Data linkages
Rationale, governance and conduct

Tim Carey MD MPH
Summer 2016
Outline

• Rationale
• Opportunities
• Governance
• Roadblocks
• Communication
• Future seminars will address important technical issues
Why are we discussing this issue now?

• Greater availability of health related information
  • Biospecimens and genetic information
  • Claims data
  • EMR based data
  • Patient generated information
  • Social Media
  • Commercially generated information (credit report, discount cards, etc)

• Data storage costs dropping

• Technical ability to perform linkage with improved privacy preservation

• Occurring extensively in the business sector
Use cases

• Examine care utilization outside of ACO network after treatment for sepsis
• After selecting patients in a health system using opiates for chronic abdominal pain, examine use of opiates for those patients with other providers
• Starting with an autism registry (greater dx accuracy), examine care utilization in a health system
• Starting with a registry of patients receiving a surgical implant, examine clinical care through EMR and insurance linkage
• Of patients with asthma in a health system, examine correlates with environmental data in their census tract
What is being linked?

- Registry with EMR or claims data
- Primary data with biologic specimens
- EMR data with patient-generated information
  - PRO’s, device information, surveys
- EMR with claims data (Medicare, Commercial, Medicaid)
- EMR data with social media information
- EMR data with purchase information
- EMR or research data with environmental information
  - Pollutant or water quality
  - Area SES derived from geocoding
Feasibility

- Adequacy of linkage will be dependent on number of identifier variables
- Adequacy of linkage will be dependent on the reliability of information collected, often by personnel who the investigator does not supervise
- To what extent are opt-out individuals different from those whose data are available?
  - Limited data is that, similar to primary data collection, individuals who opt-out of linkages are demographically and clinically different, although the magnitude and direction are difficult to predict
How good is good enough?

• Sample size increases with greater linkage success
• Potential bias due to linked and unlinked individuals having different characteristics decreases with greater linkage rate
• Linking multiple data sources may ‘stack’ errors
  • ZIP codes may be poor in source A, 4 digit SS# poor in source B, and DOB problematic in source C
• Above XX%, each increment in linkage success may come at a cost of:
  • More expense billed to analysts
  • Longer time needed
  • Need to use more personal and direct identifiers
  • Greater governance oversight
Bias resulting from linked data

• Linkage will never be complete
  • Often unreported

• Characteristics of inability to link
  • Different spelling of name
  • Missing last 4 digits of SS# (9999)
  • PO box as address or incomplete address
  • Change in ZIP code

• Is the resulting bias one of generalizability to the population or bias between groups in CER?
Relationship of data linkage governance to the IRB common rule

• The Common Rule provides guidance to all HHS and industry human subjects research
• Last revised over 20 years ago
• Draft under revision x2, final rule may be published late 2016
• Secondary data use may be less governed by IRB’s in the future
  • HIPAA oversight
  • Institutional oversight
Common Rule Revisions
Final Rule anticipated Fall 2016

• Greater regulation of “high risk” research
• Less regulation of low risk research
• Special attention to secondary use of biologic specimens
  • Global ‘opt-in’ at time of registration as a patient
• Some uncertainty regarding linkage with other data sources for research
• Current heterogeneity across IRB’s in interpretation
Common rule revisions
Why not a global opt-in or out at the first visit?

• If asked, a minority of patients will opt out of being contacted for research or opt out of secondary use of biologic specimens
• Opt-in may lead to substantial bias
• Significant data management burden of tracking both patients and specimens
• How to reverse an opt-out?
• If 15% of patients opt-out, what bias might that impose on a research result?
• Harm of secondary data use?
• Patients can always refuse to participate in primary data collection
Consent for secondary use of specimens and/or data

• Requirement (present in draft) would kick in 3 years after publication of common rule revision
• Federal DHHS would provide ‘boilerplate’ language for global consent
• Is genetic information an individual identifier similar to name, address or SS#?
• Most of the focus has been on secondary use of biologic specimens
• Rule would only govern future specimen use
• Draft rule would probably require only notification of potential secondary data use, similar to current UNC general consent. Unclear whether op-out would be required.
• HIPAA as main regulator of use of data, since risk to patients is limited to disclosure
Governance

• May be more problematic than the technical issues of linking data sources
• Circumstances of original data collection may prevent linkage without additional consent - some registry information
• Linkage may increase risk of making the individual identifiable
• Concerns regarding breach
• Health system information may raise concerns regarding competitive or embarrassing information
• Who’s at the table?
  • Researchers
  • Regulators
  • Clinicians
  • Administrators
  • Patients/public
Conduct

• Utility of ‘honest broker’ systems and secure storage environments

• While standardized methods will reduce costs, characteristics of linkages among differing sources are likely to be heterogeneous

• Use of as few personal identifiers as possible

• Charge-back to investigators

• Ad-hoc vs ongoing linkages

• Use of informatics tools to avoid risk of de-identification as much as possible
Registries

• Information system that prospectively tracks individual patients or populations of patients over time
• Often condition or treatment based
• May combine diverse data elements from EMR, provider elements, claims data, patient reported outcomes
• May be mandated by FDA, used for QI or research, developed by advocacy groups
• Multiple governance models, conditions of data collection may prohibit linkage without going back to participant. AHRQ has extensive guidance on registry governance
Insurance payers

• Comprehensive record of care for which a charge is made
• Some care utilization may not be recorded (cash for prescription meds)
• ‘Churn’ among insurance
• Considerable effort by PCORI to encourage EMR-claims linkage
• Concern from commercial insurers that their business strategy might be exposed
• Some success, but linkages are generally on a study-by-study basis
  • CMS/Medicare; commercial; Medicaid
Carolinas HealthCare, which runs more than 900 care centers, including hospitals, nursing homes, doctors’ offices, and surgical centers, has begun plugging consumer data on 2 million people into algorithms designed to identify high-risk patients so that doctors can intervene before they get sick. The company purchases the data from brokers who cull public records, store loyalty program transactions, and credit card purchases. (Bloomberg News, July 2014)
Role of the PCORnet Clinical Data Research Networks

- 13 national networks.
  - ~80 ‘nodes’, mostly integrated delivery systems
  - 70M people, diverse population
  - Mid-South: Vanderbilt, Duke, UNC, HSSC (MUSC, Palmeto, others)
- Strong desire to incorporate ‘complete data’ linking EMR with claims data
- RESDAC has agreed to study specific linkages for Medicare
  - Standard costs apply, requires DUA or data re-use agreement
- BCBSNC has a similar agreement, but approval in advance required
- Medicaid discussions ongoing
- Still sorting out the circumstances in which the linkage would be appropriate and the fixed costs of providing the service
- Similar issues regarding Carolinas Collaborative (UNC, HSSC, Duke, WF)
What do patients and the public think of these activities?

• Moderate literature

• In general, patients are supportive of use of data to improve care through QI and research

• When queried, most individuals want to be asked prior to use or linkage

• However, if such permissions would be impractical (which they generally are), most individuals are OK with the research proceeding.

• Greater concerns in the following situations:
  • Use by of sale of data to for-profit entities
  • Linkage with sources such as credit card information or credit reports
Making the case to stakeholders

• Critical to explain that UNCHCS and the UNC Chapel Hill are conducting research and QI activities to enhance science and with respect for patients and their information. Our goal should always be to improve patient care and health.

• Pro-active messaging at the point of care and through communications channels regarding what we have learned and will do.

• Collaboration with networks likely to improve acceptance

• Extreme caution regarding third party use and commodification of information
Bibliography

• Fiore LD, Lavori PW. Integrating randomized comparative effectiveness research with patient care. NEJM 2016; 374: 2152-8
• Willison DJ et al. Consent for use of personal information for health research: do people with potentially stigmatizing health conditions and the general public differ in their opinions? BMC Medical Ethics 2009; 10: 10
• Hill EM et al. “Let’s get the best quality research we can”: public awareness and acceptance of consent use existing data in health research: systematic review and qualitative study. BMC Medical Research Methodology 2013; 13:72.
Thank you

Questions?

Work supported by NIH CTSA UL1TR001111, PCORI and The Duke Endowment