ResDAC
RESEARCH DATA ASSISTANCE CENTER

Strengths, Weaknesses, and Applications of Medicare Data

Beth Virnig, Ph.D.

Associate Dean for Research and Professor

University of Minnesota

Origin of Health Services Utilization, or "Claims" Data

- Derived from reimbursement or the payment of bills
- Information needed to pay the bill will be of higher quality than other information
  - The enrollment data are the basis for determining whose bills are qualified to be paid

Medicare Data

- Medicare Administrative data are in electronic format and contain information about
  - services that are to be paid
    - » The numerator
  - information about the persons whose services are to be paid
    - » The denominator
- And can be linked to information about
  - the providers of care (who gets paid),
  - area characteristics (where the patient and care provider is located)
  - other useful information

Why use administrative data?

Clinical Validity

- Data contain information about covered services used by enrollees in program.
  - Admission and discharge dates
  - Diagnoses
  - Procedures
  - Source of care

Validity of Demographic Information

- Demographic information is largely reliable and valid
  - age/DOB, gender, race, place of residence, date (not cause) of death
  - It is possible to link the numerator to the denominator

Population Coverage

- It is estimated that over 98% of adults age 65 and over are enrolled in Medicare
- over 99% of the deaths in the US for persons age 65 and older are accounted for in the Medicare program
- But, coverage may be less due to managed care enrollment.

Large Population Base

- In Medicare we can study approximately 47 Million beneficiaries. This allows for detailed sub-group analysis with reduced concerns about loss of statistical power.

Cost Effective to Use

- Price per chart
- Access across multiple providers
- Consistency in reporting format

Can easily be combined with data from other sources*

- Census
- Cancer registries (e.g., SEER/Medicare)
- Other providers (e.g., VA, Medicaid)
- National death index/State vital statistics
- Minimum Data Set (MDS)
- Surveys (e.g,. Health and Retirement Study)
- Provider Information

* Depending on the availability of identifying/common variables

This linking can take place:

- At the group level based on geography, place of service, etc.
- At the person level with external data sources
  - SSN (plus date of birth and gender) or Medicare ID

Data are available in a timely basis

- Data files are complete and available relatively quickly
  - 20XX Beneficiary Summary file available in June 20XX+1
  - Calendar Year utilization files 98+% complete by June of the following year

What are some broad limitations of claims data?

Record of Care Received

- Conditions must be diagnosed
  - hypertension, depression, diabetes often under-diagnosed
- Record of care received not care needed
  - We don't know disease recurrence but, rather, whether they got new treatment
- Services that providers know in advance will be denied will be inconsistently submitted as bills and, therefore, inconsistently recorded

Sometimes a diagnosis code doesn't provide enough detail

- Cancer diagnosis can be found as an ICD-9 diagnosis code (e.g., lung cancer is 162.xx) but not stage or histology
- Knowing that someone has a chronic disease (e.g., congestive heart failure) doesn't provide information about how long they have had the condition (incidence vs. prevalence) or severity
- Many drugs and procedures have multiple indications. There are no diagnoses on the Part D event file

Different care settings use different coding systems for procedures

- Inpatient care is coded using ICD-9 procedure codes (4 digits)
- Carrier and DME are coded using CPT codes and HCPCS codes
- Hospital outpatient care is coded as a mix of CPT and revenue center (hospital billing center) codes
- There is a less-than-perfect cross-walk between ICD-9 codes and CPT codes

Limited Clinical Information

- Physiology lacking:
  - blood pressure
  - Pulse
  - Cardiac Ejection fraction
- Test results not included:
  - PSA
  - Angiography
  - Pathology
- Exact timing not included:
  - hours from admission to event
  - Cannot tell time of day for ER visits

Data typically limited to covered benefits and for which claims are submitted

- Prior to Part D, Medicare had no pharmacy benefit so Outpatient medications cannot be studied. With Part D, studies will have to take formularies into account.
- Covered services for which claims are not submitted are not included (immunizations provided through grocery-store immunization clinics)
- Some services are not covered (see Medicare coverage decisions)

Data typically limited to covered benefits and for which claims are submitted

- No information about Part B services for Managed Care enrollees
- Little information (of largely unknown quality) about hospitalizations for managed care enrollees
- Not all beneficiaries have Part D coverage; not all beneficiaries with Part D coverage will have Part D utilization information contained in files

Variable Quality

- General rule: if it impacts payment its quality will be better than if it doesn't.
- Different types of care may be subject to different payment rules

Implication of this rule:

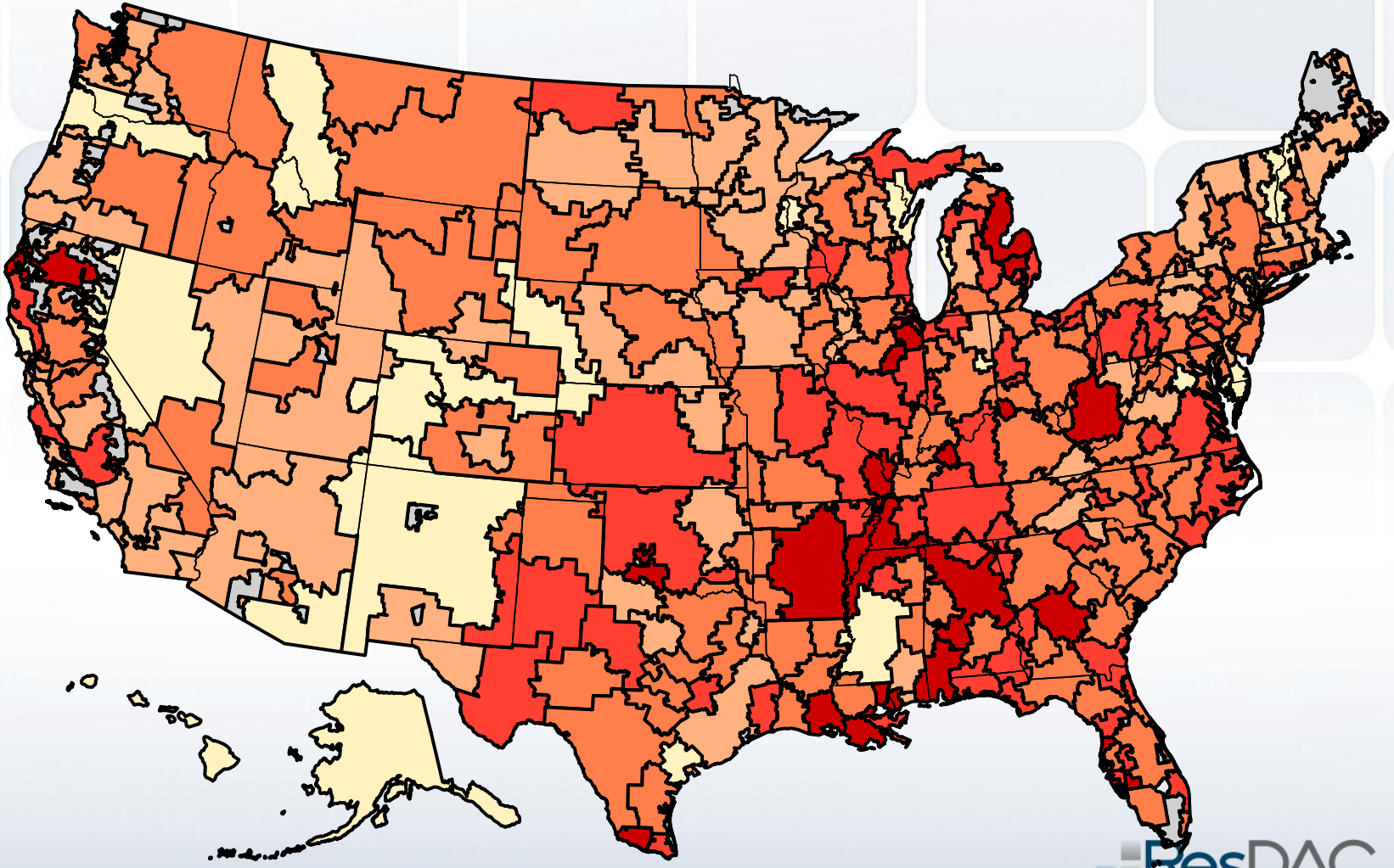
- Comorbidity and severity of illness information may be inconsistently recorded
- Some components of treatments may not be included in bills if reimbursement rates are very low (even if the treatment is provided)

Assessing Data Quality

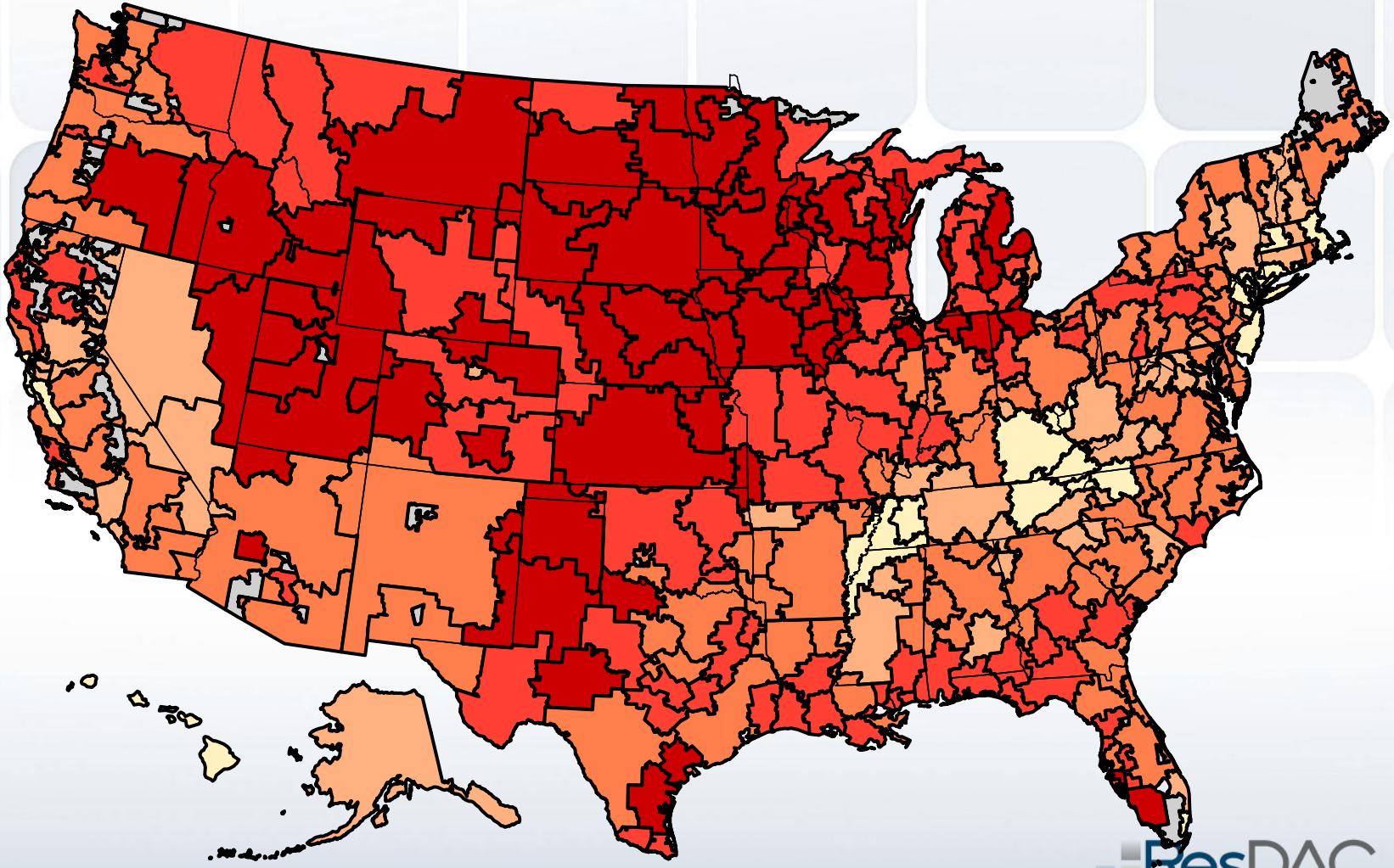
- In general, data elements provided by CMS provide consistent information
- The data dictionaries (record layouts) contain information about some assumptions, data combinations etc. These are IMPORTANT and you should have them handy while designing your study and analyzing your data

Examples of studies that could only be conducted with claims data

Ratio of HRR Rates of Coronary Artery Bypass Grafting Procedures to the U.S. Average (1994-95)



Ratio of HRR Rates of Knee Replacement Surgery to the U.S. Average (1994-95)



Other examples:

- Volume-outcome relationship
- Variation in treatment patterns
- Do racial differences in outcomes persist after treatment patterns are taken into account?

Getting ready to use Medicare data

3 broad steps must be completed whether working from population files or requesting CMS to identify a specific population:

- Define population (eligible for study)
- Define events
- Define adjusting variables

Defining the Population to be Studied

- A clear definition of the target population, whether for a study of a clinical population or an epidemiologic/population based study.
- Two general rules are:
  - all persons in the denominator must be eligible to have events
  - all persons in the numerator (events) must be eligible to be in the denominator
    - » Merging numerator and denominator into a single file will handle this

Defining the Population to be Studied

- Location of residence–continuous or at any time?
  - What about newly eligible?
  - Deaths?
  - People who move in and out of the area?

Defining the Population to be Studied

- Diagnosis or procedure
  - is there a single diagnosis that is used for all similar cases or can multiple diagnoses apply

Defining the Population to be Studied

- Demographics
- Specific coverage (Part A, Part B, MCO, State buy-in)
  - Continuous vs. ever
- Benefit program
  - OASI
  - Disability
  - ESRD

Defining the Population to be Studied

- Membership in a defined cohort which will be linked with CMS data

Define Outcomes of Interest

- Receipt of diagnosis
- Death (but, note, cause specific mortality can only be studied in some years using the death information provided by Medicare)
- Health Care Use
  - Hospitalizations
  - re-admissions
  - Clinic Visits
  - Procedures

Define Outcomes of Interest

- Duration of care or time until event
  - But, treated events will happen before untreated events
- Money spent
  - Total reimbursements
  - Co-payments
  - Deductibles

When studying time trends

- Are there changes in benefits, reporting or coding that could result in an apparent time trend?
  - Major payment policy changes
    - » e.g., DRGs (Oct 1, 1983)
  - Medicare managed care
  - New preventive services covered
  - Part D coverage
  - Coverage or payment policies for specific procedures
  - Changes in CPT codes

Define adjusting variables/covariates

- Demographics
- Location of residence
- Comorbidities

When doing the selecting

- Sample wide, analyze narrow
- It is harder to forget something and have to make a second request than it is to request more data than are needed and have to sort it out during analysis.

When it comes to actual analysis

- Spend time carefully defining population, nothing is gained by having more records that are poorly defined over fewer records that are well defined