Duke | FORGE
HEALTH DATA IN ACTION

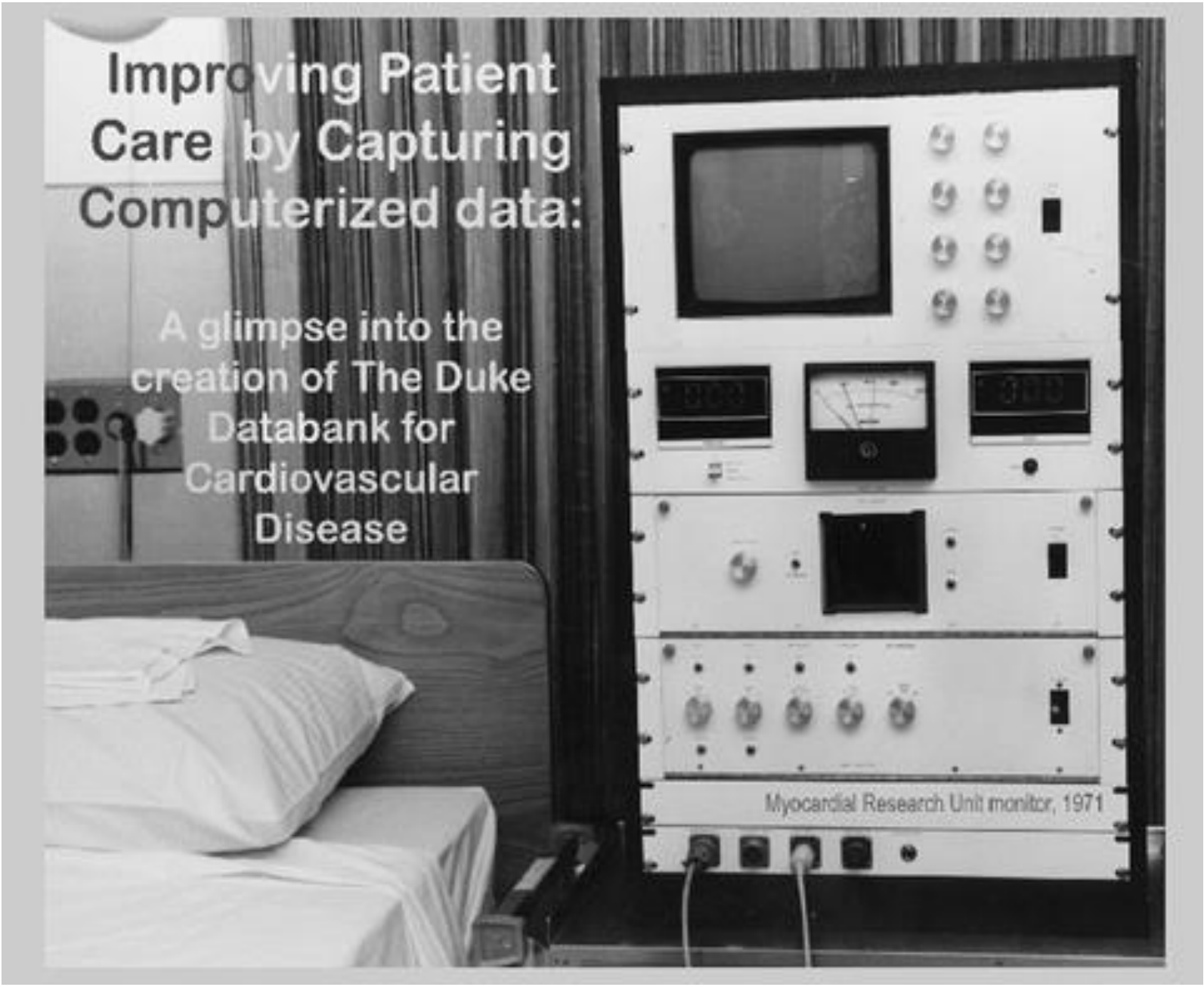
The Future of Registries in the Era of Data Ubiquity

Robert M Califf MD

Vice Chancellor for Health Data Science

Advisor, Verily Life Sciences

Duke | FORGE
HEALTH DATA IN ACTION



Duke | FORGE
HEALTH DATA IN ACTION

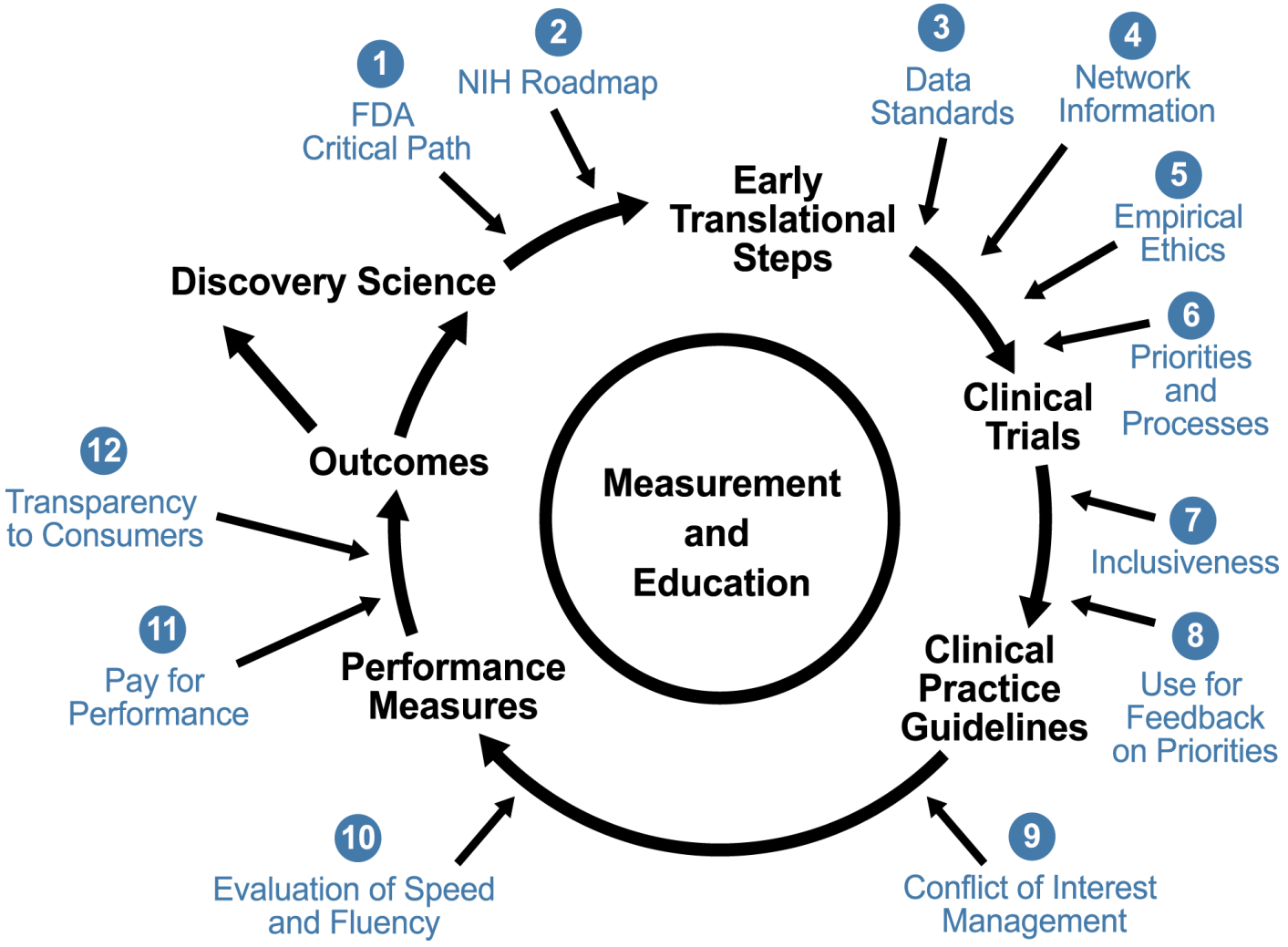
"Chronic, multifactorial disease problems can be studied, but not by the methods of the present or past. If one wishes to create useful information… computer technology must be exploited."

—Eugene Stead, MD

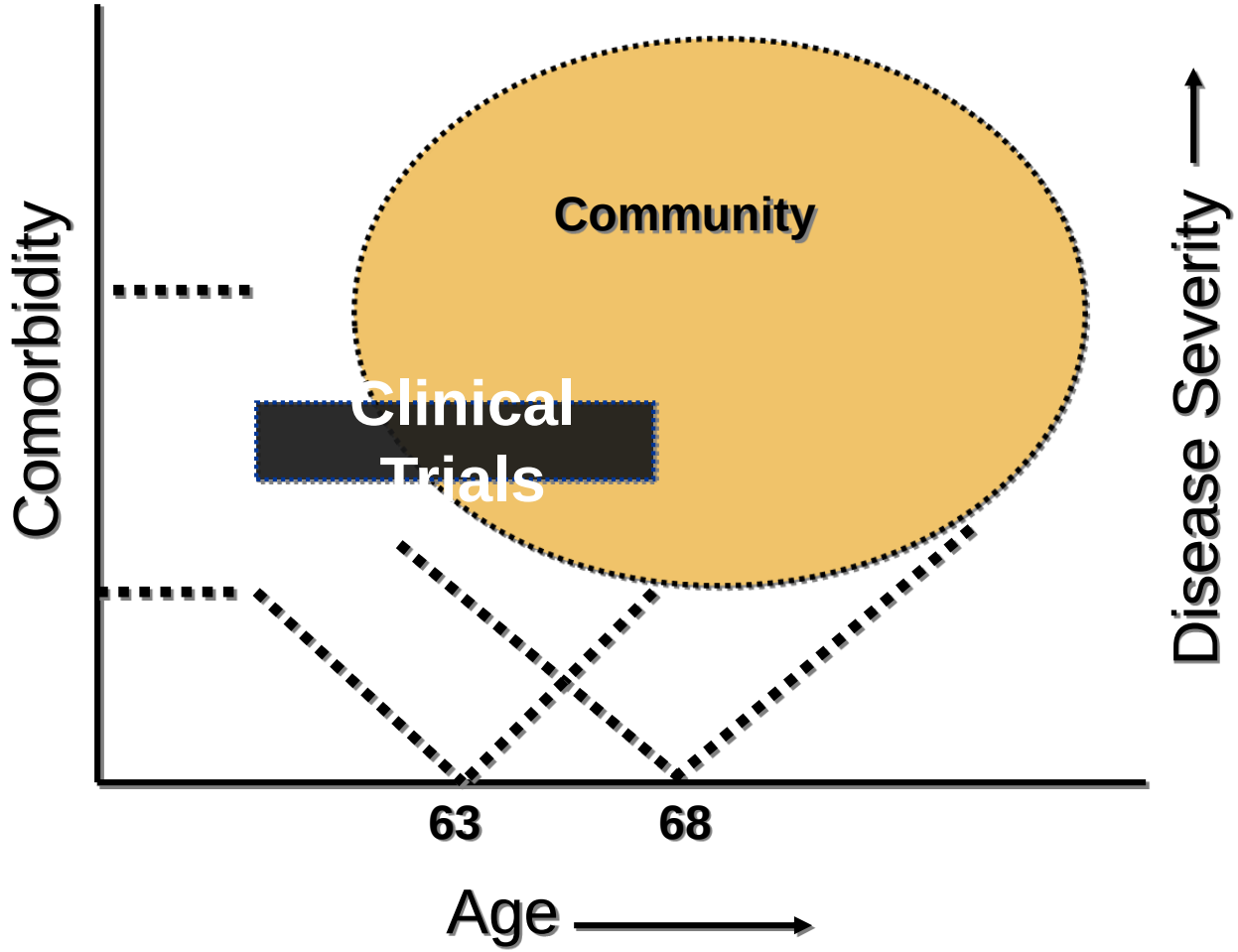
- **1960s**
  - Dr. Eugene Stead developed concept of "computerized textbook of medicine"
  - Formation of the Duke Databank for Cardiovascular Diseases

Duke | FORGE HEALTH DATA IN ACTION

Generating Evidence to Inform Decisions

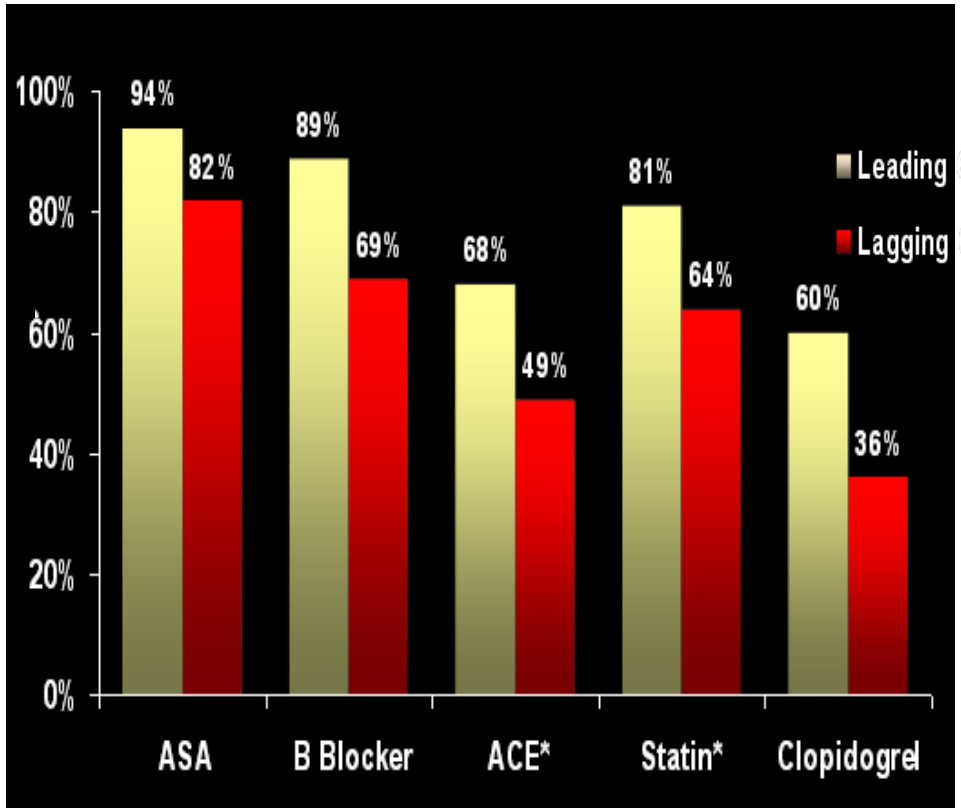


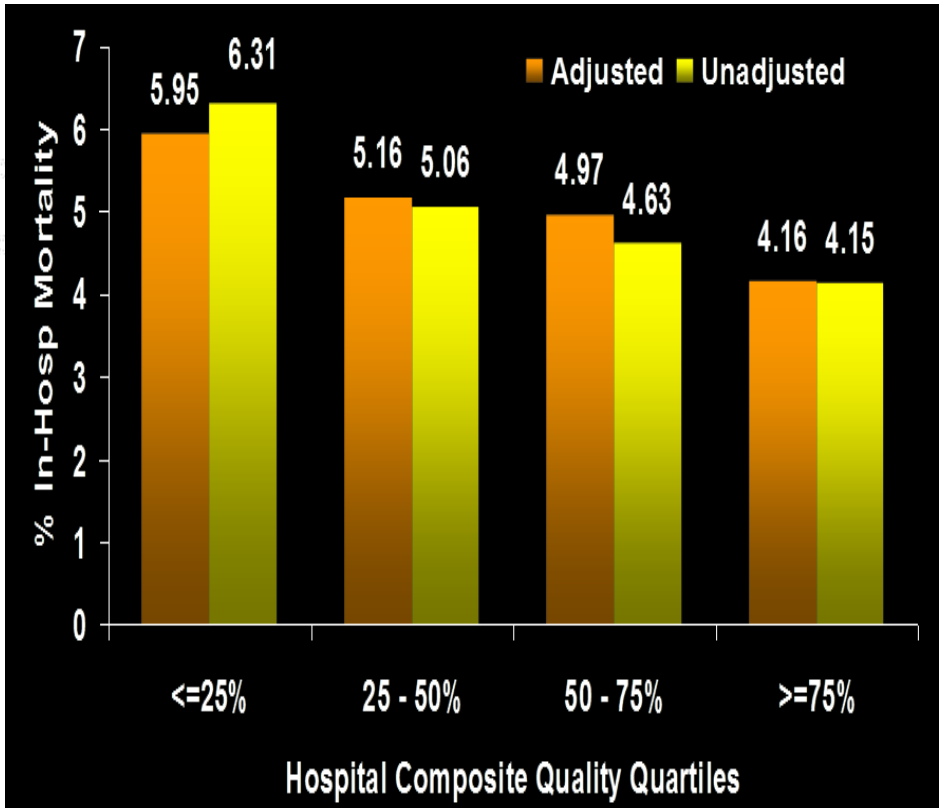
Those enrolled in CV RCTS are not representative



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**	**Beta-blocker Or anti-platelet therapy**
Performance	**>99% given At any location**	**<50% at many locations**

Duke | FORGE HEALTH DATA IN ACTION

Registries-Purpose

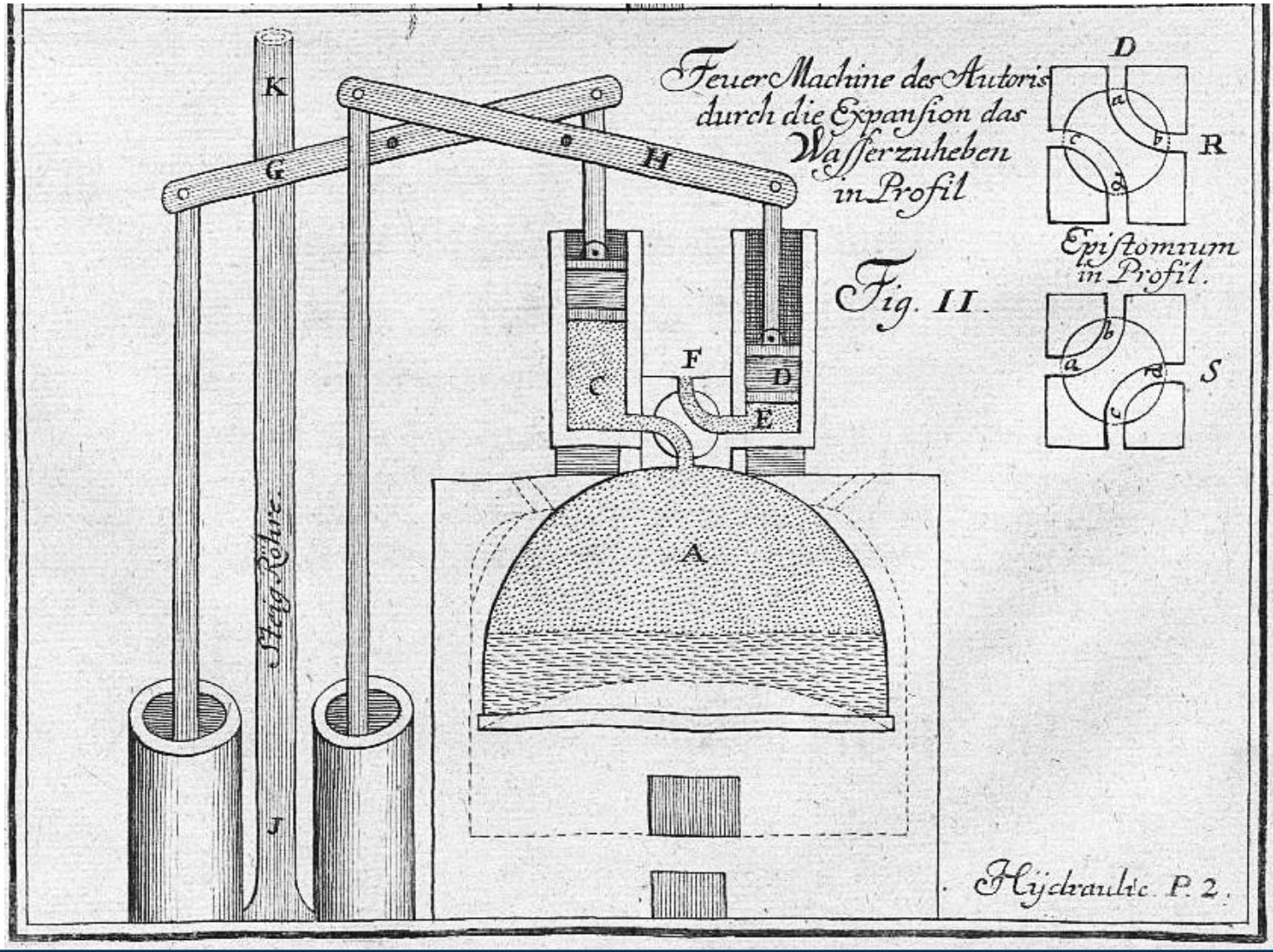
- Measurement of clinical care and outcomes
- Assessment of quality
- Implementation of quality systems/evidence based medicine
- Backbone for assessment of therapies
  - Safety
  - Observational effectiveness
  - Backbone of data for trials

Duke | FORGE
HEALTH DATA IN ACTION

Four industrial revolutions

FIRST

Water and steam power mechanize production.



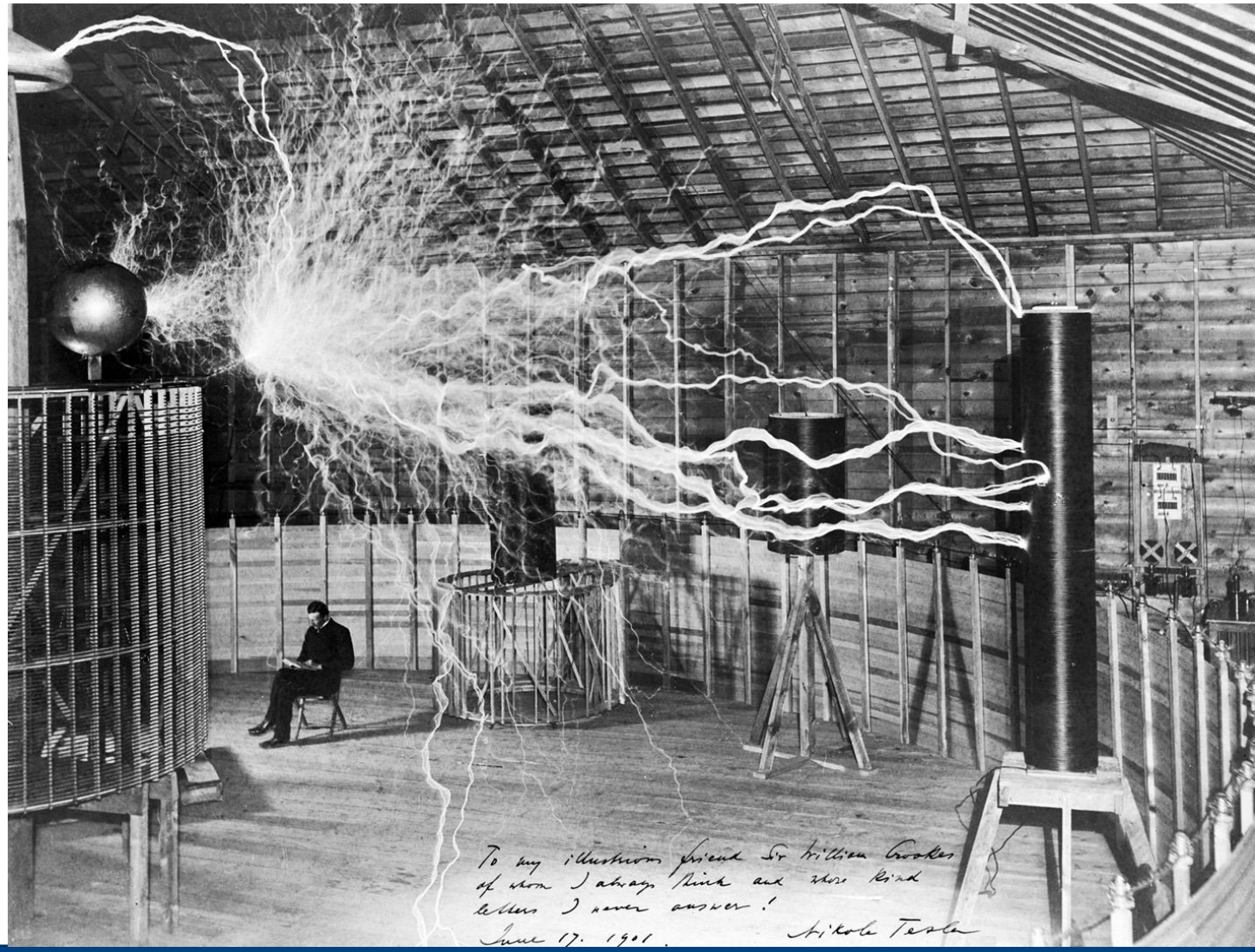
Jacob Leupold, Steam Engine, in *Theatri Machinarum Hydraulicarum II* (1720)

Duke | FORGE
HEALTH DATA IN ACTION

Four industrial revolutions

SECOND

Electric power creates mass production.

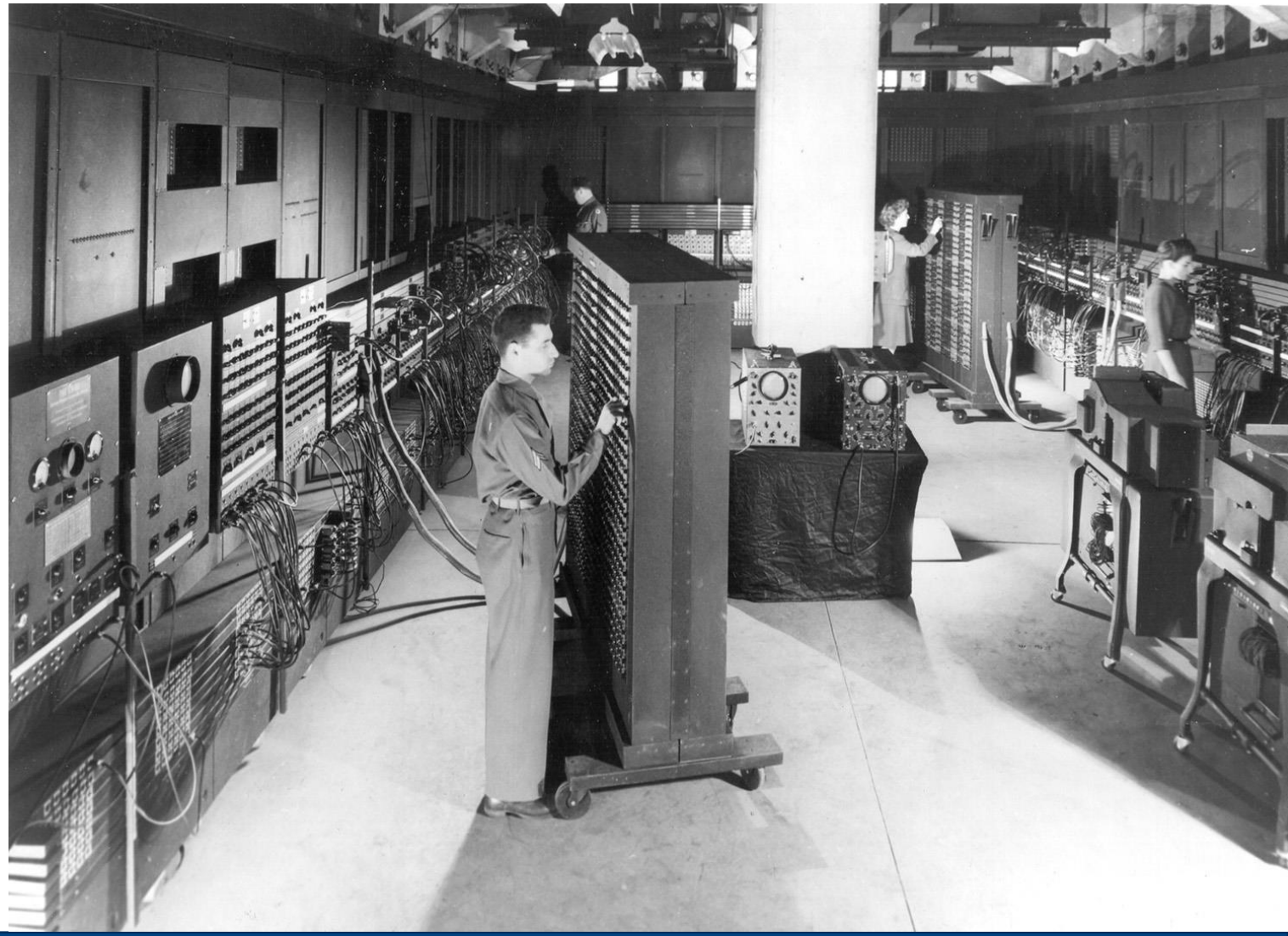


Duke | FORGE
HEALTH DATA IN ACTION

Four industrial revolutions

THIRD

Electronics and information technology automate production.



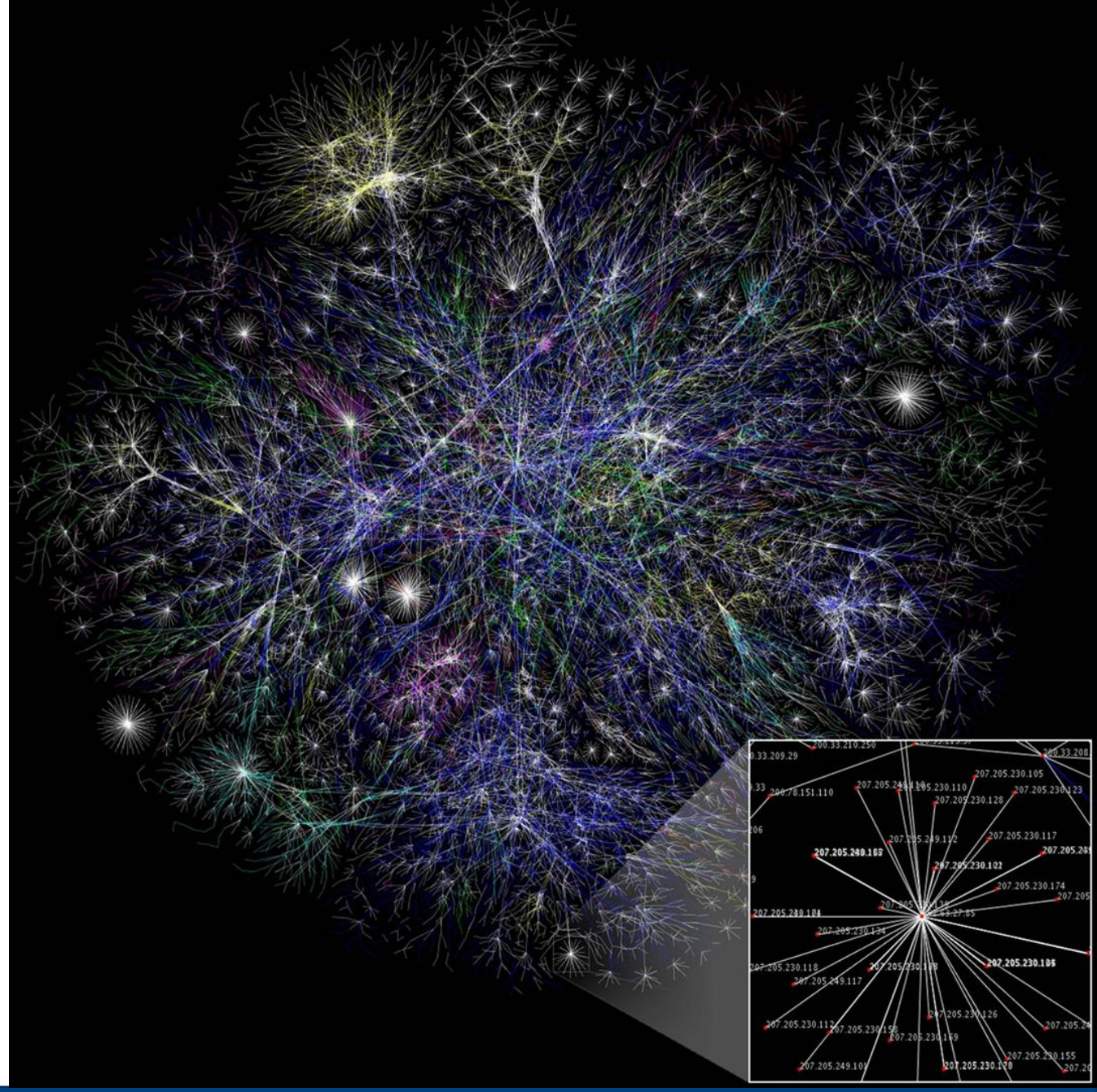
ENIAC digital computer. Unidentified U.S. Army photographer. Public Domain, https://commons.wikimedia.org/w/index.php?curid=978770

Duke | FORGE
HEALTH DATA IN ACTION

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

Duke | FORGE
HEALTH DATA IN ACTION

Project Baseline

WE'VE MAPPED THE WORLD.
NOW LET'S MAP HUMAN HEALTH.

verily

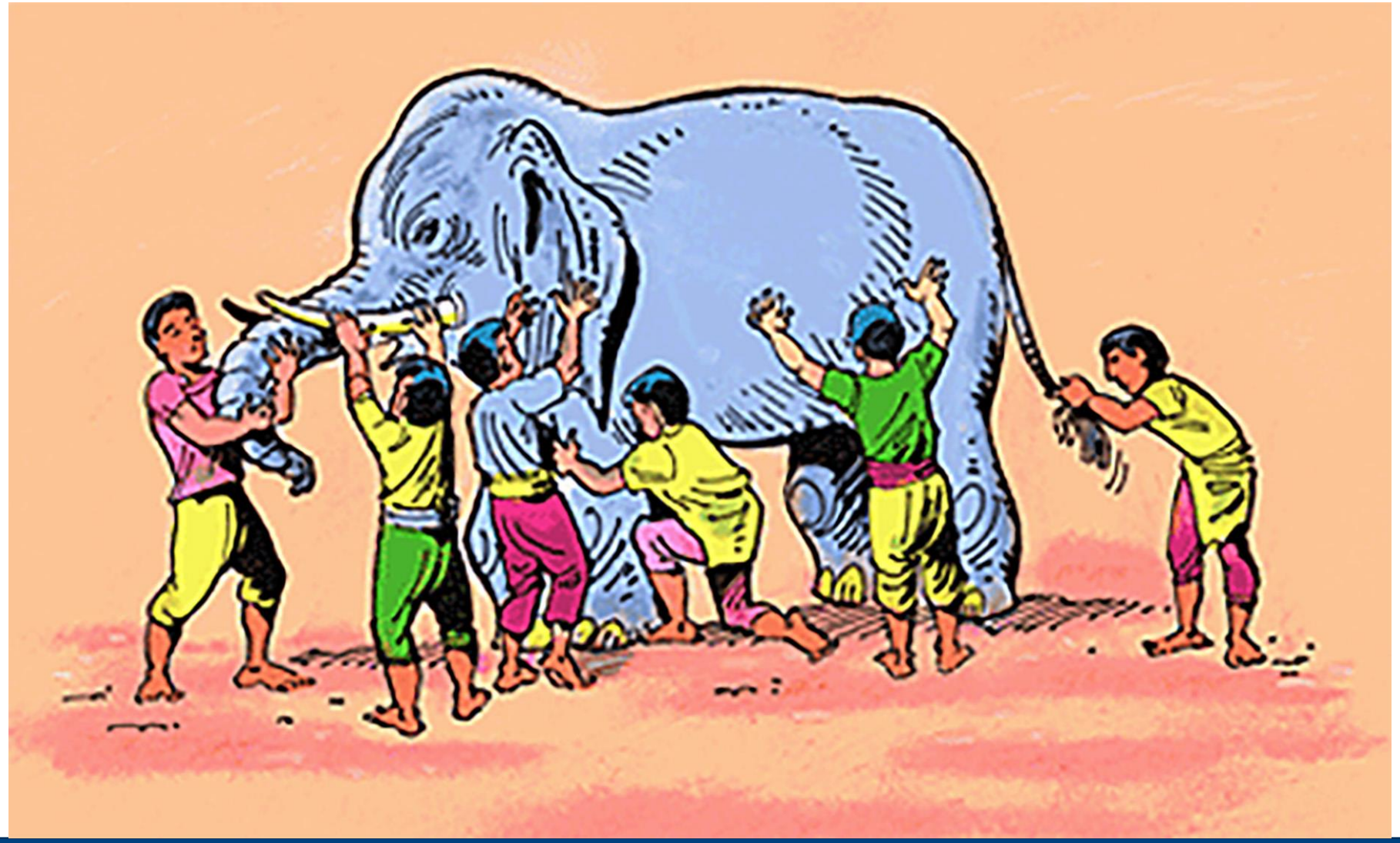
Duke University School of Medicine

Stanford MEDICINE

Google

www.projectbaseline.com

Duke | FORGE
HEALTH DATA IN ACTION



Duke | FORGE HEALTH DATA IN ACTION

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**
- **METABOLOMICS**

AUTOMATION

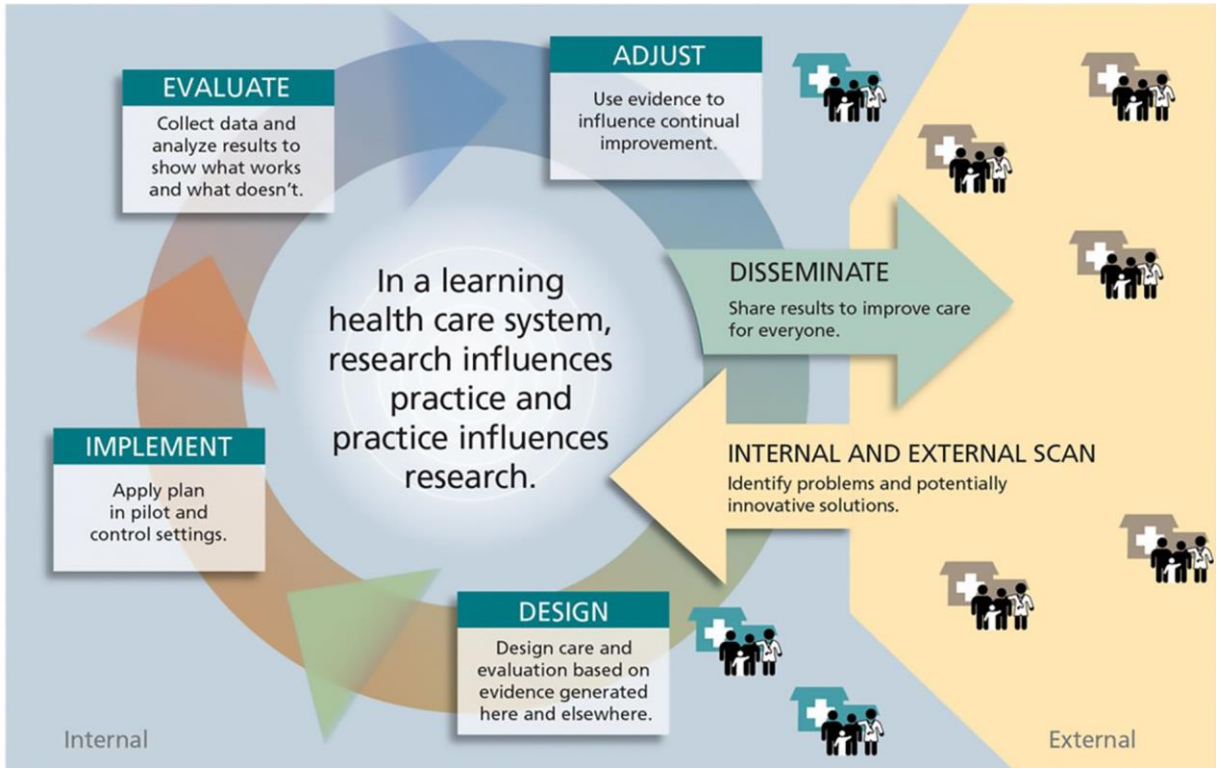
~6TB
data per subject

- External/at clinic site
- In-house
- External

Project Baseline

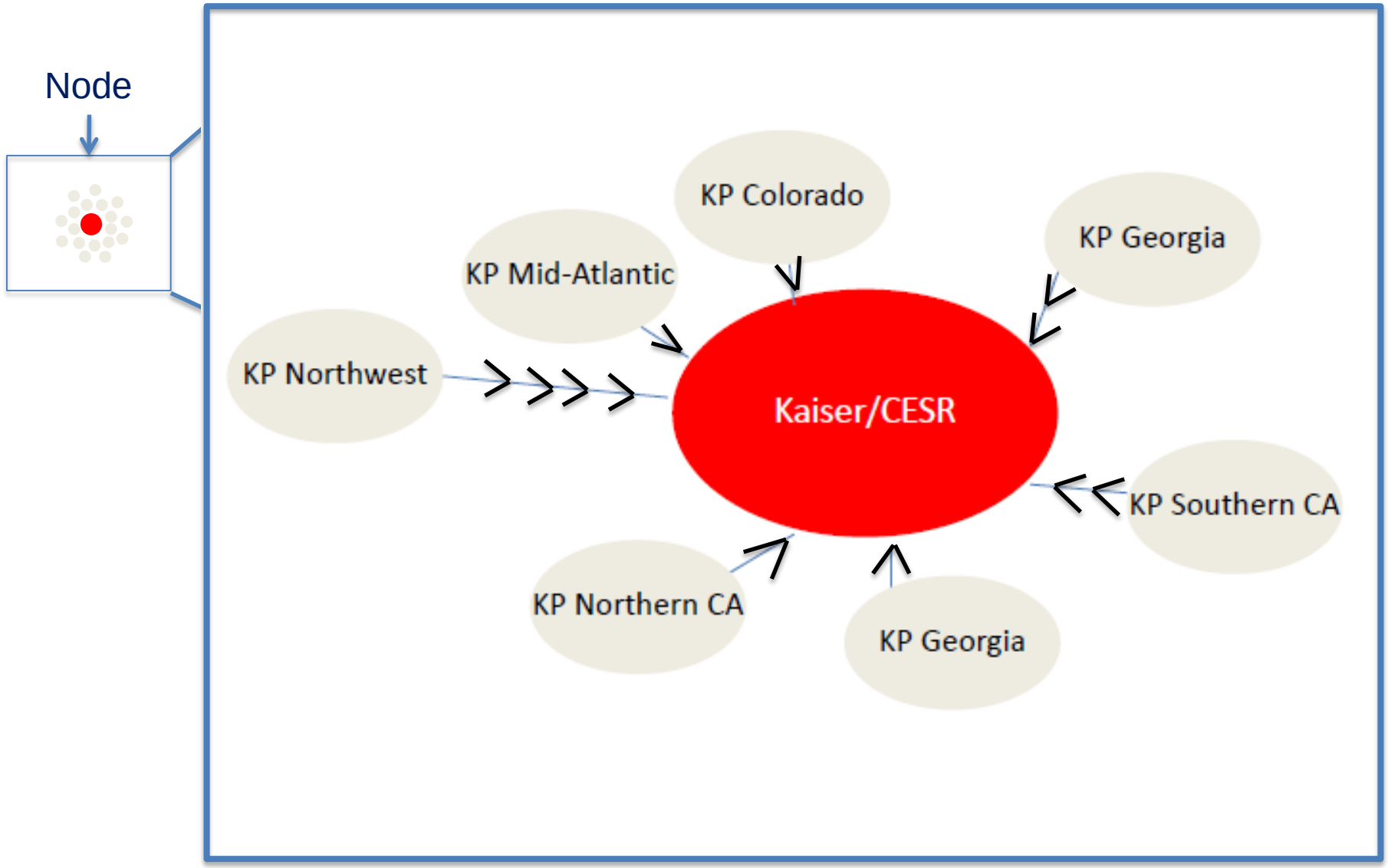
Duke | FORGE
HEALTH DATA IN ACTION

Learning health care systems

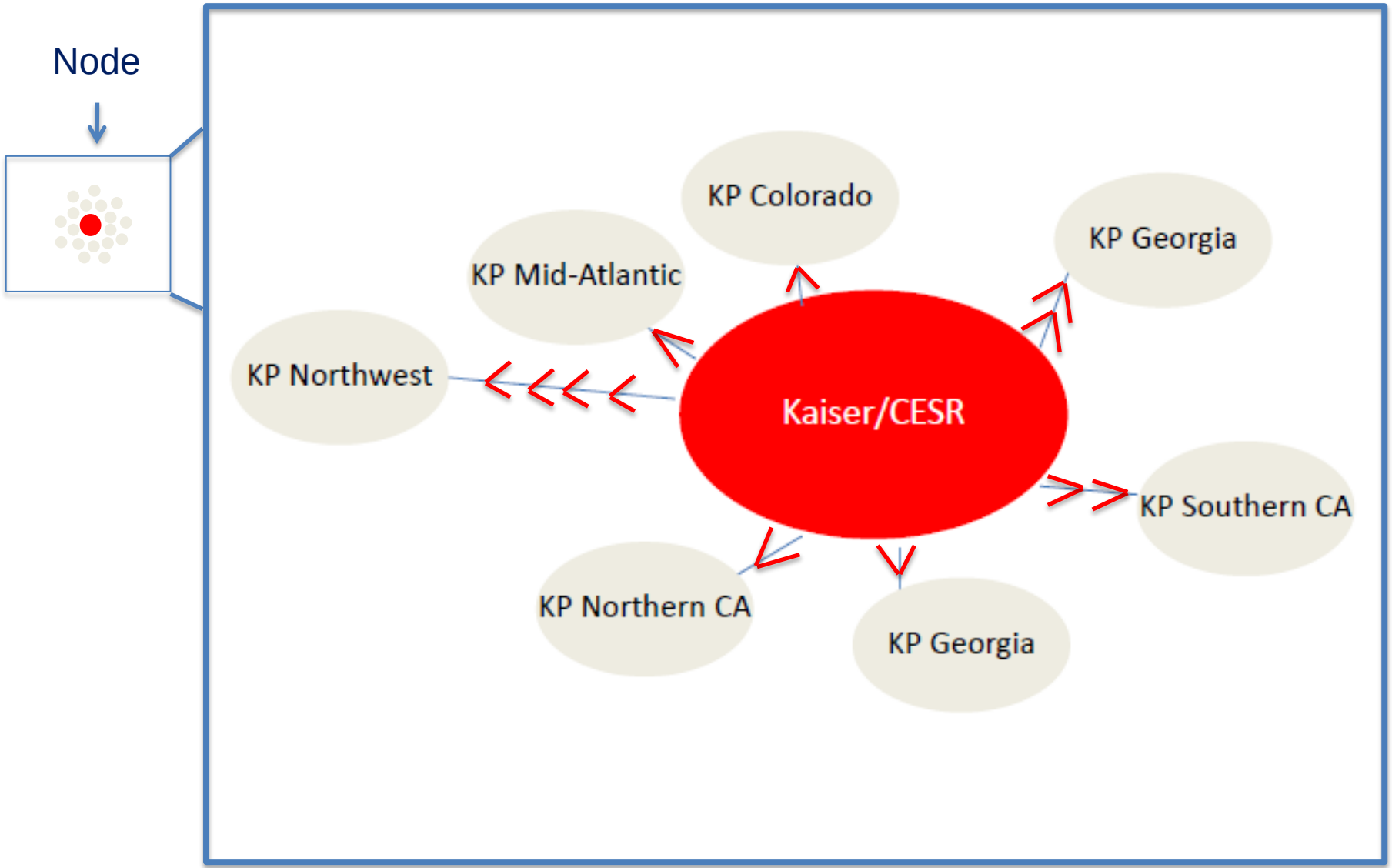


www.fda.gov

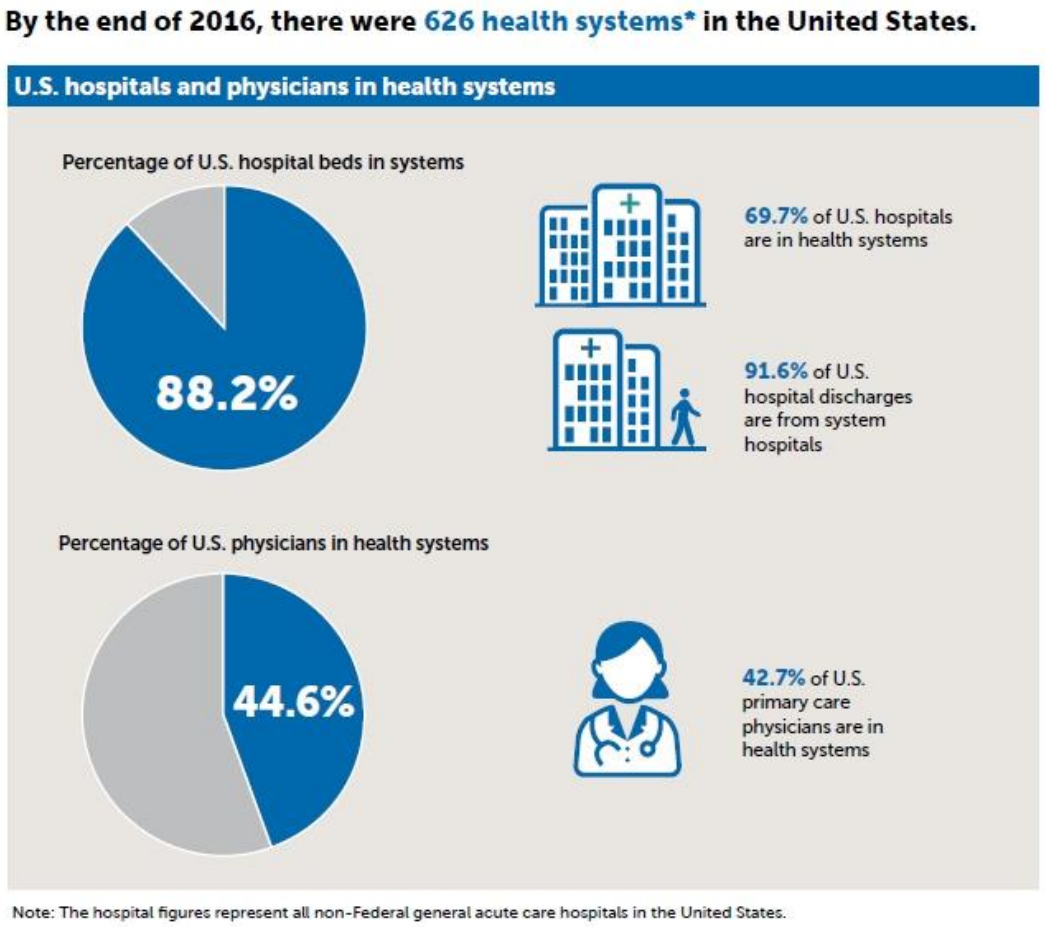
Previously Independent Sites now part of large integrated health systems
increasingly sophisticated data warehouses



Nodes are Operational Clusters Using Common Data

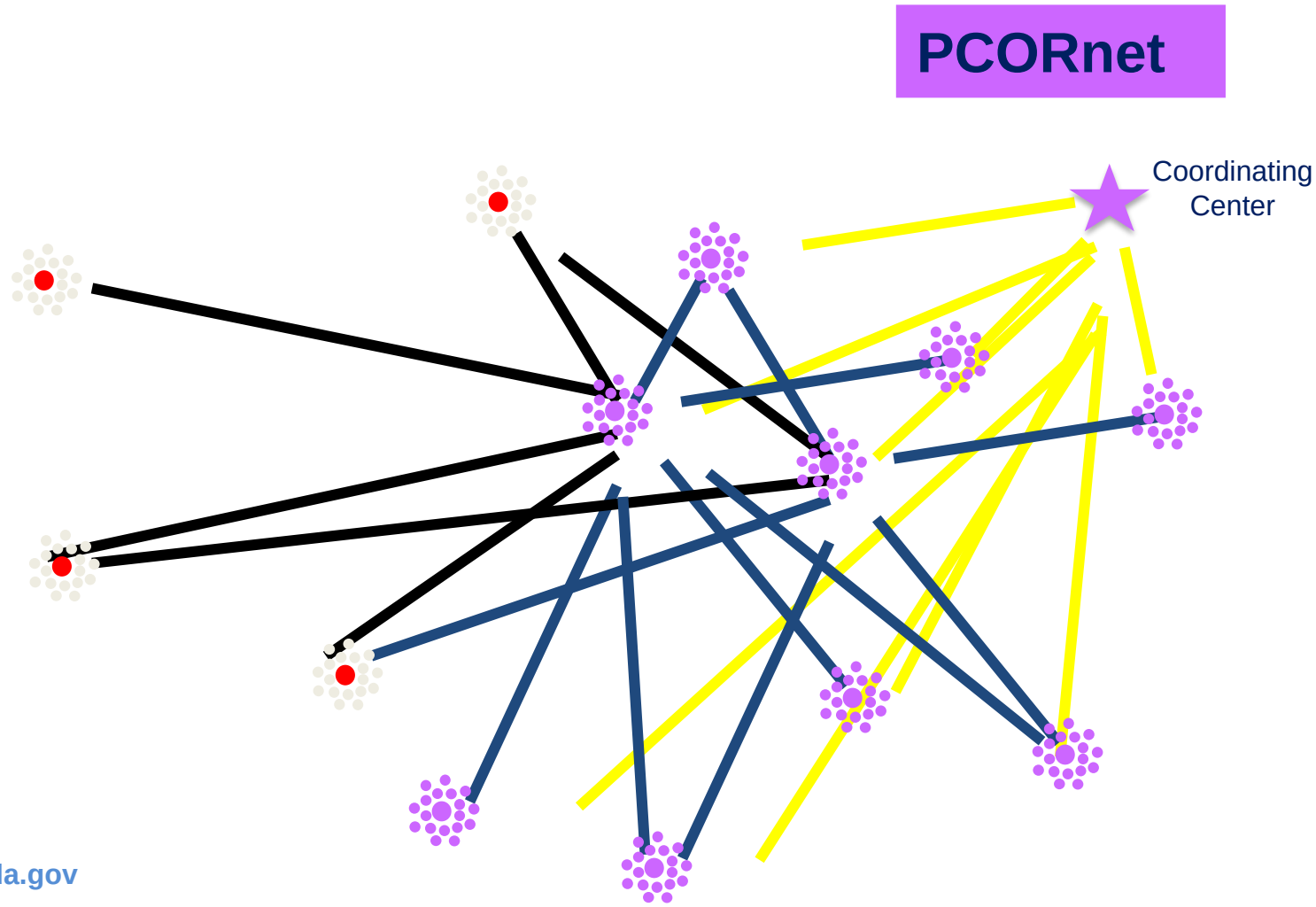


U.S. Hospitals & physicians in Health Systems

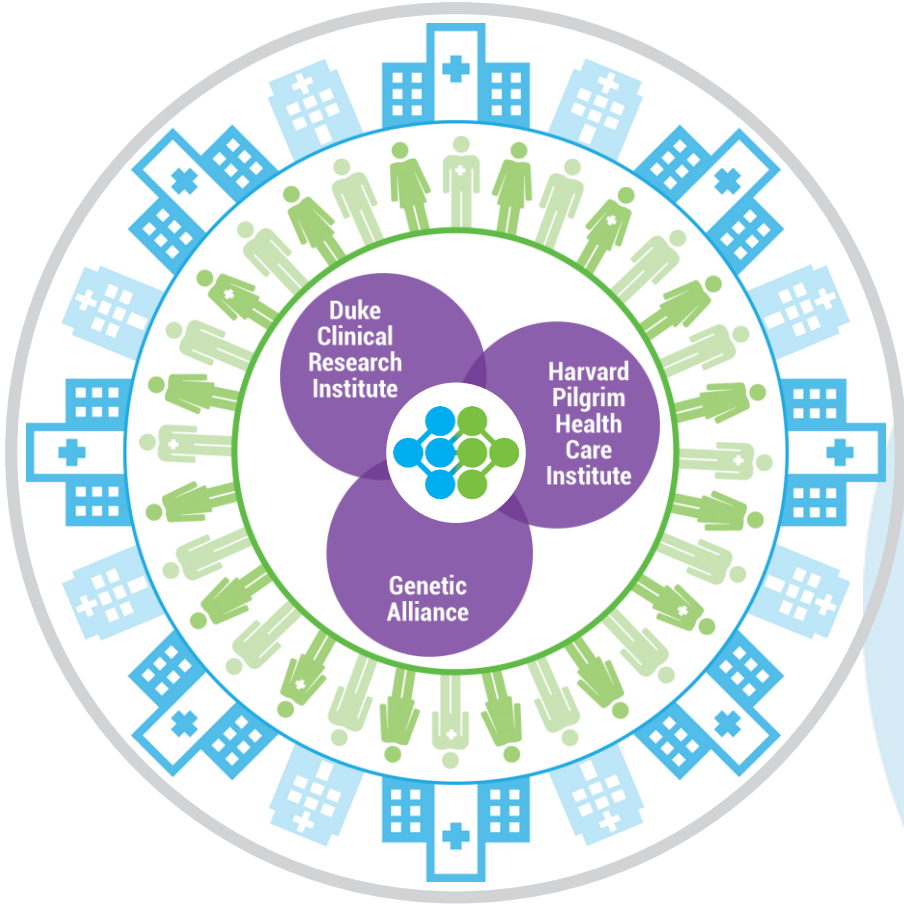


Source: Agency for Healthcare Research & Quality. Compendium of U.S. healthcare systems, 2016. Data highlights. Available at: https://www.ahrq.gov/chsp/compendium/index.html

Post Market Studies, including comparative effectiveness



PCORnet® embodies a "network of networks" that harnesses the power of partnerships



20 Patient-Powered Research Networks (**PPRNs**) + 13 Clinical Data Research Networks (**CDRNs**) + 1 Coordinating Center = pcornet®

A national infrastructure for people-centered clinical research

PPRNs

American BRCA Outcomes and Utilization of Testing Patient-Powered Research Network (ABOUT Network)
University of South Florida

ARthritis patient Partnership with comparative Effectiveness Researchers (AR-PoWER PPRN)
Global Healthy Living Foundation

CCFA Partners Patient Powered Research Network
Crohn's and Colitis Foundation of America

Collaborative Patient-Centered Rare Epilepsy Network (REN)
Epilepsy Foundation

Community and Patient-Partnered Centers of Excellence for Behavioral Health
University of California Los Angeles

Community-Engaged Network for All (CENA)
Genetic Alliance, Inc.

COPD Patient Powered Research Network
COPD Foundation

DuchenneConnect Registry Network
Parent Project Muscular Dystrophy

Health eHeart Alliance
University of California, San Francisco (UCSF)

ImproveCareNow: A Learning Health System for Children with Crohn's Disease and Ulcerative Colitis
Cincinnati Children's Hospital Medical Center

Interactive Autism Network
Kennedy Krieger Institute

Mood Patient-Powered Research Network
Massachusetts General Hospital

Multiple Sclerosis Patient-Powered Research Network
Accelerated Cure Project for Multiple Sclerosis

National Alzheimer's and Dementia Patient and Caregiver-Powered Research Network
Mayo Clinic

NephCure Kidney International
Arbor Research Collaborative for Health

Patients, Advocates and Rheumatology Teams Network for Research and Service (PARTNERS) Consortium
Duke University

Phelan-McDermid Syndrome Data Network
Phelan-McDermid Syndrome Foundation

PI Patient Research Connection: PI-CONNECT
Immune Deficiency Foundation

Population Research in Identity and Disparities for Equality Patient-Powered Research Network (PRIDEnet)
University of California San Francisco

Vasculitis Patient Powered Research Network
University of Pennsylvania

pcornet®

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

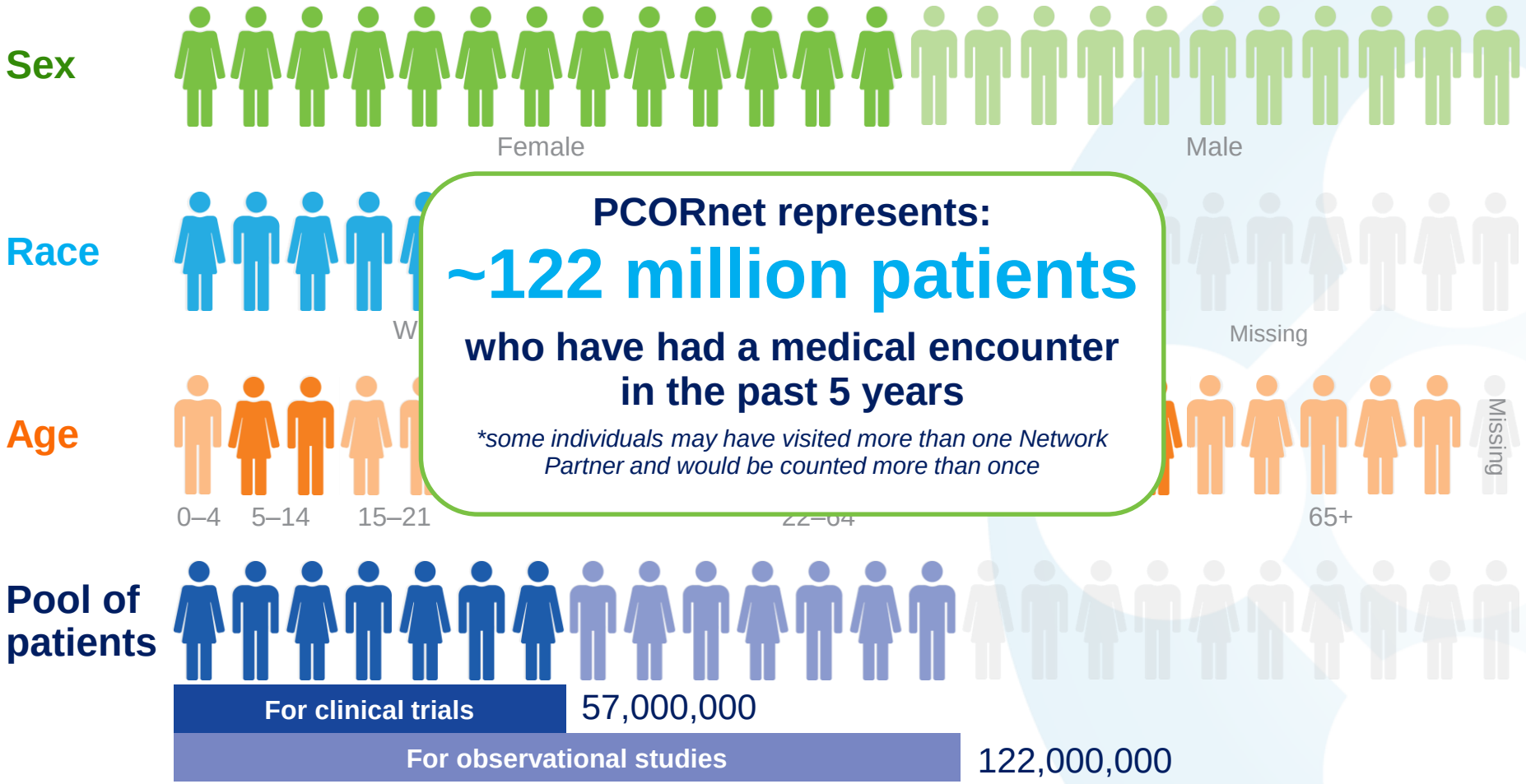
Patient-oriented SCAlable National Network for Effectiveness Research (pSCANNER)
University of California, San Diego (UCSD)

PaTH: Towards a Learning Health System
University of Pittsburgh

Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS)
Harvard University

pcornet®

Resulting in a national evidence system with “research readiness”



pcornet®

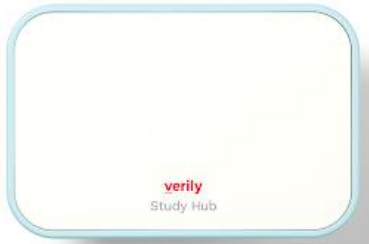
Continuous monitoring through passive sensors

Sleep sensor

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

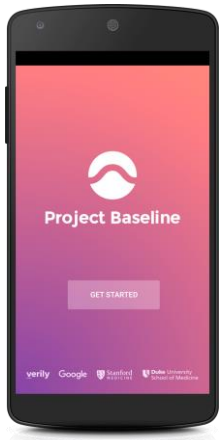
Study watch

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



Study hub

Safely sends device data to secure, encrypted Baseline database



App

Mobile interface for self-reported and passive data acquisitions

Project Baseline

Big Challenges in Biomedicine

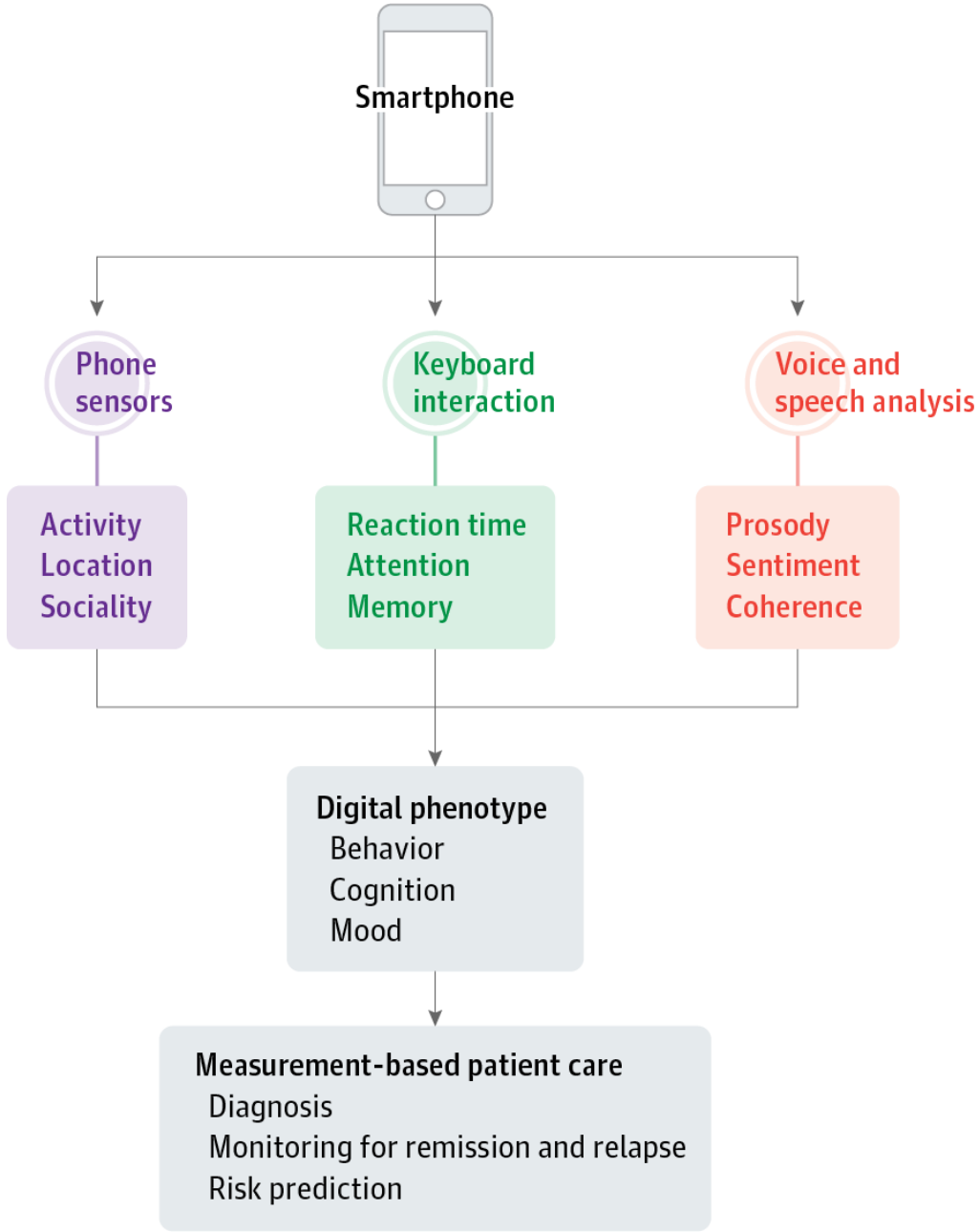
- **Lack of significant information over time dimension**
  - Measurements to assess biology and human health are made periodically in visits to healthcare or for research
- **Missing systems biology**
  - When developing concepts of human biology or drug development we make limited measurements focused 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

Duke | FORGE HEALTH DATA IN ACTION

The process of digital phenotyping

Digital phenotyping involves collecting sensor, keyboard, and voice and speech data from smartphones to measure behavior, cognition, and mood.

The JAMA Network



Insel TR. Digital phenotyping technology for a new science of behavior. JAMA. 2017;318:1215-16. doi:10.1001/jama.2017.11295

1 in 20

Google searches are health related

Our mission

Make health information universally accessible and useful.

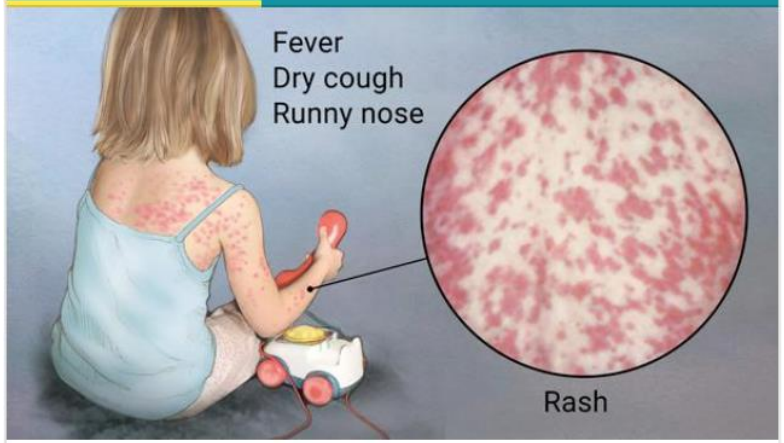
verily

Information structure

Measles

Also called: rubeola

ABOUT | SYMPTOMS | TREATMENTS



A viral infection that's serious for small children but is easily preventable by a vaccine.

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

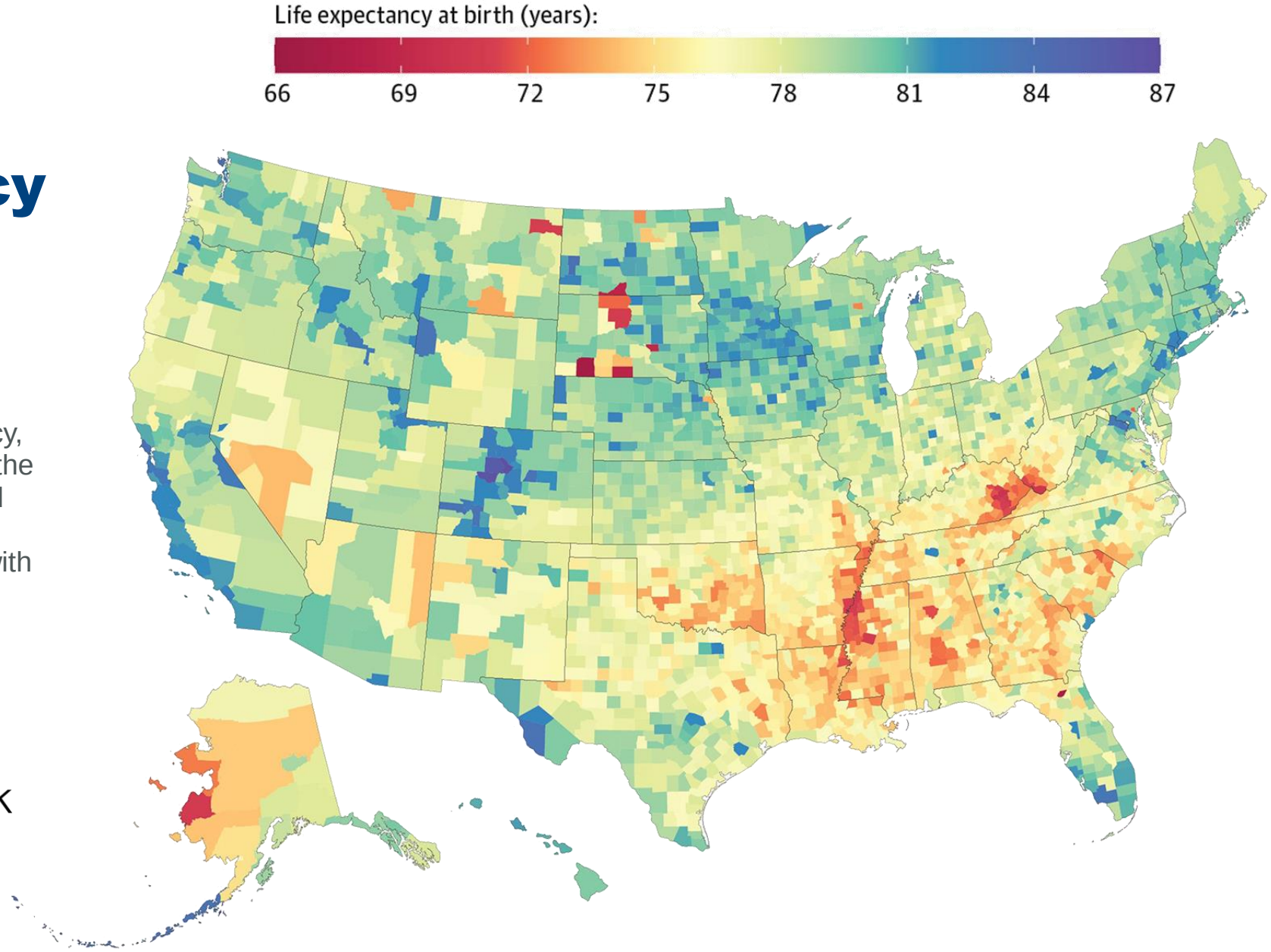
MMR vaccine

Medications

Duke | FORGE
HEALTH DATA IN ACTION

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.



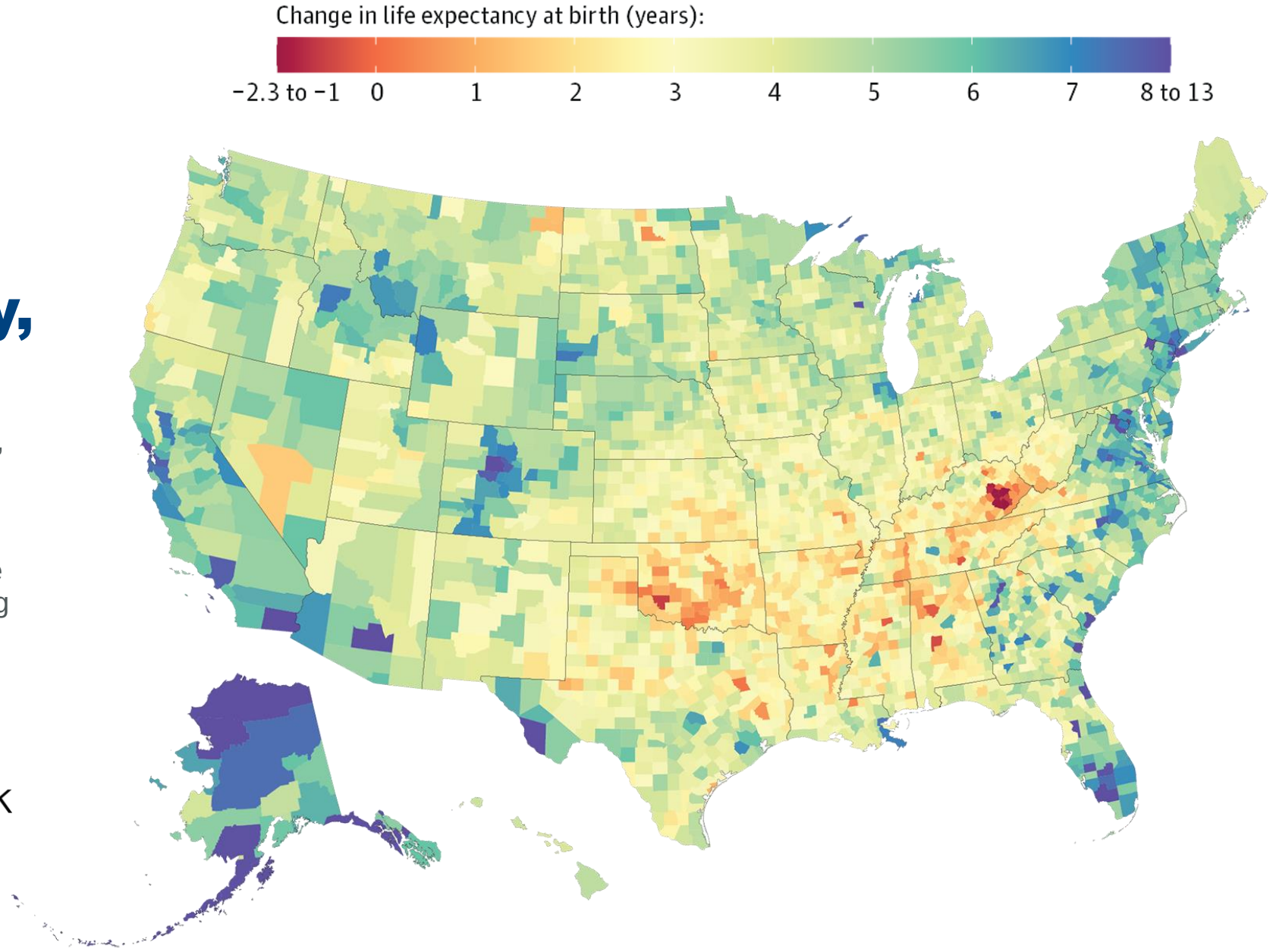
The JAMA Network

Dwyer-Lindgren L, et al. Inequalities in life expectancy among US counties, 1980 to 2014 - temporal trends and key drivers. JAMA Intern Med. 2017;177:1003-11. doi:10.1001/jamainternmed.2017.0918

Duke | FORGE
HEALTH DATA IN ACTION

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.



The JAMA Network

Dwyer-Lindgren L, et al. Inequalities in life expectancy among US counties, 1980 to 2014 - temporal trends and key drivers. JAMA Intern Med. 2017;177:1003-11. doi:10.1001/jamainternmed.2017.0918

From: **Inequalities in Life Expectancy Among US Counties, 1980 to 2014Temporal Trends and Key Drivers**

JAMA Intern Med. Published online May 08, 2017. doi:10.1001/jamainternmed.2017.0918

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

Variable	Summary Statistics, Mean (SD) [Range]	Bivariate Regression Results	
		Coefficient (SE)	R^2
Socioeconomic and race/Ethnicity factors			
Population below the poverty line, %	16.3 (6.4) [3.1-62.0]	−0.24 (0.005)	0.47
Median household income, log $	10.6 (0.2) [9.8-11.6]	6.06 (0.130)	0.41
Graduates, age ≥25 y, %			
High school	83.7 (7.2) [46.3-98.6]	0.20 (0.004)	0.42
College	19.2 (8.6) [4.2-72.0]	0.15 (0.004)	0.34
Unemployment rate, age ≥16 y, %	9.1 (3.2) [2.1-27.4]	−0.29 (0.011)	0.18
Black population, %	9.4 (14.7) [0-85.8]	−0.07 (0.002)	0.24
American Indian, Native Alaskan, and Native Hawaiian population, %	2.3 (7.9) [0-97.2]	−0.06 (0.005)	0.04
Hispanic population, %	8.1 (13.1) [0-95.9]	0.02 (0.003)	0.01
Behavioral and metabolic risk factors, %			
Obesity prevalence, age ≥20 y	37.0 (4.3) [18.0-52.0]	−0.39 (0.006)	0.54
No leisure-time physical activity prevalence, age ≥20 y	27.0 (5.2) [11.7-47.2]	−0.34 (0.005)	0.62
Cigarette smoking prevalence, age ≥18 y	24.7 (4.1) [7.7-42.1]	−0.40 (0.007)	0.54
Hypertension prevalence, age ≥30 y	39.5 (3.6) [27.9-56.4]	−0.49 (0.007)	0.62
Diabetes prevalence, age ≥20 y	14.0 (2.4) [8.1-25.5]	−0.72 (0.011)	0.59
Health care factors			
Insured population, age <65 y, %	81.7 (5.7) [57.3-96.7]	0.15 (0.007)	0.14
Quality index	70.1 (11.5) [0-100]	0.10 (0.003)	0.28
Physicians per 1000 population, No.	1.1 (1.0) [0-4.4]	0.53 (0.039)	0.06

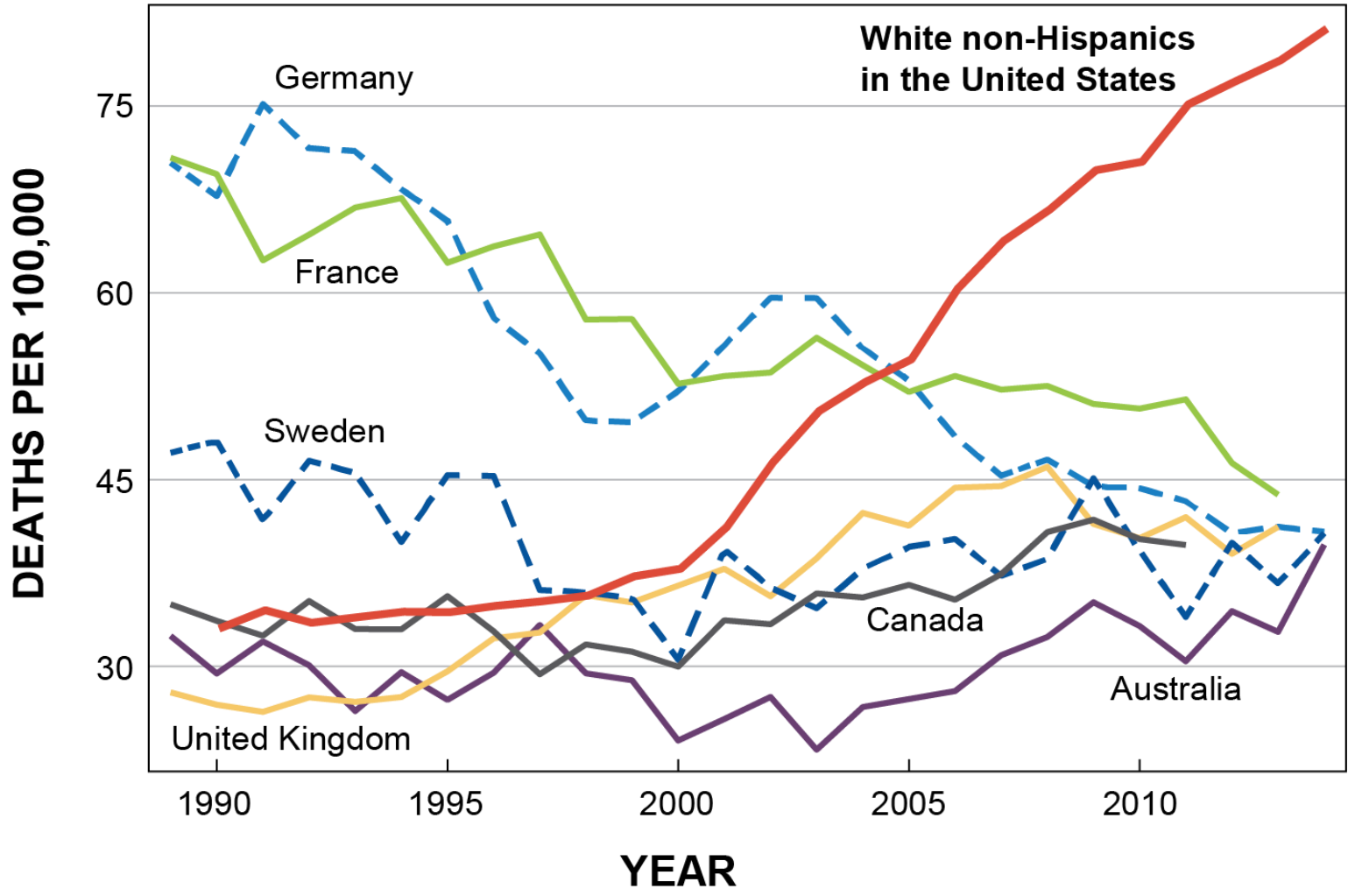
Abbreviation: SE, standard error.

Table Title:

Variables Included in the Regression Analysis With Summary Statistics and Bivariate Regression Results

Midlife mortality from “deaths of despair” across countries

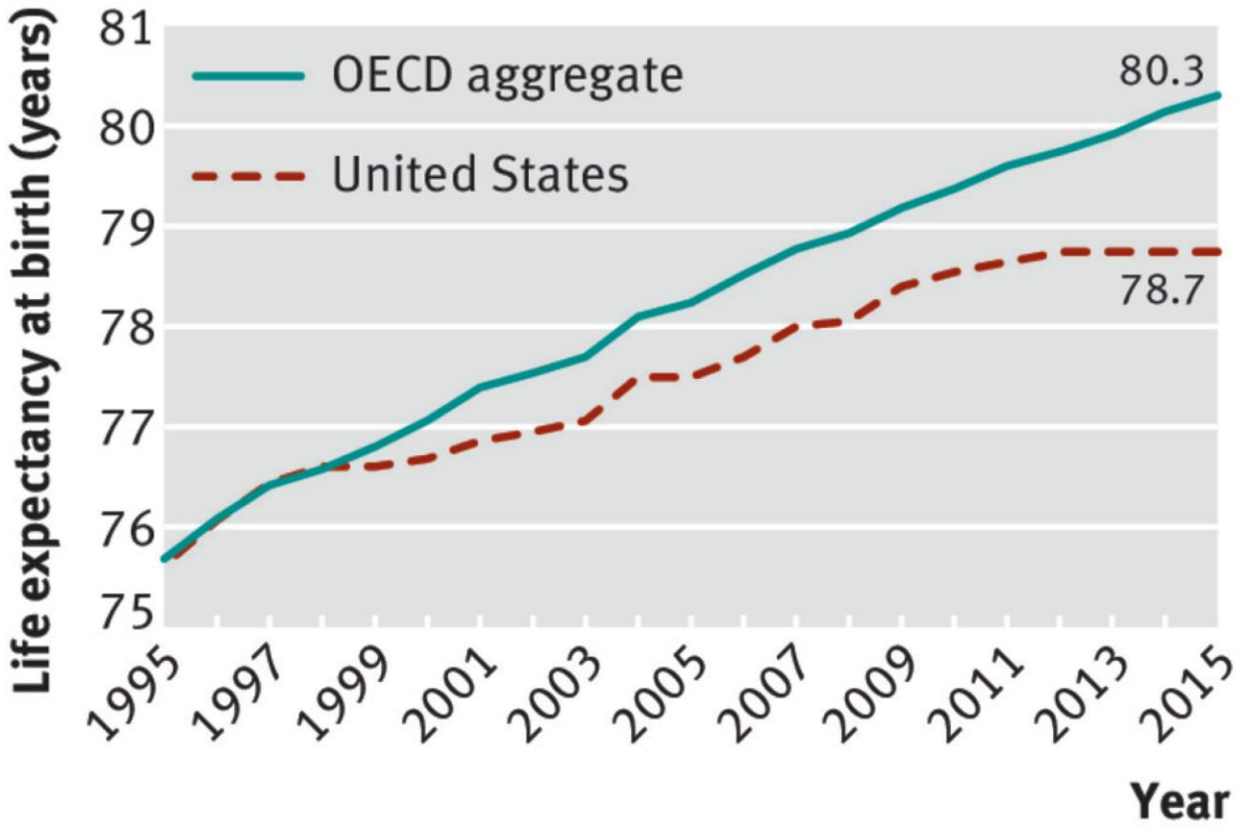
Men and women ages 50–54, deaths by drugs, alcohol, and suicide, 1989–2014



Source: “Mortality and morbidity in the 21st century” by Anne Case and Angus Deaton, Brookings Papers on Economic Activity, Spring 2017.

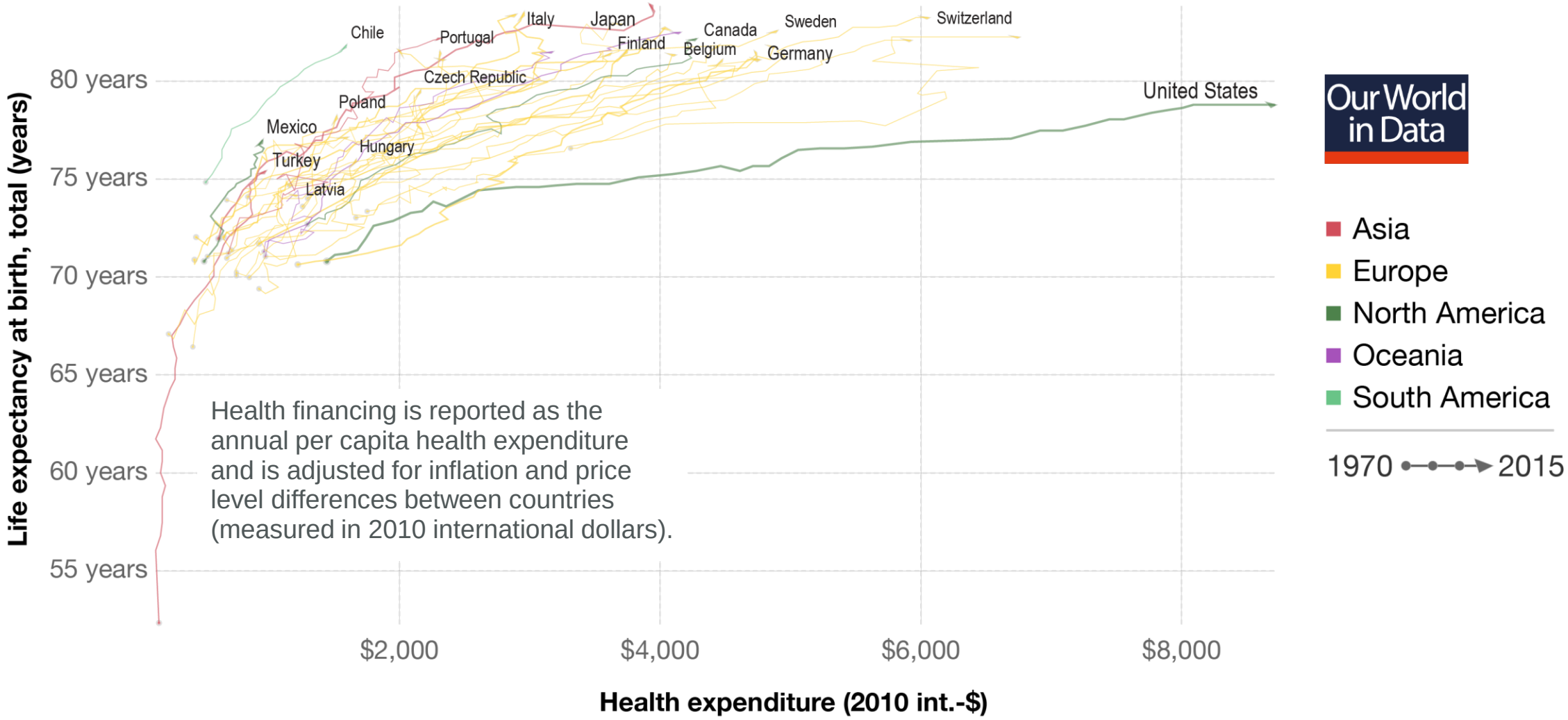
Life expectancy at birth in the US and the Organisation for Economic Cooperation and Development, 1995–2015

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.



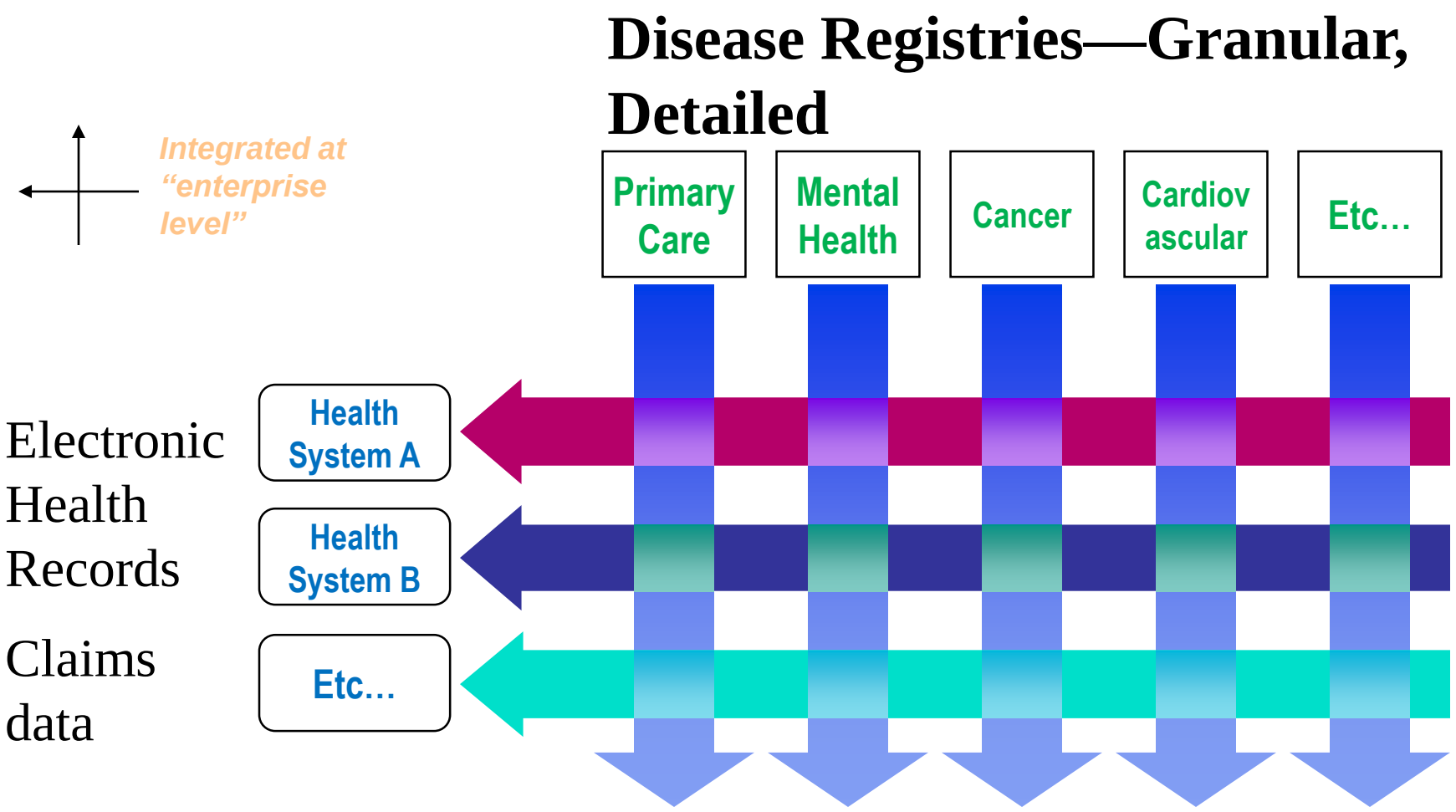
Woolf SH, Aron L. BMJ 2018;360:bmj.k496

Life expectancy vs. health expenditure, 1970 to 2015

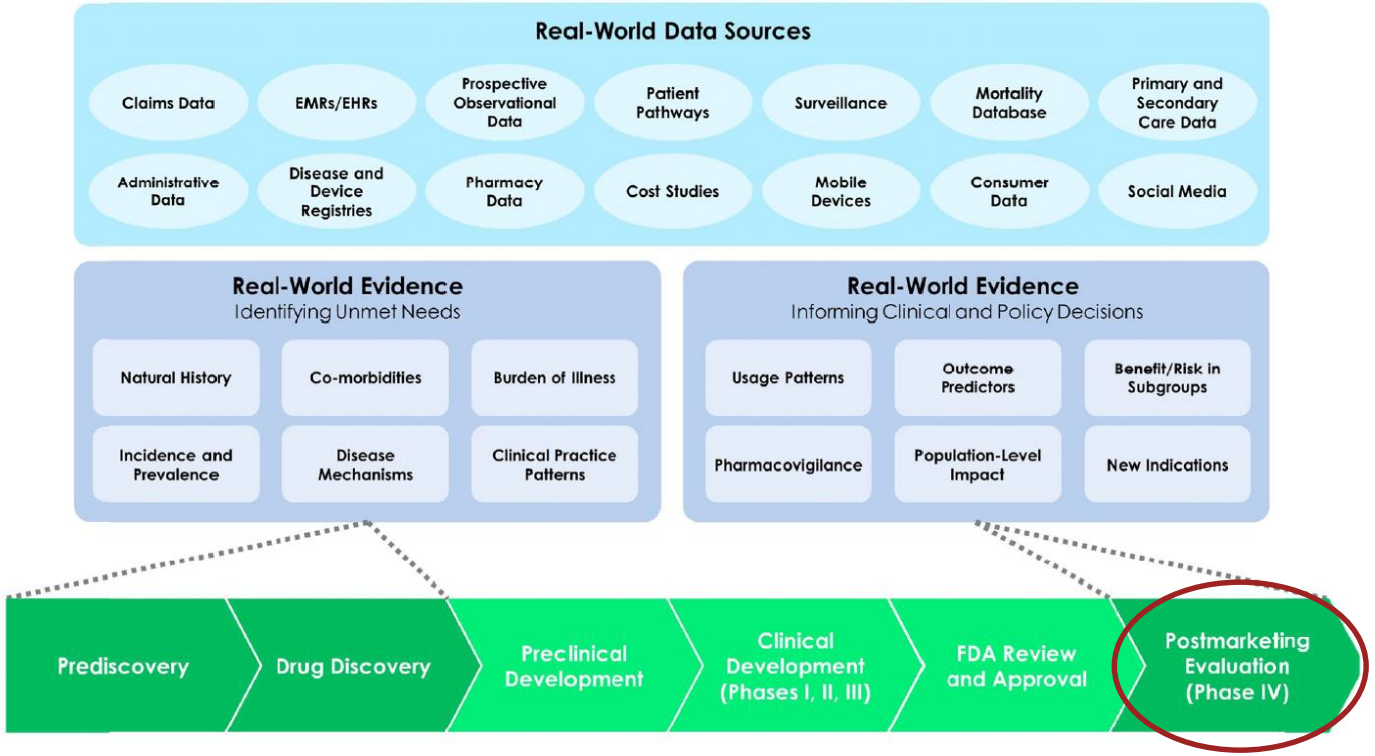


Source: World Bank—WDI, Health Expenditure and Financing – OECDstat (2017)
OurWorldinData.org/the-link-between-life-expectancy-and-health-spending-us-focus • CC BY-SA

Fundamental Informatics Infrastucture-- Matrix Organizational Structure



Real World Data vs Evidence



National Academies of Sciences, Engineering, and Medicine. 2017. *Realworld evidence generation and evaluation of therapeutics: Proceedings of a workshop*. Washington, DC: The National Academies Press. doi: 10.17226/24685

Real World Data and Efficacy

SOUNDING BOARD

Real-World Evidence — What Is It and What Can It Tell Us?

- 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

Cures provisions (Sec. 3022)

- Requires FDA to establish a program to evaluate the potential use of real world evidence to:
  - Help support the approval of new indications for an approved drug
  - Help support or satisfy post approval study requirements

PDUFA RWE provisions

- Tracks with Cures Act
- Requires FDA to establish a program to evaluate the potential use of real world evidence to:
  - Help support the approval of new indications for an approved drug
  - Help support or satisfy post approval study requirements

Reinforcing of a Learning Health Care System:

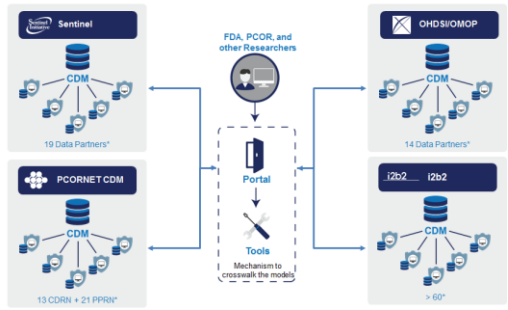
- Doesn't change approval standards, rather it better supports and enables use of data and evidence on outcomes that are hard to get from traditional RCTs (e.g., outcomes that are too costly, too small populations with particular clinical features, too long follow-up needed, diff impact in diff clinical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

Laying the Foundation

FDA

Data Standards

Stakeholder Engagement



Duke | MARGOLIS CENTER for Health Policy

CLINICAL TRIALS TRANSFORMATION INITIATIVE

Guidances

Draft

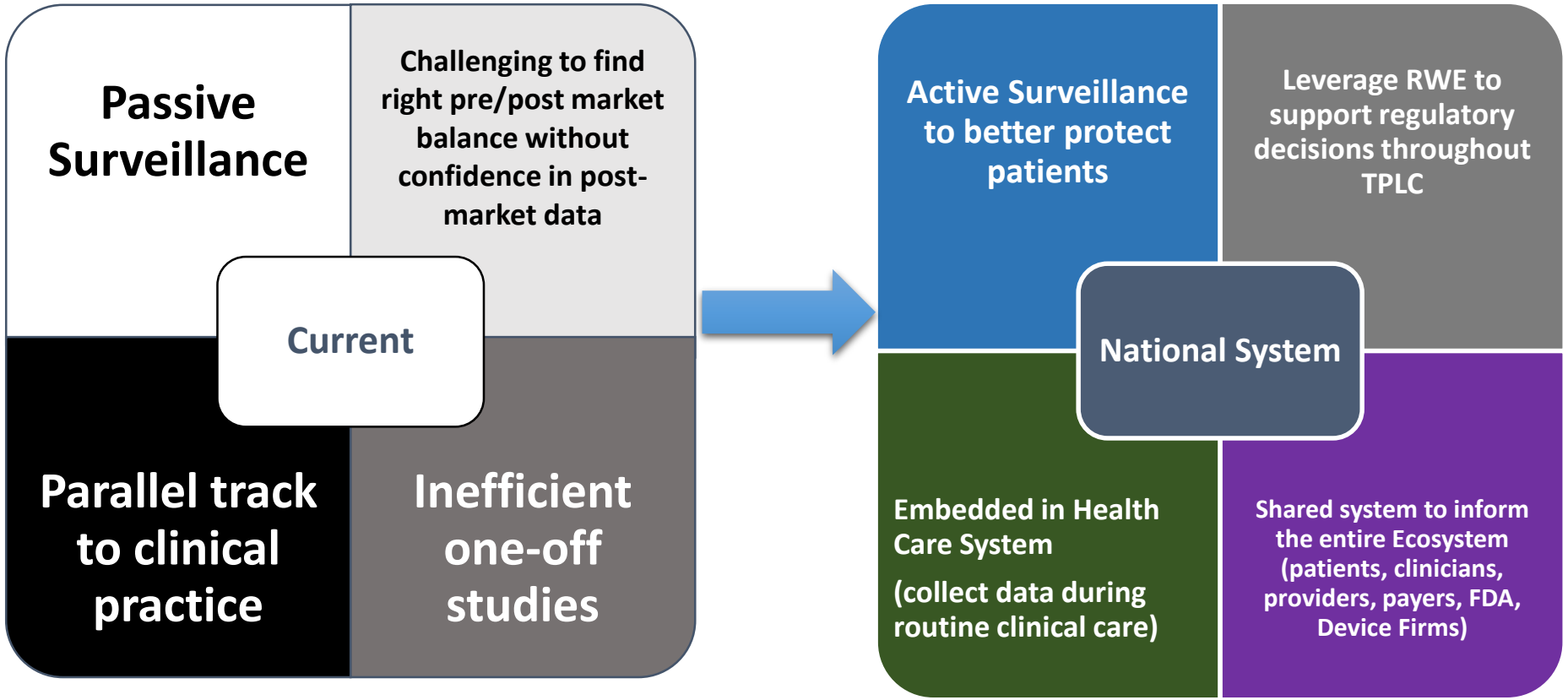
Use of Electronic Health Record Data in Clinical Investigations

Demonstratio

Electronic Source Data in Clinical Investigations

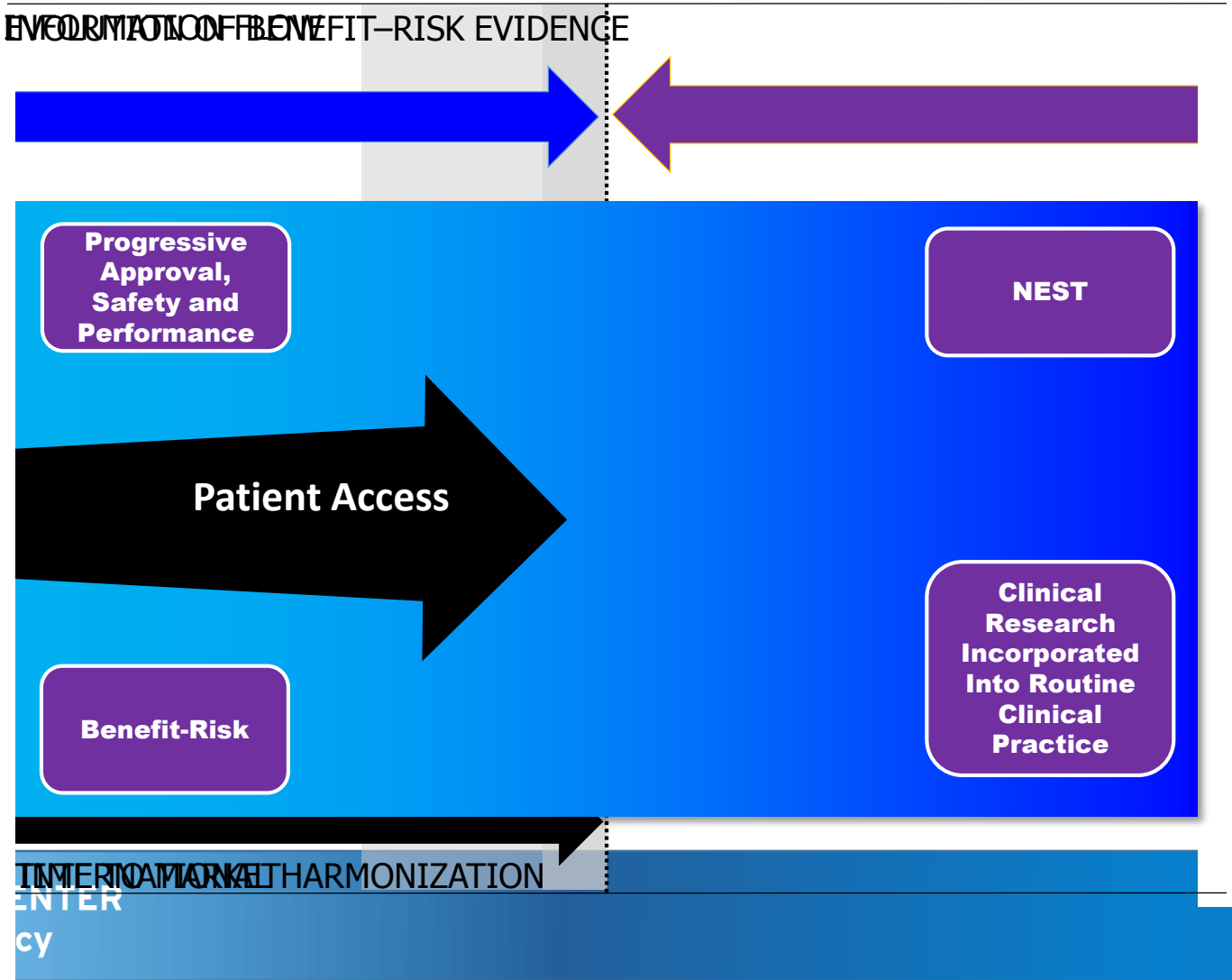
Use of Electronic Informed Consent

National System Paradigm Shift



Learning Medical Device Ecosystem

Total Product Life Cycle (TPLC) Framework



Duke | FORGE HEALTH DATA IN ACTION

WHY DEPRESSION?

DEPRESSION IS HIGHLY PREVALENT

300 M

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

7 YRS

average time from onset to treatment in the US [JAMA]

TREATMENT IS EFFECTIVE

70%

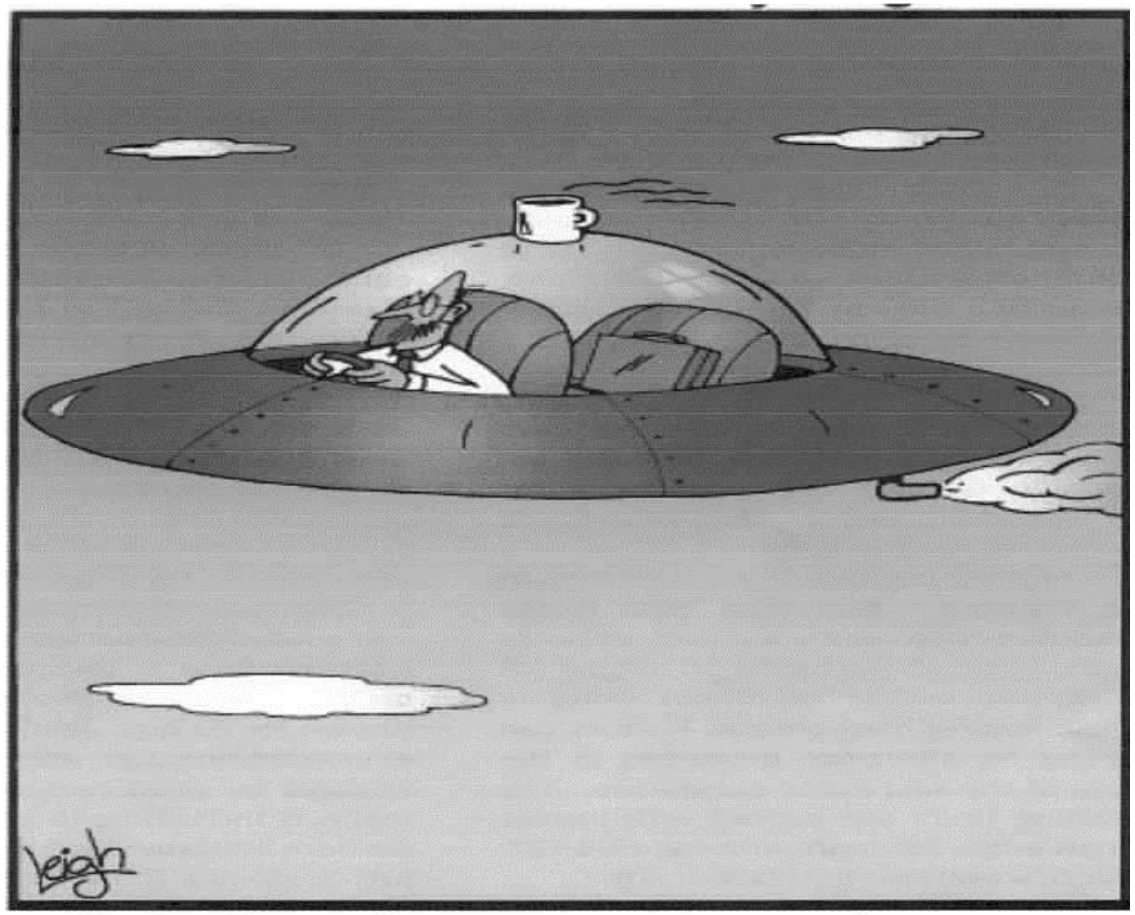
of patients can improve, often in a matter of weeks [NIMH]

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

PATIENT HEALTH QUESTIONNAIRE-9				
Over the last 2 weeks, how often have you been bothered by any of the following problems?	**Not at all**	**Several days**	**More than half the days**	**Nearly every day**
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself — or that you are a failure or have let yourself or your family down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
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	0	1	2	3
9. Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3



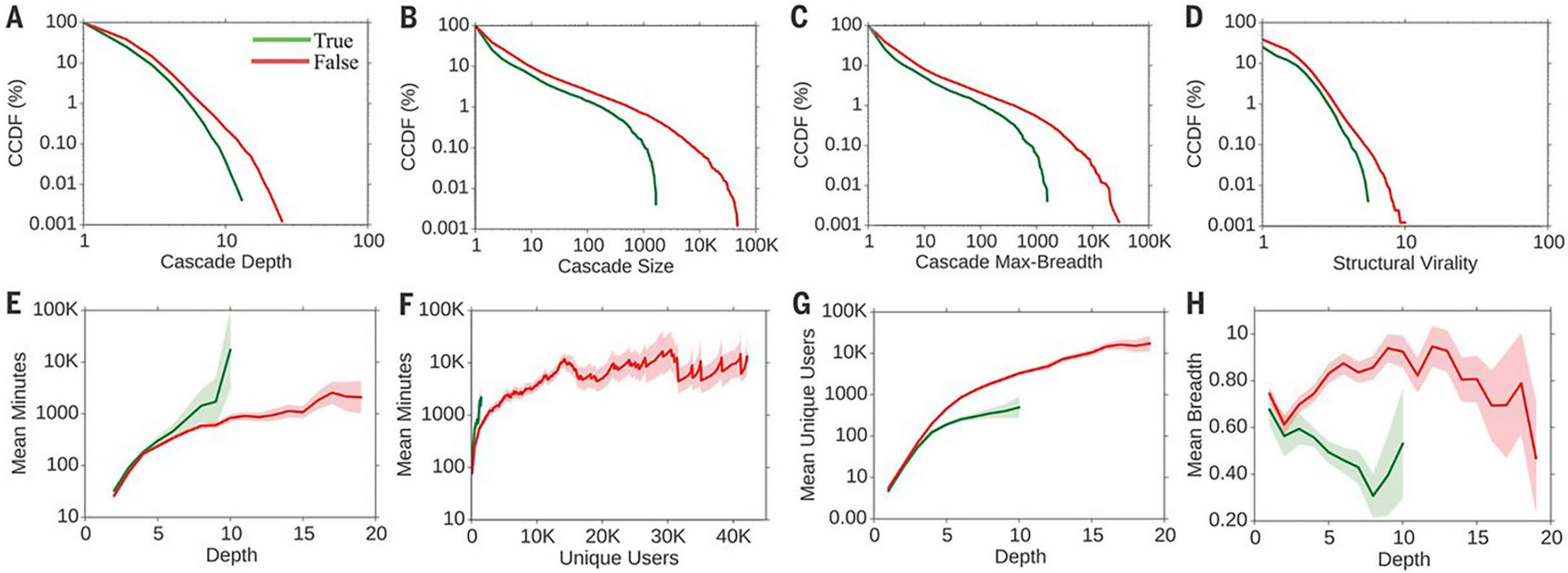
Technology advances; people stay the same.

Duke | FORGE
HEALTH DATA IN ACTION

Krista Kennell / Stone / Catwalker / Shutterstock / The Atlantic
https://www.theatlantic.com/technology/archive/2018/03/largest-study-ever-fake-news-mit-twitter/555104/

Duke | FORGE
HEALTH DATA IN ACTION

Complementary cumulative distribution functions (CCDFs) of true and false rumor cascades



Vosoughi S, et al. Science. 2018;359:1146-51.

Science
AAAS

Duke | FORGE
HEALTH DATA IN ACTION

Data Scientist:

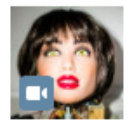
The Sexiest Job of the 21st Century

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."

Twitter's C.E.O., Dick Costolo, Is Set to Exit, Feeling Heat of Criticism



ROBOTICA EPISODE 5

Sex Dolls That Talk Back

STATE OF THE ART

For Twitter, Future Means Here and Now

Sidewalk Labs, a Start-Up Created by Google, Has Bold Aims to Improve City Living

Am
Bus
Eur
Reg

TECHNOLOGY

For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights

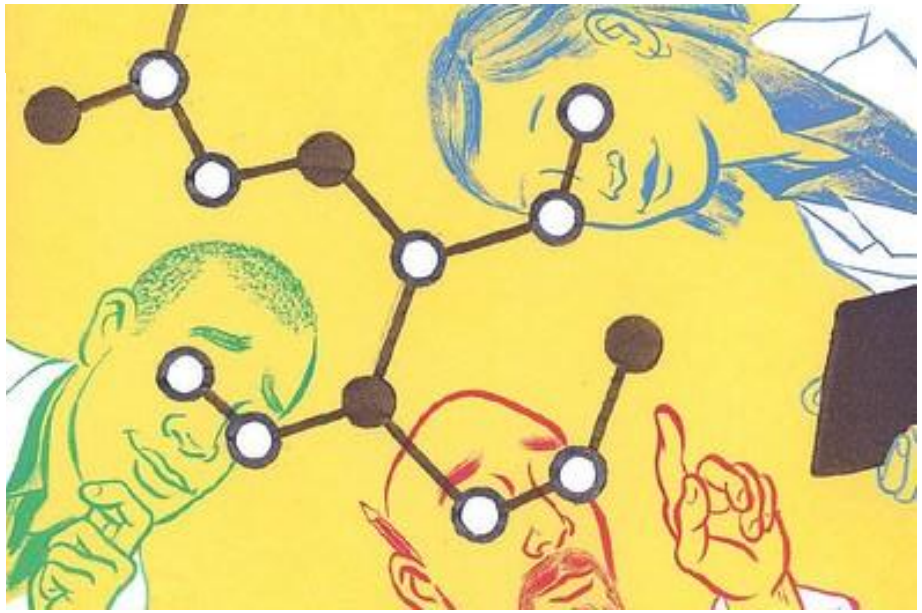
By STEVE LOHR AUG. 17, 2014

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

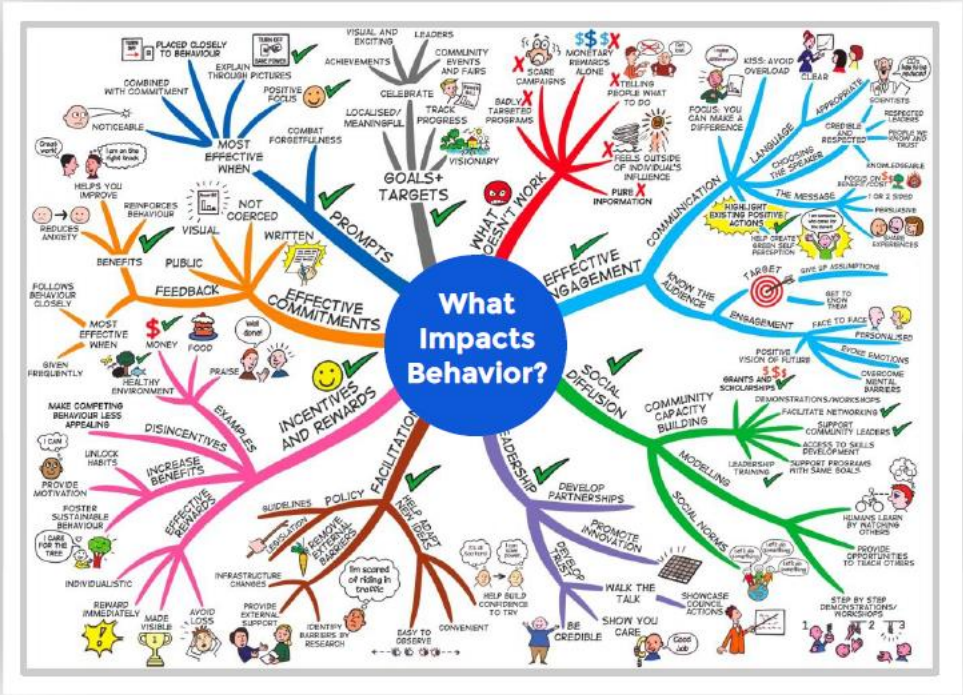
In January 2009, a mathematician at Cambridge University named Tim Gowers decided to use his blog to run an unusual social experiment. He picked out a difficult mathematical problem and tried to solve it completely in the open, using his blog to post ideas and partial progress. He issued an open invitation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

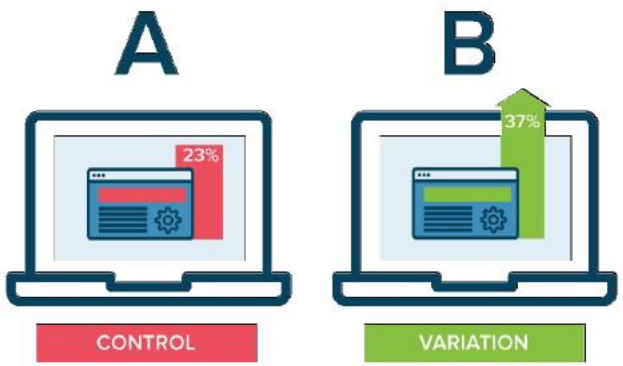


Several hours after Mr. Gowers opened up his blog for discussion, a Canadian-Hungarian mathematician posted a comment. Fifteen minutes later, an Arizona high-school math teacher chimed in. Three minutes after that, the UCLA mathematician Terence Tao commented. The discussion ignited, and in just six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

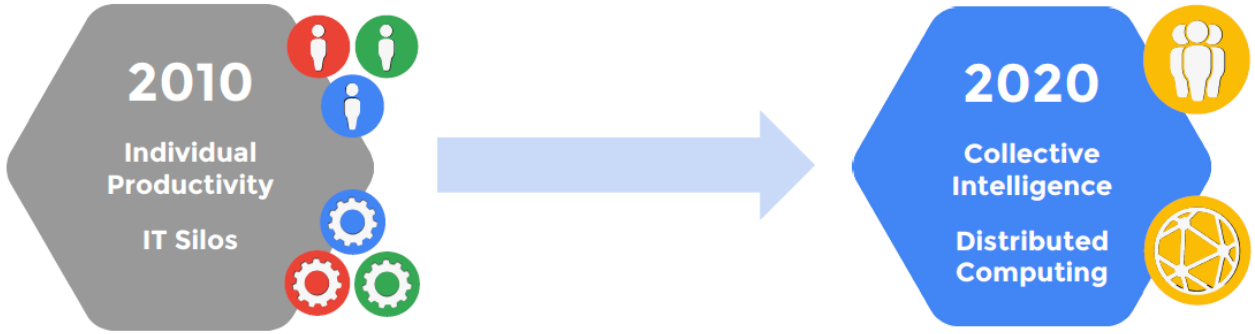
THE WALL STREET JOURNAL.

Data Activation and Testing Outcomes





Digital Transformation



- 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

Verily

Duke | FORGE HEALTH DATA IN ACTION

THE DOCTOR AND THE COMPUTER

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**