Brook I. Martin, PhD MPH

Instructor health services
Dartmouth College

Working with health claims & complex survey data

About me

• Health services PhD from University of Washington.
• Health services faculty at Dartmouth College.
• Affiliated with Dartmouth-Hitchcock Medical Center Department of Orthopaedics.
• Primary research interest is in the quality of care for musculoskeletal & spinal problems.

Goals

• Part 1: Strengths and limitations of analysis of health care claims and complex survey data.
Goals

• Part 1: Strengths and limitations of analysis of health care claims and complex survey data.

• Part 2: Identify sources of data commonly used in health services research.

• Part 3: Introduce essential Stata commands that serve as building blocks for working with claims and survey data.

Part I: Claims-based research

# of citations in PubMed from 1980 - 2010 identified with terms for claims-based or administrative data research
Claims & Survey data

• Also termed “secondary”, “administrative” data.

• “[complex] survey data” includes a sampling methodology.

• Studies reporting “population-based” estimates are often based on claims or survey data.
Claims & Survey data

- Also termed “secondary”, “administrative” data.
- “[complex] survey data” includes a sampling methodology.
- Studies reporting “population-based” estimates are often based on claims or survey data.
- By nature, these data are observational and generally retrospective.

Features of claims data

- Often used for billing purposes.
- A defined group, such as members of an insurance plan or a health network.
Features of claims data

- Often used for billing purposes.
- A defined group, such as members of an insurance plan or a health network.
- Diagnosis and procedure codes.

Features of claims data

- Often used for billing purposes.
- A defined group, such as members of an insurance plan or a health network.
- Diagnosis and procedure codes.
- Often includes other patient characteristics, comorbidity descriptions, provider information, geographic information, insurance information, reimbursement information, and quality measures.

Example – Discharge registry

WA State surgical Registry:
- Encrypted patient, hospital, surgeon identifier
- Admission & discharge dates
- Diagnosis & procedure codes
- Type of Insurance (Medicare/aid, commercial, HMO, workers’ compensation)
- Hospital charges
- Age
- Sex
- Death
- Discharge disposition
- Patient zip code
Example - Medicare

Medicare Data:
- Patient, provider & hospital identifiers
- Inpatient hospitalizations, ambulatory visit, Emergency, & pharmacy visit details.
- Diagnosis & procedure codes
- Type & duration of insurance
- Charges & reimbursements
- Patient characteristics

Features of complex survey data
- Does not have to include claims, but many complex health surveys do.

- “Complex” means sampling methodology was used so that researchers can produce unbiased estimates of the population.
Features of complex survey data

• Do not have to include claims, but many complex health surveys do.

• "Complex" means sampling methodology was used so that researchers can produce unbiased estimates of the population.

• Includes variables that describe the sampling methodology; these requires some special statistical tools.

What is the average age?

Take the average of a random sample..
More samples increase the accuracy...

I may decide to sample people within sampled “clusters”...

Example - Complex Surveys

Medical Expenditure Panel Survey (MEPS)
- Encrypted patient identifier
- Inpatient hospitalizations, ambulatory visit, Emergency, & pharmacy visit details.
- Diagnosis & procedure codes
- Insurance
- Expenditures
- Patient characteristics
- Health questionnaires
- Sampling design variables
Common uses of claims/survey data

- Describe “real world” utilization, variation, incidence, and costs.
- Informs effectiveness (versus efficacy).
- Reporting of rare outcomes.
Common uses of claims/survey data

- Describe “real world” utilization, variation, incidence, and costs.
- Informs effectiveness (versus efficacy).
- Reporting of rare outcomes.
- Reporting longer surveillance than clinical trials.
- Background data for grant proposals.
- Provides cost estimates for CEA.
Common uses of claims/survey data

- Describe “real world” utilization, variation, incidence, and costs.
- Informs effectiveness (versus efficacy).
- Reporting of rare outcomes.
- Reporting longer surveillance than clinical trials.
- Background data for grant proposals.
- Provides cost estimates for CEA.
- Monitor compliance on quality measures.

Strengths

- Less susceptible to selection and non-response bias.
Strengths

• Less susceptible to selection and non-response bias.
• Typically large sample that enhances power.

Strengths

• Less susceptible to selection and non-response bias.
• Typically large sample that enhances power.
• Representative & generalizable.

Strengths

• Less susceptible to selection and non-response bias.
• Typically large sample that enhances power.
• Representative & generalizable.
• Inexpensive & less time consuming.
Strengths

• Less susceptible to selection and non-response bias.
• Typically large sample that enhances power.
• Representative & generalizable.
• Inexpensive & less time consuming.
• Informs effectiveness.
• Generally enables long-term surveillance.
• May identify care from multiple hospitals.
Limitations

- Limited ability to make causal inference.
- Sub-groups for survey data may result in large sampling error.
- Observed variation may be due to local coding practices.
**Limitations**

- Limited ability to make causal inference.
- Sub-groups for survey data may result in large sampling error.
- Observed variation may be due to local coding practices.
- Limited data on validity & reliability of administrative data.
- Limited clinical detail and patient-reported health.
- Difficult to account for migration into or out of system.
Some practical considerations

- Human subjects review
  - Designation of “Public Use” data.
  - De-identified or limited data set.
  - waiver of consent.
  - Certificate of exemption from review for pre-existing or limited data.
  - Minimal risk review for pre-existing data.
  - Full review (especially if linking to other data).
- Compliance to confidentiality agreements, HIPAA, computer security.

Some practical considerations

- Coding and definitions may vary by source or site.
- Can often be extremely large & requires planning.
- Complex survey data requires special statistical tools to produce national estimates.

Case study of claims-based CER

Complications and Health Care Use as a Function of Type of Surgical Procedures.

Part 1: Useful Resources

- Analysis of Health Surveys (Korn & Graubard, 1999)
- Sampling: Design & Analysis. (Lohr, 2009)

Self-Check

- Explain the difference between efficacy and effectiveness.
- How does this relate to the role of claims-based and/or survey research?
- Discuss the major strength & limitations of Deyo’s (2010) study of Stenosis.

Part 2: Data sources

Where to find health services data? Will it help answer my research question?
Healthcare Cost & Utilization Project (HCUP)
- Agency for Healthcare Research and Quality (AHRQ).
- Describe rates & trends of hospital and ambulatory surgery; examine costs & quality measures.
- Includes:
  - State Ambulatory Surgery Database (SASD)
  - State Inpatient Database (SID)
  - Nationwide Inpatient Sample (NIS)
  - Kids Inpatient Database (KID)
  - Nationwide Emergency Department Sample (NEDS)
  - State Emergency Department Databases (SEDD)
  - http://www.hcup-us.ahrq.gov/overview.jsp

Example using NIS
Lumbar fusion rates by primary diagnosis.

Medical Expenditure Panel Survey (MEPS)
- Agency for Healthcare Research and Quality (AHRQ).
- Estimates of national health care utilization, insurance coverage, and expenditures.
- Includes:
  - Household surveys
  - Insurance characteristics
  - Provider components
  - http://www.meps.ahrq.gov/mepsweb/
Example using MEPS

Estimated Annual Per Capita Age- and Sex-Adjusted Health Expenditures Among US Adults With and Without Spine Problems, MEPS 1997-2005

- Adults with spine problems
- Adults without spine problems

Martin, B. I. et al. JAMA 2008;299:656-664

National Hospital Discharge Survey

- CDC’s National Center for Health Statistics (NCHS).
- National estimates of hospitalizations.
- Includes:
  - Visit & hospital information.
  - Complex sampling of hospitals to abstract visits details
  - http://www.cdc.gov/nchs/nhds/about_nhds.htm

Example using NHDS & HCUP

Estimated rates of all inpatient and outpatient lumbar spine surgery procedures, by state/data source, 1994-2000.
National (Hospital) Ambulatory Medical Care Survey (NAMCS & NHAMCS)

- Center for Disease Control
- National estimates of visits to physician offices and hospital-based ambulatory clinics.
- Includes:
  - Patient characteristics & conditions
  - Diagnosis, tests, procedures & medications.
  - Provider characteristics.
  - Complex sampling of physicians visits.

Example using NAMCS

National Health & Nutrition Examination Survey (NHANES)

- Center for Disease Control
- National estimates of health and nutritional status
- Includes:
  - Combines questionnaires with physical examinations.
  - Complex survey design.
- http://www.cdc.gov/nchs/nhanes/about_nhanes.htm
Example using NHANES

<table>
<thead>
<tr>
<th>Race/ethnicity</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Hispanic White</td>
<td>9,413</td>
<td>9,508</td>
<td>18,921</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>9,460</td>
<td>9,630</td>
<td>19,090</td>
</tr>
<tr>
<td>Non-Hispanic Asian</td>
<td>4,760</td>
<td>4,770</td>
<td>9,530</td>
</tr>
<tr>
<td>Hispanic</td>
<td>3,370</td>
<td>3,370</td>
<td>6,740</td>
</tr>
</tbody>
</table>

National Health Interview Survey (NHIS)

- CDC’s National Center for Health Statistics.
- Epidemiological trends to address health policy and public health issues.
- Includes:
  - Health questionnaires in a nationally representative sample.
  - [http://www.cdc.gov/nchs/nhis/about_nhis.htm](http://www.cdc.gov/nchs/nhis/about_nhis.htm)

Example using NHIS

Medicare data

- Centers for Medicare and Medicaid Services; Distributed by ResDac.
- Trends in utilization & costs among Medicare beneficiaries.
- Includes:
  - Inpatient claims files, outpatient claims, prescription drugs (Part D).
- http://www.resdac.org/

Example using Medicare data

Complications and Health Care Use as a Function of Type of Surgical Procedures.

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Death Rate</th>
<th>Morbidity Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open</td>
<td>0.8% (95% CI 0.5-1.3)</td>
<td>14.9% (95% CI 12.9-17.0)</td>
</tr>
<tr>
<td>Robotic</td>
<td>0.4% (95% CI 0.2-0.6)</td>
<td>9.6% (95% CI 7.9-11.4)</td>
</tr>
</tbody>
</table>


Others (1 of 3)

- American College of Surgeons National Trauma Data Bank (NTDB)
  http://www.facs.org/trauma/NTDB/index.html
- Behavioral Risk Factor Surveillance System (BRFSS; public data only)
  http://www.cdc.gov/BRFSS/
- Fatality Analysis Reporting System (FARS)
  http://www.fars.nhtsa.dot.gov/Main/index.aspx
- HIV Prevention Trials Network D01: Vaccine Preparedness Study/Uninfected Protocol Cohort
  http://www.hptn.org/network_information/public_data_sets.htm
- Inter-University Consortium for Political and Social Research (ICPSR)
  http://www.icpsr.umich.edu/icpsrweb/ICPSR/
- Longitudinal Studies of Aging (LSOA)
  http://www.cdc.gov/nchs/lsoa.htm
Others (2 of 3)

NIS: National Immunization Survey  
http://www.cdc.gov/nchs/nis.htm

NSFG: National Survey of Family Growth  
http://www.cdc.gov/nchs/nsfg.htm

SLAITS: State & Local Area Integrated Telephone Survey  
http://www.cdc.gov/nchs/slaits.htm

http://www.cdc.gov/nchs/nvss.htm

National Center for Education Statistics  
http://nces.ed.gov/

National Election Studies  
http://www.electionstudies.org/

Others (3 of 3)

National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)  
http://pubs.niaaa.nih.gov/publications/AA70/AA70.htm

National Longitudinal Survey (NLSY)  
http://www.bls.gov/nls/

Organ Procurement and Transplantation Network (OPTN)  
http://optn.transplant.hrsa.gov/

Roper Center for Public Opinion Research  
http://www.ropercenter.uconn.edu/

Survey of Consumer Finances (SCF)  
http://www.federalreserve.gov/pubs/scfindex.html

Self check

- Find the data elements (variables & coding) in the 2008 Nationwide Inpatient Sample (NIS)?
- Identify a public data source that a researcher might use to describe recent national trends in utilization of Complementary and Alternative Medicine (CAM)?
- Identify a data source that my help you answer questions related to one of your own interests.
Part 3. Building blocks for claims & complex surveys data.

- How to set up your data set for an analysis.
- Common Stata commands for claims & surveys
- Assumes: Introductory knowledge of Stata.
- Applied example: Trends in opiate medications for back pain.

Opioids in Spine

- Half of “regular” prescription opioid users have back pain.  
  (Sullivan, 2005)

- Cochrane: “…opioids for long term management of chronic LBP remains questionable”.  
  (Eriksen 2006)

- Hyperalgesia, death & side effects are significant among opioid user.  

- Recent guidelines have been developed for opioids use  
  (Chou, 2007)

Visualizing your data
Visualizing your data

MEPS Organization

- http://www.meps.ahrq.gov/mepsweb/

- File types:
  - Event Files (inpatient, outpatient, office-based, prescription, ER, home, dental)
  - Population Files
  - Condition Files

- Publicly available by year

http://www.meps.ahrq.gov/mepsweb/
### Medical Expenditure Panel Survey

#### Survey Questions

1. **What was the patient's age?**
2. **What was the patient's gender?**
3. **What was the patient's race/ethnicity?**
4. **Where did the patient receive care?**
5. **What was the patient's diagnosis?**
6. **What was the patient's treatment?**
7. **What was the patient's medication?**

#### Survey Data

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>50</td>
<td>Male</td>
</tr>
<tr>
<td>Race</td>
<td>Caucasian</td>
<td>Male</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>Caucasian</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Cancer</td>
<td>Hospital</td>
</tr>
<tr>
<td>Treatment</td>
<td>Chemotherapy</td>
<td>Outpatient</td>
</tr>
</tbody>
</table>
| Medication| Latin American | Oral

### Health Care Costs

#### Cost Breakdown

1. **Direct Costs**
   - Physician fees: $500
   - Medication: $100
   - Hospitalization: $2000
2. **Indirect Costs**
   - Lost productivity: $500
   - Transportation: $100

#### Total Costs

- **Total Direct Costs:** $2700
- **Total Indirect Costs:** $600
- **Total Costs:** $3300

### Health Outcomes

#### Measured Outcomes

1. **Health Status**
   - Overall health rating: Good
   - Physical function: 8
   - Emotional function: 7
2. **Quality of Life**
   - Satisfaction with care: 9
   - Overall health satisfaction: 8

### Health Outcomes Analysis

- **Improvement in Health Status:** 1
- **Change in Quality of Life:** 1

---

**Source:** [Medical Expenditure Panel Survey](http://www.meps.ahrq.gov/mepsweb/)
Tip: Know your variables!

**Cost data**
- Zero dollars for event
- Negative dollars (?)
- Account for inflation

**Data points**
- Unit of analysis: Patients or events (?)
- Unique or multiple?
- String, dates, or numeric formatting?

**DrugID**
- National Drug Codes
- Multum-Lexicon

**ICD-9-CM**
- Diagnosis codes

---

Loading MEPS files in Stata

**Commands:**
- `Infix` – text in fixed columns
- `Infile` – tab, comma or space delimiter
- `Insheet` – reads spreadsheets
- `FDAuse` – uses SAS or Xport data

**Syntax**
infix specifications using filename [if] [in]

---

Loading data

```
infix str dupersid 9-16 str EVNTIDX 17-31 str RXICD1X 371-373 TC151 397-399 RXXPX 512-518 PERWT 519-530 VARSTR 531-533 VARPSU 534-534 using "C:\Documents and Settings\MEPS\H85a.DAT"
format EVNTIDX %15.0g
```

**TIPS:**
- Format Identifiers
- USE "str" for identifiers and ICD codes
- Note multiple ICD code available
- Merging: common variable names across all years

---
Appending files

<table>
<thead>
<tr>
<th>Event File 2005</th>
<th>Event File combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVENTIDX</td>
<td>Dupersid</td>
</tr>
<tr>
<td>50002011</td>
<td>50002</td>
</tr>
<tr>
<td>50001042</td>
<td>50001</td>
</tr>
<tr>
<td>50001058</td>
<td>50001</td>
</tr>
</tbody>
</table>

Event File 2004

<table>
<thead>
<tr>
<th>EVENTIDX</th>
<th>Dupersid</th>
<th>TC1</th>
</tr>
</thead>
<tbody>
<tr>
<td>50002489</td>
<td>50002</td>
<td>107</td>
</tr>
<tr>
<td>50003036</td>
<td>50003</td>
<td>83</td>
</tr>
</tbody>
</table>

Appending Example

Step 1: Create a variable to indicate source year
generate float year = 2004

Step 2: Sort on unique identifier and save each data set
sort EVENTIDX
save "C:\MEPS\RXEVENT04.dta", replace

Step 3: Use append command
append using "C:\MEPS\RXEVENT05.dta"
*Also, see options, uniquemaster uniquesaving

Step 4: Save as combined file
save "C:\MEPS\RXEVENT.dta", replace

Merging files

<table>
<thead>
<tr>
<th>Population File 2005</th>
<th>Merged file 2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dupersid</td>
<td>Age</td>
</tr>
<tr>
<td>50001</td>
<td>73</td>
</tr>
<tr>
<td>50002</td>
<td>37</td>
</tr>
<tr>
<td>50003</td>
<td>56</td>
</tr>
<tr>
<td>50004</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition file 2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dupersid</td>
</tr>
<tr>
<td>50002</td>
</tr>
<tr>
<td>50002</td>
</tr>
<tr>
<td>50003</td>
</tr>
</tbody>
</table>

Warnings:
What happens to patients with no conditions?
What happens to patients with multiple conditions?
Are there conditions that don’t match to your patients?
Do you lose the unique identifier?
Merging Example

Step 1: Save each file sorted by unique identifier(s) that they are to be merged on.
   sort dupersid year

Step 2: Open one of the files ("master")
   use "C:\MEPS\POP05.dta"

Step 3: Use MERGE command to match with "using"
   merge 1:1 dupersid year using "C:\MEPS\COND05.dta"

Step 4: Save as combined file
   save "C:\MEPS\MERGE05.dta", replace

---

Tip: Always Look!

---

Reshape file (from this)

<table>
<thead>
<tr>
<th>Dupersid</th>
<th>Age</th>
<th>PERNUM</th>
<th>Year</th>
<th>ICD</th>
</tr>
</thead>
<tbody>
<tr>
<td>50001</td>
<td>73</td>
<td>26877</td>
<td>2004</td>
<td>.</td>
</tr>
<tr>
<td>50002</td>
<td>37</td>
<td>13045</td>
<td>2004</td>
<td>724</td>
</tr>
<tr>
<td>50002</td>
<td>37</td>
<td>13045</td>
<td>2004</td>
<td>724.3</td>
</tr>
<tr>
<td>50002</td>
<td>37</td>
<td>13045</td>
<td>2004</td>
<td>V02.6</td>
</tr>
<tr>
<td>50003</td>
<td>56</td>
<td>14186</td>
<td>2004</td>
<td>847</td>
</tr>
<tr>
<td>50005</td>
<td>18</td>
<td>13464</td>
<td>2004</td>
<td>786.5</td>
</tr>
<tr>
<td>50005</td>
<td>18</td>
<td>16544</td>
<td>2005</td>
<td>724</td>
</tr>
</tbody>
</table>
Reshape Example

Step 1: sort by unique identifier(s)

```
sort dupersid year
```

Step 2: generate a count variable by reshape group

```
gen num = _n, over(dupersid year)
```

...
Reshape Example

Step 3: Use reshape command
   reshape wide ICD, i (dupersid year) j (num)

Reshape file (to this)

<table>
<thead>
<tr>
<th>Dupersid</th>
<th>Age</th>
<th>PERFWT</th>
<th>Year</th>
<th>ICD1</th>
<th>ICD2</th>
<th>ICD3</th>
</tr>
</thead>
<tbody>
<tr>
<td>50001</td>
<td>73</td>
<td>28877</td>
<td>2004</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>50002</td>
<td>37</td>
<td>10046</td>
<td>2004</td>
<td>724</td>
<td>724.3</td>
<td>702.8</td>
</tr>
<tr>
<td>50003</td>
<td>66</td>
<td>14188</td>
<td>2004</td>
<td>847</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>50004</td>
<td>26</td>
<td></td>
<td>2006</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>50005</td>
<td>18</td>
<td>13464</td>
<td>2004</td>
<td>786.5</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>50005</td>
<td>18</td>
<td>16544</td>
<td>2005</td>
<td>724</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

Warnings:
Non-reshaped variables can’t differ within unique combination.

For each unique combination of dupersid*year, Stata creates a variable:
   ICD+num1, ICD+num2, ICD+num3

Using ICD9 codes

What is icd-9 code 724?
   icd9 lookup 724
      1 match found:
       724  back disorder nec & non*

What icd-9 codes is sciatica?
   icd9 search sciatica
      1 match found:
       724.3  sciatica
Using ICD9 codes

```
ic check icd1, gen(newvar)
1. Invalid placement of period 0
2. Too many periods 0
3. Code too short 21,356
4. Code too long 0
5. Invalid 1st char (not 0-9, E, or V) 0
6. Invalid 2nd char (not 0-9) 0
7. Invalid 3rd char (not 0-9) 0
8. Invalid 4th char (not 0-9) 0
9. Invalid 5th char (not 0-9) 0
10. Code not defined 0

Total 21,356
```

Formatting an ICD variable

```
icd9 clean ICD1, pad
```

```
Name of ICD-9 diagnosis variable in my dataset
```

Using icd-9 variables

Make dichotomous variable by searching for several ICD-9 codes.

```
icd9 generate BACKPT = ICD1, range(720 721 722 723 724 737 805 806 839 846 847)
```

Then label variable

```
label define BPPT
0"NOT spine patient" 1"spine patient"
variable label BACKPT BPPT
```
Looping & Foreach

Use loops to perform the same operation on multiple variables. Here I search three ICD codes for each event for those that are back related.

local vars = "ICD1 ICD2 ICD3"
foreach var of varlist `vars` {
icd9 gen IN`var' = `var', range(720 721 722 723 724 737 805 806 839 846 847)
}
egen BACKPT = anymatch(INICD1 INICD2 INICD3), values (1)

e-gen (see help: egen)

local vars = "ICD1 ICD2 ICD3"
foreach var of varlist `vars` {
icd9 gen IN`var' = `var', range(720 721 722 723 724 737 805 806 839 846 847)
}
egen BACKPT = anymatch(INICD1 INICD2 INICD3), values (1)

Collapse

collapse (sum)sum=RXCOST [pw=PERWTF] if BACKPT == 1, by(year narc)

<table>
<thead>
<tr>
<th>PERWTF</th>
<th>Year</th>
<th>RXCOST</th>
<th>Narc</th>
</tr>
</thead>
<tbody>
<tr>
<td>12345</td>
<td>2004</td>
<td>128</td>
<td>1</td>
</tr>
<tr>
<td>21221</td>
<td>2004</td>
<td>238</td>
<td>1</td>
</tr>
<tr>
<td>23296</td>
<td>2004</td>
<td>654</td>
<td>0</td>
</tr>
<tr>
<td>19732</td>
<td>2003</td>
<td>346</td>
<td>0</td>
</tr>
<tr>
<td>20023</td>
<td>2003</td>
<td>654</td>
<td>0</td>
</tr>
<tr>
<td>12311</td>
<td>2003</td>
<td>346</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>sum</th>
<th>Narc</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>6.6e+06</td>
<td>1</td>
</tr>
<tr>
<td>2004</td>
<td>1.5e+07</td>
<td>0</td>
</tr>
<tr>
<td>2003</td>
<td>2.0e+07</td>
<td>0</td>
</tr>
<tr>
<td>2003</td>
<td>4.3e+06</td>
<td>1</td>
</tr>
</tbody>
</table>
Analysis with survey

In Stata:
Statistics> Survey data analysis or help svy

Regression with survey variables:

“Survey” or “Complex Survey” Biostatistics courses


Using Survey Design Variables

Tell Stata you are using survey data:
svyset VARPSU [pweight=PERWTF], strata(VARSTR) single(centered)

Use survey commands to produce summary statistics:
svy, subpop(if narc == 1 & BACKPT == 1) vce(linearized):
total sum, over(year)

Or use survey regression techniques:
xi: svy linear, subpop(if narc == 1): regress sum age sex
i.year, over(year)

Graph

twoway (connected sum year if narc ==1, sort),
title(National Expenditures for Spine related opioid prescriptions ) subtitle(Medical Expenditure Panel Survey, 1997-2008)
ytitle("Billions,USD", orientation(vertical))
legend (off) caption("Inflation adjusted to 2008 equivalents, based on Medical CPI")
Finally!

Resources

- http://www.stata.com/
- Getting started with Stata (StataCorp)
- Data Analysis Using Stata (Kohler)
- http://www.meps.ahrq.gov/mepsweb/

Self check

- Find the MEPS codebook and data set for the 2008 prescription medication event.
- What measures of health status are available in MEPS? (look for these in the population files)
Thanks!

Brook.L.Martin@Dartmouth.edu

References & Resources