

# Large Databases: Pros, Cons, and Limitations

Joachim H. Ix, MD, MAS, FASN

Professor of Medicine

Chief; Division of Nephrology-Hypertension

University of California San Diego



# Why consider work in large datasets?

- Think broadly about your research topic.
- Your first paper will probably not be the definitive RCT.
- The perfect study is likely many years away.
- The academic “credit” for enrolling patients into someone else’s study, or writing review papers summarizing the work of others is less than original research.
- Building a research program is a series of small steps. Observational work in large datasets can:
  - Determine the prevalence of the condition of interest.
  - Define the importance of the topic, for example by linking the disease to unique health outcomes.
  - Identify potential therapeutic targets.

# Why consider work in large datasets?

- Research productivity
- Collaboration and mentorship
- Statistical power
- High quality data
- Potential for ancillary studies

# Research Productivity

- Perhaps the single most important component of building a successful research career is early productivity.
- Using existing data that can answer your question(s) is much more efficient than collecting data first-hand.
- Often one question leads to another.
- Projects have natural ebbs and tides, so observational work may compliment productivity with other ongoing projects.

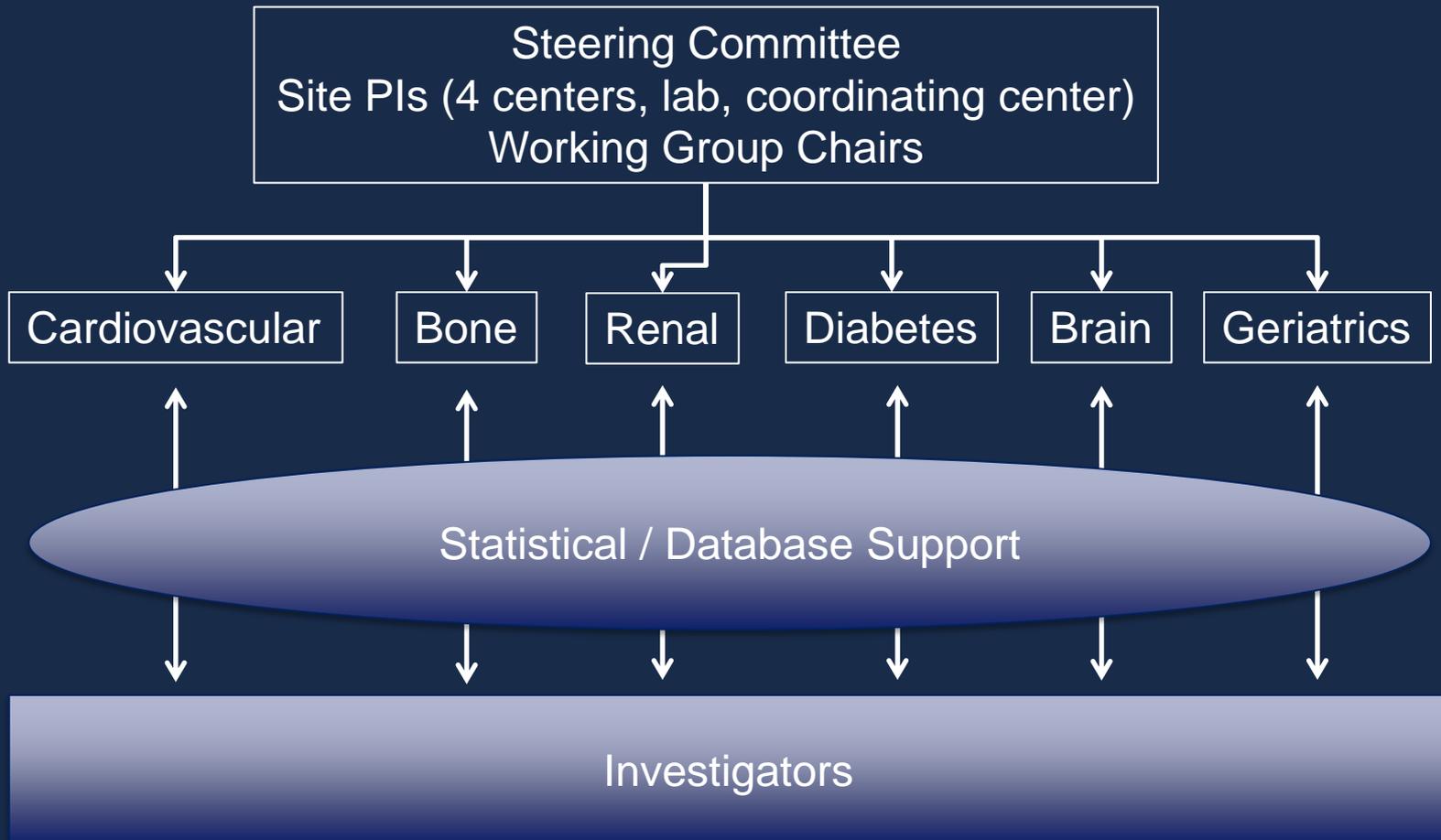
# Collaboration and Mentorship

- Senior leadership promotes engagement of new investigators.
- Many large epidemiologic studies have established, productive, topic-oriented working groups that meet monthly by teleconference.
- Not uncommonly, statistical support is available for new investigators, free of charge.
- Outstanding environment to get advice on ideas, papers, and career advancement.

# Case Study: The Cardiovascular Health Study

- 5,888 persons aged > 65 years from 4 US communities enrolled in 1989-90.
- Major focus was on utility of risk factors on cardiovascular disease endpoints in “real life”.
- All CVD events were adjudicated by panel of experts.
- Multiple follow-up exams to see participants again and repeat blood and urine (and other) measures.
- About 800 participants are still living, so large number of events available.

# Cardiovascular Health Study - Structure



# Some Success Stories – Renal Working Group



## *The* NEW ENGLAND JOURNAL *of* MEDICINE

ESTABLISHED IN 1812

MAY 19, 2005

VOL. 352 NO. 20

### Cystatin C and the Risk of Death and Cardiovascular Events among Elderly Persons

Michael G. Shlipak, M.D., M.P.H., Mark J. Sarnak, M.D., Ronit Katz, Ph.D., Linda F. Fried, M.D., M.P.H.,  
Stephen L. Seliger, M.D., Anne B. Newman, M.D., M.P.H., David S. Siscovick, M.D., M.P.H.,  
and Catherine Stehman-Breen, M.D.

*The* NEW ENGLAND JOURNAL *of* MEDICINE

### ORIGINAL ARTICLE

### Cystatin C versus Creatinine in Determining Risk Based on Kidney Function

Michael G. Shlipak, M.D., M.P.H., Kunihiro Matsushita, M.D., Ph.D.,  
Johan Ärnlöv, M.D., Ph.D., Lesley A. Inker, M.D., Ronit Katz, D.Phil.,  
Kevan R. Polkinghorne, F.R.A.C.P., M.Clin.Epi., Ph.D.,  
Dietrich Rothenbacher, M.D., M.P.H., Mark J. Sarnak, M.D.,  
Brad C. Astor, Ph.D., M.P.H., Josef Coresh, M.D., Ph.D., Andrew S. Levey, M.D.,  
and Ron T. Gansevoort, M.D., Ph.D., for the CKD Prognosis Consortium\*

# Some Success Stories – Renal Working Group



## **Epidemiology/Population**

### **Systolic and Diastolic Blood Pressure, Incident Cardiovascular Events, and Death in Elderly Persons The Role of Functional Limitation in the Cardiovascular Health Study**

Carmen A. Peralta, Ronit Katz, Anne B. Newman, Bruce M. Psaty, Michelle C. Odden

Hypertension, 2014; 64: 472-80

- PI of R01AG046206
- Leading efforts to see remaining CHS participants in a follow-up visit.
- Pending R21.
- Associate Professor

# Some Success Stories – Renal Working Group



American Journal of Epidemiology  
© The Author 2014. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Vol. 180, No. 1  
DOI: 10.1093/aje/kwu102  
Advance Access publication:  
May 20, 2014

## Original Contribution

### **Kidney Function and Cognitive Health in Older Adults: The Cardiovascular Health Study**

**Brendan Darsie\*, Michael G. Shlipak, Mark J. Sarnak, Ronit Katz, Annette L. Fitzpatrick, and Michelle C. Odden**

Three academic generations of mentoring and career development.

# Some Success Stories – Diabetes Working Group



David Siscovick



K. Mukamal



J. Kizer



L. Djousse



S. Ziemann



J. Ix

- R01HL09455
- Funded 49 peer reviewed papers.
- R01AG053325 grant funded 9/2016

# Some Success Stories – Brain Working Group



## **Association Between Left Atrial Abnormality on ECG and Vascular Brain Injury on MRI in the Cardiovascular Health Study**

Hooman Kamel, MD; Traci M. Bartz, MS; W.T. Longstreth Jr, MD, MPH; Peter M. Okin, MD; Evan L. Thacker, PhD; Kristen K. Patton, MD; Phyllis K. Stein, PhD; Rebecca F. Gottesman, MD, PhD; Susan R. Heckbert, MD, PhD; Richard A. Kronmal, PhD; Mitchell S.V. Elkind, MD, MS; Elsayed Z. Soliman, MD, MS

Stroke, 2015; 46: 711-16

- R01NS097443: Race Differences in LA Abnormalities on ECG and contribution to Black / White differences in stroke risk.



- Funded clinical trial evaluating anticoagulation in patients with LA abnormalities (but without afib) for stroke prevention.

# Other Studies with Similar Structures

- Health Aging and Body Composition (Health ABC):
  - 3075 adults aged 70-79 recruited between 1997-98.
  - Detailed measures of glucose metabolism & body composition (OGTT, DEXA, CT), strength (grip strength, sit-stand-sit), at baseline. Most available annually for 10 years.
- Study of Osteoporosis in Men (MrOs)
  - 5994 men aged > 65 years recruited from 6 centers between 2000 and 2002.
  - Main focus is on bone disease.
  - Multiple subsequent visits.
    - Ancillary studies on prostate disease, sleep, dental health, and others.
- Multi-Ethnic Study of Atherosclerosis (MESA)
  - 6814 persons 45-84 years, 6 centers, multi-ethnic (white, black, Hispanic, Chinese).
  - Excluded prevalent CVD.
  - Extensive measures of subclinical CVD (CT for CAC, FMD, carotid U/S, cardiac MRI).
  - Multiple subsequent visits, visit 6 occurring now.
- Many others

# Papers, Ancillary Studies, Grants

- Papers:
  - Use data that is already available within the cohort, thus not requiring any participant contact or use of biospecimens.
  - Usually the best way to start to learn about the available data, other investigators, and structure.
- Ancillary Studies:
  - A more extensive program of research within the cohort.
  - May require contact with participants (ex. a new questionnaire, an additional clinical measure at the exam), or use of the biorepository (measurement of a new biomarker, etc.).
  - Often require extramural funding, and approval is required first, principally to avoid participant burden or depletion of the repository.
- Grants:
  - Are actively encouraged.
  - Typically are submitted to support an (already approved) ancillary study.
  - Useful, when possible, to use centers and central statisticians to help support the infrastructure.

# Cons: Downsides to Using Large Datasets

- Formal processes take some time (paper proposal approval, internal review, NIH sign-off).
- Difficult to test interventions.
- Limited by available measures.
- Lack of sufficient sub-samples with the spectrum of disease of interest.
- Temporal trends in medicine change over time.
- Major focus on conserving the biorepository.
- Concern about sharing your ideas. Typically overblown.

# A few words about etiquette

- Almost any data you read about is within reach. However, how you approach investigators is important. Leverage the help of colleagues if possible.
- Be respectful for others' time and contributions they've made for generating these resources.
- Don't expect to co-author every project.
- Need to respect intellectual property and ideas of others.
- You need to be okay to share your ideas freely, suppressing the concern of being "scooped".
  - Sometimes people really are already doing what you had hoped to do.
- Make a commitment to join conference calls and share your expertise to help others.
- If you volunteer, follow through.

# Summary

- Using large existing datasets is an efficient way to answer important questions, particularly early in your career.
- Datasets are generally eager to make data available and to assist junior investigators.
- An existing dataset that can answer (or help answer) your research hypothesis probably exists, you just need to find it.
- The main cost of the high quality data is additional administrative steps. It is well worth it.

# Questions?

- [joeix@ucsd.edu](mailto:joeix@ucsd.edu)