

2018 Nursing Knowledge: Big Data Science Conference

Pre-Conference Data Science Workshop: Understanding the Research Question

June 13, 2018 University of Minnesota

Thompson Forbes III, PhD, RN, NE-BC, Assistant Professor, College of Nursing,
East Carolina University

Kenrick D Cato, PhD, RN, CPHIMS, Assistant Professor Columbia University
School of Nursing, Nurse Researcher, New York-Presbyterian Hospital

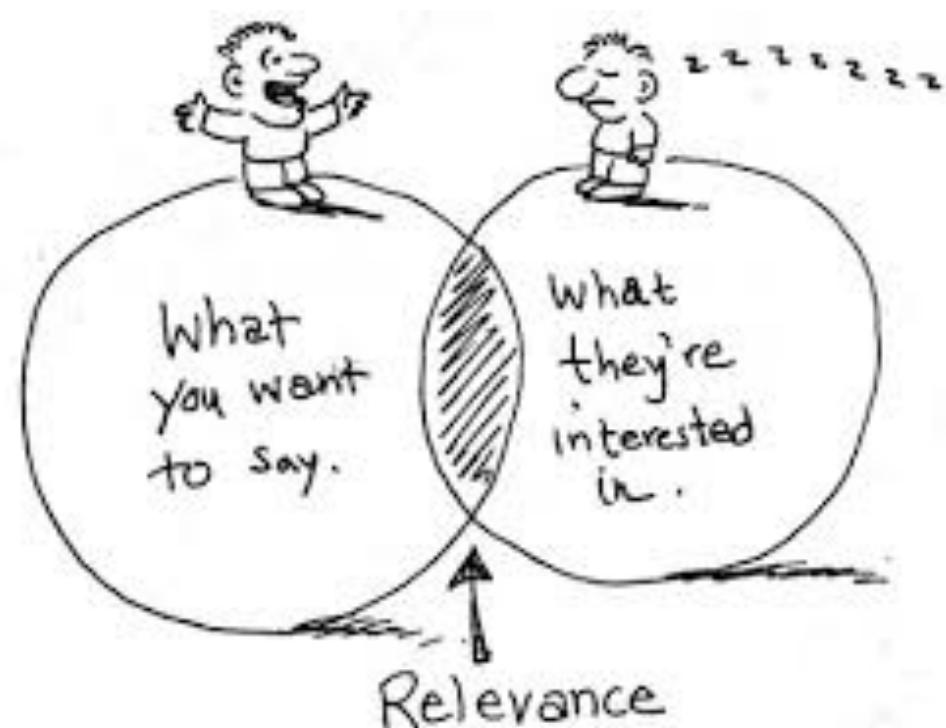
Relevance of the Research Question

Clinical Relevance

Funding Alignment

Nursing vs. Non-Nursing Data
Science Applications

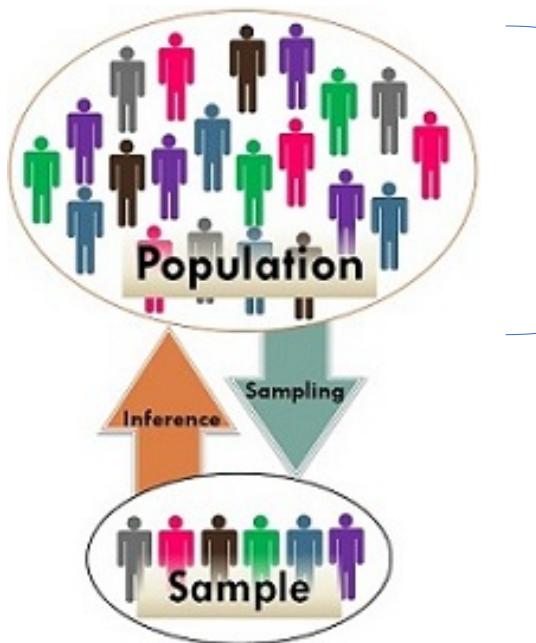
Domain Expertise on the Research
Team



Our Research Question Today

Can we predict which patients might
overdose on opioids?

Evolution of the Research Question – Defining Sample/Population



Data may contain an entire population!!

What potential sampling biases may be present when using large or population based data sets?

Kaplan, R.M., Chambers, D.A., & Glasgow, R.E. (2014). Big data and large sample size: A cautionary note on the potential for bias. *Clinical and Translational Science*, 7(4), 342-346.

Evolution of the Research Question – Modifying as Research Progresses

- May begin with general insight that is needed.
- As new data is analyzed/acquired research question may evolve
- Mini-analyses may assist with refining



Data Exploration – Data Understanding

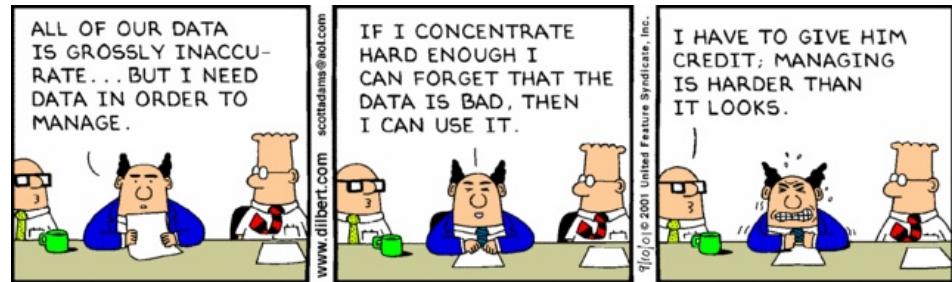
- Understanding the Data is Imperative!!
- Number and Types of Variables
- Format of Data and Implications on Analysis
- Volume of Data (File Size, Population Size, Timespan)



Data Exploration

– Data Quality

- Collection Process
- Alignment with Clinical Workflow
- Data Source
- Missing Data



Data Exploration – Data Gaps

- Are Variables Aligned with Research Question?
- Determine if Other Data Sources are Needed
- Access to Additional Data Sources
- Competency on the Team to Combine Datasets



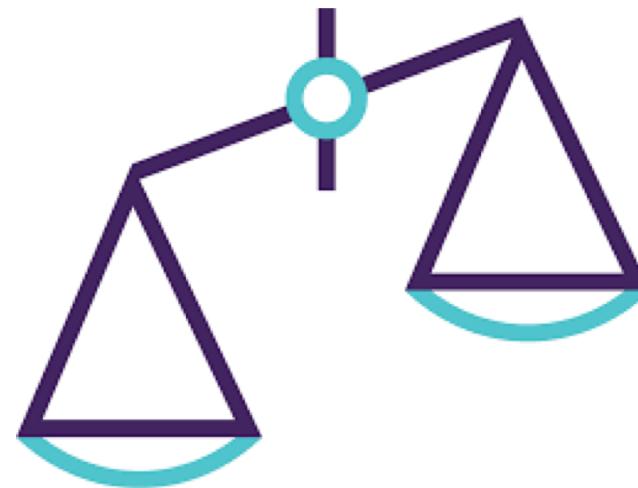
Operationalizing the research question

- Is the research question feasible
- “Can my team do it?”



Feasibility of research questions

- Is this a research question that I have the capability to answer ?
- Do we have:
 - Access to the relevant data
 - The resources
 - The appropriate tools
 - The time
 - The funding



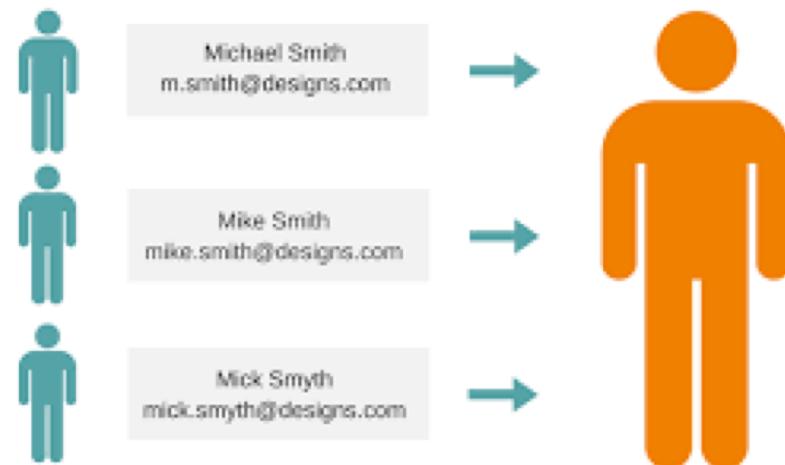
Availability of Data sources

- Organizational ability limitations
 - Will all of the owners of those data share their data?
- Time limitation: How long has those data existed and has it changed over time?
- Do those data even exist?
 - e.g location information

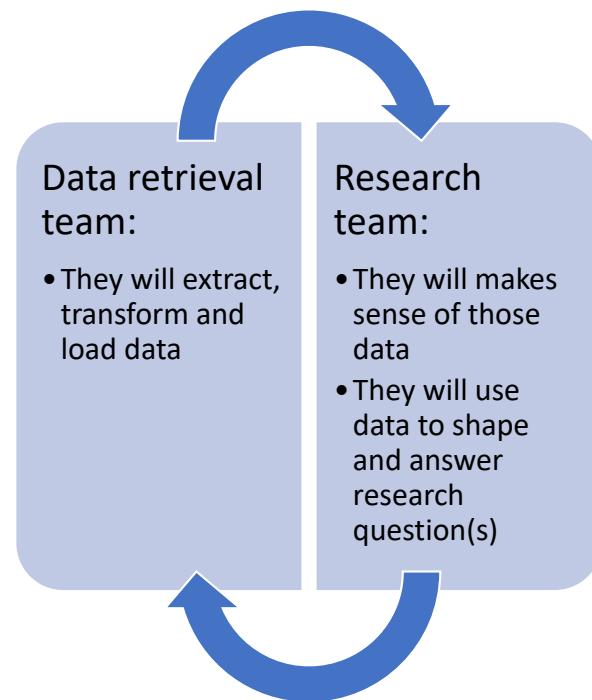


Linking of Data sources

- Will you be able to link patients, clinicians, activities across data sources?
- For the opioid research question may have to do fuzzy matching (e.g. name, zip, phone, email)



Resources Needed



Technical resources/roles

Essential members/roles of the team

- Statistical
- Programming
- Database
 -  Other department's database resources
- Information technology(IT)
- Domain expertise
 - Clinicians(that understand the opioid related data)
 - Nurses
 - Physicians
 - Researchers



Hardware

- Storage
 - Question: Will you need to store data vs metadata?
 - Local vs cloud storage
- Processing Power
 - Tip: Unit testing can help you forecast.
 - For example, after database optimization (e.g. indexing) initial 100K rows takes 6 seconds to search. With same configuration 1million rows will take a minute.
 - For the opioid research question you may want to pull all medications and nursing assessments. That will be a large sum of data.
- There should be a team member that understands how your software utilizes RAM and CPUs (i.e IT person).

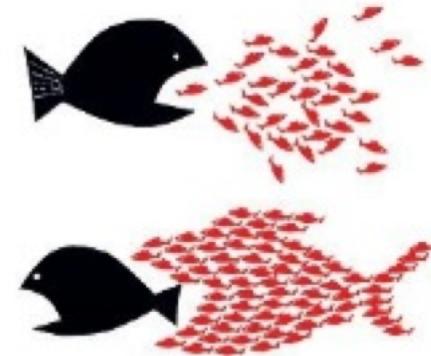
Software

- Do we have the appropriate analytical tools
- Consider:
 - Data types
 - Data volume
 - Visualizations
 - Required analysis
- Often used tools: R studio, R Packages, Python Packages, Tableau, Amazon Web Services(AWS) software



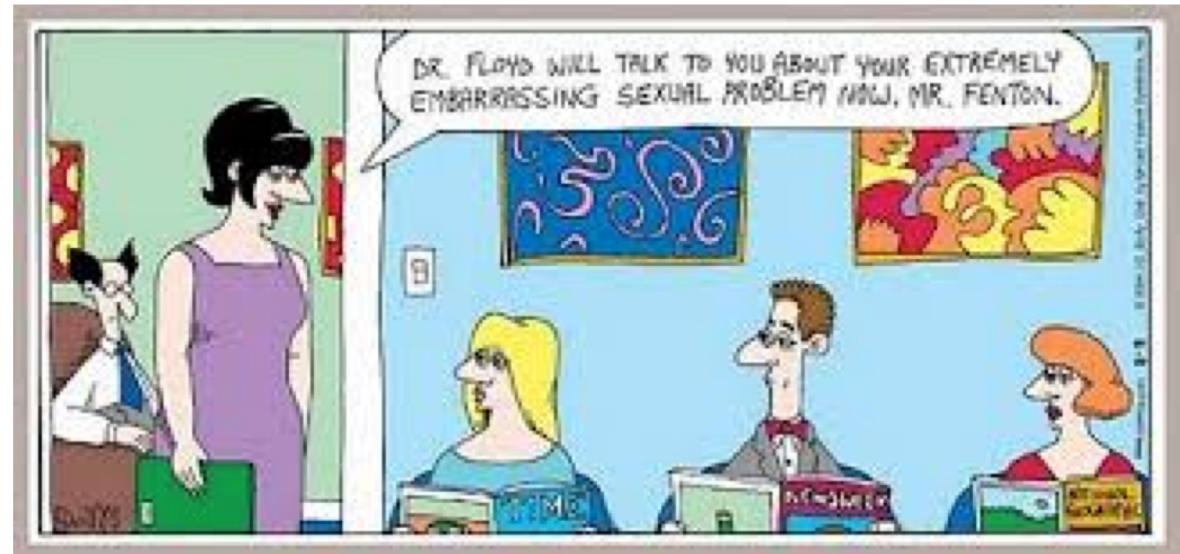
Ethical concerns- IRB

- Knowledge gap of traditional research and data science
 - e.g. cohort definition might need to be well thought out
- Sample vs entire population
 - Questions regarding sampling method



HIPAA

- Will you be able to keep your data secure?



Questions

