

BIG DATA AND THE FUTURE OF CV MEDICINE

Robert M Califf MD

Vice Chancellor for Health Data Science

Duke University

Advisor, Verily Life Sciences

Inova

April 27th, 2019

CONFLICTS OF INTEREST

- Employment
 - Duke University
 - Verily Life Sciences
- Corporate Board
 - Cytokinetics
- Consulting
 - Merck
 - Boeringer Ingelheim
 - Amgen
 - Biogen
 - Genentech

Alphabet

verily

Verily partners closely with Google teams on commercial tools and applications across the healthcare vertical.

Google



User experience



Science



Holistic health platforms



Tools and medical Devices



Regulatory expertise



Software and machine learning



Google Cloud Platform



Search, Advertising & Maps



DeepMind



YouTube



Research



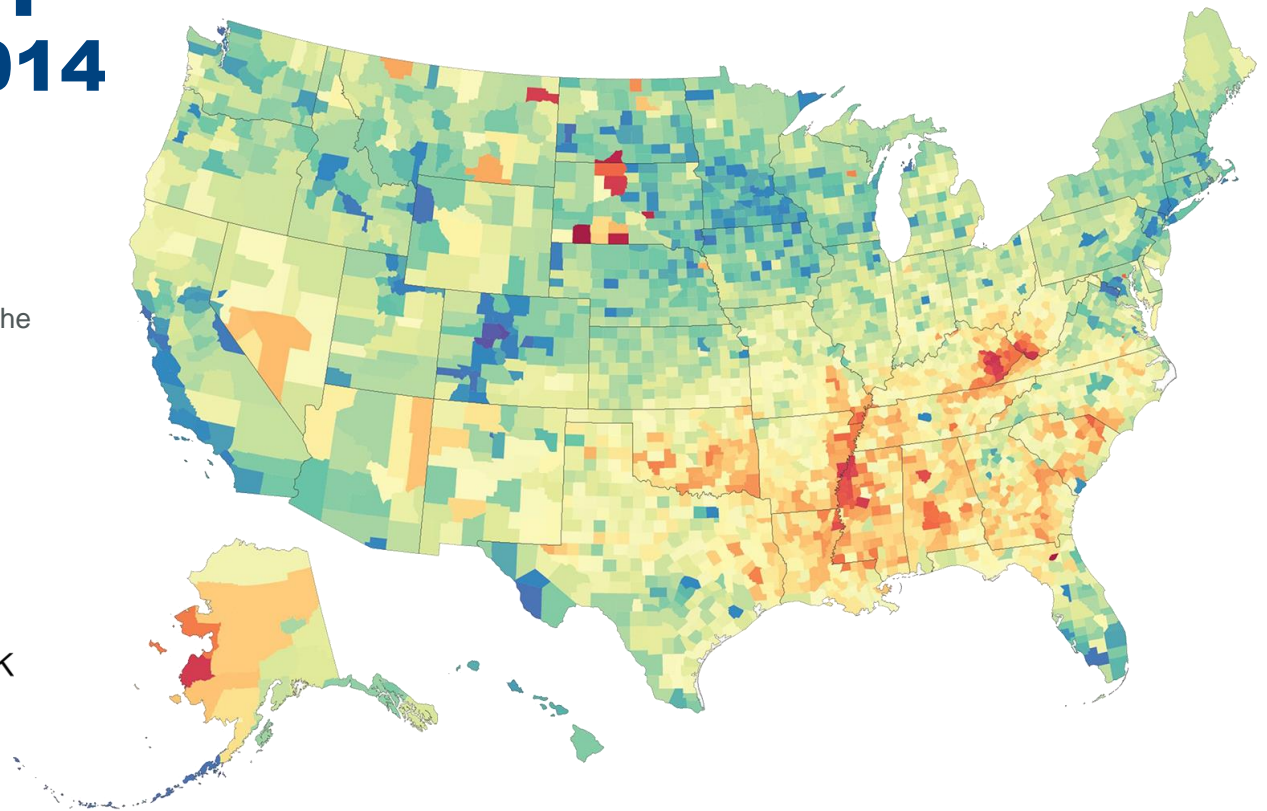
Android

LIFE EXPECTANCY AT BIRTH BY COUNTY, 2014

Life expectancy at birth (years):



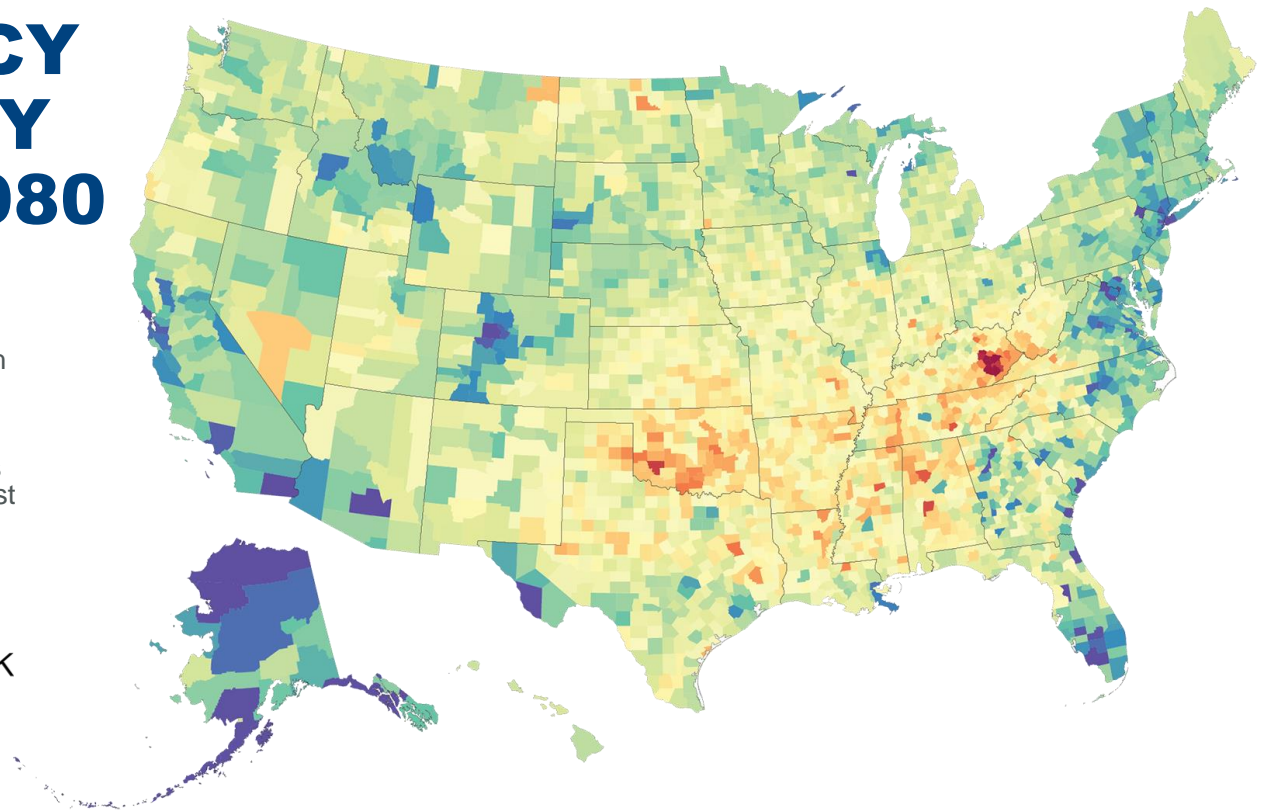
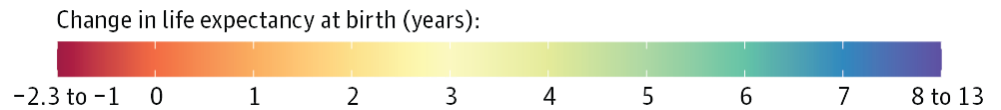
- Counties in South Dakota and North Dakota had the lowest life expectancy, and counties along the lower half of the Mississippi, in eastern Kentucky, and southwestern West Virginia also had very low life expectancy compared with the rest of the country. Counties in central Colorado had the highest life expectancies.



The **JAMA** Network

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

From: Inequalities in Life Expectancy Among US Counties, 1980 to 2014 Temporal 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

From: **Trends and Patterns of Geographic Variation in Cardiovascular Mortality Among US Counties, 1980-2014**

JAMA. 2017;317(19):1976-1992. doi:10.1001/jama.2017.4150

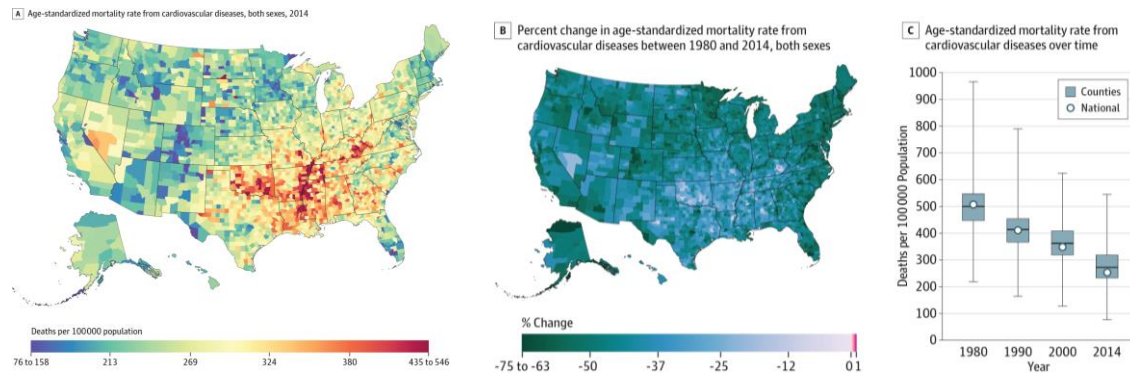
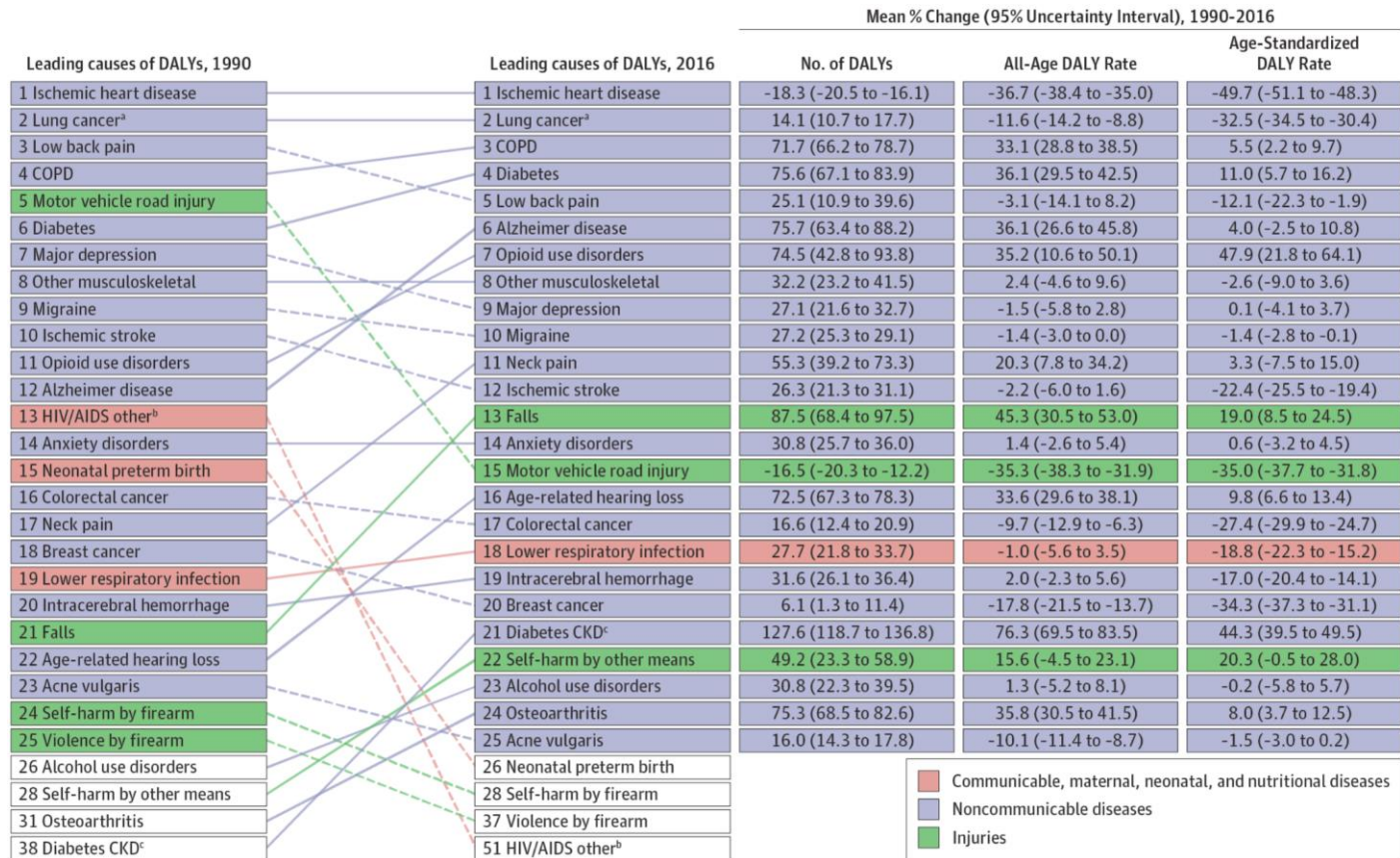


Figure Legend:

US County-Level Mortality From Cardiovascular Diseases A, Age-standardized mortality rate for both sexes combined in 2014. B, Percent change in the age-standardized mortality rate for both sexes combined between 1980 and 2014. In panel A, the color scale is truncated at approximately the 1st and 99th percentiles as indicated by the range given on the scale. In panel B, the color scale is similarly truncated at the 1st percentile but not at the 99th percentile to avoid combining counties with decreases in the mortality rate and counties with increases in the mortality rate into a single group. C, Age-standardized mortality rate in 1980, 1990, 2000, and 2014. The bottom border, middle line, and top border of the boxes indicate the 25th, 50th, and 75th percentiles, respectively, across all counties; whiskers, the full range across counties; and circles, the national-level rate.

Top 25 Causes of Disability-Adjusted Life-Years (DALYs) and % Change in Number of DALYs, All-Age DALYs, and Age-Standardized DALYs, 1990-2016

Dotted lines: leading cause decreased in rank between 1990-2016; solid lines: cause maintained/ascended to higher ranking.

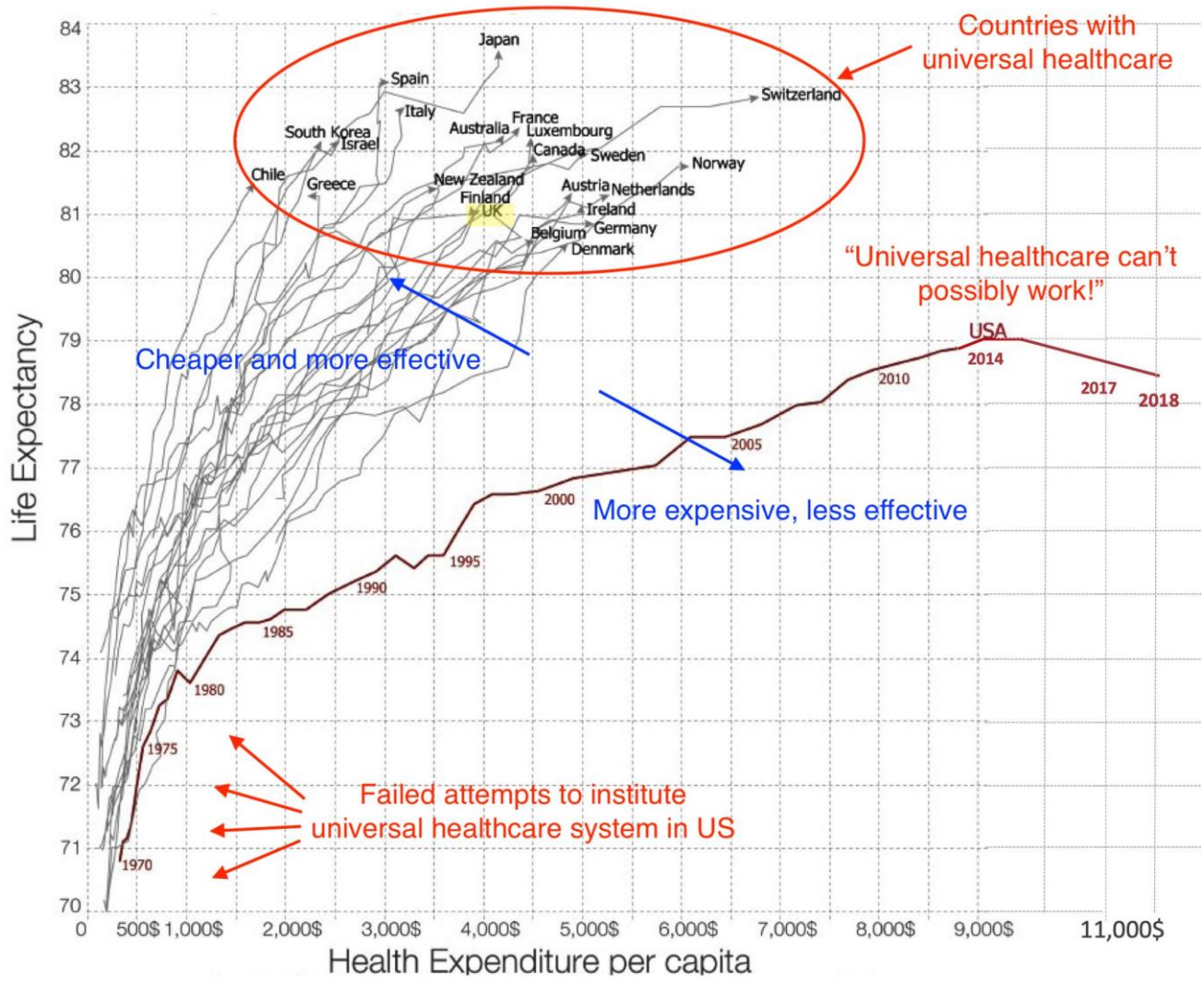


Life expectancy at birth (years) in 18 high income countries for women and men during 2010-16 and 1990-2015.



Jessica Y Ho, and Arun S Hendi BMJ 2018;362:bmj.k2562

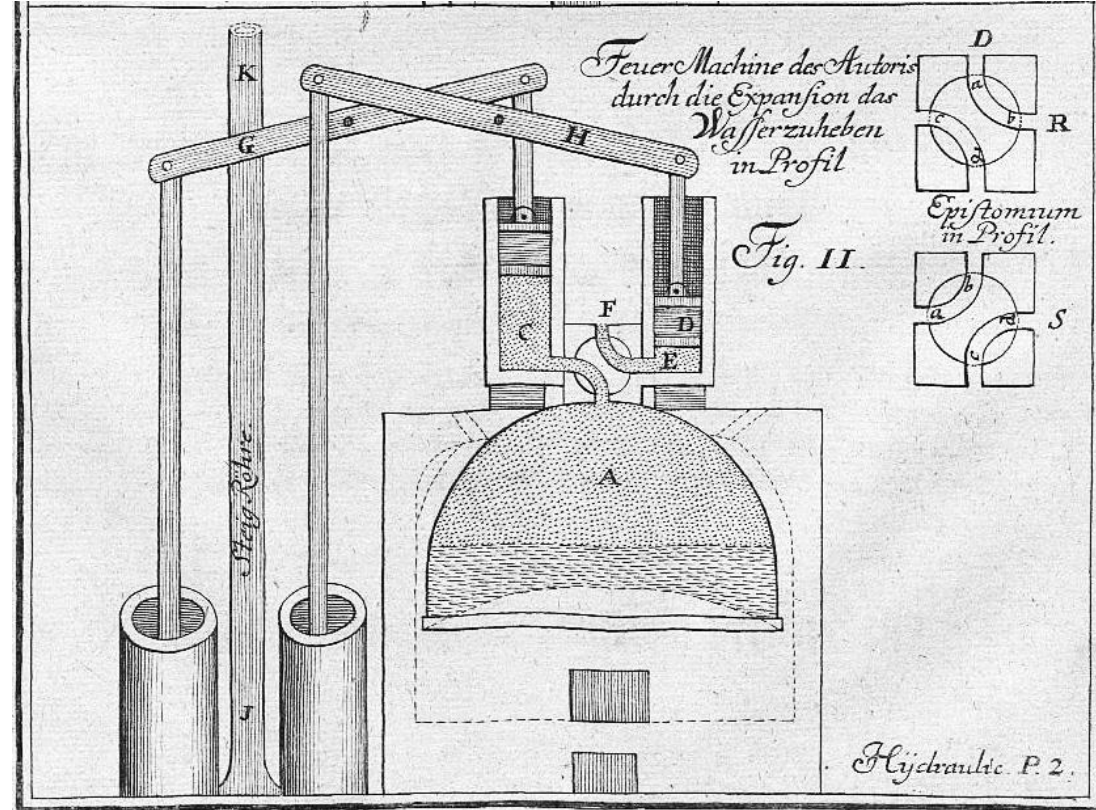




Four industrial revolutions

FIRST

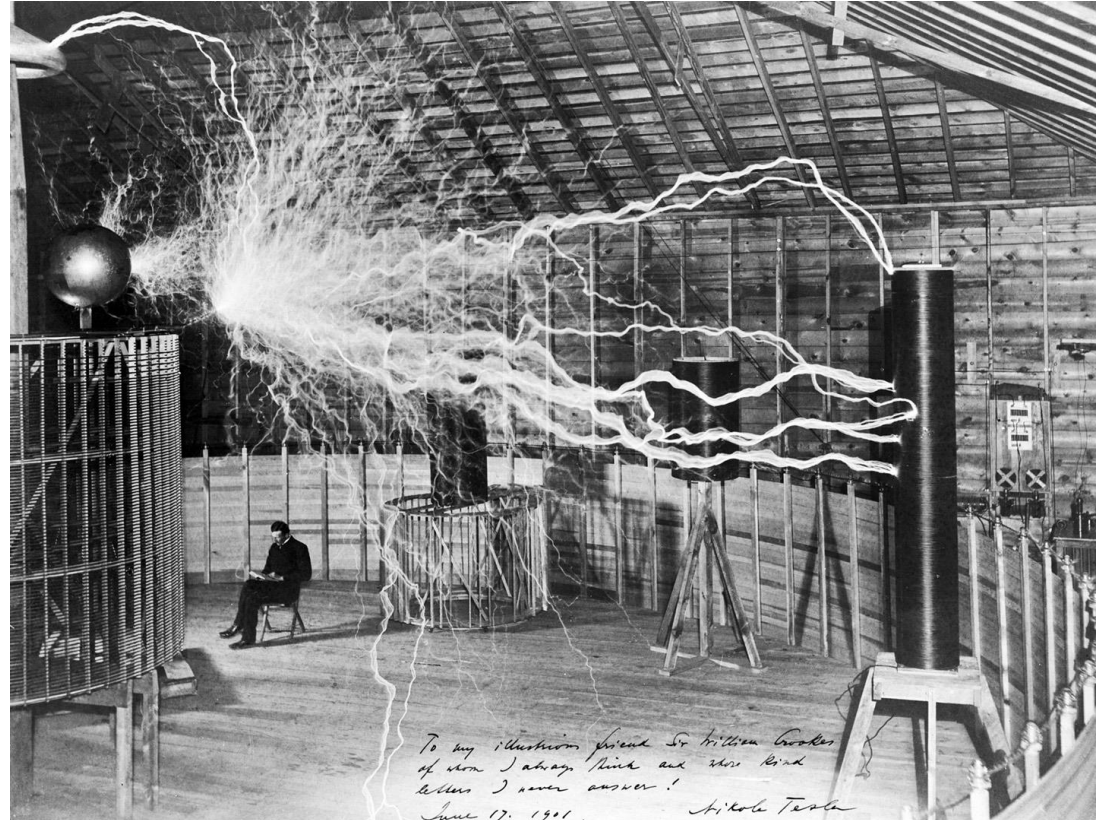
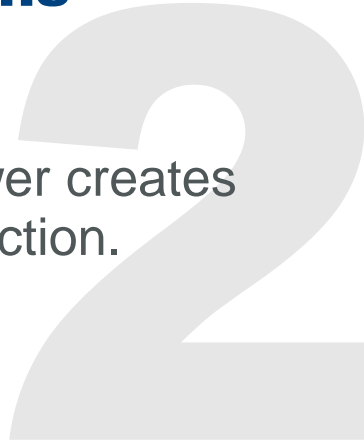
Water and steam power mechanize production.



Four industrial revolutions

SECOND

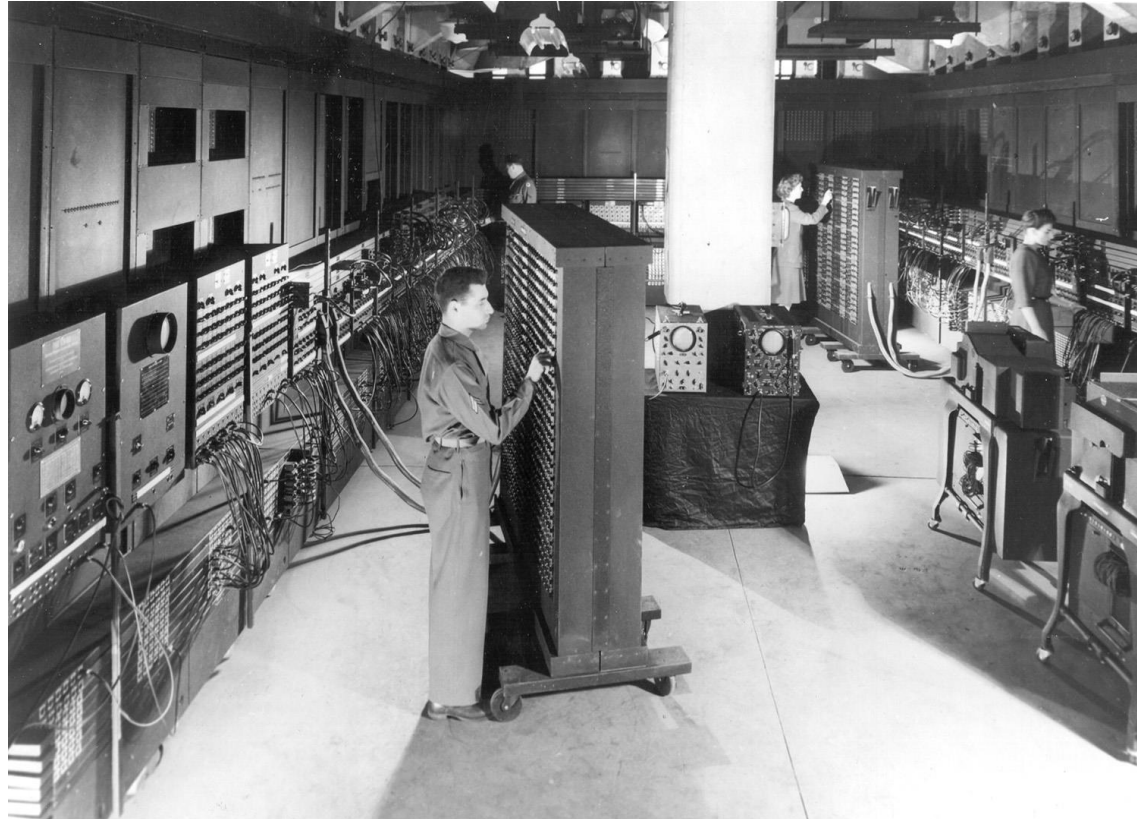
Electric power creates mass production.



Four industrial revolutions

THIRD

Electronics and information technology automate production.

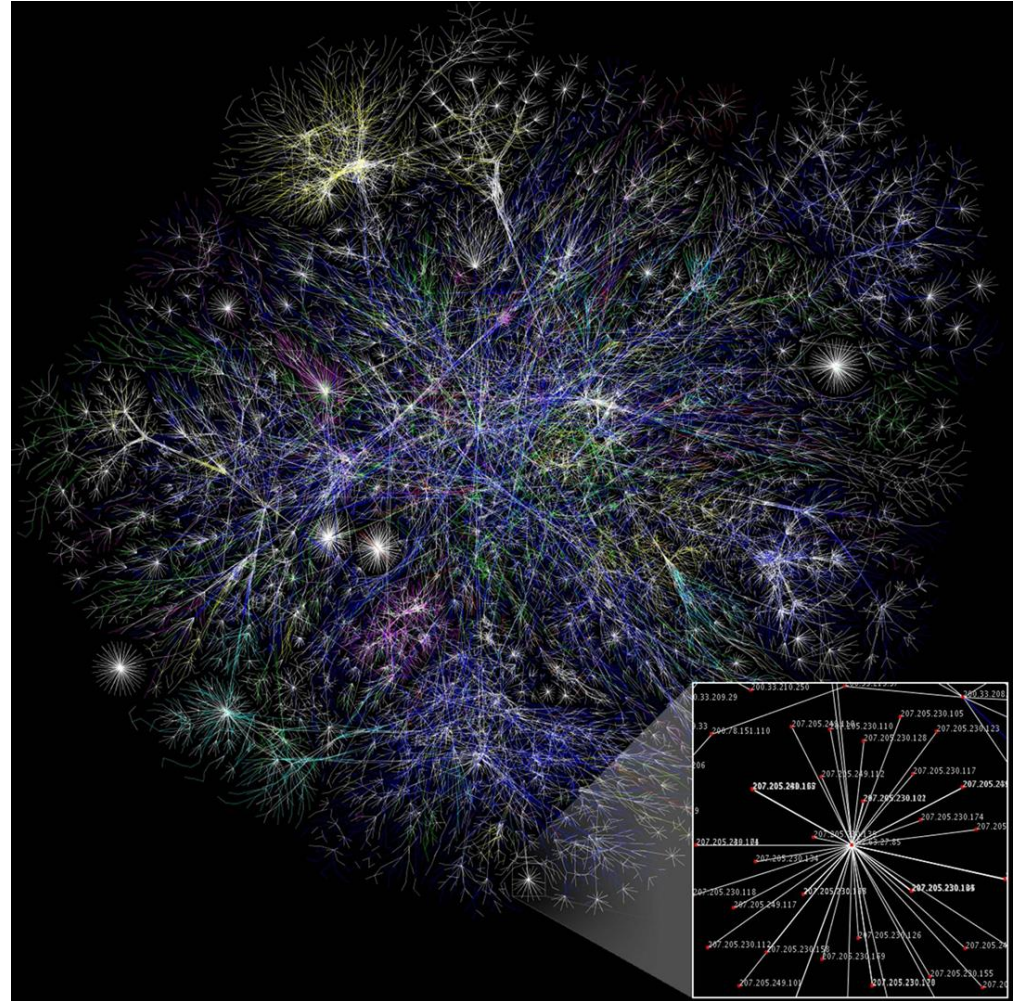


ENIAC digital computer. Unidentified U.S. Army photographer. Public Domain, <https://commons.wikimedia.org/w/index.php?curid=978770>










Four industrial revolutions

FOURTH

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



HISTORY OF DIGITAL DISRUPTION

Company /Industry	Core Business	Transformational Change	Digital Disruption Enhance Existing Income Model?	Successful Internal Transformation?	Digital Disrupter
	Photographic Film & Paper	Digital Photography	NO	X	
	Selling Books from Stores	Online Book Orders	NO	X	
	Lending Money	ATMs and Online Banking	YES	✓	All Modern Banks
	Video Rental	Digital Streaming	NO	X	
	Fee-for-Service Health Care	Value-based, Digitally Enabled Medicine	NO	?	?
	Sell more drugs at higher prices	Value based reimbursement	NO	?	?

Qualities of the New Data Environment

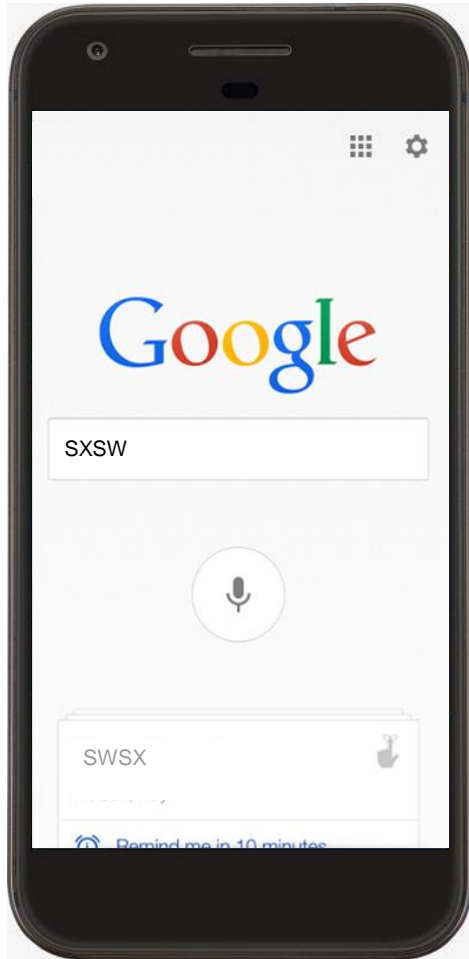
- **Volume**
 - New methods of data storage allow access to huge amounts of data
- **Ubiquity/Liquidity**
 - Data are available anywhere across geography, social and economic classes
- **Latency**
 - There is no delay in access to data inherent in the technology
- **Analysis**
 - Data, information, knowledge, wisdom continuum is being shifted to the right



“To learn the truth, we must put all the parts together.”



















16.3M results

in 0.57 second

The cost of a smartphone in 1985: \$32M



today
\$250
(with contract)

Mobile Phone \$9,000 (DynaTAC) 	Text Messaging \$1,105 (fax machine) 	GPS \$6,630 (Magellan GPS) 	Voice Recorder \$110 (Realistic CTA) 
Digital Watch \$45 (Casio DBC) 	Music Player \$400 (Sony Discman) 	Video Camera \$3,745 (Sony V8) 	Video Player \$1,105 (Sony VCR) 
Encyclopedia \$2,200 (Encyclopedia) 	Processor \$32M (Cray) 	Portable TV \$665 (Casio Mini TV) 	Video Conference \$110,520 (Future Sys) 



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

www.projectbaseline.com



verily

 **Duke University**
School of Medicine

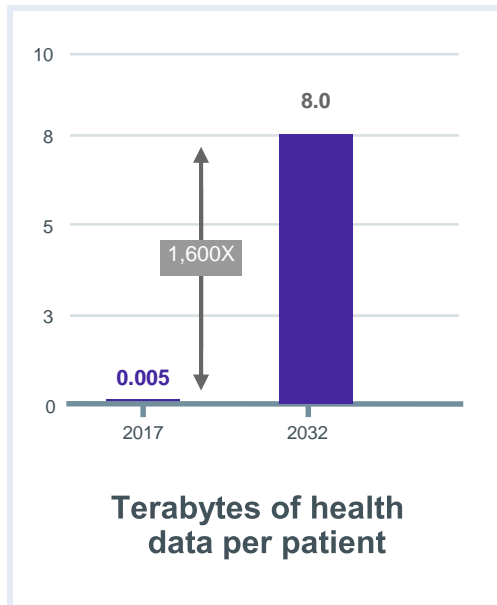
 **Stanford**
MEDICINE

Google

Collecting comprehensive health data

Generate tools and technologies to collect diverse, comprehensive health data in-clinic, at-home, and remotely.

6 TB of collected data on each Project Baseline participant per year



In Clinic



Epic

EMR

Medicare.gov



Claims



Lab



Imaging

At Home



Sleep



Behavior



Medication



Weight

Remote



Mobile



Behavioral



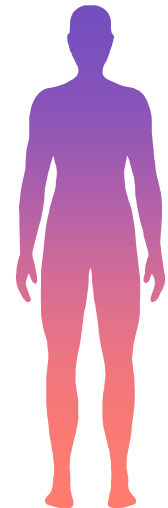
Socio-economic



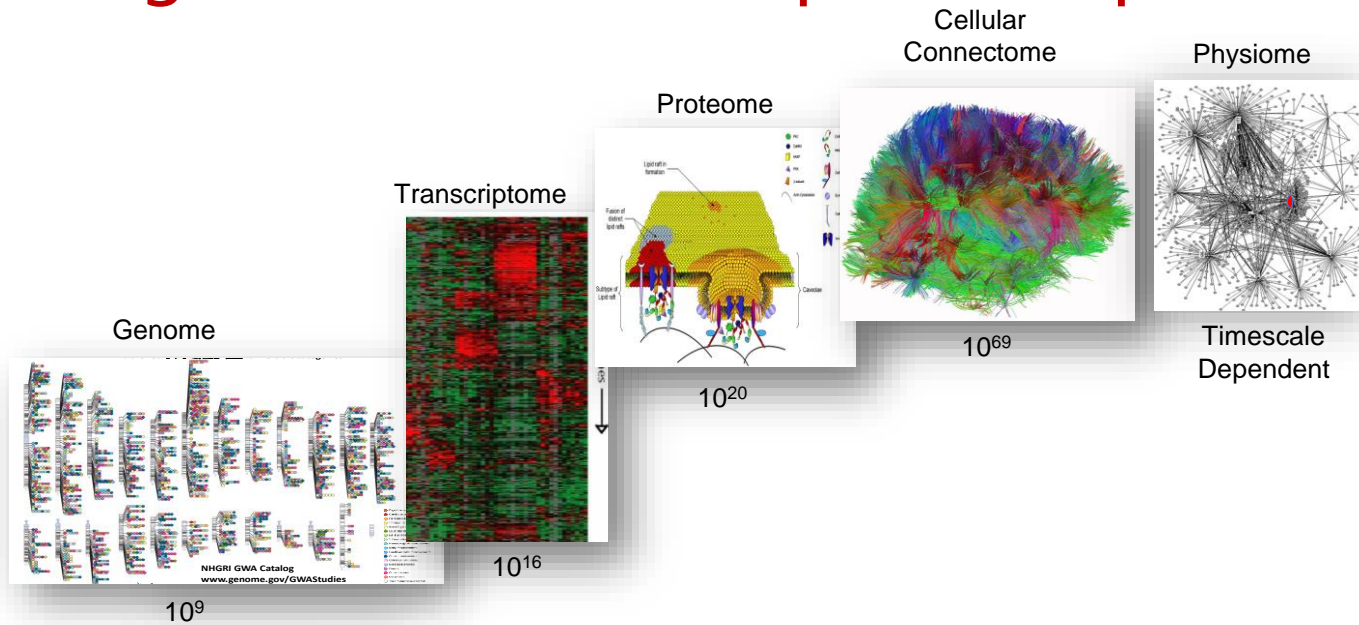
Wearable



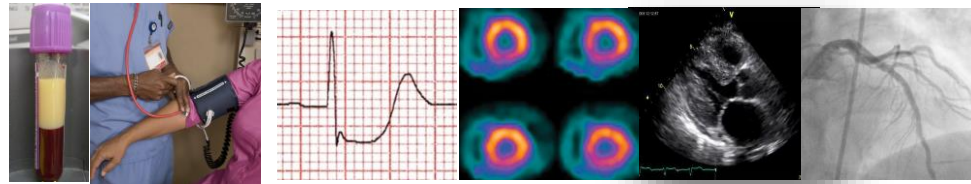
Location



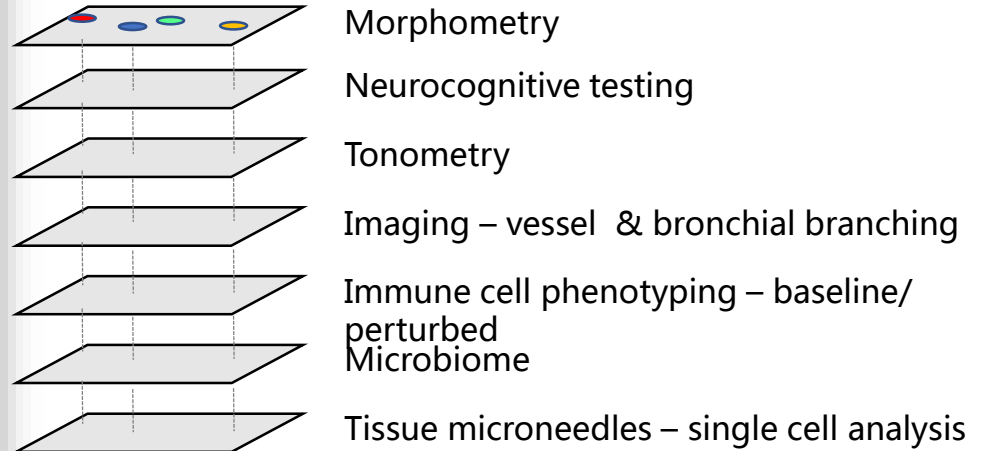
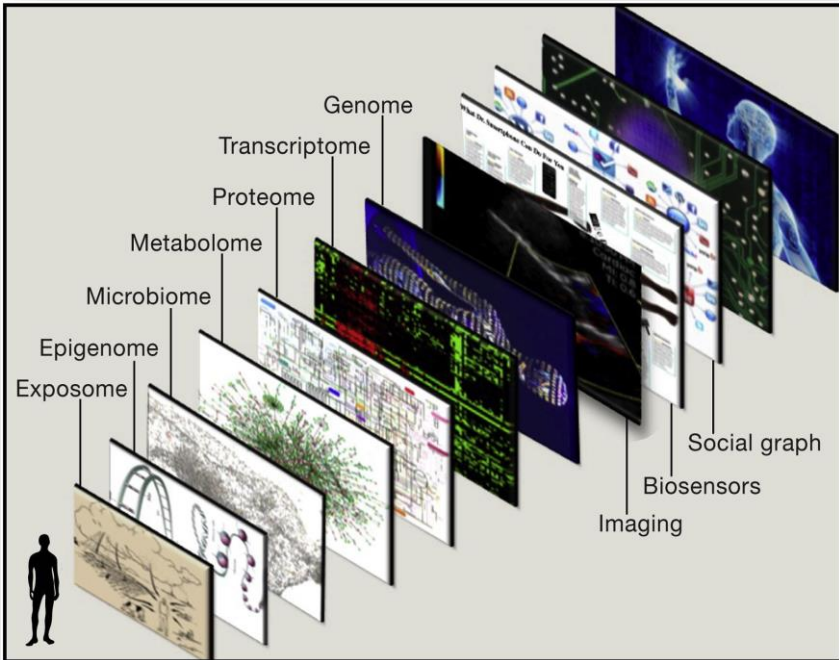
Creating a 'stack' of novel personal phenotypes



All clinical phenotypes 10^4

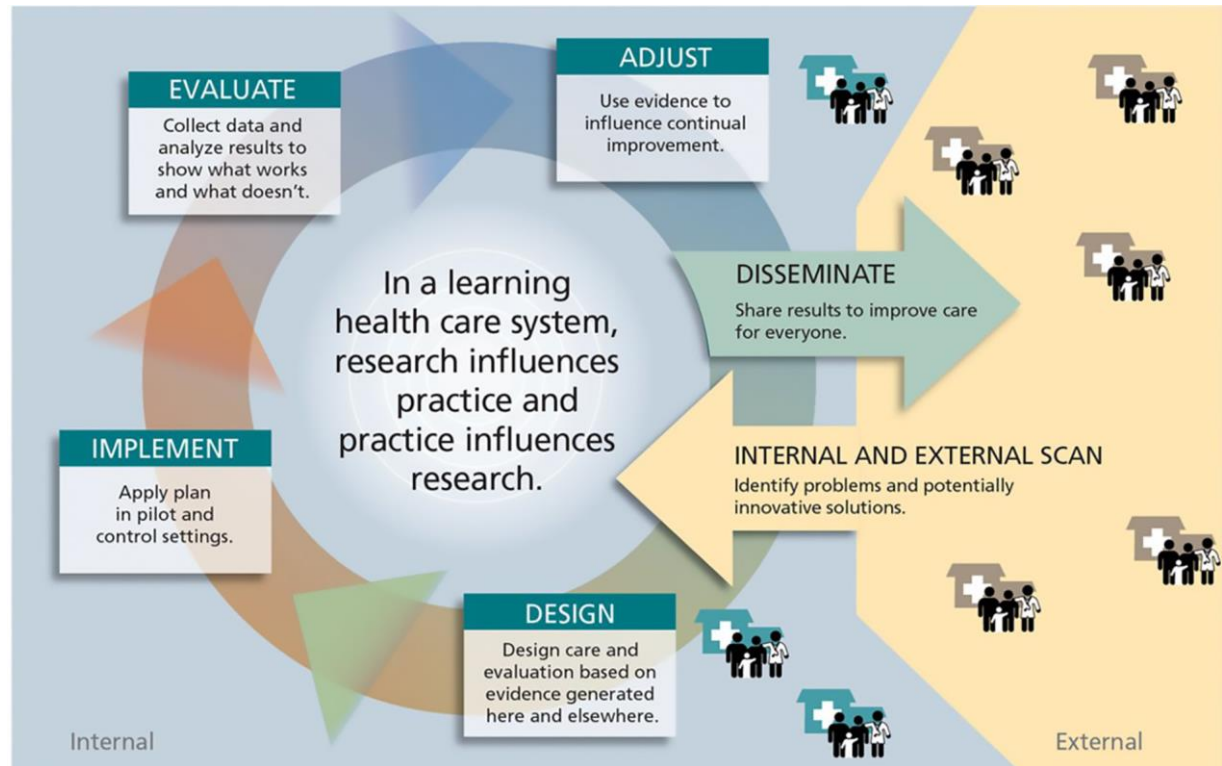


Phenotype stack

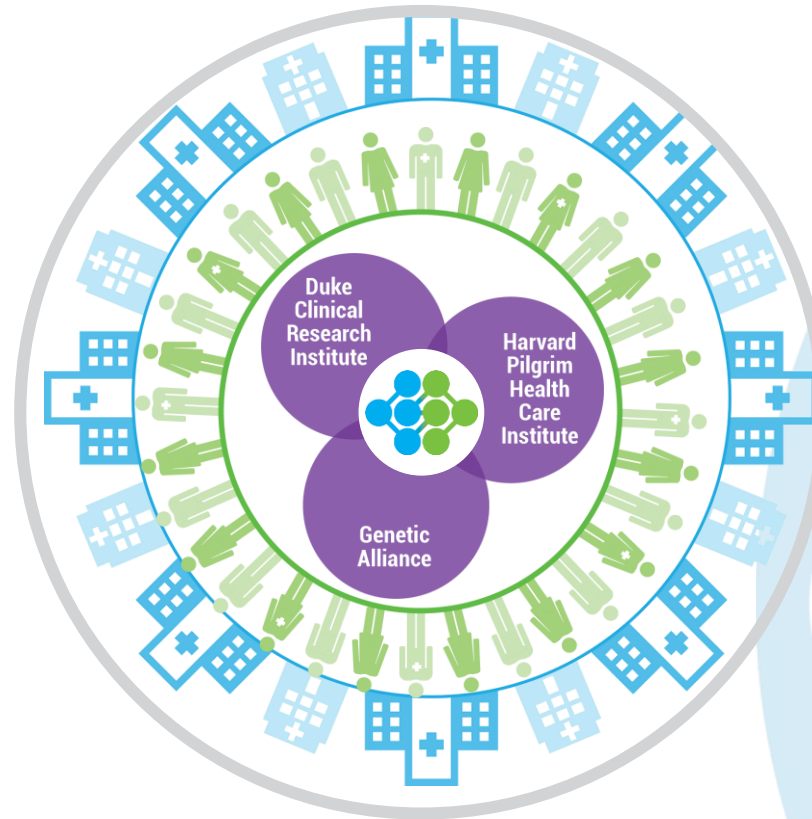


- Integration across scales requires new data, new tools, new taxonomy
- Unifying metadata: **small molecules, biophysical stimuli**
- Breadth vs Depth
- Co-clinical modeling

Learning health care systems



PCORnet[®] embodies a “network of networks” that harnesses the power of partnerships



9
Clinical Research
Networks (CRNs)



2
Health Plan
Research Networks
(HPRNs)



Patient Partners



1
Coordinating
Center



A national
infrastructure for
people-centered
clinical research



CRNs

ADVANCE

[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



[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



[PaTH: Towards a Learning Health System](#)
University of Pittsburgh

HPRNs



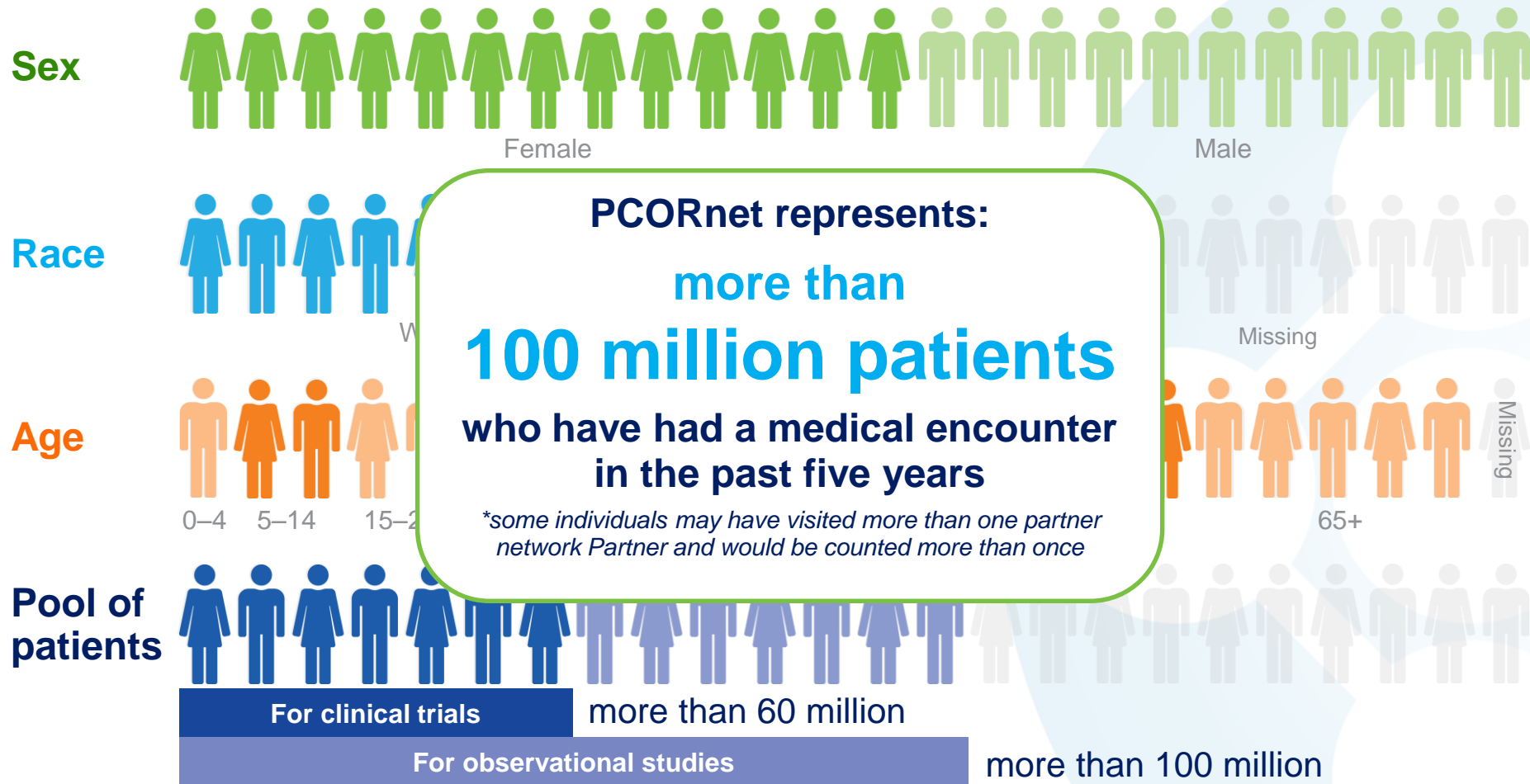
HealthCore (a subsidiary of Anthem)



Humana – Comprehensive Health Insights
(CHI; a subsidiary of Humana Pharmacy Solutions)



Resulting in a national evidence system with unparalleled research readiness



ADAPTABLE Study Design

15,000 patients with known ASCVD + ≥ 1 “enrichment factor”

Patients identified by research networks through EHR searches
Computable phenotype identifies pts via inclusion/exclusion criteria

Patients able to learn, provide e-consent, and self randomize on a web portal
Treatment assignment provided directly to patient

ASA 81 mg QD

RANDOMIZATION

ASA 325 mg QD

Electronic patient follow-up: Every 3 or 6 months
Supplemented with EHR, health plans, Medicare

Primary endpoint:

Composite of all-cause mortality, hospitalization for MI, or hospitalization for stroke

Primary safety endpoint:

Hospitalization for major bleeding

ClinicalTrials.gov:
NCT02697916

🏠 To develop and refine the infrastructure for PCORnet to conduct multiple comparative effectiveness trials in the future

Site Approach and Enrollment

CDRN	Total Number Eligible	Total Number Approached	% of Eligible Approached	Golden Tickets Entered	% Golden Tickets entered per Approached	Total Enrolled	# Non-internet Enrolled	% Enrolled Per Approached	% Enrolled Per Golden Ticket Entered
CAPriCORN	18,389	12,251	67%	821	7%	516	203	4%	63%
GPC	92,053	62,365	68%	3594	6%	1690	119	3%	47%
HPRN	160,914	160,914	100%	1,551	1%	358	2	0%	23%
LHSNet	128,981	35,342	27%	1493	4%	865	115	2%	58%
Mid-South	92,714	43,629	47%	7,283	17%	3942	491	9%	54%
NYC-CDRN	22,141	6,575	30%	1339	20%	710	253	11%	53%
OneFlorida	59,373	5,220	9%	749	14%	593	154	11%	79%
PaTH	47,594	41,187	87%	3682	9%	1279	58	3%	35%
pScanner	15,669	6,855	44%	253	4%	131	8	2%	52%
REACHnet	33,299	20,583	62%	1801	9%	773	240	4%	43%
TOTAL	671,133	394,921	59%	22,566	6%	10,857	1,643	3%	48%

What if a choice made over the counter
prevented...

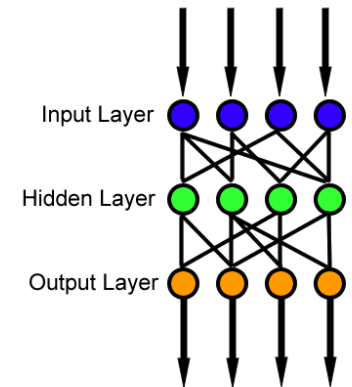


19,000 Deaths/Heart Attacks
Or
Thousands of Bleeds
Annually in the United States

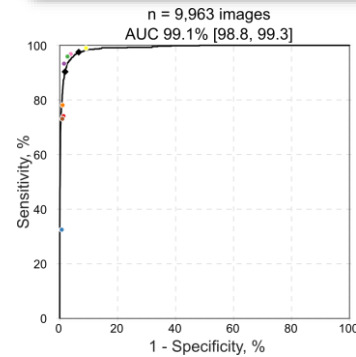
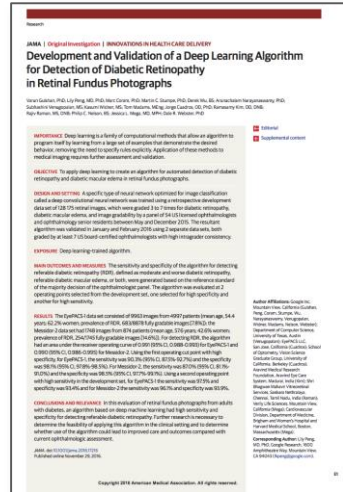


definitions

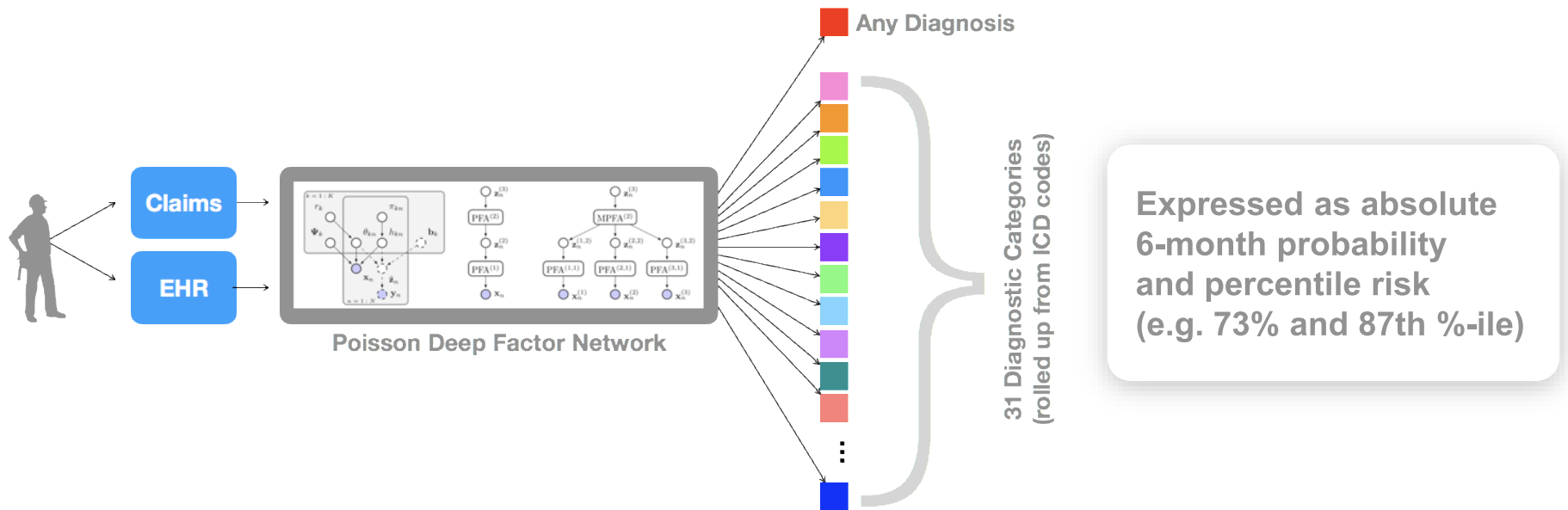
- Machine learning is a subset of artificial intelligence in the field of computer science that often uses statistical techniques to give computers the ability to "learn" (i.e., **progressively improve performance**) on a **specific task**) with data, without being explicitly programmed.
- Artificial neural networks (ANNs) or connectionist systems are computing systems vaguely inspired by the biological neural networks that constitute animal brains. Such systems "**learn**" to perform tasks by considering **examples**, generally without being programmed with any task-specific rules.
- Natural language processing (NLP) is an area of computer science and artificial intelligence concerned with the **interactions between computers and human (natural) languages**, in particular how to program computers to process and analyze large amounts of natural language data.



Innovations jointly deployed by Google + Verily

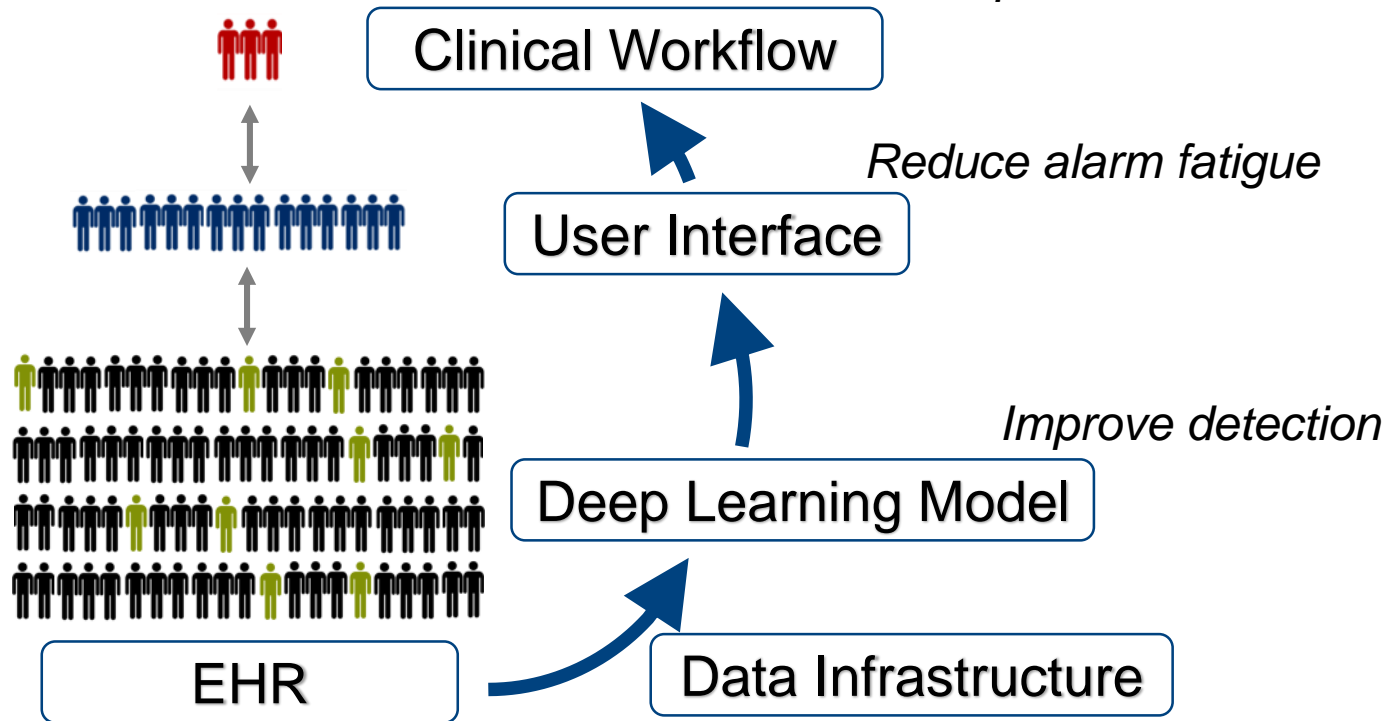


Admission Risk Prediction Model



FLAGGING ACUTE INPATIENT ISSUES

*Shorten time to treatment
and better patient outcomes*



**“The Boeing 737 Max and the Problems
Autopilot Can’t Solve”
- *New York Times***



**“Trump Laments Modern Airplanes as ‘Too Complex to Fly’ in
Wake of Deadly Crashes” - *Chicago Tribune***

1 in 20

Google searches
are health
related

Google



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?

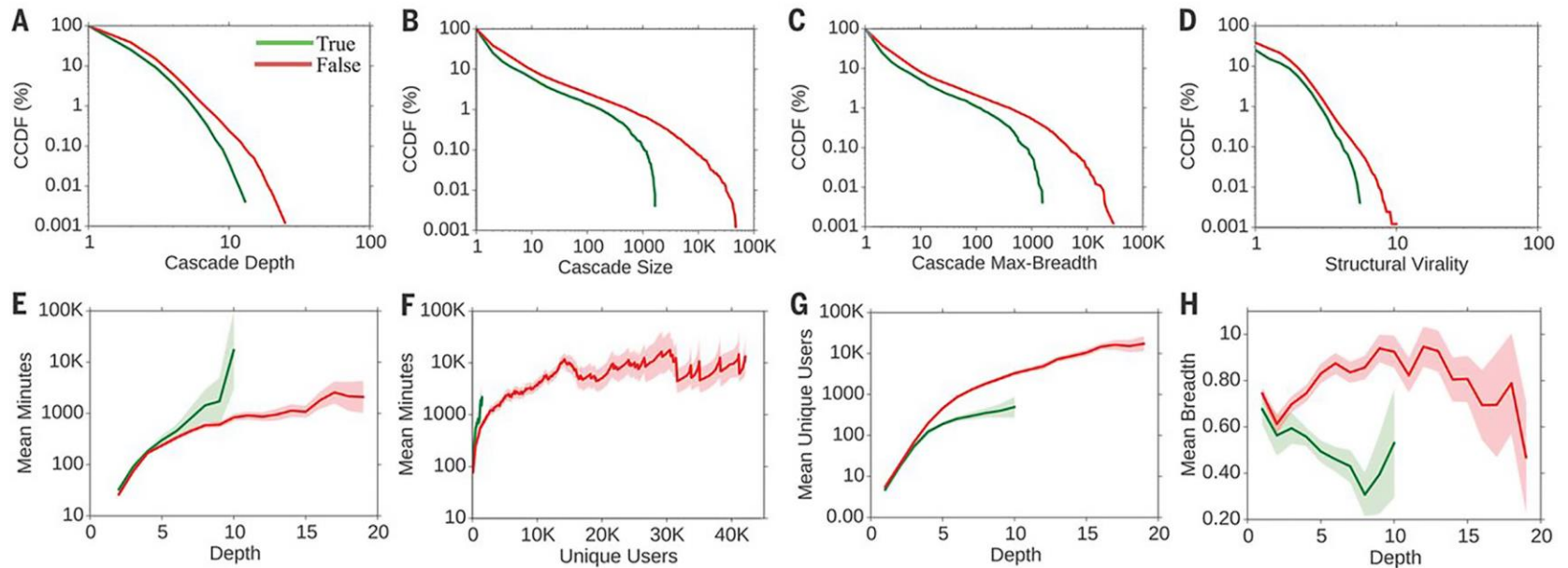
PHQ-9 is a Patient Health Questionnaire, with 9 questions, that is used to measure depression severity

PATIENT HEALTH QUESTIONNAIRE - 9				
Over the <u>last 2 weeks</u> , 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



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

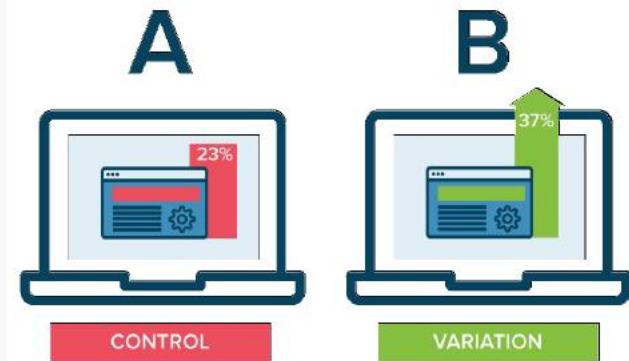
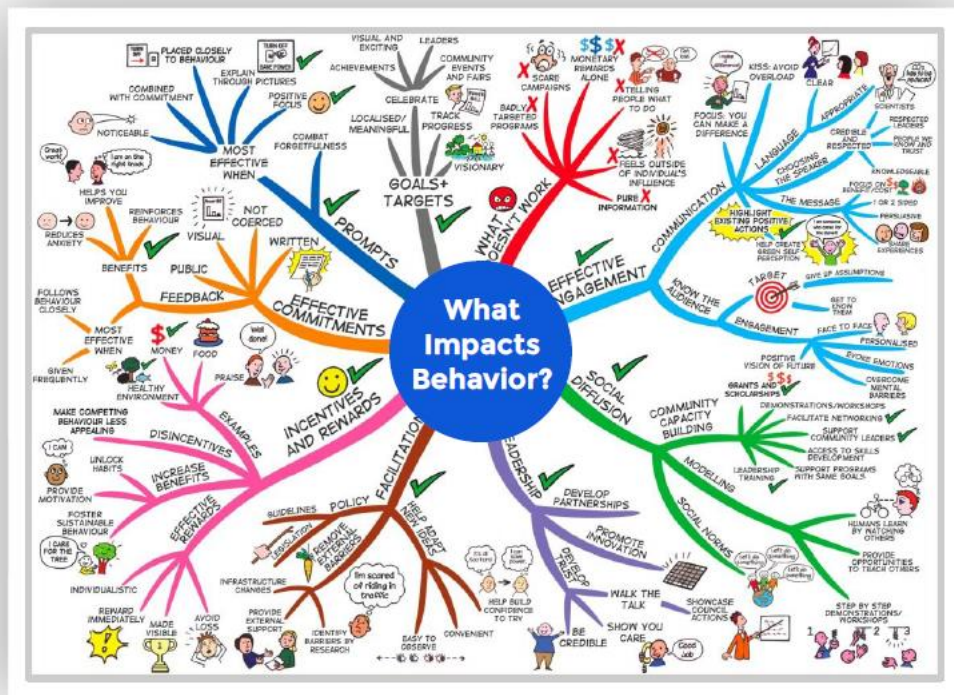
Complementary cumulative distribution functions (CCDFs) of true and false rumor cascades



Science

AAAS

Data Activation and Testing Outcomes



Digital transformation

2010

Individual Productivity and IT Silos

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

—————→
Mapping Human Health

2020

Collective Intelligence and Distributed Computing

- Data stored in cloud, simple to query
- Collaborative, cloud based productivity
- Machine learning drives deep, actionable insights
- IT changing **how** it computes

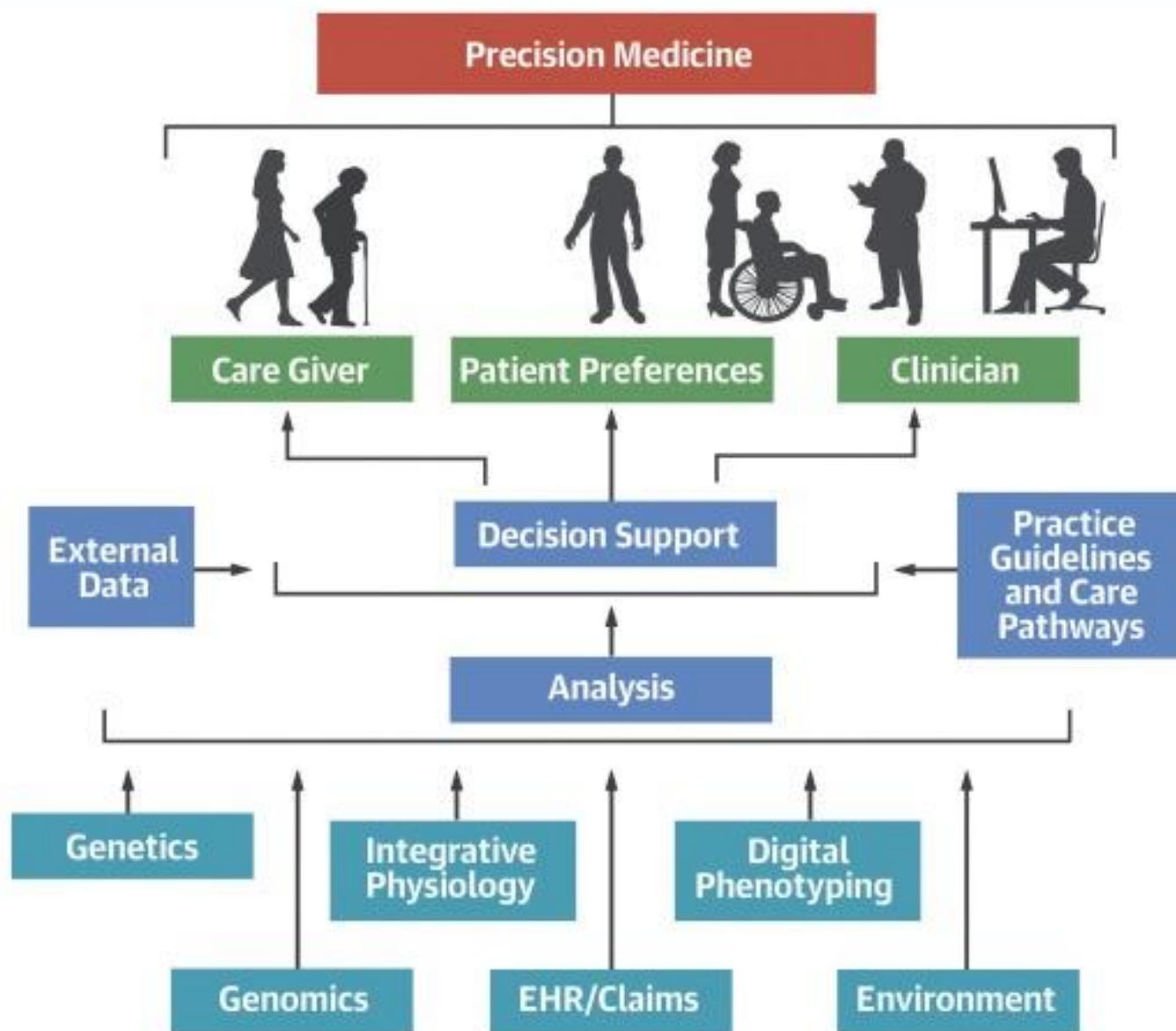
CONCRETE CHANGES DUE TO THE 4TH INDUSTRIAL REVOLUTION

1. **”Home inversion”**—prevention and chronic care will move to the home, school, workplace and neighborhood
2. Health care team—function will be optimized to enable members to function at the top of their capability with task shifting to community health workers and nurses for many activities in the home
3. Clinic visit preparation—will set up using interactive system at home using sensors, cell phones and chatbots
4. **Clinic visits**—will be virtual in many cases, but when human interaction useful, clinician and patient will talk and interact physically (clinic notes done using NLP and AI)
5. Post-clinic visit—information will be available on an as-needed basis, tailored to the needs, health literacy and numeracy of the patient

CONCRETE CHANGES DUE TO THE 4TH INDUSTRIAL REVOLUTION

6. **Behavior change**—will be reinforced by the digital environment
7. Procedures and surgeries—will be monitored by a digital environment in which ML will be used to guide procedures;
8. Ingestion of data across the spectrum of biology, clinical, imaging, sensors, behavior, social interactions and environment will be routine
9. **Precision medicine will stratify people based on risk and knowledge of effective interventions, and personalized medicine will tailor actions to the needs and values of people and families**
10. **Population health will use the same information aggregated at the level of families, neighborhoods, precincts, towns, counties, states and regions**

CENTRAL ILLUSTRATION: Precision Medicine



Population health



Actionable Information



Analysis



Individuals



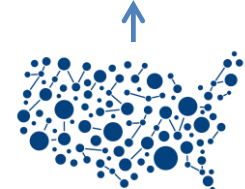
Households



Neighborhoods



Precincts



States

