Statewide Health Datasets

February 24, 2017

Zuber D. Mulla
Professor of OB/GYN
Assistant Dean for Faculty Development
Director of Research, ATACS Center
PLFSOM, TTUHSC EP
Objectives

- Define the FINER approach to clinical research

- List 2 sources of statewide (Texas) health data that can facilitate rapid research projects
Identifying a Research Question

• Asking a meaningful research question can be tough but it is a very important aspect of research.

• No ideas yet? Address a gap in knowledge. Ask mentors for advice.
Clinical Research the UCSF Way

*Designing Clinical Research,*
*Fourth Edition by*

Stephen B. Hulley, MD, MPH
Steven R. Cummings, MD
Warren S. Browner, MD, MPH
*et al.*
Clinical Research the UCSF Way

- Feasible
- Interesting
- Novel
- Ethical
- Relevant
What if you don’t want to collect data?

• Collecting data (e.g., reviewing charts or interviewing subjects) is a valuable experience. This leads to data entry. Entering data is also a good experience (helps teach the value of “clean” dataset)

• However, if you absolutely do not want to do this, use an existing dataset
Using Existing Data

- Some call this a “secondary data analysis”
Two Approaches to Secondary Data Analyses

- Generate research question and then find a suitable database
- Find a database, examine it, form a suitable research question

Hulley et al. 2001
Framework for Secondary Data Analysis

- View dataset: is a study possible?
- Review of literature: gap in knowledge?
- Consult experts
- Organize team
- Draft protocol with detailed lit review
- Submit to IRB(s)
- Analyze data
- Draft manuscript
- Submit to one journal at a time
Two major health datasets

- Texas hospital inpatient discharge data (statewide data from many hospitals, millions of records, not necessarily Texas residents)
  - Public Use Data File (PUDF)
  - Research Data File

- Texas Cancer Registry incidence file (limited data on every reported case from 1995 onwards, only Texas residents – legal or illegal)
Selected Variables Found in the Texas PUDF

• Demographic: age group, race, Hispanic ethnicity

• Financial: Insurance payor, total charges in $

• Length of stay

• Diagnosis and procedure variables: ICD-9-CM through 3rd quarter of 2015 (then ICD-10)

• Hospital mortality
# Line Listing of Three Fictitious Patients

<table>
<thead>
<tr>
<th>THCIC_ID</th>
<th>PAT_AGE</th>
<th>RACE</th>
<th>ETHNICITY</th>
<th>PRINC_DIAG_CODE</th>
<th>OTH_DIAG_CODE_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>263000</td>
<td>09</td>
<td>4</td>
<td>2</td>
<td>5400</td>
<td>25000</td>
</tr>
<tr>
<td>130000</td>
<td>00</td>
<td>4</td>
<td>2</td>
<td>4280</td>
<td>769</td>
</tr>
<tr>
<td>000105</td>
<td>21</td>
<td>3</td>
<td>1</td>
<td>41010</td>
<td>4019</td>
</tr>
</tbody>
</table>
# Line Listing of Three Fictitious Patients

<table>
<thead>
<tr>
<th>THCIC_ID</th>
<th>PAT_AGE</th>
<th>RACE</th>
<th>ETHNICITY</th>
<th>PRINC_DIAG_CODE</th>
<th>OTH_DIAG_CODE_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>263000</td>
<td>09</td>
<td>4</td>
<td>2</td>
<td>5400</td>
<td>25000</td>
</tr>
<tr>
<td>130000</td>
<td>00</td>
<td>4</td>
<td>2</td>
<td>4280</td>
<td>769</td>
</tr>
<tr>
<td>000105</td>
<td>21</td>
<td>3</td>
<td>1</td>
<td>41010</td>
<td>4019</td>
</tr>
</tbody>
</table>
Some of our publications using Florida and Texas hospital discharge datasets...

- Correlates of length of stay, cost of care, and mortality among patients hospitalized for necrotizing fasciitis.

Mulla ZD, Gibbs SG, Aronoff DM.
Hyperbaric Oxygen Treatment (HBOT)

- 216 cases of necrotizing fasciitis

- Overall mortality: $24 / 216 = 11.1\%$

- 19 had HBOT (used procedure codes)

- Adjusted relative risk for hospital mortality for HBOT (Y / N): $0.48$, 95% CI: $0.09 – 2.56$

- $P$ value $= 0.39$
• Anaphylaxis in the obstetric patient: analysis of a statewide hospital discharge database.

Mulla ZD, Ebrahim MS, Gonzalez JL.
Maternal Anaphylaxis During Pregnancy

• 705,183 deliveries in Texas, years 2004-05

• 19 maternal cases

• Prevalence = 2.7 cases per 100,000 deliveries (95% confidence interval: 1.7 per 100,000 – 4.2 per 100,000)
Maternal Anaphylaxis During Pregnancy

- Penicillins and cephalosporins were the anaphylactic trigger in 11 of the patients.

- Five patients were emergent admissions.

- There were no maternal deaths.

- The majority of the patients (73.7%) delivered by cesarean section.
Strengths

• Serious conditions

• Rare and serious conditions

• Unique identifier available in the Research Data File
Limitations

• Accuracy of diagnosis codes

• Yasmeen et al. found that the ICD-9-CM coding of preeclampsia (any type) in hospital discharge data was accurate: sensitivity = 88\%, PPV = 91\% 
  
  *Am J Obstet Gynecol* 2006; 194: 992-1001
Limitations

- It appears that ICD-9 codes for obesity, tobacco use, and BMI category are underutilized
- Cross-sectional
- Data suppression in the PUDF
Limitations

• Not all hospitals required to report their data to the Texas Health Care Information Collection (military and VA hospitals don’t report)

• Codes change over time

• Length of stay is not a bell-shaped distribution
Histogram of Length of Stay (LOS)
Histogram of LOS with Normal, Log Normal, & Gamma
Major potential limitation (consult statistician or epidemiologist)

- Nesting / clustering of pts by facility. Most statistical tests and models assume independence.

- Solutions: reduce to one record per patient, or keep all of the records and use special techniques such as GEE or a mixed model (fixed plus random effects)
Major potential limitation

• *In the PUDF, there may be multiple records per patient*
Selected Variables Found in Limited Use Incidence File from Texas Cancer Registry

- Primary site and histology
- Stage and grade
- Demographics
- Treatments
Two publications using Texas tumor registry data...
Season of birth and the risk of uveal melanoma.

Mulla ZD, Margo CE.
Season of Birth and Melanoma of the Eye

- Season of birth and brain tumors

- Years 1995 - 2004

- 743 cases and 781,029 controls

- Autumn (Sep, Oct, Nov) vs. Spring (Mar, Apr, May) adjusted odds ratio = 0.82, $P = 0.07$

- Descriptive epidemiology of gastric adenocarcinoma in the state of Texas by ethnicity: Hispanic versus White non-Hispanic.

Rajabi B, Corral JC, Hakim N, Mulla ZD.
Gastric Adenocarcinoma

• Several research questions

• What factors are associated with a late stage at diagnosis (dx)?

• 8412 cases: 515 residents of El Paso County, 7897 residents of the remaining counties of Texas
Adjusted Prevalence Ratios (PR) for Late Stage at Dx in El Paso County

<table>
<thead>
<tr>
<th>Ethnicity/Race</th>
<th>PR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic (any race)</td>
<td>1.00</td>
<td>0.87 – 1.15</td>
</tr>
<tr>
<td>White non-Hispanic</td>
<td>1</td>
<td>(Referent)</td>
</tr>
</tbody>
</table>
### Adjusted Prevalence Ratios (PR) for Late Stage at Dx in El Paso County

<table>
<thead>
<tr>
<th>Age at Dx (y)</th>
<th>PR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 40</td>
<td>1.71</td>
<td>1.24 – 2.37</td>
</tr>
<tr>
<td>40 - 44</td>
<td>1.65</td>
<td>1.19 – 2.29</td>
</tr>
<tr>
<td>≥ 85</td>
<td>1</td>
<td>(Referent)</td>
</tr>
</tbody>
</table>
Links

- Texas Cancer Registry:
  
  http://www.dshs.state.tx.us/tcr/
Links

• Free Texas hospital discharge data (Public Use Data File [PUDF]) for the years 1999 through 2009:

• http://www.dshs.state.tx.us/THCIC/Hospitals/Download.shtm
Public Use Data File
Hospital Inpatient Discharge Data Form

Payment accepted by check only
Please complete packet and mail to the address below.

<table>
<thead>
<tr>
<th>Type of Business</th>
<th>2013 - **2015 purchased by calendar year</th>
<th>2013 - **2015 purchased by calendar quarter</th>
<th>2010 - 2012 per purchased by calendar quarter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texas State Agencies and State Universities</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
</tr>
<tr>
<td>Texas City/County/Local Government Health Departments</td>
<td>$1,500</td>
<td>$438</td>
<td>$156.25</td>
</tr>
<tr>
<td>Texas Reporting Hospitals*</td>
<td>$3,000</td>
<td>$875</td>
<td>$312.50</td>
</tr>
<tr>
<td>Texas In-State Media</td>
<td>$3,000</td>
<td>$875</td>
<td>$312.50</td>
</tr>
<tr>
<td>Out of State Health Departments</td>
<td>$3,000</td>
<td>$875</td>
<td>$312.50</td>
</tr>
<tr>
<td>Texas Private universities/colleges</td>
<td>$6,000</td>
<td>$1,750</td>
<td>$625</td>
</tr>
<tr>
<td>Out of State Media</td>
<td>$6,000</td>
<td>$1,750</td>
<td>$625</td>
</tr>
<tr>
<td>Out of State Agencies</td>
<td>$6,000</td>
<td>$1,750</td>
<td>$625</td>
</tr>
<tr>
<td>Out of State Universities</td>
<td>$6,000</td>
<td>$1,750</td>
<td>$625</td>
</tr>
<tr>
<td>Out of State Hospitals</td>
<td>$6,000</td>
<td>$1,750</td>
<td>$625</td>
</tr>
<tr>
<td>All other businesses or consumers, including hospital or ASC affiliates, organizations, institutions, corporate</td>
<td>$6,000</td>
<td>$1,750</td>
<td>$625</td>
</tr>
</tbody>
</table>
Conclusions

• Find a research mentor

• Consult an epidemiologist or biostatistician

• Statewide data such as electronic hospital discharge records and tumor registry data can aid in elucidating the descriptive, analytic, and clinical epidemiology of various conditions