Data Sharing for Research (& Commercialization?)
The Opportunity and the Problem

Real World Data is now more available than ever
Consortia to share data across institutions are forming
Claims, social determinants, and other data sets can be combined
Non-research data sharing is already widespread
  Revenue Cycle
  ePrescribing
  Purchasing Records
  23&me / Ancestry.com
Researchers follow ethical codes that corporations do not
Sharing for research gets far more scrutiny than commercial sharing
Some Questions

• How do we share data to do good while protecting patients, providers, staff, and others?
• What is compensation to support the mission vs buying data?
• How do operate within our missions without succumbing to commercial incentives that extend beyond our mission?
• How do you value the data from patients in the “ownership” of intellectual property created with that data?
• How do you share value among all the contributors to a data set?
One Proposal

• Develop a set of shared principles
• Try to achieve broad agreement across stakeholders
• Seek a national entity to endorse the principles as standards
• Use the national standards as leverage in negotiations
Core Principles: Alignment with Mission

• Data sharing should always be consistent with the core missions of clinical care, research discovery, and education.
• Payment should be for academic work or invention – not just for data alone.
  – Faculty should be key intellectual contributors in any sharing agreement that involves compensation
• Work should always be bound by national standards
  – E.g. Belmont Report
Core Principles: Limited Contracts

- Sharing, even if de-identified, should be bound by a contract that limits the types of data, data use, and duration of sharing.
- Sharing should only support activities consonant with the missions.
- Recipients must never re-identify deidentified data.
- Association with other data sets must not be done outside of the agreed upon plan within the contract.
- All sharing agreements should be revocable at any time.
Core Principles: Transparency

• All sharing agreements should be transparent to patients and other stakeholders

• Algorithms themselves should be transparent in terms of
  – Intent
  – Value
  – Limitations
  – Context for use
  – Validity
  – Reliability

• All conflicts of interest should be transparent and managed
Core Principles: Limitations on Custody and Scope

• Data sharing should be limited to the minimum necessary to achieve the intended goal
• Data should be de-identified to the extent possible within the clinical or scientific purpose
• Whenever possible, data should not leave control of the original custodian
  – External parties should use a local enclave for their analysis
Core Principles: Oversight

• Data sharing agreements should be overseen by stakeholders with expertise
  – Human subjects research
  – Clinical care
  – Education
  – Patients’ rights
  – Law and regulatory compliance
  – Informatics
  – Medical ethics
Logical Tests

- Can stakeholders agree to these principles?
- Do existing agreements follow these principles?
- Would these principles help partners come together?
- Do the principles create problematic barriers to critical work?
- Could the principles help deflect pressure from mandated monopolies?
Rainu Kaushal, MD, MPH

Nanette Laitman Distinguished Professor
Chair, Healthcare Policy & Research
PCORnet: The National Patient-Centered Clinical Research Network

**PCORnet** is a network of networks with access to secure, curated data from millions of patients across the largest health systems in the United States. Today, the People-Centered Research Foundation (PCRF) is a non-profit organization that serves as the Central Office for PCORnet.

With 66 million patients accessible for observational studies and 30 million accessible for clinical trials, PCORnet offers:

- Access to electronic health record (EHR) data
- Exceptional research teams
- Expertise in integrating research with clinical care
- Streamlined administrative processes
- Partnered with patients

Weill Cornell Medicine
Healthcare Policy & Research
PCORnet: Three Key Assets

**Patient-Partnered**
Patients, and those who love and care for them, are integral to ALL PHASES OF PCORNET-ENABLED RESEARCH.

**Data-Driven**
Data accessible via PCORnet draws from millions of EHRs with growing links to patient-reported and payor data, all consolidated using a common data model.

**Broadly Connected**
PCORnet connects you to thousands of clinicians and researchers who are committed to answering important questions that will improve patient lives.
The PCORnet solution starts with real-world data. Users can access data from everyday medical encounters from more than 66 million people across the United States.
The PCORnet Common Data Model standardizes data into a single language, enabling fast insights, including:

<table>
<thead>
<tr>
<th>Ready for Research</th>
<th>Available, But Still Evolving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death Data</td>
<td>Social Determinants of Health</td>
</tr>
<tr>
<td>Diagnoses</td>
<td>Tumor Registry</td>
</tr>
<tr>
<td>Medication Orders</td>
<td>Biosamples</td>
</tr>
<tr>
<td>Procedures</td>
<td></td>
</tr>
<tr>
<td>Claims</td>
<td>Genomic Results</td>
</tr>
<tr>
<td>Patient-Reported Outcomes</td>
<td>Natural Language Processing Derived Concepts</td>
</tr>
<tr>
<td>Geocodes</td>
<td></td>
</tr>
<tr>
<td>Labs</td>
<td></td>
</tr>
<tr>
<td>Demographics</td>
<td></td>
</tr>
<tr>
<td>Data available from several Clinical Research Networks, in the PCORnet Common Data Model and ready for use in research.</td>
<td></td>
</tr>
<tr>
<td>Data available at some Clinical Research Networks, may or may not be in the PCORnet Common Data Model and require additional work for use in research.</td>
<td></td>
</tr>
</tbody>
</table>
INSIGHT Network: A Collaborative Partnership

Central Database of 12m Unique Patients

Clinical Data
EHR: 300m clinical encounters

Claims Data:
Medicare: 1m patients
Medicaid: 1.4m patients
1199: 400k patients
Partnership with HealthCore
Data and Services

- **Data**
  - Clinical Data (5 health systems)
  - Medicaid and Medicare Data
  - Social Determinants
  - Demographic
  - Bio Specimen data

- **Services**
  - Registry
  - Patient Reported Outcome Measures
  - Administrative efficiency and simplicity
  - Stakeholder Engagement, including patients
  - Multi-Site Trial Support
  - Study Specific Analytical Data Registries
  - Patient Enrollment
  - Project Development and Coordination
  - Linked EHR and Claims Data

---

Weill Cornell Medicine
Healthcare Policy & Research
Data Quality and Security

AQUA sub-committee focused on rules to identify, revise, and correct errors that affect data accuracy, completeness and consistency.

Third party hosting vendor that is HIPAA compliant, and complies with the national security protocols set by the Federal Government. Completes regular security audits.

CDM Expansion: Claims, PROs, Addresses and Immunization Records.

Weill Cornell Medicine
Healthcare Policy & Research
INSIGHT’s Evolution from a De-Identified to a Limited Dataset

De-identified dataset
• Dates shifted
• No zip codes

Limited dataset
• Real appointment, procedure, prescription, diagnoses, birth dates
• Zip codes

Key Success Factors:
• Secure infrastructure
• Established governance and trust
• Data use agreements
• Common Data Model
• Driven by researchers’ needs
Data Sharing for Research (And Commercialization?)

David Vawdrey, PhD
Chief Data Informatics Officer
dvawdrey@geisinger.edu
GEISINGER
Integrated Health System
$8.0B Combined Revenues

We care for patients
- 11 hospital campuses, 253 clinic sites
- 3,000+ providers
- 106K Inpatient, 6.4M Outpatient, 350K Emergency Visits/year
- Electronic health records for 2M individuals

We provide quality, affordable healthcare coverage
- 578,000 members
- 55,000 contracted providers/facilities

We teach, research and innovate
- 523 MBS/MD students; 505 residents/fellows
- 48 GLH School of Nursing, 2,000+ nursing students
- MyCode® 250,000+ participants
- Fresh Food Farmacy, Geisinger @ Home

Note: Combined revenues include intercompany transactions among Geisinger providers and insurers.
MyCode® scorecard

- Total consented participants: 258,121
- Samples provided: 176,330
- DNA sequences available for research: 92,455
- DNA sequences eligible for clinical review: 65,012
- DNA sequences analyzed for clinical relevance: 65,012
- Participants with clinical result reported: 1,490

As of November 1, 2019

2 million Geisinger patients
David –

I hope you’re doing well. I’m not sure if you remember me, but I met with you while I was working at [redacted] a few years back. I was hoping we could set up a few minutes to talk about [redacted] ability to provide unique Social Determinants of Health data? We are just rolling this out to payors and providers, but it’s been very well received and I would appreciate the chance to get your opinion and feedback. There is a long list of data elements that we can provide to enhance your analytic capabilities to support VBC and Population Health, but a few include:

- Court and Judgement data
- Criminal history
- Properties
- Assets
- Motor Vehicle Records
- Voter data
- Social Media
- Liens
“We at Apple believe that privacy is a fundamental human right,” Apple’s CEO, Tim Cook, said in a privacy-conference keynote last year in Brussels. “But we also recognize that not everyone sees things as we do.” Cook was making an impassioned plea to end the technology industry’s collection and sale of user data. “This is surveillance,” he continued. “And these stockpiles of personal data serve only to enrich the companies that collect them.” Cook called for a comprehensive U.S. data-privacy law focused on minimizing data collection, securing that data, and informing users about its nature and use.
The EU General Data Protection Regulation (GDPR) is the most important change in data privacy regulation in 20 years. The regulation will fundamentally reshape the way in which data is handled across every sector, from healthcare to banking and beyond.
Mortal! I come offering a deal—

Read the sign.

"By entering this room, you agree to forfeit your own soul rather than negotiate with the mortal residing therein..."

Wait, you can't--

Too late.

Mephistopheles encounters the EULA.
“It’s a good bet that the fine print of the consent form you signed before your latest test or operation said that all the data or tissue samples belong to the doctor or institution performing it. They can study it, sell it or do whatever they want with it, without notifying or compensating you, although the data must be depersonalized in their best effort to make sure you are anonymous.”

“In 2017, the legal buying, selling and trading of our personal medical data totaled $14 billion.”

“This is all perfectly legal, and courts have upheld it even if you didn’t quite know what you were agreeing to.”

Opinion

LETTERS

Sharing the Profits From Cell Research

Jan. 11, 2016

To the Editor:

Re “Your Cells, Their Research, Your Permission?,” by Rebecca Skloot (Op-Ed, Dec. 30):

Both institutions and individuals profited greatly from replication of the HeLa cell line. The family of Henrietta Lacks, whose cancer cells were taken and studied without her knowledge, did not. If ethics dictate that they should not, why should others profit?

Given that patients “own” their cells and tissue, I agree with Ms. Skloot that they should at a minimum be informed that others may profit from the research or products that derive from their tissue as part of the consent process.

GILBERT LITALIEN

Deep River, Conn.

The writer is a researcher for a pharmaceutical company.
Data Sharing for Research (And Commercialization?)

David Vawdrey, PhD
Chief Data Informatics Officer
dvawdrey@geisinger.edu