Big Data: Research Ethics, Regulation and the Way Forward

Tia Powell, MD
AAIC
Washington, DC, 2015

- Well-known security flaws
- All applicants for background check 15 years
- Vast trove of data
  - Name, DOB
  - Social security numbers
  - Fingerprints
- 21.5 million individuals’ records hacked
Ethics question:

• How do we devise policies that facilitate the harvest of knowledge from big data while protecting individual privacy, and respecting individual preferences?
What are challenges to finding right answer?

1) inherent tension between easy flow of data for research and restrictions that protect privacy

2) Regulation is blunt instrument
   - Broad concerns about privacy, data use
   - Research context swept into larger problems
   - May lead to wrong answer for research
What’s different in today’s data?

• Big Data:
  – volume
  – storage
  – analysis

• Multiple sources
  – Electronic Medical Record (EMR)
  – Data bases: ADNI, Biobanks, RHIOs
  – Social Networks, Market Data, Websearch
  – Financial, Personnel, Government
How is big data used?

- Multiple purposes
  - Medical/public health research
  - Marketing
  - Insurance
  - Government surveillance (Snowden)
  - Financial risk assessment (credit score)
Uneasy Beginnings: big data, privacy and regulation

- 1997 release of de-identified Massachusetts employee records to facilitate research
- Latanya Sweeney re-identifies using voter registration data
- Presents Governor Weld his personal health record
- Shapes HIPAA legislation
US Protections for PHI: HIPAA and HITECH

- HIPAA 1996:
  - First major federal legislation to specify protections for sharing personal health information
  - De-identification of individual information: removal of 18 specific items
- HITECH: 2009 addition
  - Required reporting of data breaches >500 pts
Security breaches of EMR

- Using required public report of PHI data breaches for > 500 Patients:
- 29 million medical records breached 2010-13
- 6 breaches >1 million records
- Trend is toward more, not fewer breaches
- Most common mode: inappropriate storage of patient data on personal laptop

Liu, JAMA 2015
Could re-identification of individuals within large data set happen today?

• Much more difficult
  – HIPAA precludes use of name, DOB, full zip code, address
  – 2010 HHS study: 2/15,000 patients re-identified

• But still possible
  – Social media: DOB, schools, jobs,
  – SS#: not randomly generated
Participant preferences for data use:

- Vary widely by region, nation
- Many willing to give data for med research
- Strong wish for consent, privacy protections
- Consent varies by:
  - who holds data: hospital, corporation, govt
  - Purpose: cure dx, marketing, surveillance

- Kim, *JAMIA* 2015
General Data Protection Regulation

- Pending EU legislation, hotly debated
- Added restrictions for big data collection
- Without strong exemptions for medical research
- May require re-consent for many legacy repositories
- Possible impediment to international collaboration
- Significant implications for genetic, big data medical research
Scientists should not neglect privacy, security, consent
Global Alliance for Genomics and Health

- International group of researchers, patient advocates, bioethicists, privacy experts
- Goal is to develop best practices for both sharing and protecting research data, especially within genetics
- Publically available consent policies, forms, white papers
Best practices and recommendations for big data and research:

- Build public trust, respect values
- Focus on consent, controlled access, privacy
- Rigorous security
- Distinguish medical research repositories, purpose, from other uses
References and contact info

- Global Alliance for Genomics and Health: genomicsandhealth.org

- Contact Info: Tpowell@montefiore.org