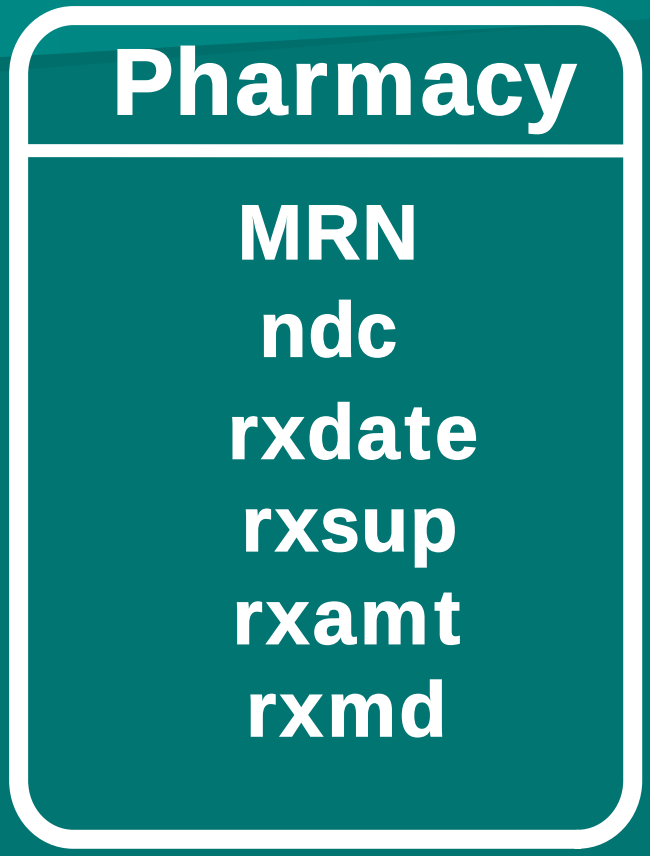
- Follows SEER coding for race
- Standardized macro for calculating age

Tumor

Tumor
MRN
dxdate
staging vars...
tumor vars...
treatment vars
etc.

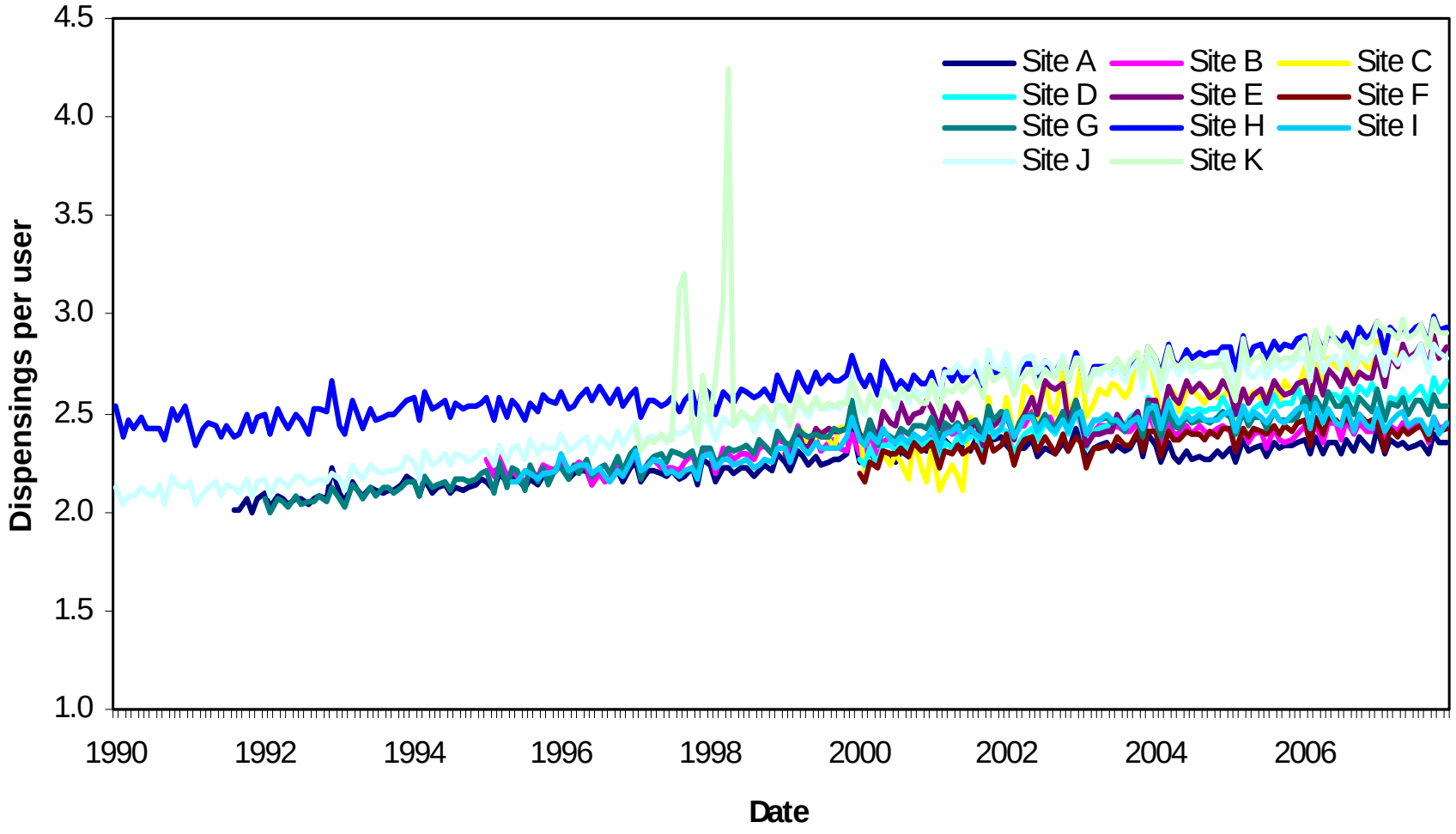
- Generally a subset of NAACCR standard variables.
- Standard Macro for drawing samples of women w/an Invasive Breast Cancer between specified dates.

Outpatient Pharmacy

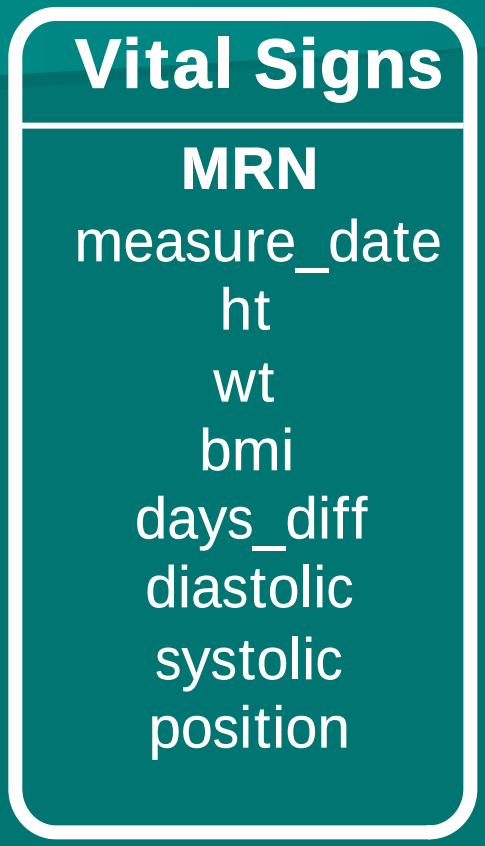


- Macros for:
  - Pulling all fills for a given sample of people.
  - Pulling all fills for a given list of National Drug Codes
  - Producing counts of fills for a given list of NDCs

Monthly dispensings per user, by site, 1990-2007

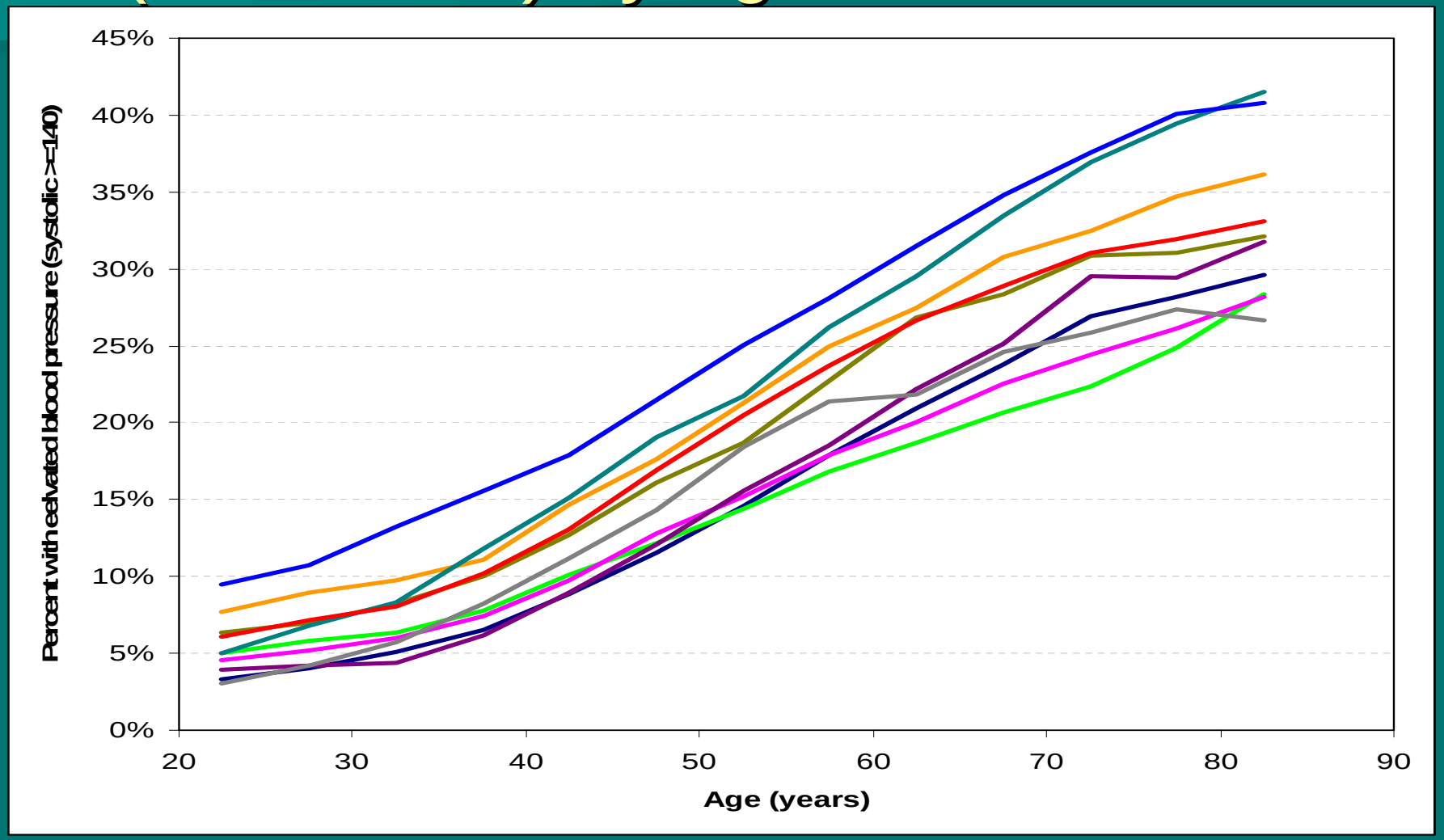


Vital Signs



- No need to impute weights for the BMI calculations—it's already done!
- Standard macro for pulling all vital sign measures for a given sample of people.

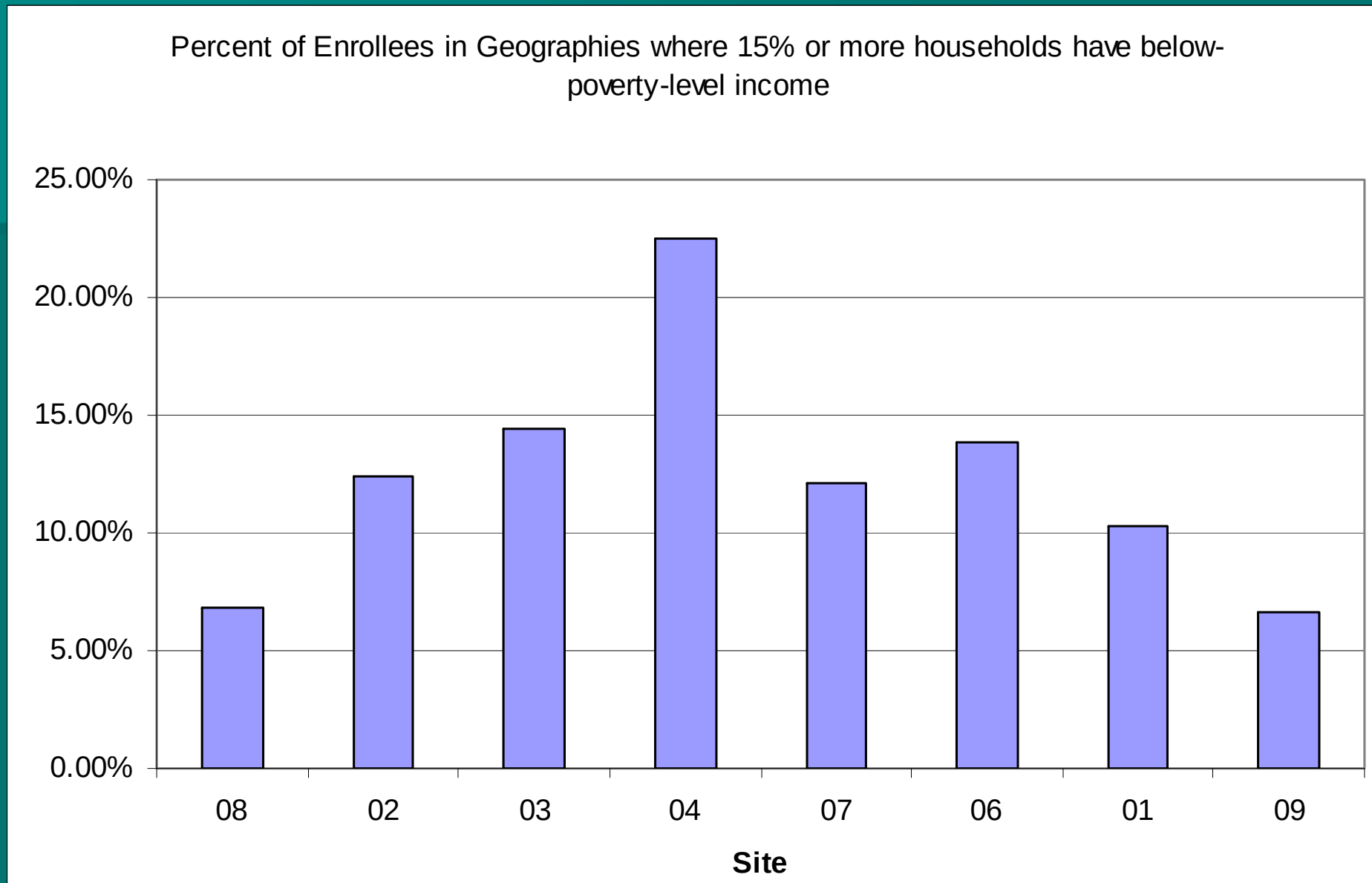
Percent Of VDW Adults With Elevated Systolic Blood Pressure (SBP>=140) By Age & Site In 2007



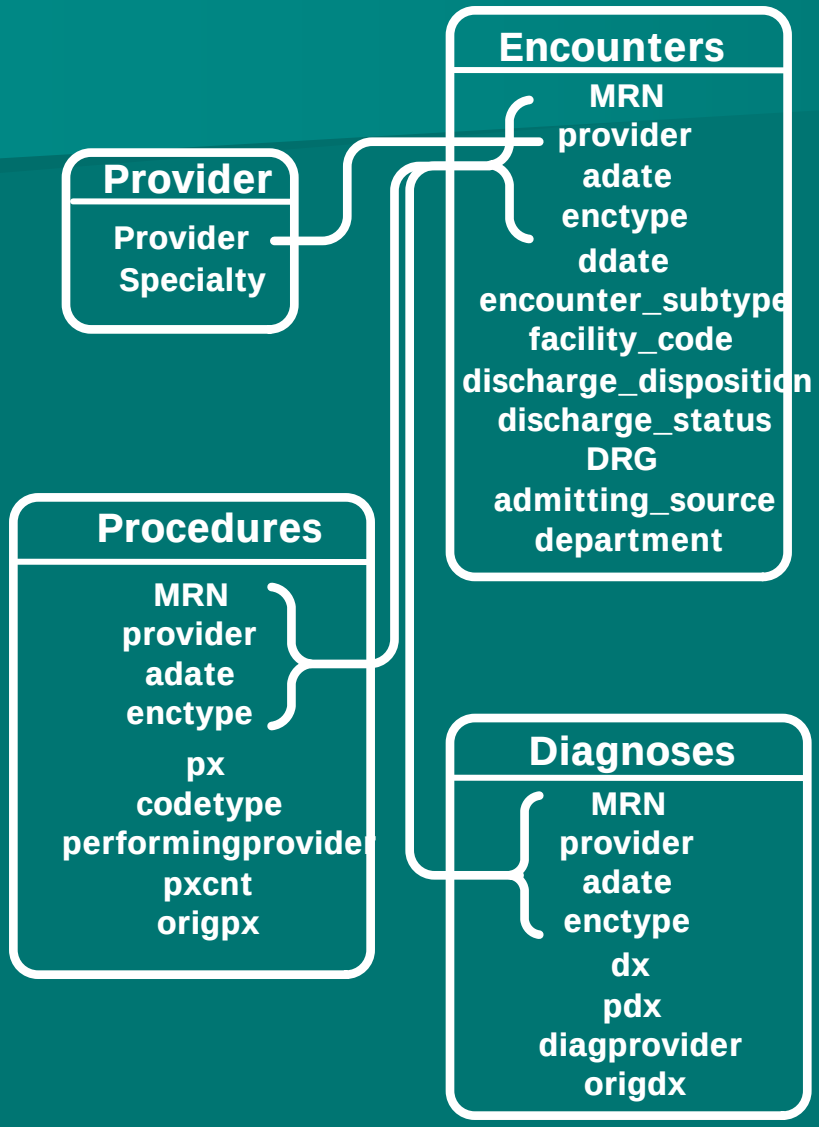
Census

Census
MRN
block
blockgp
county
state
tract
zip
education vars…
income vars…
race vars…

- Census Bureau reported demographics tied to individuals via geocoded addresses.
- Standard macro for pulling census for a given sample of people.


Percent of Enrollees in Geographies where 15% or more households have below-poverty-level income
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
08
02
03
04
07
06
01
09
Site

Encounters



- In- and Out-patient encounters
- Tons of macros for pulling people, procedures & diagnoses.

Figure 1. Rates of Patients with Diabetes Diagnoses per 1000 Members By Site and Year

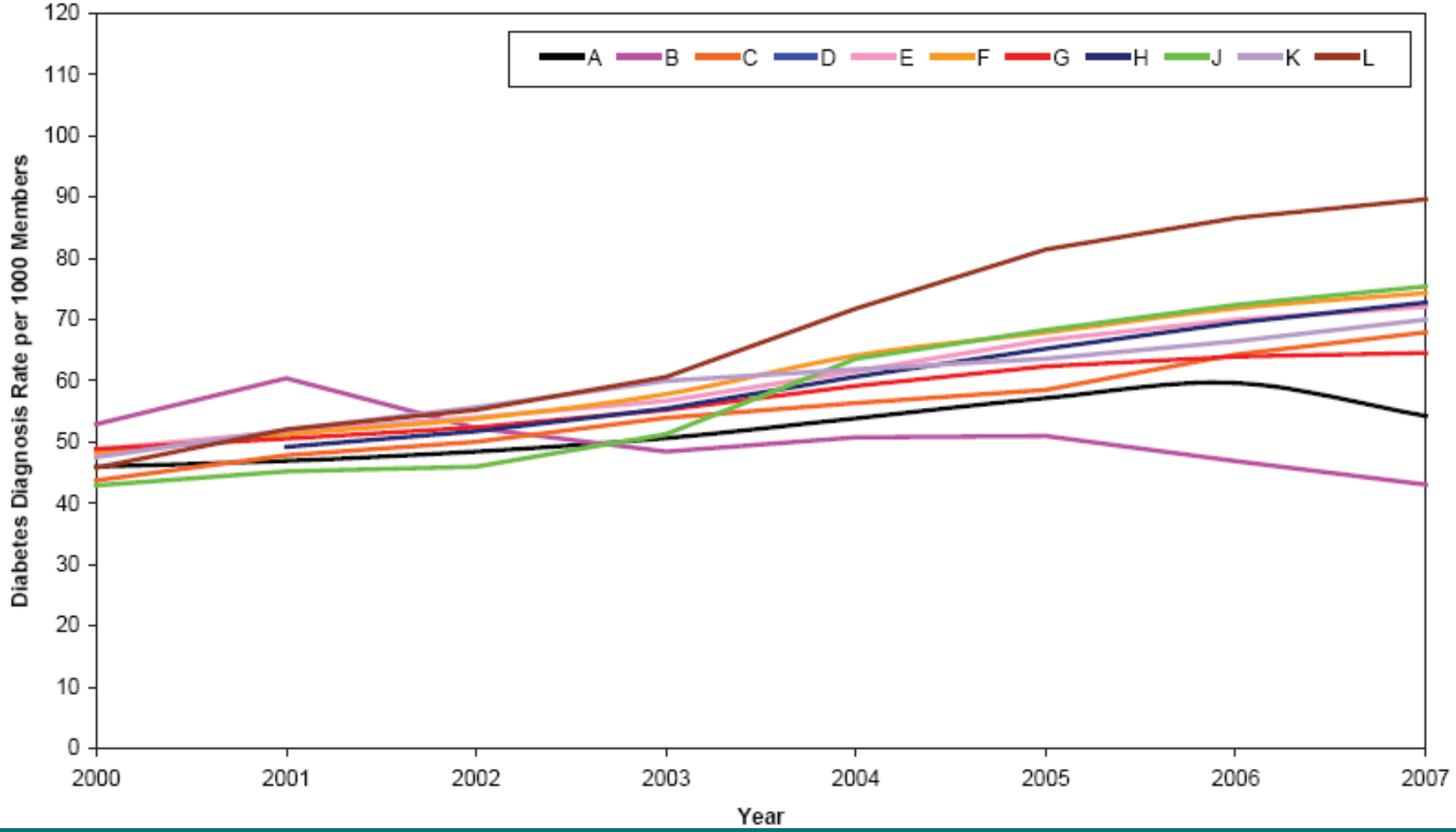


Figure 2. Rates of Female Patients Aged 75+ with Hip Fracture Diagnoses per 1000 Members By Site and Year

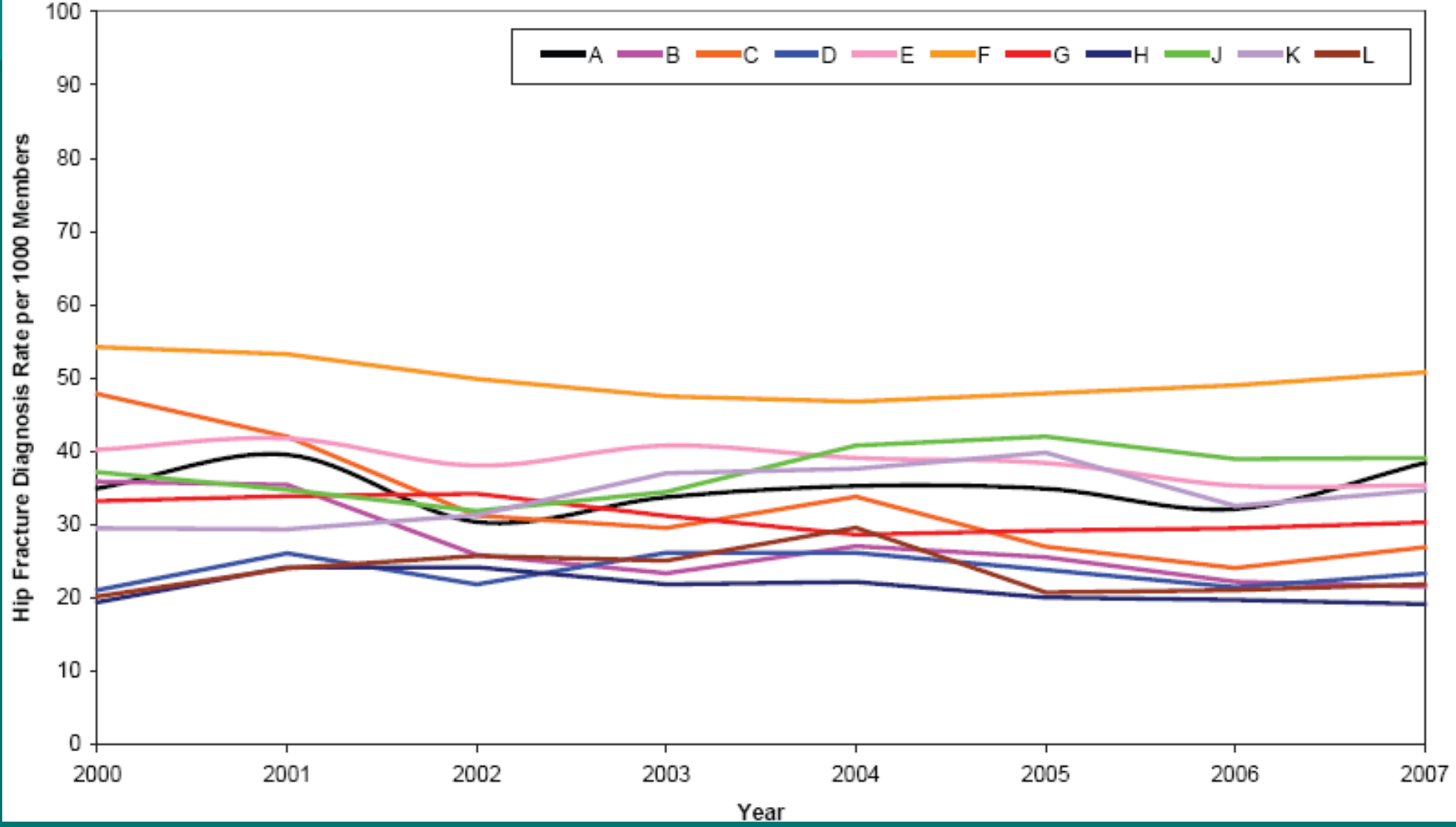
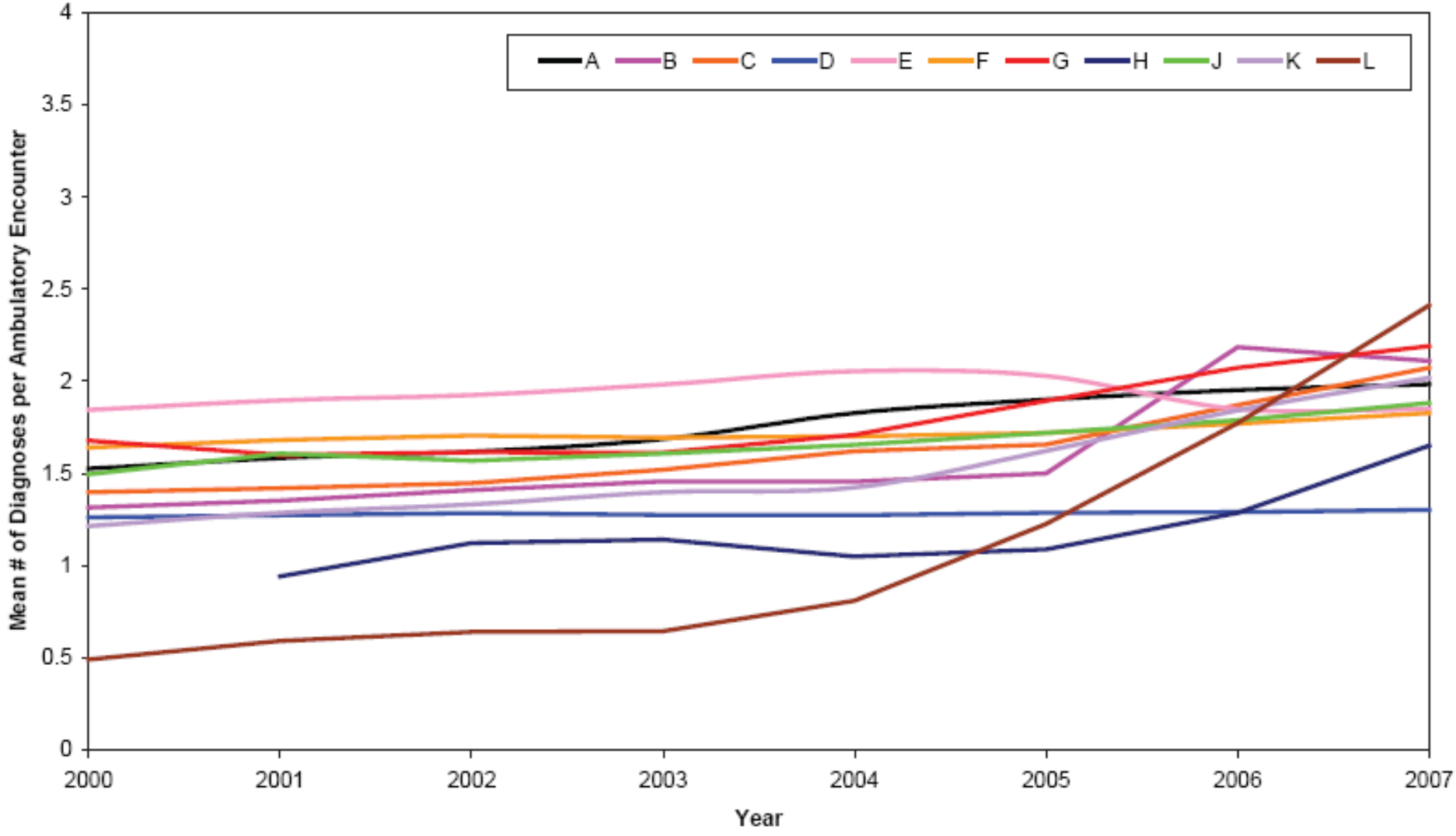


Figure 3. Mean Number of Unique Diagnoses per Ambulatory Encounter By Site and Year



Future Content Areas

- Infusion data – initially starting with Chemotherapy.
- Pathology data
- Benefit data
- Cost data
- Other areas related to CVRN & CERT initiatives

Using the VDW: Approval Steps

- Find collaborating investigator at each interested site
- At each HMO the local site investigator gets or confirms IRB approval
- Explore necessity of a DUA with each site providing data
  - Consider who will be 1st authors of papers.
  - Where will analysis be done?
  - Will the final analytic dataset be shared at all sites?

Finding Collaborators

- Talk with local investigators already involved in multi-site projects
- Attend the annual HMORN and meet investigators with similar interests
- Review the HMORN PI Directory
- Use the search tool on the HMORN website: http://www.hmoresearchnetwork.org

Making Collaboration Work

- Be proactive
- Avoid second guessing
- Build trust
- Capitalize on efficiency
- Clarify roles
- Discuss papers
- Don’t strain resources
- Empower success
- Maximize face time
- Share opportunities

Hints from the Collaboration Toolkit – HMORN website

Budgeting for VDW Use

- FTEs for each site should include investigator, project manager, programmer
- Use of the VDW should lead to much lower programmer FTEs at the sites
- Include face to face meetings in the budget

Does the VDW solve all our research problems?

- Data quality checks a work in progress
  - Usual scrutiny and suspicions apply
  - Incorporate data quality checks
  - Validation of diagnoses varies
- Site-to-site variations in implementation
  - Dates of coverage
  - Some components not available at all sites
  - Access and turn-around times differ
- Costs are still an issue

Conclusions

- Standardization of content areas across health plans enables sharing compatible data in multi-site studies - improves programming efficiency, accuracy and completeness of data.
- Each content area requires upfront effort from the lead site for each variable that is mapped.
- Participating sites must invest time determining local codes and issues with legacy systems.
- The aggregate information from all sites participating in a study will enhance the usefulness of the VDW to epidemiological, observational, and interventional research.
- More Q/A and Q/C effort needs to be done!
- Continued internal support is needed in order to maintain and expand the VDW

Where to go for Information

- Site Data Manager at your site
- Online resources:
  - Programmer's Wiki
  - Cancer counter
  - Other counters
    - Diagnosis
    - Procedure
    - Pharmacy (coming soon?)
  - CCSN handout?