Precision Medicine:
Building a Large U.S. Research Cohort

Workshop Planning Team:

Electronic Health Records and Informatics

Co-chairs: Dan Masys, MD; Rex Chisholm, PhD
Group Members: Josh Denny MD, MS; Doug Fridsma, MD, PhD; Sachin Kheterpal, MD; Lucila Ohno-Machado, MD, PhD

Current Landscape

- Electronic Health Records (EHRs) offer the promise of supporting both discovery science and improved healthcare process and outcome
- Strengths
  - Longitudinal (increasingly, lifelong) observations
  - Document each individual as ‘an experiment of Nature’ (and experiment of personal choices, environment, health decisions and interventions)
  - Now past the ‘tipping point’ in adoption: ~100% of hospitals; 85+% outpatient practices
  - Federal certification process for EHRs includes incentives for interoperability

Challenges and Opportunities

1. Human and business factors: Individual and institutional motivation to participate, and integrated consent management systems
2. Technical issues in integrating and analyzing data from heterogeneous systems
3. Putting patients in control: Enhancing “Blue Button” functionality for research
4. Industry engagement
5. Cybersecurity

Challenge: Human and Business Factors

Creating and Articulating a Compelling Value Proposition for participation by Organizations and Individuals

- Three pathways for access to clinical data:
  - Union of existing NIH-sponsored cohorts
  - New organizational relationships with healthcare entities
  - Direct submission by individuals in the cohort
- Alignment of Incentives
  - Financial incentives small per participant due to cohort size
  - Quid pro quo models that recognize value of returned research information to participating organizations and individuals should be considered
- Federated vs. centralized models of operation affect willingness to participate
- Electronic systems infrastructure for interactive, fine-grained consent is feasible

Challenge: Integrating and Analyzing Data from Heterogeneous Systems

- “Research grade” phenotypes can be extracted from routine clinical data in EHR systems
- Requires both structured data (billing codes, lab values, medications) and analysis of unstructured text (H&P, procedure and discharge summaries, progress notes, etc.)
- Depends critically upon linking data to the correct individual
- Data is rich in features that support re-identification of individuals: no technology-only solution to ensuring privacy
- Required expertise in Natural Language Processing of clinical text is a scarce resource currently

Opportunity: enhancing "Blue Button" functionality for Research

- HIPAA/HITECH gives individuals rights to electronic copies of their EHR data.
- ONC "Blue Button" campaign to encourage individuals to exercise this right, and EHR system builders to implement it
- Currently there are technical specifications for clinical summaries and insurance benefits.
- Vision: a "Synch for Science" (S4S) button that enables an individual to download their clinical data and transmit it to a research data center.
- Getting there: work with ONC to add additional data types and formats to Blue Button, and with EHR vendors to implement

Challenge: Industry Engagement

- A national scale cohort will depend upon engagement and support from commercial EHR vendors
- To date, research has not been a prominent ‘use case’ for EHRs
- Industry engagement needs to be based on:
  - Practical, specific, and certifiable functionality
  - Leveraging existing government-supported EHR requirements
  - Create transparent, objective measures of success
  - Technology-agnostic approaches that do not favor one vendor over another
- Industry goals include customer “delighters”, not just $$

Challenge: Cybersecurity

- Data for research will arise in part as Protected Health Information (PHI) from HIPAA covered entities: highly sensitive
- A national cohort will depend critically upon digital telecommunications via Internet, smartphones, and other network-connected devices
- Cyberattacks on healthcare data are increasing and will remain a persistent threat
- No novel project-specific data and communications security technologies likely to be needed, but maintenance of state-of-the-art cybersecurity will be essential