Data Sharing in an Integrated Health Delivery System

Stephen K. Van Den Eeden, PhD
Division of Research
Kaiser Permanente Northern California
Oakland, CA

Presented at:
Data Sharing Strategies for Environmental Health Science Research Workshop
NIEHS
7 February 2012
Overview

• Describe basic relationship between researchers and data in an integrated system.

• Working examples
  – BEST – biomonitoring study
  – Autism studies
  – Research Program on Genes, Environment & Health (RPGEH)

• Opportunities and Challenges
Why Talk about Integrated Health Systems?

• KP alone represents the largest defined healthcare population covered by a single electronic health record in the US - >8.2 million. HMORN has > 12 million members.

• Approximately 20% of US population in integrated systems (VAs, KP, GHC, HFHS, GHS, etc.)
  – Expected to rise
    • health care reform
    • less integrated systems will exist as ‘local’ EHRs expand data collection and merge with others to create ‘networks’.
Admin data: name, address, medical record number
Inpatient data
Outpatient data
Pathology data
Laboratory data
Biospecimens
Radiology
Mortality Data
Cancer Registry
Outside KPNC data
Outside specific biospecimens and/or data
Held by numerous investigators
Environmental Data
Networks creating data resources to address a wide range of research questions

Cancer, cardiovascular disease, mental health, others
BEST Study

• Collaboration of Biomonitoring California (BC) & Kaiser Permanente N. Calif. (KPNC)
• Conducting biomonitoring project in Central Valley of California
• Recruitment, contact, consent & sample ascertained by KP research staff
BEST Study

• Individuals consented to complete questionnaire on exposures and provide blood and urine specimens.
• Provides permission to share contact info & biospecimens with BC
• Biospecimens split between groups.
• Joint effort to return results to participants
Autism Portfolio
Lisa Croen, PhD
Autism Studies

• Multi-site studies
• One is ‘virtual’ – creating a database at each center than is uniform in data elements & can be queried quickly.

• Two share data & contact info
  – one where contact only by individual center
  – Another where coordinating center to conduct future contact & share data
RPGEH

• To build a comprehensive resource for research on genetic and environmental influences on health and disease.

• The resource links:
  – Clinical data from detailed EHR & legacy clinical data systems
  – Questionnaire data (410,000)
  – Biospecimens (190,000 w/ DNA, 8,000 w/serum)
RPGEH GERA Cohort

• ARRA GO Grant to KPNC and UCSF
• 118,226 adults
  – Genetic data – genome-wide scan on custom arrays
  – Telomere length – in lab of E. Blackburn
  – Linkage to:
    • EHR for clinical data
    • Linkage to available geospatial data - partnerships
      – Census data
      – built environment- food, green space, recreational, safety
      – chemical/toxin/air pollution – air quality, air toxins, hazardous chemicals
      – water quality
Diverse Population

Map of California with various counties labeled. Bars on the right side of the image show the percentage of the population living in poverty for different racial groups, comparing the general population to Kaiser Permanente. The bars indicate that Kaiser Permanente has a lower percentage of the population living in poverty compared to the general population across all racial groups.
Data Sharing / Access

- Access Committee (AC) is established
- AC reviews, prioritizes applications from external scientists for use of the resource
- High priority given to proposals with an internal collaborator
- Policies anticipate outside requests will merit access and support
Structured application & agreement

- Investigator(s) description
- Background & significance
- Aims
- Methods (analytic & statistical)
- Data required – survey, genetic, clinical.
- Timelines
- Funding
- IRB
- Signed agreement
  - Restricted use
  - Limited time
  - No transfer of data
  - Return any ‘new’ data to RPGEH to incorporate into available data
Considerations

• Consent tiered
  – Broad use
  – Original consent did not mention dbGaP
  – GERA cohort reconsented
  – New consent added a dbGaP tier

• Biospecimen use anticipated to be behind firewall (in-house or 3rd party).
• Developing remote access for outside analysis
The Future is (almost) here

• Increasing EHR data
  – Fully integrated systems (VA, KP, GHC, etc.)
  – Networks of providers
• NIH Health Care Systems Research Collaboratory - Coordinating Center (U54)
• RFA-RM-11-021
Opportunities & Challenges in this Type of Setting

• Maintaining confidentiality
  – Individually consented projects
    • CARDIA type multi-site study
    • Single or limited site study
  – IRB considers waiving consent for data only studies

• It’s complicated, the data that is.

• Misuse or misinterpretation of data

• Support –
  – On-going follow-up
  – preparation of materials
  – Analyses

• Large, diverse population

• Residential histories

• Longitudinal & detailed health data