The Future of Registries in the Era of Data Ubiquity

Robert M Califf MD
Vice Chancellor for Health Data Science
Advisor, Verily Life Sciences
Improving Patient Care by Capturing Computerized data:

A glimpse into the creation of The Duke Databank for Cardiovascular Disease
“Chronic, multifactorial disease problems can be studied, but not by the methods of the present or past. If one wishes to create useful information, computer technology must be exploited.”

—Eugene Stead, MD

1960s
- Dr. Eugene Stead developed concept of “computerized textbook of medicine”
- Formation of the Duke Databank for Cardiovascular Diseases
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

Discovery Science

Outcomes

Performance Measures

Measurement and Education

Early Translational Steps

Clinical Trials

Clinical Practice Guidelines
Those enrolled in CV RCTS are not representative

- Comorbidity
- Community
- Clinical Trials
- Age
- Disease Severity

- 63
- 68
Slow Evidence Adoption Means Lost Lives

ACS Care at 430 US Hospitals

Peterson et al, JAMA 2006;295:1863-1912
McDonalds vs the US Healthcare Profession?

Provider

Quality Metric

Napkin and Ketchup in Bag
>99% given At any location

Beta-blocker Or anti-platelet therapy
<br><50% at many locations
Registries-Purpose

• Measurement of clinical care and outcomes
• Assessment of quality
• Implementation of quality systems/evidence based medicine
• Backbone for assessment of therapies
  — Safety
  — Observational effectiveness
  — Backbone of data for trials
Four industrial revolutions

FIRST
Water and steam power mechanize production.

Jacob Leupold, Steam Engine, in *Theatri Machinarum Hydraulicarum II* (1720)
Four industrial revolutions

SECOND
Electric power creates mass production.
Four industrial revolutions

THIRD
Electronics and information technology automate production.

Four industrial revolutions

FOURTH

The digital revolution—characterized by a fusion of technologies—blurs the lines between physical, digital, and biological spheres.

Opte Project. Internet map. https://commons.wikimedia.org/wiki/File:Internet_map_1024.jpg
WE’VE MAPPED THE WORLD.
NOW LET’S MAP HUMAN HEALTH.

www.projectbaseline.com
## Deep molecular profiling

### SAMPLES
- SERUM
- WHOLE BLOOD
- PBMCS
- PLASMA
- STOOL
- SALIVA
- URINE

### CORE PLATFORMS
- **CLINICAL LABS**
- GENOMICS (WGS, DNA arrays)
- EPIGENOMICS (Methyl arrays)
- TRANSCRIPTOMICS (RNA-seq)
- IMMUNOPHENOTYPING (CyTOF)
- MICROBIOME (16S rRNA)
- PROTEOMICS
- METABOLICOMICS

### AUTOMATION

![Automation Image]

~6TB data per subject
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.

IMPLEMENT
Apply plan in pilot and control settings.

ADJUST
Use evidence to influence continual improvement.

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.

www.fda.gov
Previously Independent Sites now part of large integrated health systems
increasingly sophisticated data warehouses
Nodes are Operational Clusters Using Common Data

- KP Northwest
- KP Mid-Atlantic
- KP Northern CA
- KP Colorado
- KP Georgia
- KP Southern CA
By the end of 2016, there were 626 health systems* in the United States.

- 69.7% of U.S. hospitals are in health systems
- 91.6% of U.S. hospital discharges are from system hospitals
- 42.7% of U.S. primary care physicians are in health systems

Post Market Studies, including comparative effectiveness

PCORnet

Coordinating Center
PCORnet® embodies a "network of networks" that harnesses the power of partnerships.
CDRNs

Accelerating Data Value Across a National Community Health Center Network (ADVANCE)
Oregon Community Health Information Network (OCHIN)

Chicago Area Patient Centered Outcomes Research Network (CAPriCORN)
The Chicago Community Trust

Greater Plains Collaborative (GPC)
University of Kansas Medical Center

Kaiser Permanente & Strategic Partners Patient Outcomes Research To Advance Learning (PORTAL) Network
Kaiser Foundation Research Institute

Research Action for Health Network (REACHnet)
Louisiana Public Health Institute (LPHI)

Mid-South CDRN
Vanderbilt University

National PEDSnet: A Pediatric Learning Health System
The Children’s Hospital of Philadelphia

New York City Clinical Data Research Network (NYC-CDRN)
Weill Medical College of Cornell University

OneFlorida Clinical Data Research Network
University of Florida

Patient-Centered Network of Learning Health Systems (LHSNet)
Mayo Clinic

Patient-oriented SCAlable National Network for Effectiveness Research (pSCANNER)
University of California, San Diego (UCSD)

PaTH: Towards a Learning Health System
University of Pittsburgh

Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS)
Harvard University
Resulting in a national evidence system with “research readiness”

PCORnet represents:

~122 million patients

who have had a medical encounter in the past 5 years

*some individuals may have visited more than one Network Partner and would be counted more than once
Continuous monitoring through passive sensors

**Sleep sensor**
Commercially available, placed under mattress to passively monitor multiple physiologic data parameters

**Study hub**
Safely sends device data to secure, encrypted Baseline database

**Study watch**
Investigational wrist-worn sensor for continuous recording of physiological and environmental data

**App**
Mobile interface for self-reported and passive data acquisitions
Big Challenges in Biomedicine

• Lack of significant information over time dimension
  — Measurements to assess biology and human health are made periodically in visits to healthcare or for research

• Missing systems biology
  — When developing concepts of human biology or drug development we make limited measurements focused on specific mechanisms — we look “under the lamppost”

• Missing the opportunity to measure 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
The process of digital phenotyping

Digital phenotyping involves collecting sensor, keyboard, and voice and speech data from smartphones to measure behavior, cognition, and mood.
1 in 20 Google searches are health related
Our mission
Make health information universally accessible and useful.
Information structure

Measles
Also called: rubeola

ABOUT
SYMPTOMS
TREATMENTS

Requires a medical diagnosis
Measles symptoms don't appear until 10 to 14 days after exposure. They include cough, runny nose, inflamed eyes, sore throat, fever, and a red, blotchy skin rash.

People may experience:
Pain areas: in the muscles
Whole body: fever, malaise, fatigue, or loss of appetite

Measles
Also called: rubeola

ABOUT
SYMPTOMS
TREATMENTS

Treatment consists of preventative measures
There's no treatment to get rid of an established measles infection, but over-the-counter fever reducers or vitamin A may help with symptoms.

Preventative
MMR vaccine

Medications
Life expectancy at birth by county, 2014

Counties in South Dakota and North Dakota had the lowest life expectancy, and counties along the lower half of the Mississippi, in eastern Kentucky, and southwestern West Virginia also had very low life expectancy compared with the rest of the country. Counties in central Colorado had the highest life expectancies.
Change in life expectancy at birth by county, 1980 to 2014

Compared with the national average, counties in central Colorado, Alaska, and along both coasts experienced larger increases in life expectancy between 1980 and 2014, while some southern counties in states stretching from Oklahoma to West Virginia saw little, if any, improvement over this same period.

### Table 1. Variables Included in the Regression Analysis With Summary Statistics and Bivariate Regression Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Summary Statistics, Mean (SD) [Range]</th>
<th>Bivariate Regression Results</th>
<th>Rsquared</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Socioeconomic and race/Ethnicity factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population below the poverty line, %</td>
<td>16.3 (6.4) [3.1-62.0]</td>
<td>-0.24 (0.005)</td>
<td>0.47</td>
</tr>
<tr>
<td>Median household income, log $</td>
<td>10.6 (0.2) [9.8-11.6]</td>
<td>6.06 (0.130)</td>
<td>0.41</td>
</tr>
<tr>
<td>Graduates, age ≥25 y, %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>83.7 (7.2) [46.3-98.6]</td>
<td>0.20 (0.004)</td>
<td>0.42</td>
</tr>
<tr>
<td>College</td>
<td>19.2 (8.6) [4.2-72.0]</td>
<td>0.15 (0.004)</td>
<td>0.34</td>
</tr>
<tr>
<td>Unemployment rate, age ≥16 y, %</td>
<td>9.1 (3.2) [2.1-27.4]</td>
<td>-0.29 (0.011)</td>
<td>0.18</td>
</tr>
<tr>
<td>Black population, %</td>
<td>9.4 (4.7) [0-85.8]</td>
<td>-0.07 (0.002)</td>
<td>0.24</td>
</tr>
<tr>
<td>American Indian, Native Alaskan, and Native Hawaiian population, %</td>
<td>2.3 (7.9) [0-97.2]</td>
<td>-0.06 (0.005)</td>
<td>0.04</td>
</tr>
<tr>
<td>Hispanic population, %</td>
<td>8.1 (13.1) [0-95.9]</td>
<td>0.02 (0.003)</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Behavioral and metabolic risk factors, %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obesity prevalence, age ≥20 y</td>
<td>37.0 (4.3) [18.0-52.0]</td>
<td>-0.39 (0.006)</td>
<td>0.54</td>
</tr>
<tr>
<td>No leisure-time physical activity prevalence, age ≥20 y</td>
<td>27.0 (5.2) [11.7-47.2]</td>
<td>-0.34 (0.005)</td>
<td>0.62</td>
</tr>
<tr>
<td>Cigarette smoking prevalence, age ≥18 y</td>
<td>24.7 (4.1) [7.7-42.1]</td>
<td>-0.40 (0.007)</td>
<td>0.54</td>
</tr>
<tr>
<td>Hypertension prevalence, age ≥30 y</td>
<td>39.5 (3.6) [27.9-56.4]</td>
<td>-0.49 (0.007)</td>
<td>0.62</td>
</tr>
<tr>
<td>Diabetes prevalence, age ≥20 y</td>
<td>14.0 (2.4) [8.1-25.5]</td>
<td>-0.72 (0.011)</td>
<td>0.59</td>
</tr>
<tr>
<td><strong>Health care factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insured population, age &lt;65 y, %</td>
<td>81.7 (5.7) [57.3-96.7]</td>
<td>0.15 (0.007)</td>
<td>0.14</td>
</tr>
<tr>
<td>Quality index</td>
<td>70.1 (11.5) [0-100]</td>
<td>0.10 (0.003)</td>
<td>0.28</td>
</tr>
<tr>
<td>Physicians per 1000 population, No.</td>
<td>1.1 (1.0) [0-4.4]</td>
<td>0.53 (0.039)</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Abbreviation: SE, standard error.
Men and women ages 50–54, deaths by drugs, alcohol, and suicide, 1989–2014

Midlife mortality from “deaths of despair” across countries


Members of the Organisation for Economic Coordination and Development include Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the UK, and the US.
Life expectancy vs. health expenditure, 1970 to 2015

Health financing is reported as the annual per capita health expenditure and is adjusted for inflation and price level differences between countries (measured in 2010 international dollars).

OurWorldinData.org/the-link-between-life-expectancy-and-health-spending-us-focus • CC BY-SA
Fundamental Informatics Infrastructure—Matrix Organizational Structure

Integrated at "enterprise level"

Disease Registries—Granular, Detailed

- Primary Care
- Mental Health
- Cancer
- Cardiovascular
- Etc...

Electronic Health Records

Health System A

Health System B

Etc...

Claims data
Real World Data vs Evidence

Real-World Data Sources
- Claims Data
- EHRs/EMRs
- Prospective Observational Data
- Patient Pathways
- Surveillance
- Mortality Database
- Primary and Secondary Care Data
- Administrative Data
- Disease and Device Registries
- Pharmacy Data
- Cost Studies
- Mobile Devices
- Consumer Data
- Social Media

Real-World Evidence
- Identifying Unmet Needs:
  - Natural History
  - Co-morbidities
  - Disease Mechanisms
  - Burden of Illness
  - Clinical Practice Patterns

Real-World Evidence
- Informing Clinical and Policy Decisions:
  - Usage Patterns
  - Outcome Predictors
  - Population-Level Impact
  - Clinical Review
  - Benefit/risk in Subgroups
  - New Indications

Prediscovery → Drug Discovery → Preclinical Development → Clinical Development (Phases I, II, III) → FDA Review and Approval → Postmarketing Evaluation (Phase IV)

Real-World Evidence can be used across a wide spectrum of research, ranging from observational studies to studies that incorporate planned interventions, whether with or without randomization at the point of care. Incorrect to contrast the term “real-world evidence” with the use of randomization in a manner that implies that they are disparate or even incompatible concepts. Must consider the components of such trials that are critical to obtaining valid results and minimizing bias.
Policy efforts underpinning RWE push

<table>
<thead>
<tr>
<th><strong>Cures provisions (Sec. 3022)</strong></th>
<th><strong>PDUFA RWE provisions</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Requires FDA to establish a program to evaluate the potential use of real world evidence to:</td>
<td>Tracks with Cures Act</td>
</tr>
<tr>
<td>• Help support the approval of new indications for an approved drug</td>
<td>• Help support the approval of new indications for an approved drug</td>
</tr>
<tr>
<td>• Help support or satisfy post approval study requirements</td>
<td>• Help support or satisfy post approval study requirements</td>
</tr>
</tbody>
</table>

**Reinforcing of a Learning Health Care System:**

- Doesn’t change approval standards, rather it better supports and enables use of data and evidence on outcomes that are hard to get from traditional RCTs (e.g., outcomes that are too costly, too small populations with particular clinical features, too long follow-up needed, diff impact in diff clinical settings, etc.)
- Learning from real-world patient experiences can support better informed health care decision-making by a range of stakeholders
Laying the Foundation

Stakeholder Engagement

Data Standards

Guidances

Duke MARCOLIS CENTER for Health Policy

Use of Electronic Health Record Data in Clinical Investigations

Electronic Source Data in Clinical Investigations

Use of Electronic Informed Consent
National System Paradigm Shift

Passive Surveillance
- Parallel track to clinical practice
- Inefficient one-off studies
- Challenging to find right pre/post market balance without confidence in post-market data

Active Surveillance to better protect patients
- Embedded in Health Care System (collect data during routine clinical care)

Leverage RWE to support regulatory decisions throughout TPLC
- Shared system to inform the entire Ecosystem (patients, clinicians, providers, payers, FDA, Device Firms)

Current State: Passive Surveillance
- Parallel track to clinical practice
- Inefficient one-off studies

Future State: Active Surveillance
- Embedded in Health Care System

Decision Support: Leverage RWE to support regulatory decisions throughout TPLC
- Shared system to inform the entire Ecosystem

Ecosystem: Patients, clinicians, providers, payers, FDA, Device Firms
Learning Medical Device Ecosystem

Total Product Life Cycle (TPLC) Framework

Information Flow

Benefit-Risk Evidence

Patient Access

Progressive Approval, Safety and Performance

NEST

Clinical Research Incorporated Into Routine Clinical Practice

INTERNATIONAL HARMONIZATION

Patient Access
Depression is highly prevalent: 300 million people suffer from depression globally, WHO has declared it a leading cause of disability [WHO].

Many people don’t get treatment: 50% of people with depression in the US did not get any treatment [JAMA].

Treatment is often delayed: On average, 7 years from onset to treatment in the US [JAMA].

Treatment is effective: 70% of patients can improve, often in a matter of weeks [NIMH].

Google has the reach, scale and technology to help.
PRODUCT OVERVIEW: What is PHQ-9?

PHQ-9 is a Patient Health Questionnaire, with 9 questions, that is used to measure depression severity.

<table>
<thead>
<tr>
<th>Over the last 2 weeks, how often have you been bothered by any of the following problems?</th>
<th>Not at all</th>
<th>Several days</th>
<th>More than half the days</th>
<th>Nearly every day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Little interest or pleasure in doing things</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2. Feeling down, depressed, or hopeless</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3. Trouble falling or staying asleep, or sleeping too much</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4. Feeling tired or having little energy</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5. Poor appetite or overeating</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6. Feeling bad about yourself — or that you are a failure or have let yourself or your family down</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7. Trouble concentrating on things, such as reading the newspaper or watching television</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8. Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9. Thoughts that you would be better off dead or of hurting yourself in some way</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
Technology advances; people stay the same.
Complementary cumulative distribution functions (CCDFs) of true and false rumor cascades
Meet the people who can coax treasure out of messy, unstructured data.
by Thomas H. Davenport and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, “It was like arriving at a conference reception and realizing you don’t know anyone. So you just stand in the corner sipping your drink—and you probably leave early.”
For Big-Data Scientists, ‘Janitor Work’ Is Key Hurdle to Insights

By STEVE LOHR  AUG. 17, 2014
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.
2010
Individual Productivity
IT Silos

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

2020
Collective Intelligence
Distributed Computing

- Data stored in cloud, simple to query
- Collaborative, cloud based productivity applications
- Machine learning drives deep, actionable insights
- IT changing how it computes
In summary, the Seattle project represents an implementation of an approach that illustrates how doctors and patients can gain from carefully collected and computerized clinical experience. Predictions were that many such projects would be flourishing by 1980. The time course has been slower because of the difficulty of characterizing the complexity of chronic illness rather than because of problems with computer technology. In the future, data banks will provide a reference library for each patient with chronic disease. Proper interpretation and use of computerized data will depend as much on wise doctors as any other source of data in the past.

ROBERT M, CALIFF, MD
ROBERT A. ROSATI, MD
Cardiovascular Division, Department of Medicine
Duke University Medical Center
Durham, North Carolina

THE WESTERN JOURNAL OF MEDICINE  October 1981