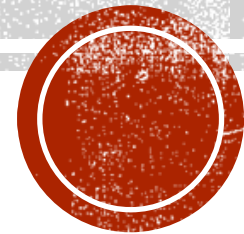


# Database Research

Drowning in Information, but starved for Knowledge. John Naisbitt

Michael DeDonno, Ph.D., MBA

[www.michaeldedonno.com](http://www.michaeldedonno.com)





A

G

E

N

D

A

By the end of this session, you should be able to:

- Define and differentiate database research from research databases
- Understand the value of database research
- Recognize the general steps in a database research study
- Explore a popular database
- Get started in a successful database research project

# What is Database Research?

- **Research Databases** – Subject-specific, typically limited to scholarly journals, books, proceedings, etc.
  - PsychInfo, MEDLINE, Google Scholar
- **Search Engines** – Search across the Internet (the good, bad, and the ugly)
  - Google
- **Database Research** – The use of existing data (secondary data) to investigate a topic of interest.
  - Dataset(s) of data containing specific variables



# What's the Value of Database Research?

- Secondary datasets provide a mechanism for researchers to address high impact questions that would otherwise be expensive and time-consuming to study.
  - Typically less time and resource needs than required for most studies involving primary data collection. No time spent collecting data, not restricted by data collection periods.
- For fellows, residents and junior faculty who need to demonstrate productivity in a timely manner, secondary data can be a key foundation to success.

Smith, Alexander K., et al. "Conducting high-value secondary dataset analysis: An introductory guide and resources." *Journal of general internal medicine* 26.8 (2011): 920-929.



# What's the Value of Database Research? (cont.)

## Surgical Research Using National Databases

Alluri, Leland & Heckmann, 2016

**Abstract:** Recent changes in healthcare and advances in technology have increased the use of large-volume national databases in surgical research.... The use of large-volume national databases is increasing in popularity, and thorough understanding of these databases will allow for a more sophisticated and better educated interpretation of studies that utilize such databases.



# Are Researchers Publishing via Database Research?

Google Scholar (Since 2020)

- National Health & Nutrition Examination Survey (NHANES): 29,300
- National Trauma Databank: 15,900
- National Electronic Injury Surveillance System (NEISS): 14,700

*Note: Results represent number of hits on the Internet and include duplicates and references, but not actual publications.*



# What are General Steps in a Database Research Study?

- Consider topic of interest
- Find an existing dataset
- Study the variables within the dataset
- Generate a research question based on the variables available
- Import data into a statistical software package
- Clean and prepare the data
- Conduct analyses
- Generate output
- Share findings (poster, presentation, manuscript)

*Notice there is  
no reference  
to obtaining  
IRB approval!*



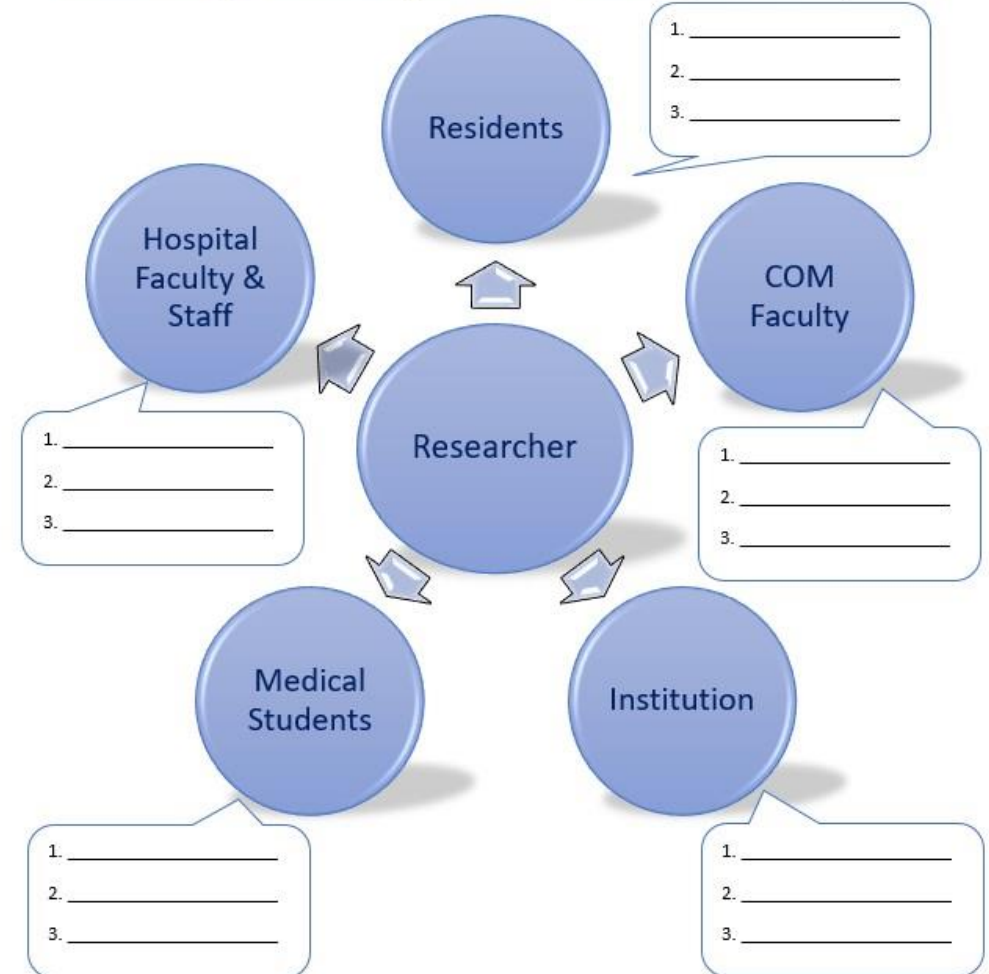
# How do I Get Started

## ➤ Collaborate!!



## Collaborate360

Collaborate360 - A tool that can increase the breadth of knowledge and understanding of different approaches to answering a research question.



Consider collaborations with College of Medicine faculty, hospital faculty, residents, medical students, and institution faculty in other colleges such as the College of Education, College of Nursing, College of Engineering.

# How do I Get Started? (cont.)

- Review articles available at [www.michaeldedonno.com](http://www.michaeldedonno.com)
- Search for databases in your field of interest
  - National Trauma Data Bank (NTDB)
  - National Center for Health Statistics (NCHS)
  - FAU COM STAR Database
- Investigate any potential costs associated with database access
  - Costs may vary based on the age of the dataset
- Explore variables contained in database
  - Data dictionary



# How do I Get Started? (cont.)

14,800 results

- Explore recent publications based on your interests and the data in the database
  - e.g., Google Scholar – Since 2020, NHANES GI

Google Scholar nhanes gi

Articles About 14,800 results (0.29 sec)

Any time  
Since 2024  
Since 2023  
Since 2020  
Custom range...

Sort by relevance  
Sort by date

Any type  
Review articles

include patents  
 include citations

Create alert

[HTML] Bowel health in US shift workers: insights from a cross-sectional study (NHANES)  
MA Storz, M Lombardo, G Rizzo, A Müller... - International Journal of ..., 2022 - mdpi.com  
... It is now widely accepted that the **gastrointestinal (GI)** tract has a circadian rhythm that influences bowel movements, secretion of gastric juices, and appetite regulation [8,9,10]. ...  
☆ Save 📄 Cite Cited by 6 Related articles All 12 versions 🔗

[HTML] The associations between dietary fibers intake and systemic immune and inflammatory biomarkers, a multi-cycle study of NHANES 2015–2020  
X Qi, Y Li, C Fang, Y Jia, M Chen, X Chen... - Frontiers in Nutrition, 2023 - ncbi.nlm.nih.gov  
... the epithelial cells and immune cells in the **gastrointestinal** tract. In vitro studies show dietary ... A cross-section study based on **NHANES** data conducted in adults indicates that high-fiber ...  
☆ Save 📄 Cite Cited by 2 Related articles All 6 versions 🔗

[HTML] Systemic immune-inflammation index is associated with hepatic steatosis: Evidence from NHANES 2015-2018  
Y Song, W Guo, Z Li, D Guo, Z Li, Y Li - Frontiers in Immunology, 2022 - frontiersin.org  
... Among individuals from the 2011–2016 **NHANES**, a cross-sectional study discovered a statistically significant correlation between SII and higher BMI (17). Other **NHANES** studies ...  
☆ Save 📄 Cite Cited by 23 Related articles All 5 versions 🔗

# How do I Get Started? (cont.)

- Use recent publications to serve as direction for your research question(s)
  - Add a variable, a novel methodology, a unique sample, also future research...

## Example

Anker & Krill (2021). Stress, drink, leave: An examination of gender-specific risk factors for mental health problems and attrition among licensed attorneys *PLoS ONE*.

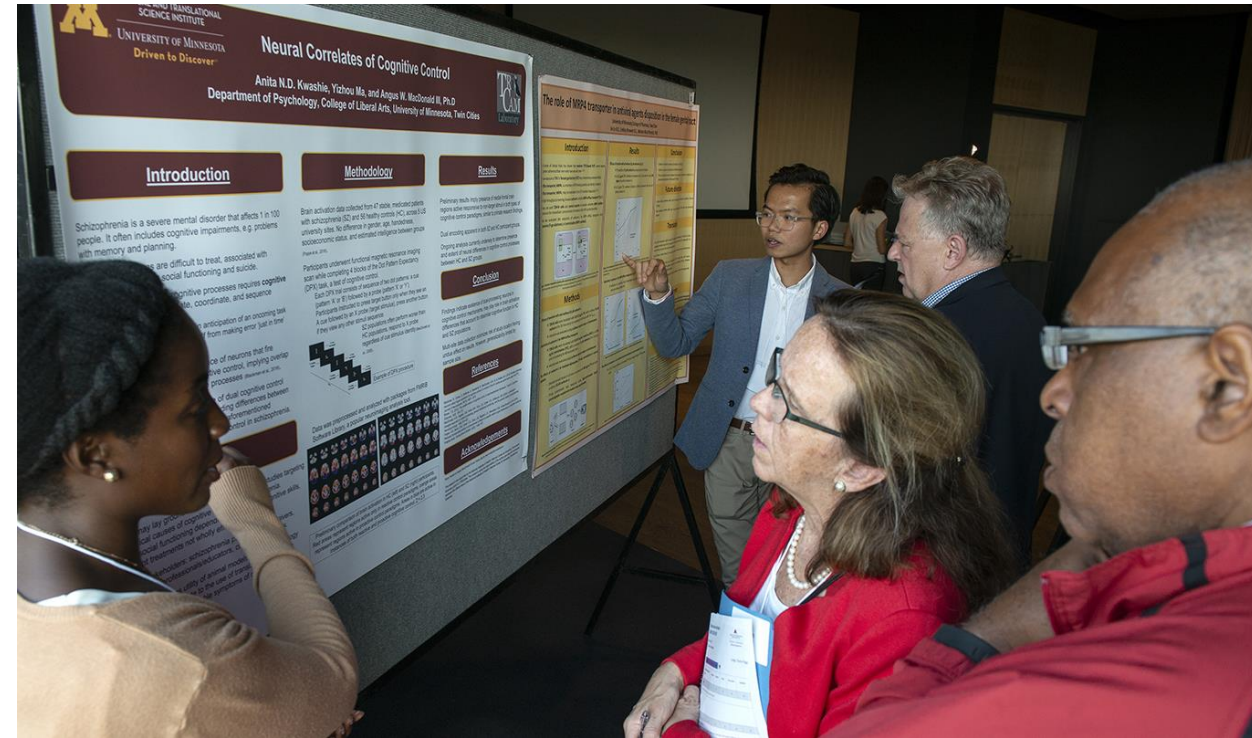
Rates of mental illness and heavy alcohol use are exceedingly high in the legal profession, while attrition among women has also been a longstanding problem.... Findings indicated prevalence and severity of depression, anxiety, stress, and risky/hazardous drinking were significantly higher among women. Further, **one-quarter of all women contemplated leaving the profession due to mental health concerns**, compared to 17% of men...

What differentiates the 24% from the 76% of women attorneys?



# How do I Get Started? (cont.)

- Access and download the data
  - Follow detailed steps
  - Familiar program (e.g., Excel)
  - Be sure to download the data dictionary
- Data cleaning
  - Labeling values (e.g., 1= Strongly disagree)
  - Decide what to do with missing values
- Analysis
  - Consider working with a statistician
- Write-up
  - Poster → Manuscript





## The Good

- No IRB
- No data collection
- Numerous examples of Method section
- Can be done anywhere and at anytime
- Great flexibility in time and effort

## The Bad

- Takes time to learn how to maneuver within the dataset and generate exports
- Takes time to learn the variables
- Limited to what is in the dataset(s)
- Need to have proficiency in exporting, importing, labeling, and possibly computing new variables



SEQN	HSD010	HSQ500	HSQ510	HSQ520	HSQ571	HSQ580	HSQ590	HSAQUEX
93703	.	.	.	.	.	.	.	1
93704	.	2	2	2	.	.	.	1
93705	3	2	2	2	2	.	1	2
93706	2	2	2	2	1	5	2	2
93707	3	2	2	2	.	.	.	2
93708	3	2	2	2	2	.	2	2
93709	.	.	.	.	.	.	.	2
93710	.	.	.	.	.	.	.	.
93711	2	2	2	2	2	.	2	2
93712	3	2	2	2	2	.	2	2
93713	2	2	2	2	2	.	2	2
93714	3	2	2	2	2	.	2	2
93715	4	2	2	2	2	.	1	2
93716	2	2	2	2	2	.	2	2
93717	2	1	2	1	2	.	1	2
93718	3	2	2	2	2	.	9	2
93719	1	2	2	2	.	.	.	2
93720	.	1	2	1	.	.	.	1
93721	3	2	2	2	2	.	2	2
93722	3	2	2	2	2	.	2	2
93723	2	2	2	2	2	.	2	2
93724	.	2	2	2	.	.	.	1
93725	.	.	.	.	.	.	.	1
93726	2	2	2	2	2	.	1	2
93727	1	2	2	2	2	.	2	2

The Ugly

# An example NHANES

- [www.michaeldedonno.com](http://www.michaeldedonno.com)
- Select Database Research
- Scroll to the bottom
- Select NHANES
  - Select Questionnaires, Datasets, and Related Documentation
  - Readily available datasets from 1999-2016
  - Select NHANES 2017-2018
  - Select Questionnaire Data
  - Select Hospital Utilization & Access to Care (HUQ\_I.Doc)
  - Review General Health Condition Question



Check out the NHANES story

[https://youtu.be/EXXTQlju\\_Nc](https://youtu.be/EXXTQlju_Nc)



# Presentation Summary

- Dealing with large datasets can be intimidating
- Documentation may be limited or difficult to understand

However,

- Existing data eliminates time consuming data collection, risk of low response rates
- Typically, the data is readily available with no need to complete formal requests to receive the data
- Typically, no need for IRB applications
- Database research can be done at the researchers own pace



# Start a Database Research Project

- Follow the Getting Started slides
- Consider creating a Database Research Special Interest Group (DB-SIG)
- Contact me to learn more about how database research can be of value to your research endeavors



# References

1. Muñoz E, Muñoz W 3rd, Wise L. National and surgical health care expenditures, 2005-2025. *Ann Surg* 2010;251:195-200.
2. Alluri, R. K., Leland, H., and Heckmann, N. "Surgical research using national databases." *Annals of translational medicine* 4.20 (2016).
3. Clements MA, Foster NC, Maahs DM, et al. Hemoglobin. A1c (HbA1c) changes over time among adolescent and young adult participants in the T1D exchange clinic registry. *Pediatr Diabetes* 2016;17:327-36.
4. Cafri G, Paxton EW, Love R, et al. Is There a Difference in Revision Risk Between Metal and Ceramic Heads on Highly Crosslinked Polyethylene Liners? *Clin Orthop Relat Res* 2016. [Epub ahead of print].

