

Socio-technical Data Analytics (SODA)

Catherine Blake, Ph.D

GRADUATE SCHOOL OF LIBRARY AND
INFORMATION SCIENCE

The  School AT ILLINOIS

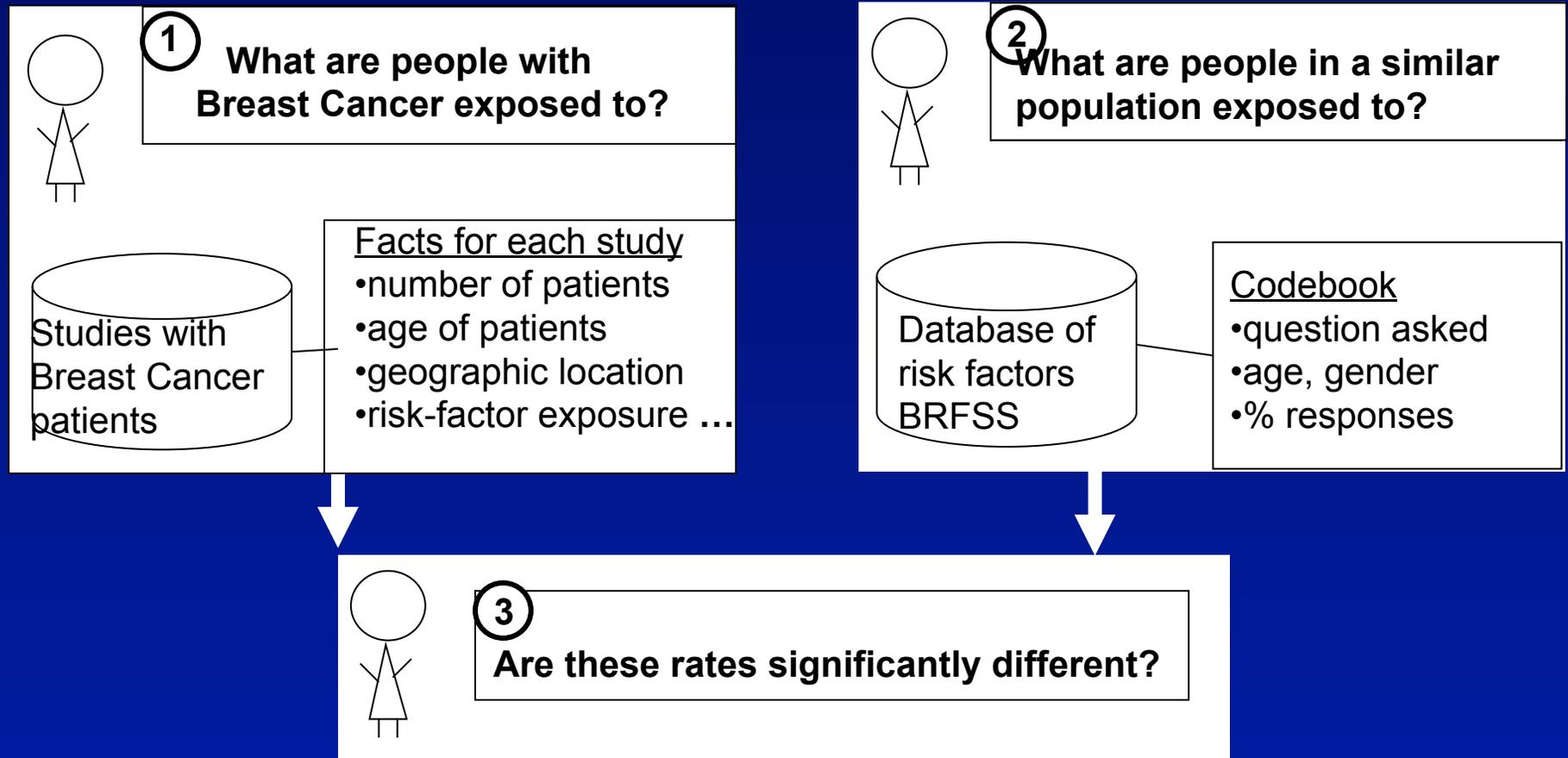
GSLIS Research Showcase, 2014



Socio-technical Data Analytics

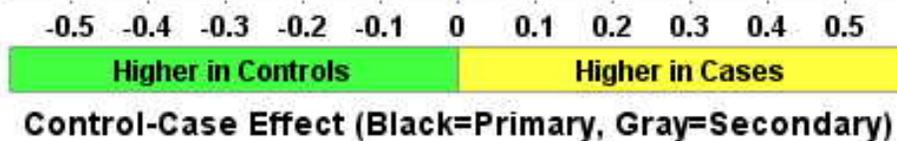
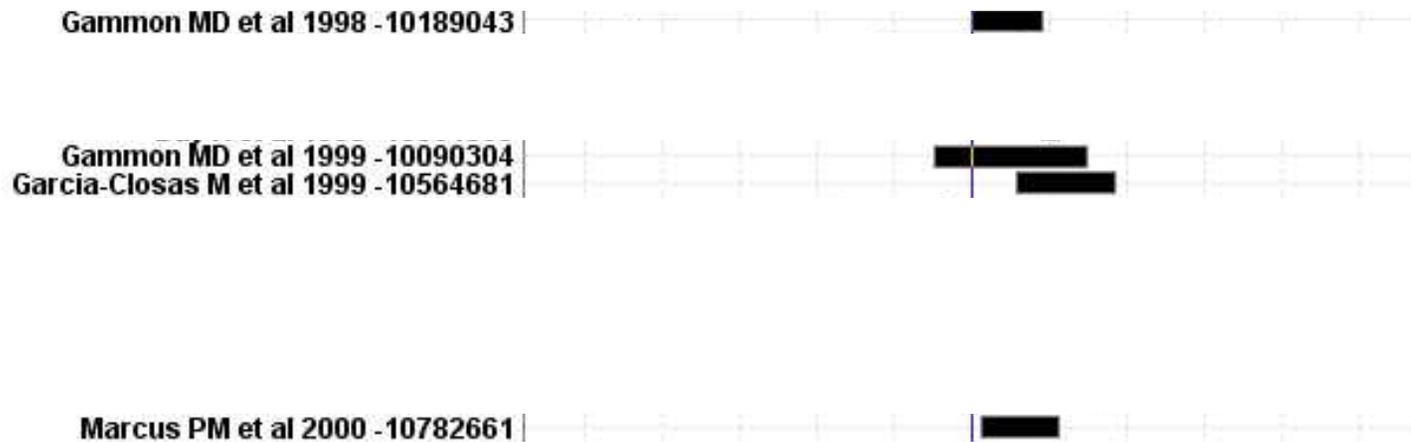
- Two core components:
 - Social (economic, ethical, policy, and political)
 - Technical (databases, social networking, data mining, text mining)
- Train leaders at both the MS and PhD levels
 - SODA curriculum
 - hands-on experience necessary to navigate both social and technical challenges involved with real data
 - Share case studies with LIS community
 - Attract diverse, top-level students
- Promote SODA best practice

Breast Cancer Example



Ever-Smoking Estimated Effect Size 0.013 more in Cases than Controls
Cases 0.456 versus Controls 0.443 (35 articles, n=42055)

Article



Course Examples

- Cultural informatics
 - Cultural uses of negation using in the New York Times and Xinhua News
 - Social roles in fan-translated Japanese visual novels
- Medical Informatics
 - Children with Asthma in east St Louis visit hospitals in Missouri → Illinois data does not capture this population
 - Using local data to understand possible connection between ozone and asthma

SODA can fill this gap

- Industry
 - “The United States alone faces a shortage of 140,000 to 190,000 people with deep analytics skills as well as 1.5 million managers and analysts to analyze big data and make decisions based on their findings” (Manyika et al., 2011, pg 104).
- Academia
 - “gap between the potential value of analytics and the actual value achieved” (Kohavi et al, 2004).
 - “careful consideration of the basic factors that govern how data is generated (and collected) can lead to significantly more accurate predictions.” (Smyth & Elkan, 2010).
 - data preparation stage of the knowledge discovery process typically requires up to 60% of the effort (Cabena, 1998)