

Population-based Health Services Research in the Era of Big Data

Leighton Chan, MD, MPH

Chief, Rehabilitation Medicine Department

Clinical Center

NIH

Disclosures

No financial disclosures

A “new” era: BIG DATA

Advances in computing power-IBM Watson

Increasing data capture- EMR

We can use these data to:

Provide “decision support” to providers

Inform policy

Types of Data

Structured: highly organized data containing pre-defined elements with standardized relationships to one another (e.g., data in a database)

Unstructured: data that is not structured in a pre-defined manner

Clinical notes, Faxed documents

Anything written in narrative

Majority of medical data is unstructured

Where Does Data Come From?

Primary Data – generated for research purposes, including national surveys and disease registries

Secondary Data – Secondary data is administrative/billing/encounter data

Often generated with utilization in mind

Enough data to make meaningful population-base conclusions

Common Sources of Secondary Data

Populations:

Medicare (54 million)

Integrated health systems (e.g., Kaiser, 10 million)

Medicaid (55 million)

Pharmacies

Billing /encounter/administrative data

Electronic Medical Record (RMR)

Why Study Medicare Patients?

Largest purchaser of health care in the world

54 million enrollees

\$613 billion in expenditures in 2014

The percentages

Almost 16% of U.S. budget outlays

22% of all health care dollars in US

26% of hospital spending

22% of nursing home spending

22% of physician billings

Single data system

Who are Medicare patients?

83% age > 65

<1% ESRD

16% disabled

20% in an HMO

Medicare

FFS beneficiaries health care costs ~\$8,000/yr

Costs/beneficiary rise ~7 percent/yr

Part A trust fund depletion projection: 2030

Organization of Medicare Data: Files

Hospital stays

Physician visits (inpt & outpt)

specialty

experience

SNF, DME, & Home Health

Hospital level data: size, non-profit status, staffing, etc.

Organization of Medicare Data: Variables

Patient demographics:

age, sex, race, zip code

Primary Diagnosis(ICD-9)

Associated co-morbidities

Procedure codes - CPT-4 & HCPCs

Organization of Medicare Data: Variables

Hospitalization and Rehospitalization

Sentinel Events

Length of Stay

Disposition

Mortality

Pharmacy use

Costs

Benefits of Medicare Data

Pre-existing data

less expensive

less time

large numbers of cases

Generalizability

Links to other data

Zip codes, SSN

Accurate measure of resource use

Can measure “effectiveness”

Limitations of Medicare data

Lots of limitations...

Limited data on severity of illness

Not generalizable to the US working population

Coding and billing errors/bias

Limited outcome measures of interest

No QOL, Patient Satisfaction, functional assessment, illness severity

Limitations

Studies limited to non-experimental design (observational studies)

Difficult to avoid selection bias

Impossible to control for all possible confounders (eg. severity of illness & functional status)

HMO patients are excluded

Cost of obtaining the data

Administrative overhead

How can I get data access?

Medicare administrative/billing/encounter data

<http://www.resdac.org/about-resdac/our-services>

Available through CCW Data Enclave

Types of Studies using Medicare data

Monitoring secular trends

Measuring disparities

Race, ethnicity, SES, geographic variation

Supporting the evaluation of specific conditions, treatments or procedures

Monitoring Secular Trends

Examine changes in health care over time

Take advantage of “Natural Experiments”

Examine the impact of policy changes

e.g.. Epidural Steroid Injections for Back Pain

14,000 recently exposed to contaminated steroids

Percent of US Population with LBP

Prevalence of Low Back Pain (>age 65)

Medicare patients 2002-2006

Nonspecific backache 60.3%

Degenerative changes 14.7%

Sciatica 11.8%

Spinal Stenosis 7.3%

LBP Costs

Deyo, MEPS

Costs are very high (\$86B)

Mean adjusted costs

1997 \$4,695 (95% CI, \$4,181-\$5,209)

2005 \$6,096 (95% CI, \$5,670-\$6,522)

30% increase in costs

Self-reported measures of mental health, physical functioning, work or school limitations, and social limitations among adults with spine problems were worse in 2005 than in 1997.

Definitions

Methods

Retrospective Cohort Study

Medicare claims data

1995-2006, 5%-20% sample of physician bills

Cohorts defined by CPT and ICD-9 codes

62311 Caudal or interlaminar

64483 Transforaminal

64475 Facet Injection

LBP dx from ICD-9 codes

Physician specialty from UPIN

How much has this
increase cost Medicare?

Total Estimated Costs

Physician Professional Fees \$175 million

+

Facility Fees \$275 million

Total Cost to Medicare = \$450 million

Where ESIs are Performed?

Outpatient Hospital Clinics

Physician Offices

Ambulatory Surgical Centers

Ambulatory Surgery Centers (ASCs)

Most Medicare certified ASCs are: (n=5,000)

privately owned

for profit

urban locations

Ambulatory Surgery Centers (ASCs)

Supposed to reduce costs by avoiding hospital overhead

Majority owned by local physician investors

The Stark self-referral law (1989 Social Security Act) does not apply to ASCs

MDs can invest in ASCs and increase revenue by receiving ASC facility payments

ASC Advantages

More convenient locations, shorter wait times

Medicare coinsurance is lower than in hospitals

(\$9 difference in 2004)

Customized environments, specialized staffing

Customer friendly

ESIs at ASCs: 1995-2005

1995: 13% of ESIs performed at ASCs

2005: 29% of ESIs performed at ASCs

Results

ASC ESI Facility Payment*

1995: \$7.5 million

2005: \$101 million

>1200% increase

Conclusions

Lots of growth in ESI

Growth associated with shift in “injectionists”

Growth associated ASC growth

Significant cost increases for Medicare

Do patients benefit?

Measuring geographic variation

Tom Wennberg & Alan Gittelsohn

Examine procedure rates in different geographic areas

If rates differ this suggests inequity or inefficiency in practice

Geographic Variations

Health Referral Regions (HRR):

Smaller geographic regions

Defined by Dartmouth's Atlas for Health Care (<http://www.dartmouthatlas.org/>)

306 HRRs across the country

Defined by where most of the cardiovascular and neurosurgery is performed

Geographic Variation

2005 Geographic Variations:

Health Referral Regions:

9-fold difference in ESIs/1000 patients

7.9/1000 in Honolulu, HI

103.6/1000 in Palm Springs, CA

Geographic Variations

Summary

Large geographic variations in ESI use

High ESI rates are not associated with lower surgery rates

High ESI rates are moderately associated with “injectionist” supply

Limitations

Only study those over age 65 in Medicare

No young active workers

No HMO patients

Possible errors in diagnosis/billing codes

Which ESI rate is right?

No data on pain relief

No data on return to work

No data on functional improvement

Unanswered Questions

Are ESIs effective?

How do we select the ideal patients for ESIs?

How many should we be doing?

How often should we be doing them?

Should we be doing them with other treatments? (i.e. multidisciplinary approach)

Analyzing a specific procedure

Bariatric surgery for obesity

Bariatric Surgical Procedures Mortality Study

Dramatic growth in bariatric surgery in obese adults

the number between 1998 and 2004 from 13,386 to 136,000

No national coverage decision or consensus regarding efficacy and safety in older adults

Bariatric Surgical Procedures Mortality Study

Objective:

Evaluate the risk factors of early mortality among Medicare beneficiaries (age, gender, surgeon experience)

Determine relative risk of death among older patients

Retrospective cohort design using Medicare physician bills, (1996-2002), 16,155 cases

Limitations

CPT codes not precise enough to determine exact procedure

Analysis of missing data?

Only those over age 65

Surgeon volume related to Medicare pts only

Bariatric Surgery Mortality Study

Conclusions:

Medicare beneficiaries ≥ 65 nearly 3 fold increase in risk of early mortality

Post-operative mortality rates associated with;

Advancing age

Male sex

Lower surgeon volume bariatric procedures

Medicare allowed those over 65 to have surgery, but only at those sites that did more than 125/year and by surgeons doing more than 50/yr

2010 follow-up study showed a reduction in LOS and complications after the NCD

Conclusions

Medicare Claims data is widely used in outcomes research

The data has significant advantages and disadvantages

The importance of this data will increase as the US demographics change and Medicare enrollment accelerates

CCW Data Enclave

CMS developed virtual data access for investigators through a new data enclave

Cost determined by the number of licenses (seats) in the enclave, and not by the amount of data requested

Users are assigned a dedicated workspace within the CCW Virtual Data Enclave where they can directly access approved CMS data and run analyses in SAS

Users may:

Upload external files to their Data Enclave workspace for use with CMS data

Download aggregate, statistical files to their workstations

NLP: The Next Big Thing

Unstructured data

There is a lot!

Medical history

Large numbers of patients

Machine learning methods: identify patterns, trends, and long-term changes

Need input from clinician and a linguist

Successful pre-processing critical

NLP Application to HF

Identification of Framingham HF criteria in PCP notes

Based on Unstructured Information Management Architecture (UIMA) framework

Partnership between IBM T.J. Watson Research Center, Geisinger Medical Center, and Sutter Health

NLP Application to HF

Iterative process: cardiologist, linguist, and coders in partnership

Cardiologist and linguist jointly review case files for key words and linguistics

Linguist builds NLP tools

Joint review of outcomes; extraction improvement

Coders create a “gold standard” for comparison

Promising Results

High accuracy in identification of Framingham HF criteria

Few false negatives: successfully identified 90% of true positives

Few false positives: >92% of cases labeled positive were true positives

Demonstrates PC notes can be successfully extracted

Shows potential for early identification methods