Thinking through an Administrative Data Research Project

Developing Your Idea / Brainstorming:
- Clinical issues relevant to providers and/or patients
- Variation in care/practice patterns
- Rare diseases or outcomes that require large initial population
- Policy relevant issues
- Keep in mind that you will be using administrative data, therefore:
  - Relying on ICD-9-CM diagnosis or procedure, HCPCS, and DRG codes
  - Diagnostic tests may be poorly coded, invasive procedures better
  - Diagnoses that affect management and reimbursement coded better

Background / Literature review (in addition to framing the significance/background of study)
- LOOK AT PUBLICATIONS using administrative data. Most important first step to find out what others have done and to develop strategies for classifying diagnoses and procedures
- Look for publications validating codes of interest
- If using HCUP data look at statistical briefs to give examples of how analyses done

Study Design
- State your hypothesis in the beginning
- Select population of interest to guide your database selection
  - Ages, Demographics
  - Clinical diagnosis or procedures
  - Need for clinical detail
  - Inpatient, outpatient, all spectrum of care
  - State-level or national
  - One year or across time
  - Perspective of interest – facilities, insurance providers, physicians, patients
- Availability and costs of databases
- Narrow your database options
- Define cases of interest and independent variables of interest – BE EXPLICIT!
  - Learn to love codes! ICD and CPT/HCPCS codes are the clinical backbone of administrative data
  - Detailed codes review and classification important ESSENTIAL
    - Must review CODEBOOKS in depth
    - Review the ICD coding manual / CPT and HCPCS coding manual – don’t just rely on the index
    - Look at exclusions and inclusions, read the definitions carefully
    - Understand the codes you are using – look at coding guidelines if necessary
    - Talk to providers familiar with the codes you want to use
    - Always check coding guidelines when doing trends
    - Key resource: www.cdc.gov/nchs/icd/icd9cm_addenda_guidelines.htm
  - Avoid double-counting cases
    - Similar numeric and E codes – many cases may have both
    - Only count cases once, don’t count diagnoses
  - Often use hierarchy or combination of diagnoses and procedure codes
  - When using coding algorithms from others, always review them critically
  - Consider whether you are interested in principal/primary or secondary diagnoses codes
  - Think carefully about inclusion and exclusion criteria (e.g., transfer cases, age, gender)
  - Consider potential confounders and whether you can adequately control for them
    - Clinical data elements
    - Severity adjustments – use carefully, not always a good alternative to clinical diagnoses
KNOW YOUR DATABASE .... READ the documentation to be sure your database can support your analysis AND how you need to define your case
  o Summary statistics (missing observations, distribution of values)
    ▪ Presence of data element does not mean it has been validated
  o Data elements definitions and exceptions
  o Survey questionnaires
  o Data elements may or may not be present in all databases or years
• Develop research protocol
• BE EXPLICIT about methods

Data Acquisition
• CADR processes or other means to acquire data

Data Analysis
• Obtain appropriate programming and statistical support given your expertise
  o Do not underestimate the time for data cleaning and data analysis
  o May require vast majority of programmer time
• Be aware of tools available to assist with analyzing data
• Look at methods/statistical reports available through various database projects
  o For HCUP data read Methods Series – methodology and programming codes
  o Medicare/Medicaid data – ResDAC website has link to statistical resources and technical publications
  o SEER/Medicare data – NCI Health Services and Economics website – analytic support
  o MarketScan Data – publications in literature
• Think about population-based rates and how to obtain
  o Age, gender, rural/urban location, region, median household income
  o Payer (have enrollment files for Medicare and Medicaid but not HCUP)
• Think about how you will handle missing data – imputation or drop encounter/person from analysis
  o If dropping person/encounter consider potential bias if information is not missing at random
• Create tables that you will populate to help guide your analysis and clarify what you are trying to obtain from the data
• Remember with these large databases, statistical significance is frequently obtained – clinical significance is not
• Consider analytic techniques that can address some (not all) of the limitations of administrative data analysis
  o Multivariable analysis, propensity scores, instrumental variable analysis
• DOCUMENT all analytic decisions regarding coding of variables and process of selecting population

Results
• Be very careful in the interpretation and write-up of the results
• Do not over-interpret the results
• The interpretation needs to reflect the level of analysis performed
  o Person level
  o Encounter level
• Remember that data was not originally collected for research purposes
• Carefully consider the limitations of the data and research design

Publishing
• Be sure to follow DUA policies
• Check data project publication guidelines