Big Data to Knowledge: Much at Stake for FDA and its Constituents

Robert M Califf MD
Commissioner of Food and Drugs
NIH Big Data to Knowledge
November 29th, 2016
Bottom Line

• The FDA makes decisions
  – When the data and information high quality, things go well
  – When decisions must be made with poor data....

• Feels like a transition point in information that can inform decisions
  – Precision decisions for individual patients
  – Choices by consumers
  – Policy decisions

• The technology is moving quickly
• The duration of transition vs reality will depend upon culture and workforce
FDA Regulates a Spectrum of Health Products: 20-25 cents of every GDP dollar
FDA Mission

FDA is responsible for protecting the public health by assuring the safety, efficacy and security of human and veterinary drugs, biological products, medical devices, our nation’s food supply, cosmetics, and products that emit radiation.
FDA Mission

FDA also has responsibility for regulating the manufacturing, marketing, and distribution of tobacco products to protect the public health and to reduce tobacco use by minors.
FDA Mission

- FDA is also responsible for **advancing the public health** by helping to speed innovations that make medical products more effective, safer, and more affordable and by **helping the public get the accurate, science-based information** they need to use medical products and foods to maintain and improve their health. FDA also has responsibility for regulating the manufacturing, marketing and distribution of tobacco products to protect the public health and to reduce tobacco use by minors.
FDA Mission

Finally, FDA plays a significant role in the Nation’s counterterrorism capability. FDA fulfills this responsibility by ensuring the security of the food supply and by fostering development of medical products to respond to deliberate and naturally emerging public health threats.
Personal Perspective

- I have witnessed a revolution in cardiovascular care
- During my professional career
  - Insights into public health measures like blood pressure and fundamentals of diet have advanced dramatically
  - Amazing drugs and devices have been developed and deployed
  - We have taken on the menace of tobacco and we’re winning the battle
  - Age specific mortality has been reduced by over 50% with resulting increases in American and global longevity and functional status
The Next Revolution

• Will result from the transformation of information
• In order to apply this fundamental revolution to improve health and quality of life, we must:
  – Learn how to share and create business models that work to improve sharing
  – Invest heavily in curating information
  – Work together to develop social and ethical constructs to deal with privacy, confidentiality and security
  – Create a workforce that can both create new methods and integrate information into practice
  – Give the workforce time to invest in knowledge generation as a routine part of practice
  – Work with the public to gain support and understanding
Precision Therapeutics

• Not a new concept, but new technology is making it happen
• Tremendous opportunity to improve health or to do harm
• Integration of testing, preferences and therapies-how do we know when it is correct?
• Regulatory/public health scheme is challenging
  – Protect public without “stifling innovation”
  – Will require a different ecosystem
  – Data sharing, transparency, standards and quality systems will be key
Possible Functional Definition of Precision Medicine

• Defining and implementing
  – The right strategy, including behavioral and technological (drugs, devices and biologics) interventions
  – At the right time
  – In the right intensity (dose, duration, invasiveness)

• Based on
  – A person’s biology (genes, proteins),
  – Integrative physiology (blood pressure, heart rate)
  – Environmental and social exposures (where people live and with whom they interact)
  – Preferences and beliefs
The U.S. Precision Medicine Initiative
THE PRECISION MEDICINE INITIATIVE® COHORT PROGRAM

• One million or more volunteers, reflecting the broad diversity of the U.S.

• Opportunities for volunteers to provide data on an ongoing basis

• Data shared freely and fast to inform a broad variety of research studies
Patient Partnerships

EHRs

Technologies

Genomics

Data Science
Big Challenges in Biomedicine

- Lack of significant information over the time dimension — Measurements made to assess biology and human health are made periodically in visits to healthcare or research.
- Missing systems biology — When developing concepts of human biology or drug development we make limited measurements focused on specific mechanisms—we’re looking “under the lamppost”.
- Missing the ability to measure the interactions of biology, sociology, environment and decision-making that could enable optimization of individualized and population health — Although we know that health and disease are the product of the interactions of genes, multiple derivative biological systems, environment, social context and personal decisions, we tend to look at one part of the time.
“My hope is that this becomes the foundation, the architecture, whereby in 10 years from now we can look back and say that we have revolutionized medicine.”

—Barack Obama
Many tools to dissect individualized health

- Health records
- Poverty
- Genomics
- Metabolomics
- Images
- mHealth
The challenge: integrating multiple datasets for discovery and implementation.
Moving discovery to practice

"Here's my sequence...

New Yorker, 2000
Precision medicine for the population, and the patient

It is more important to know what sort of person has a disease than to know what sort of disease a person has.

Hippocrates
A Critical Conceptual Issue in Precision Medicine

• Personalized medicine was sold in many quarters as an approach in which deep study of a few people could be generalized to entire populations.

• Precision medicine is evolving with the realization that due to the multidimensional heterogeneity of the problem, in order to devise effective therapies we need deep study of the entire population to find predictive relationships that can be applied to individuals based on what they have in common with others (“patients like us”).
Evidence Based Practice

If we optimize the evidence we can do much better with the other parts!
Generating Evidence to Inform Decisions

1. FDA Critical Path
2. NIH Roadmap
3. Data Standards
4. Network Information
5. Empirical Ethics
6. Priorities and Processes
7. Inclusiveness
8. Use for Feedback on Priorities
9. Conflict of Interest Management
10. Evaluation of Speed and Fluency
11. Pay for Performance
12. Transparency to Consumers

Measurement and Education

Discovery Science

Outcomes

Performance Measures

Early Translational Steps

Clinical Trials

Clinical Practice Guidelines
Which Treatment is Best for Whom?
High-Quality Evidence is Scarce

< 15% of Guideline Recommendations Supported by High Quality Evidence

Scientific Evidence Underlying the ACC/AHA Clinical Practice Guidelines

Context The joint cardiovascular practice guidelines of the American College of Cardiology (ACC) and the American Heart Association (AHA) have become important documents for guiding cardiology practice and establishing benchmarks for quality of care.

Objective To describe the evolution of recommendations in ACC/AHA cardiovascular guidelines and the distribution of recommendations across classes of recommendations and levels of evidence.

Data Sources and Study Selection Data from all ACC/AHA practice guidelines issued from 1984 to September 2008 were abstracted by personnel in the ACC Science and Quality Division. Fifty-three guidelines on 22 topics, including a total of 7196 recommendations, were abstracted.
Fundamental Flaw in System of the Past

• Ability to measure, compare and draw conclusions was limited to an individual lab, clinic or system

• Developing scale could only happen in expensive stand alone systems due to cultural and technical limitations

• Changes in information and data science and computation have allowed us to overcome the technical limits

• Now the limits are cultural!
Fundamental Transformation

Old System
- Generate idea
- Obtain funding
- Build research system
- Conduct research
- Decide whether to disseminate and if so, how
- Tear it down
- Start over again with next project
- Learning (research) is separate from practice

New System
- Develop system
- Allocate funding to best ideas across system
- Use existing resources with common standards and information
- Information accrues through transparency and sharing
- Start next project taking advantage of standing infrastructure
- Learning is a part of practice
Learning health care systems

In a learning health care system, research influences practice and practice influences research.

**EVALUATE**
Collect data and analyze results to show what works and what doesn’t.

**ADJUST**
Use evidence to influence continual improvement.

**IMPLEMENT**
Apply plan in pilot and control settings.

**DESIGN**
Design care and evaluation based on evidence generated here and elsewhere.

**DISSEMINATE**
Share results to improve care for everyone.

**INTERNAL AND EXTERNAL SCAN**
Identify problems and potentially innovative solutions.

Internal

External
Precision Medicine Initiative: Modernizing FDA Regulation of Genomic Laboratory Tests

Traditional testing

Next generation
Modernizing FDA Regulation of Genomics

- Develop and implement **standards** to assure quality
- Develop **open-source tools** to help test developers meet standards (*precisionFDA*)
- Support the development of a **data commons** for evidence on the clinical relevance of genetic variation

Develop and implement an adaptive standards-based regulatory approach
A community platform for NGS assay evaluation and regulatory science exploration.

precisionFDA demonstrates FDA's commitment to innovating the regulatory science needed to advance the growing era of precision medicine

ROBERT CALIFF
FDA
Whole Genome Sequencing Program (WGS)

GenomeTrakr

- State and Federal laboratory network collecting and sharing genomic data from foodborne pathogens
- Distributed sequencing based network
- Partner with NIH
- Open-access genomic reference database
- Can be used to find the contamination sources of current and future outbreaks

http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/default.htm#trakr
Digital Transformation

2010
- Individual Productivity
- IT Silos

2020
- Collective Intelligence
- Distributed Computing

- Data on premise, hard to access, analyze and use
- Productivity tools built for individual, local usage
- IT focusing on where it computes

- Data stored in cloud, simple to query
- Collaborative, cloud based productivity applications
- Machine learning drives deep, actionable insights
- IT changing how it computes
DRUG DISCOVERY AND DEVELOPMENT TIMELINE

Drug Discovery

Preclinical

Clinical Trials

FDA Review

Scale-Up to Mfg.

Post-Marketing Surveillance

Pre-Discovery

~ 5,000 – 10,000 COMPOUNDS

250

5

PHASE 1

IND SUBMITTED

PHASE 2

IND SUBMITTED

PHASE 3

IND SUBMITTED

One FDA-Approved Drug

IND SUBMITTED

NUMBER OF VOLUNTEERS

20-100

100-500

1,000-5,000

3 – 6 YEARS

6 – 7 YEARS

0.5 – 2 YEARS

INDEFINITE

IND SUBMITTED

6 – 7 YEARS

0.5 – 2 YEARS

INDEFINITE

www.fda.gov
Relative Complexity of Therapies

One subunit of a protein
L-tryptophan
Small Molecule Drug
$10^2$ Atoms

Protein composed of about 1100 subunits
IgG antibody molecule
Protein Biologic
$10^5$ Atoms

Cell composed of about 3.6 x $10^6$ proteins
Mesenchymal stem cell
Cellular Biologic
$10^{14}$ Atoms
CRISPR/Cas9 Gene Editing

• Cas9 nuclease can be directed to cut at specific locations designated by guide RNAs

• Though there is some concern for off-target effects, CRISPR/Cas9 is a powerful technique for altering genes
Short Introduction to Genome Editing

Three major forms of genome editing:

– Zinc Finger Nucleases (mid-2000s)
– TALENs (Transcription activator-like effector nucleases) (late 2000s)
– CRISPR (clustered regulatory interspersed short palindromic repeats and associated enzymatic activities (e.g., Cas9) (2011-2012 depending on whom you ask)

Until these three forms of editing, alteration of genomic DNA could control the nature of the change (i.e., sequence-specific alterations), but except for the technically very difficult homologous recombination, neither:

• the specific location (i.e., site-specific alterations), nor
• the exact nature of the change
  – Deletion of specific nucleotides
  – Substitution of nucleotide/s
  – Addition of sequences by insertion at a specific site
National System Paradigm Shift

Passive Surveillance
- Challenging to find right pre/post market balance without confidence in post-market data

Active Surveillance
- To better protect patients

Leverage RWE to support regulatory decisions throughout TPLC

National System
- Embedded in Health Care System (collect data during routine clinical care)
- Shared system to inform the entire Ecosystem (patients, clinicians, providers, payers, FDA, Device Firms)

Parallel track to clinical practice
- Inefficient one-off studies

Current

www.fda.gov
In a learning health care system, research influences practice and practice influences research.

**Evaluate**
Collect data and analyze results to show what works and what doesn’t.

**Implement**
Apply plan in pilot and control settings.

**Design**
Design care and evaluation based on evidence generated here and elsewhere.

**Adjust**
Use evidence to influence continual improvement.

**Disseminate**
Share results to improve care for everyone.

**Internal and External Scan**
Identify problems and potentially innovative solutions.
Historical model of clinical research: Many recruitment sites and a coordinating center

- Hub & spoke model
- Top-down decision-making
- Sites operated independently
Modified Model
Data Shared, Sites owned by Health Systems
Previously Independent Sites now part of large integrated health systems increasingly sophisticated data warehouses
Nodes are Operational Clusters Using Common Data
Drug Surveillance and Trials
Post Market Studies, including comparative effectiveness
Multiple Developing Efforts

- FDA
  - Sentinel, National Evaluation System for health Technology (NEST), MDUFA data standardization
- NIH
  - CTSA, HCS Collaboratory, Multiple institute/Center Networks
- CDC Vaccine Surveillance Network
- ASPE—PCOR-Trust Fund
- PCORI-PCORnet
- CMS
  - Enclave, Coverage with Evidence Development
- Million Veterans’ Program (MVP)
- Precision Medicine Initiative (PMI)
- Professional society quality registries
Call to Action

• Organize operational systems that bring together research networks embedded in practice
  – to enable patients, consumers, clinicians, industry, government, and health care systems to participate in prospective trials and observational studies
  – Develop operational/regulatory approaches to facilitate practice-based systems for therapeutic research, safety surveillance, public health, and quality improvement.
  – Support adequate time commitment for clinicians to engage with patients to ensure mutual understanding and appropriate consent
  – Efficient systems for contracting and liability
  – Clinical care and research closely aligned in “learning health system” supported by education and training
Our National Clinical Research System is Well-intentioned But Flawed

• High percentage of decisions not supported by evidence*
• Health outcomes and disparities are not improving
• Current system is great except:
  • Too slow, too expensive, and not reliable
  • Doesn’t answer questions that matter most to patients
  • Unattractive to clinicians & administrators

We are not generating the evidence we need to support the healthcare decisions that patients and their doctors have to make every day.

*Tricoci P et al. JAMA 2009;301:831-41
PCORnet embodies a “community of research” by uniting people, clinicians & systems

20 Patient-Powered Research Networks (PPRNs)

+ 13 Clinical Data Research Networks (CDRNs)

PCORnet = A national infrastructure for people-centered clinical research
Resulting in a national evidence system with unparalleled research readiness

Number of people with data available in PCORnet to date:

~145 Million

*Based on data from 57 DataMarts as of July 15, 2016
Call to Action

• Establish a robust framework for privacy, confidentiality, and security
  • endorsed by patients and consumers to ensure the trust a learning health system will require,
  • Robust procedures that ensure data security and protect confidentiality
  • Efficient and thorough digital system of education and research permissions for patients
  • Balance of individual autonomy and public health needs
  • Great start: Precision Medicine Initiative: Privacy and Trust Principles
Call to Action

• Adopt a common approach to configuring, storing, and re-using digital health care data to enable use in care, research, safety surveillance, and public health
  – As called for in the Nationwide Interoperability Roadmap published by the Office of the National Coordinator for Health Information Technology.
  – Common standards and terminology for prospective data collection
  – Continuous effort to curate data to produce high quality data sets for analysis using common data models
  – Leverage existing digital health/healthcare data to create efficiencies (registries, claims data, EHR data, personal devices)
Sentinel Distributed Analysis

1- User creates and submits query (a computer program)
2- Data partners retrieve query
3- Data partners review and run query against their local data
4- Data partners review results
5- Data partners return results via secure network
6- Results are aggregated
Sentinel Distributed Database*

- Populations with well-defined person-time for which most medically-attended events are known
  - 193 million members**
  - 351 million person-years of observation time
  - 39 million people currently accruing new data
  - 4.8 billion dispensings
  - 5.5 billion unique encounters
    - 51 million acute inpatient stays
  - 33 million people with ≥1 laboratory test result

* As of August 2015, excludes HCA and BCBS of Massachusetts
** Double counting exists for individuals who change health plans
For Big-Data Scientists, ‘Janitor Work’ Is Key Hurdle to Insights
The New Einsteins Will Be Scientists Who Share

From cancer to cosmology, researchers could race ahead by working together—online and in the open

By MICHAEL NIELSEN

In January 2009, a mathematician at Cambridge University named Tim Gowers decided to use his blog to run an unusual social experiment. He picked out a difficult mathematical problem and tried to solve it completely in the open, using his blog to post ideas and partial progress. He issued an open invitation for others to contribute their own ideas, hoping that many minds would be more powerful than one. He dubbed the experiment the Polymath Project.

Several hours after Mr. Gowers opened up his blog for discussion, a Canadian-Hungarian mathematician posted a comment. Fifteen minutes later, an Arizona high-school math teacher chimed in. Three minutes after that, the UCLA mathematician Terence Tao commented. The discussion ignited, and in just six weeks, the mathematical problem had been solved.
Call to Action

• Develop and test new methods to reliably answer research questions
  – more efficient RCTs,
  – Novel designs such as cluster-randomized trials, basket trials
  – And more reliable observational studies aimed at assessment of interventions
  – “Meta-knowledge” on which methods are best for which types of questions
  – By leveraging data already collected by health information technology and other electronic sources to answer research questions or facilitate the conduct of new trials.
Data Activation and Testing Outcomes

What Impacts Behavior?

A
CONTROL

B
VARIATION

37%
Call to Action

• Ensure the development of novel approaches focusing on streamlining and harmonizing processes in ways that eliminate barriers that promote unnecessary complexity, while ensuring safeguards that are truly needed.
  – Streamlined and harmonized processes eliminate barriers to efficient research while ensuring needed safeguards
  – Systems for high quality and efficient ethics review and contracting
  – Development of approaches to assuring quality systems through better use of analytics
Trial Hyperinflation

Figure 3. Mean Total Grant Cost per Patient Index, Biomedical R&D Price Index, and pooled hedonic indexes, 1989-2011

Source: Authors’ calculations based on Medidata Solutions, Inc.’s, PICAS™ database.

Berndt E, Cockburn I. Monthly Labor Review, June 2014

www.fda.gov
We believe several EvGen use-cases can be launched in the short and medium term making incremental changes to today’s infrastructure.

<table>
<thead>
<tr>
<th>Current state (U.S.)</th>
<th>Data Creation</th>
<th>Aggregation</th>
<th>Utilization</th>
<th>Dissemination</th>
<th>Key requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Use cases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1A Real world data for more efficient clinical trials</td>
<td><img src="#" alt="Least advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td>Maturity of methods, Regulatory pathways</td>
</tr>
<tr>
<td>1B Learning from historical clinical trial failures</td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Least advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td>Aggregation of specialized data, Sophisticated collaboration/ usage models</td>
</tr>
<tr>
<td>1C Accelerating innovation in rare and ultra-rare diseases</td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Least advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td>Specialized datasets with patient privacy challenges, Maturity of methods and decision making</td>
</tr>
<tr>
<td>2A Real world insights to evaluate efficacy of generics and biosimilars</td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Least advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td>Maturity of methods, Regulatory pathways</td>
</tr>
<tr>
<td>2B Adapting the regulatory paradigm for rare and ultra-rare diseases</td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Least advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td>Maturity of methods, Regulatory pathways</td>
</tr>
<tr>
<td>2C Enhancing active safety surveillance and information availability</td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td>High quality data, Methods and regulatory pathways</td>
</tr>
<tr>
<td>2D Early warnings of epidemics and drug dev. for infectious diseases</td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td>High quality data to act in a timely fashion</td>
</tr>
<tr>
<td>3A Patient-level coverage and pricing, based on real world outcomes</td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td>High quality data, Maturity of methods</td>
</tr>
<tr>
<td>4A Insights for physicians to provide better care for patients</td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td>High quality longitudinal data, Maturity of methods and decision making</td>
</tr>
<tr>
<td>4B Enabling precision medicine through big data</td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td>Needs more sophisticated analytics and furthering of science to understand links between genotypes and phenotypes</td>
</tr>
<tr>
<td>4C Population-level insights into antibiotic usage and resistance</td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td><img src="#" alt="Most advanced" /></td>
<td>Require large, interconnected data sets to make population-level correlations</td>
</tr>
</tbody>
</table>
World Population Dominated by Asia

- China
- India
- USA
- Indonesia
- Brazil
- Pakistan
- Russia
- Bangladesh
- Nigeria
- Japan

Source: GeoHive
The Global Drug Manufacturing Supply Chain

Illustration of drug manufacturing supply chain: A U.S. finished drug may be produced using an active pharmaceutical ingredient (API) made in China and ingredients made in Europe, Japan, or the U.S. These components may be shipped to India where the finished drug is manufactured and then imported into the U.S. for distribution.
If We Had An Efficient Evidence Generation System....

• Translation of discovery to useful medical products would accelerate
• Much more of clinical practice could be guided by high quality evidence
• Policy decisions would be more rational
• Clinicians and their practice organizations could focus on interpreting the evidence and applying it
• The role of opinion and expertise would be at least as important, but it would be put to a much higher purpose—providing precision healthcare