Duke | FORGE
HEALTH DATA IN ACTION

DATA SCIENCE IN THE ERA OF DATA UBIQUITY

Robert M. Califf, MD
Vice Chancellor for Health Data Science
Director, Duke Forge
Duke University School of Medicine
Advisor, Verily Life Sciences

NIH HCS Collaboratory Grand Rounds
March 23rd, 2018

Duke | FORGE
HEALTH DATA IN ACTION

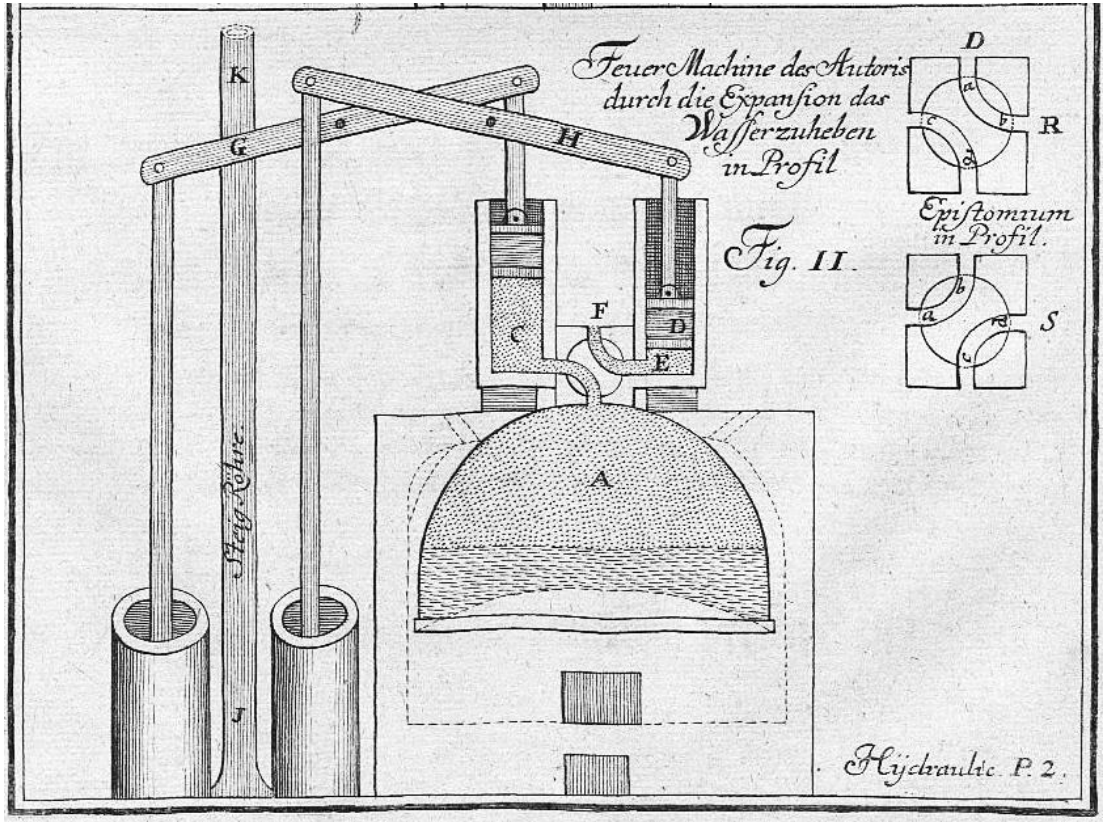
CONFLICTS OF INTEREST

- Employment
  - Duke University
  - Verily Life Sciences
- Corporate Board
  - Cytokinetics
- Consulting
  - Merck
  - Boeringher Ingelheim

Four industrial revolutions

FIRST

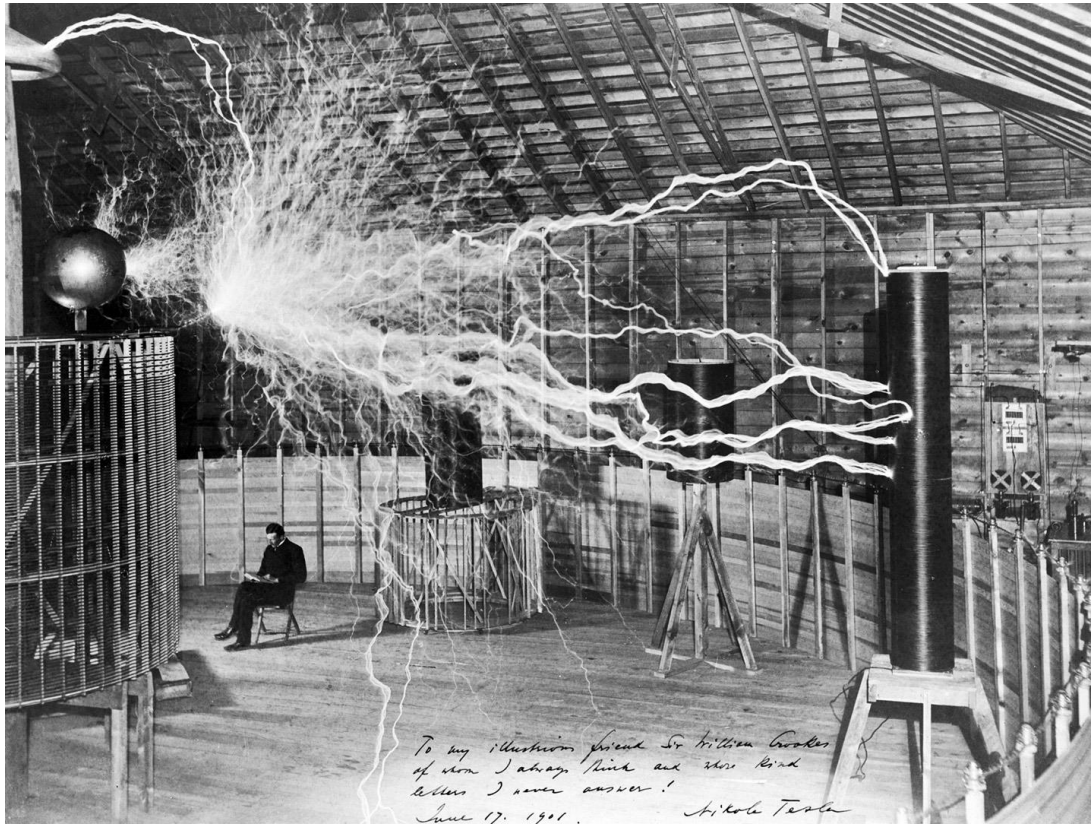
Water and steam power mechanize production.



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

Four industrial revolutions

SECOND
Electric power creates mass production.

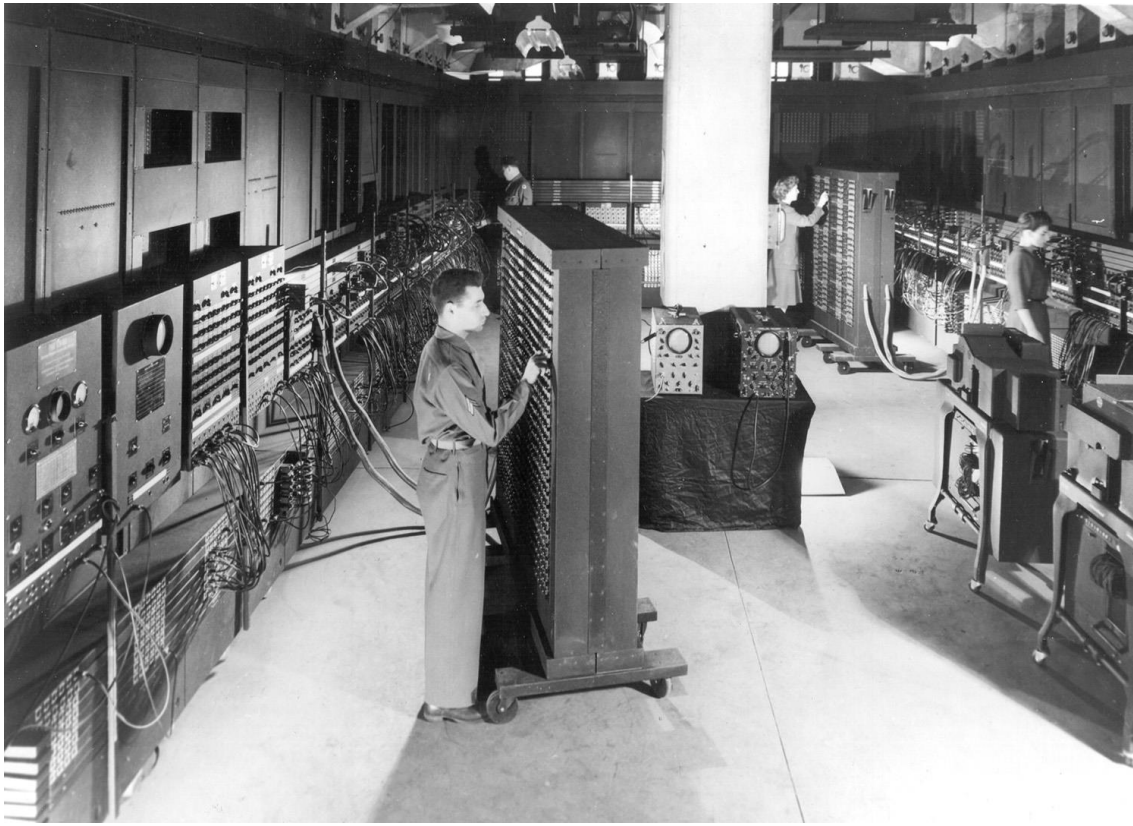


Photographer: Dickenson V. Alley, CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=36367226

Duke | FORGE
HEALTH DATA IN ACTION

Four industrial revolutions

THIRD
Electronics and information technology automate production.

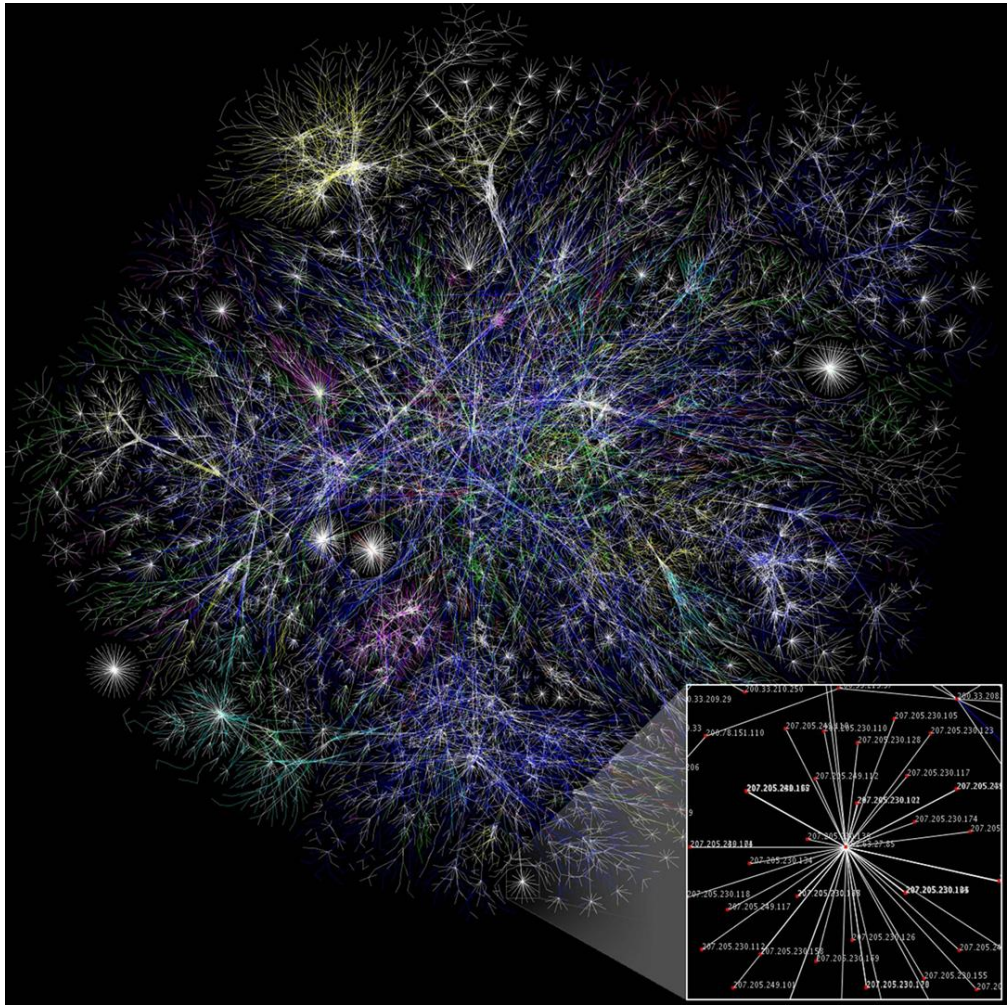


ENIAC digital computer. Unidentified U.S. Army photographer. Public Domain, https://commons.wikimedia.org/w/index.php?curid=978770

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

DATA DELUGE

Illustration credit: Brett Ryder. https://www.economist.com/node/15579717

Duke | FORGE
HEALTH DATA IN ACTION

PROGRESS ON MULTIPLE DIMENSIONS

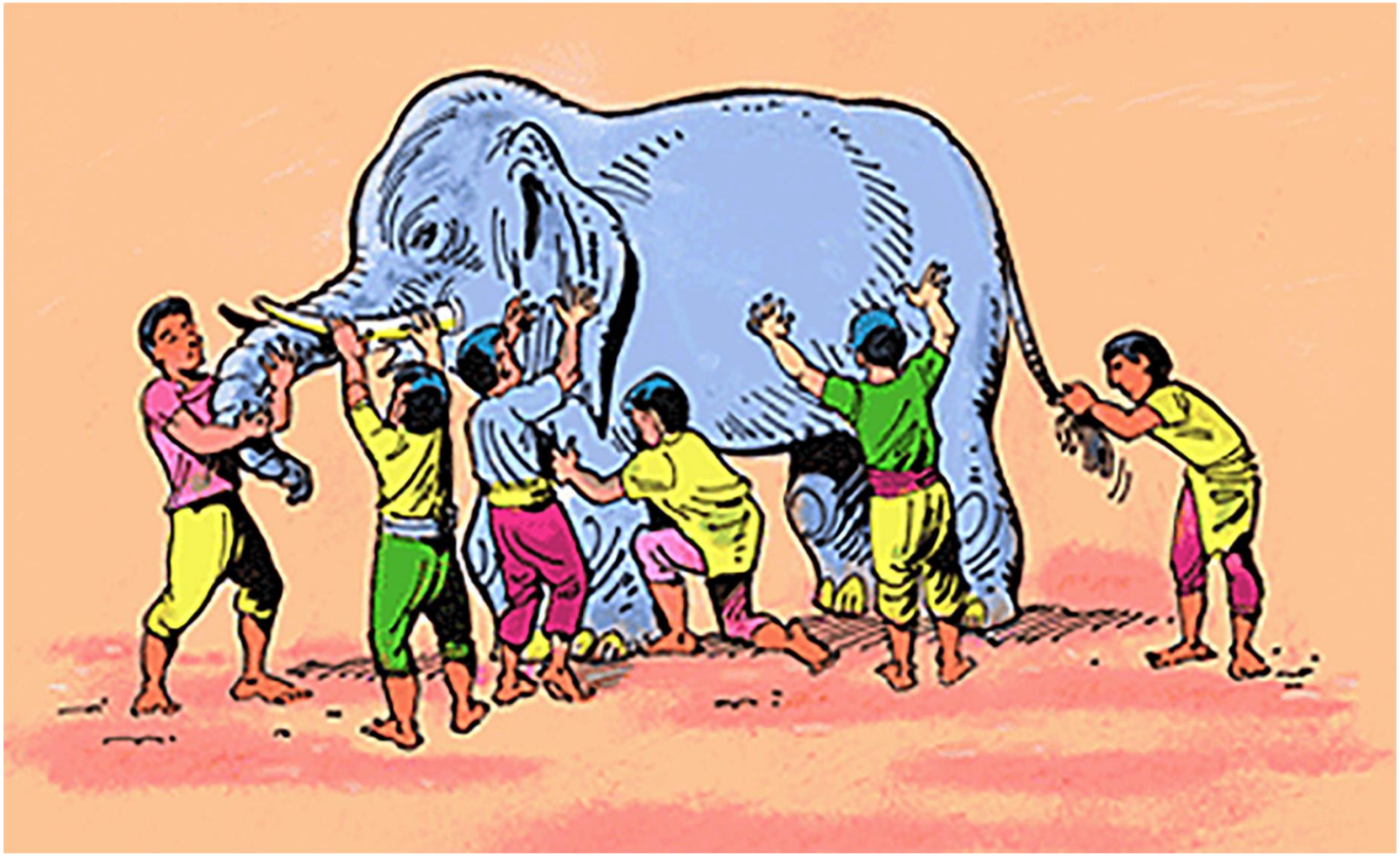
- Biomolecular and integrative physiology
- Current version of electronic health records
  - —Claims
  - —Clinical records
- Behavioral phenotypes and social interactions
- Geospatial
- Time

Duke | FORGE
HEALTH DATA IN ACTION

CRITICAL ISSUES IN NEED OF PROGRESS

- Continued focus on curating the data (”data janitorship”)
- Connecting clinicians and people (consumers, patients, carers) in functional networks to advance knowledge
- Culture and ethics
  - Veracity (synonyms: truthfulness, truth, accuracy, correctness, faithfulness, fidelity)
  - Privacy and confidentiality
  - Monitoring the function of the system
  - Data sharing
  - Priority setting

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

WHAT IS THE PROJECT BASELINE STUDY?

A coalition to develop **gold standard data, tools and technologies** to provide a holistic view of human health and more efficiently and effectively conduct clinical research.

verily

Duke University School of Medicine

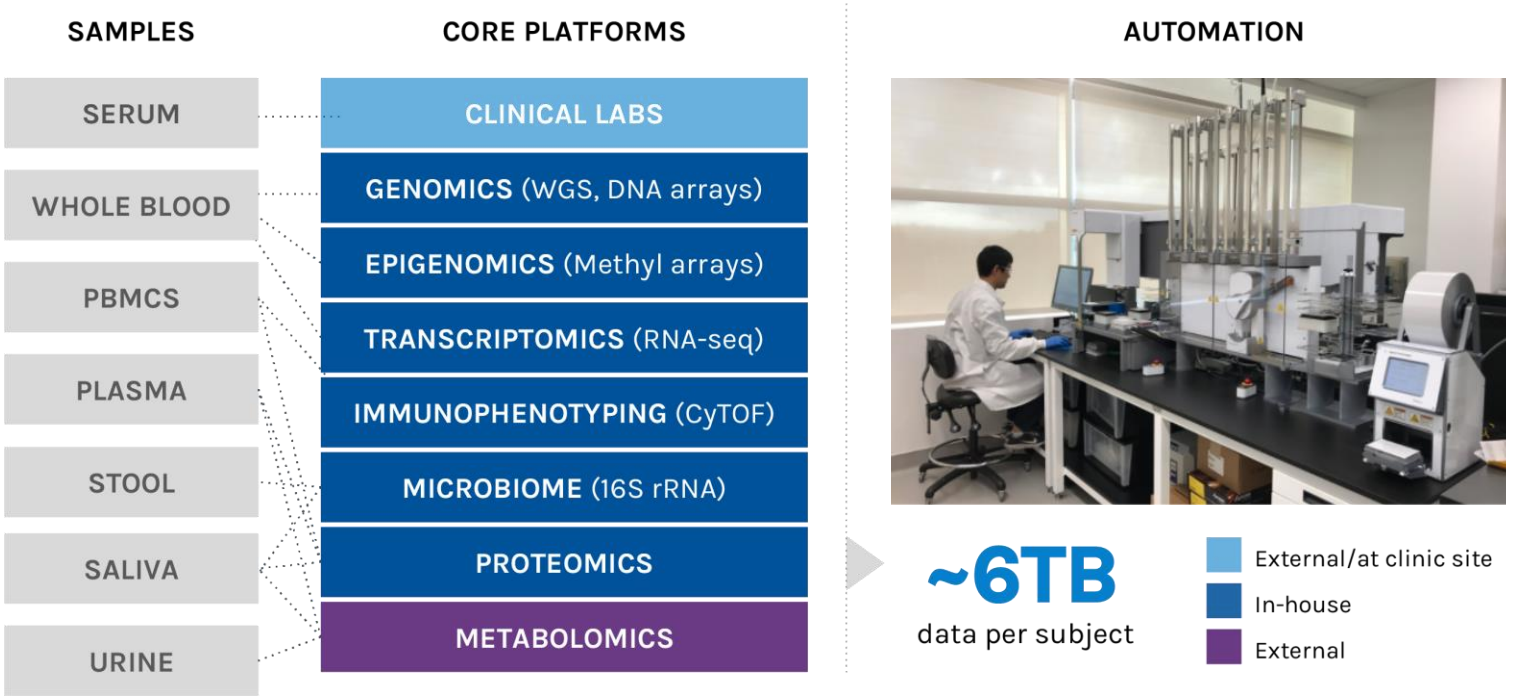
Stanford MEDICINE

Google

American Heart Association.

Duke | FORGE
HEALTH DATA IN ACTION

Deep molecular profiling



Project Baseline

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

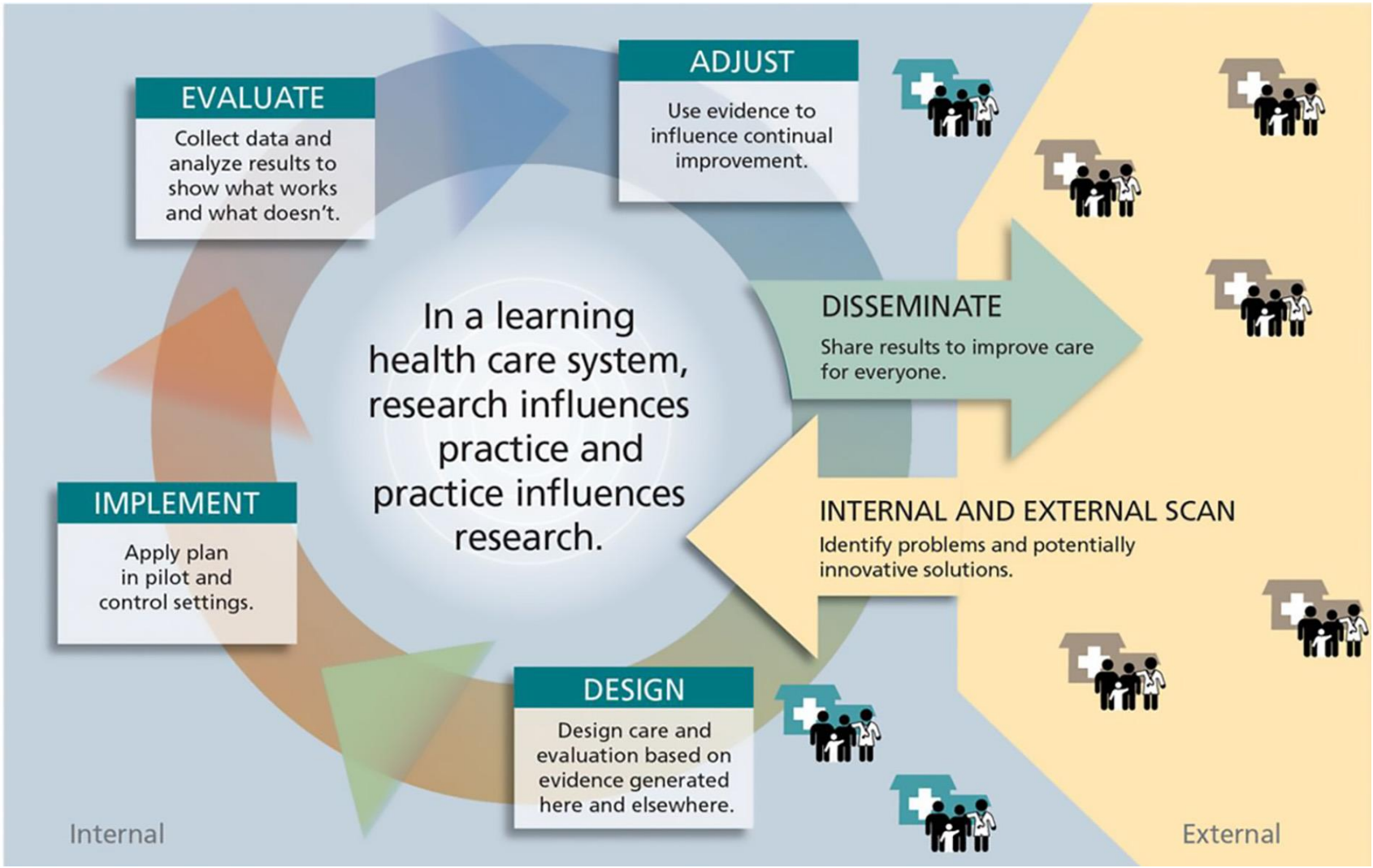
Safely sends device data to secure, encrypted Baseline database

App

Mobile interface for self-reported and passive data acquisitions

Project Baseline

Learning health care systems



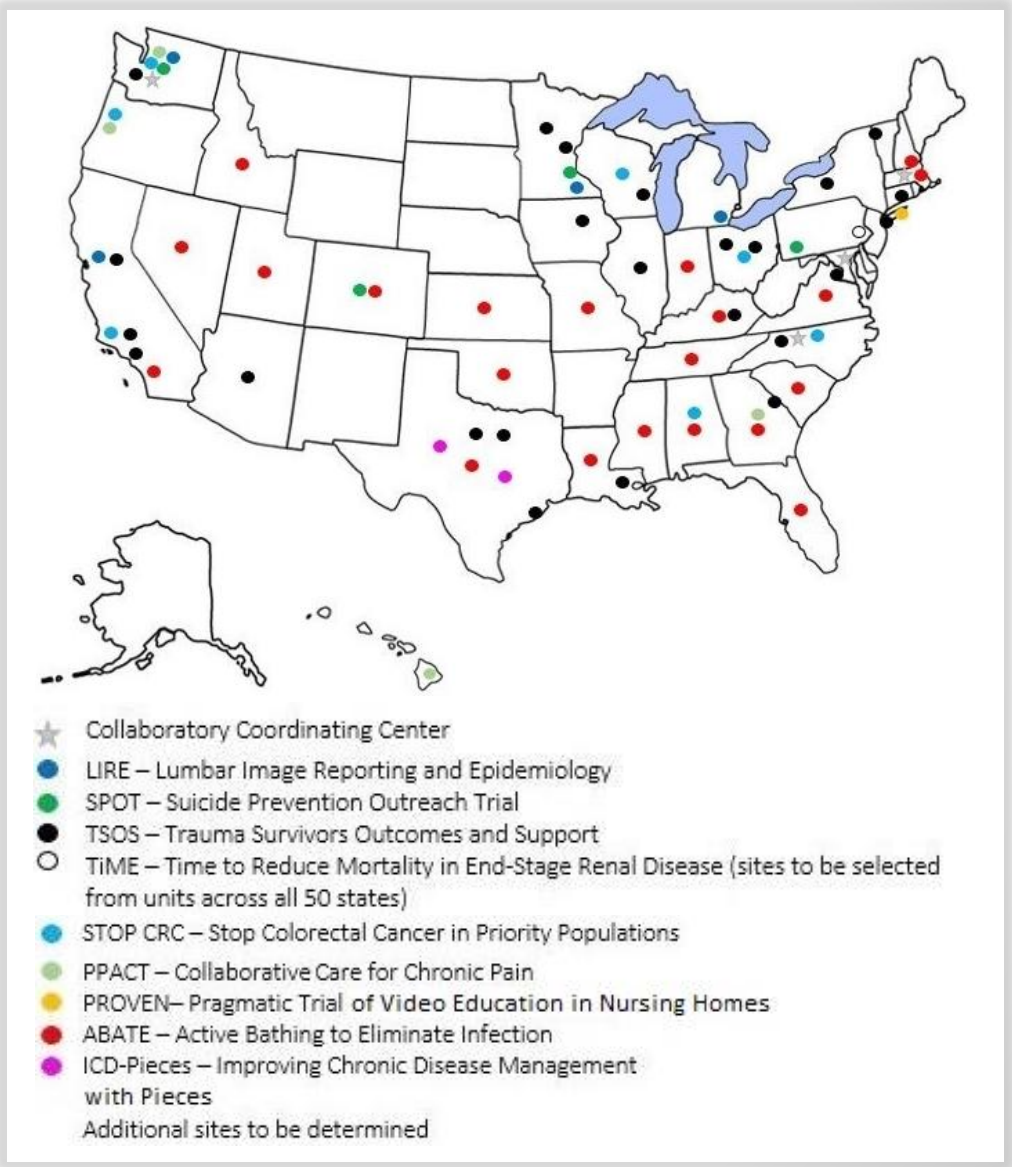
www.fda.gov

Demonstration Project Overview

10 Demonstration Projects spanning 12 NIH institutes and centers

1-year planning phase (UH2)

Implementation phase (UH3)



Duke | FORGE
HEALTH DATA IN ACTION

We face many important, unanswered **questions** about health care.

But our current research system is not set up to answer these questions in the most **useful** and **efficient** way.

We need to link **patients**, **clinicians**, **health systems**, and **researchers** as **partners**.

And harness health **data** to foster **knowledge** that can lead to **better care**.

Together, **partners** can decide what questions to study and how to use the data.

And get **answers** more quickly to the health and healthcare questions that matter most.

A collaborative **national resource** using the power of **partnerships** and **health data** for better research.

That's the vision of

pcornet

The National Patient-Centered Clinical Research Network

With sites and partners in every state...

...And protection of patients' privacy and data security.

pcori

PCORnet is an initiative of the Patient-Centered Outcomes Research Institute

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

Duke | FORGE
HEALTH DATA IN ACTION

The ADAPTABLE Aspirin Study

pcornet
The National Patient-Centered Clinical Research Network

A national resource for conducting patient-centered research harnessing the power of health data.

THE QUESTION

Clinicians often prescribe aspirin to prevent strokes and heart attacks in people living with heart disease. Research has yet to determine the best dose to use, since aspirin can cause serious side effects – like bleeding – in some people.

THE PROBLEM

Heart disease is the No. 1 killer in the U.S.

611,000
people in 2013, one death in 4; accounting for 1 in every 6 healthcare dollars.

Cardiovascular disease (heart attack and stroke) is the most common form of heart disease.

Heart disease strikes someone in the U.S. about once every
43 seconds.

Aspirin is widely prescribed to prevent heart attacks and strokes in people living with heart disease.

60%
of patients with heart disease take a 325 milligram dose each day while 36% take 81 milligrams (or baby aspirin).

pcori

PCORnet is an initiative of the Patient-Centered Outcomes Research Institute.

THE STUDY

The ADAPTABLE trial will compare two common aspirin dosages.

325 mg

81mg

The study will be large and will involve patients across the U.S.

20,000
patients living with heart disease will use a daily aspirin dose of either 81 mg or 325 mg.

ADAPTABLE will use PCORnet to conduct the study and disseminate results. Patients will be partners at every stage of the trial, which will collect data using tools with state-of-the-art security.

ANSWERS FOR BETTER CARE

Results of this study will help patients and their caregivers answer questions like:

- How much aspirin should I take each day to reduce my risk of another heart attack or stroke?
- Do the benefits of taking aspirin every day differ based on the dose?
- Do the risks differ based on the dose?
- Based on my health, age, and other circumstances, what's the best dose to protect my health?

This study will use the power of PCORnet to seek answers to these questions and improve patient care and outcomes.

DATA — KNOWLEDGE — CARE

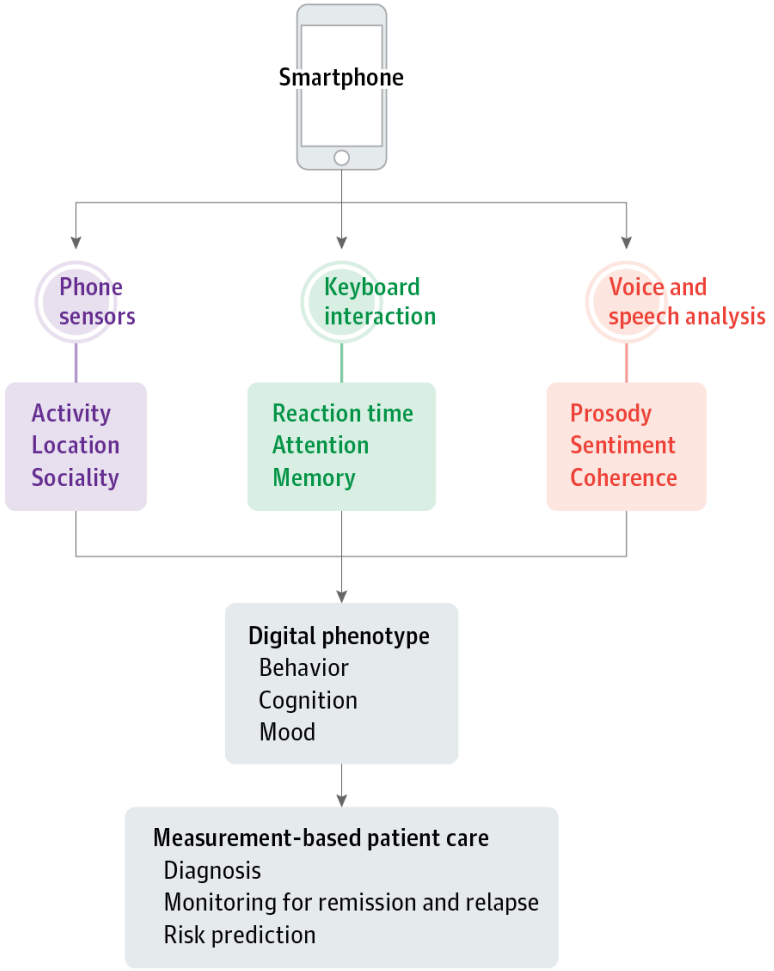
Identifying the aspirin dose that works best could prevent as many as **88,800** deaths per year worldwide

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

Duke | FORGE
HEALTH DATA IN ACTION

1 in 20

Google searches are health related

Duke | FORGE
HEALTH DATA IN ACTION

Our Mission:

Make health information universally accessible and useful.

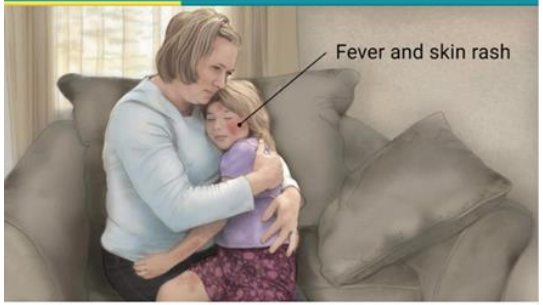
Duke | FORGE
HEALTH DATA IN ACTION

Card Creation

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.

Very rare
645 US cases in 2014

- Preventable by vaccine
- Treatable by a medical professional
- Requires a medical diagnosis
- Lab tests or imaging often required
- Spreads easily
- Short-term: resolves within days to weeks

Measles symptoms don't appear until 10 to 14 days after exposure. They include cough, runny

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

Cough: can be dry or severe

Whole body: fatigue, fever, loss of appetite, or malaise

Nasal: redness, runny nose, or sneezing

Skin: rashes or red spots

Also common: eye irritation, headache, pink eye, sensitivity to light, sore throat, or swollen lymph nodes

Measles
Also called: rubeola

ABOUT | SYMPTOMS | TREATMENTS

Treatable by a medical professional
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.

Prescription
Vitamin A

Self-treatment
Pain reliever: Acetaminophen (Tylenol)

Also common
MMR vaccine

Specialists
Primary care provider (PCP): Prevents, diagnoses, and treats diseases.

Infectious disease doctor: Treats infections, including those that are tropical in nature.

Pediatrician: Provides medical care for infants, children, and teenagers.

Consult a doctor for medical advice
Sources: Mayo Clinic and others.

Duke | FORGE
HEALTH DATA IN ACTION

PRODUCT OVERVIEW: What is PHQ-9?

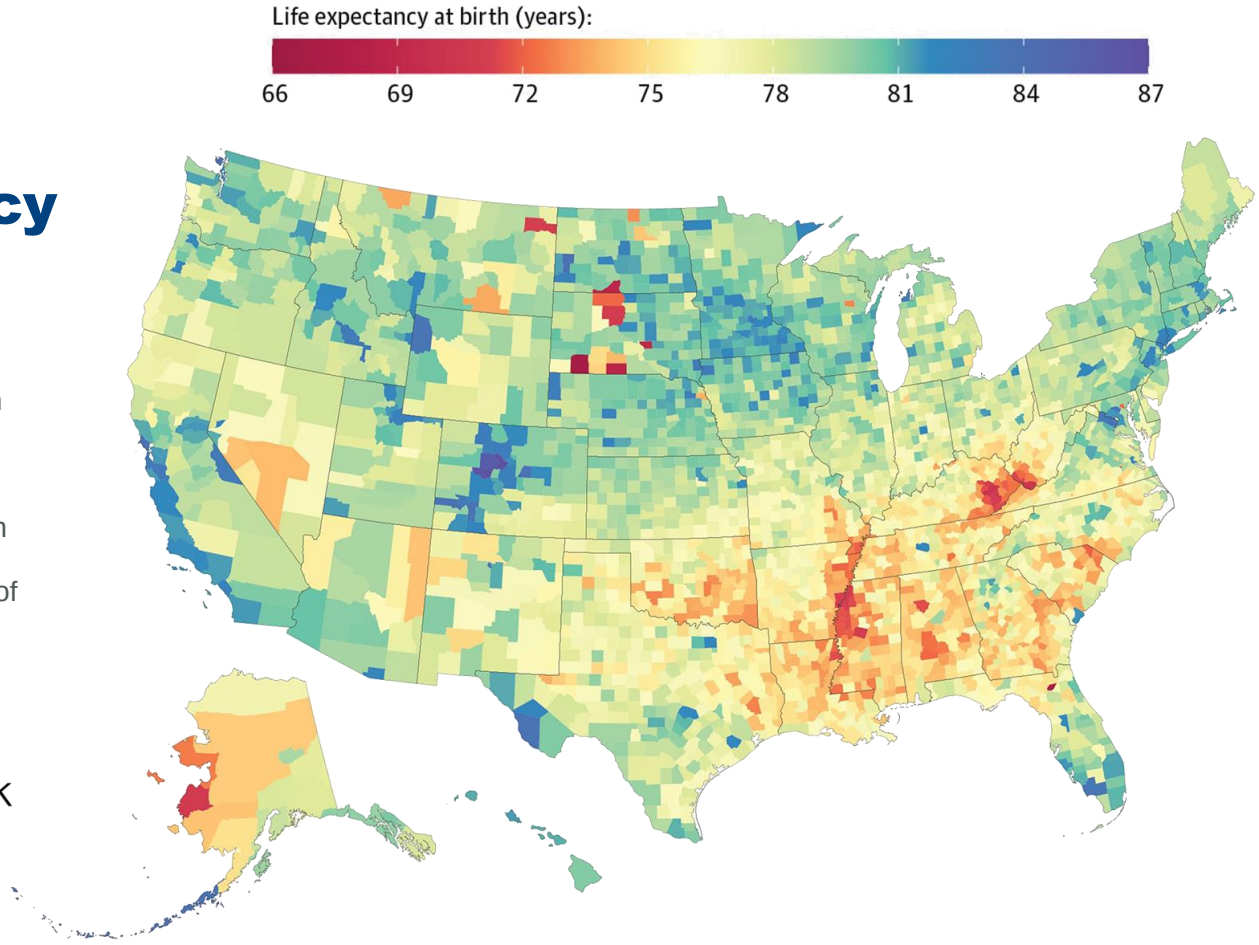
PHQ-9 is a Patient Health Questionnaire, with 9 questions, that is used to measure depression severity

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

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.



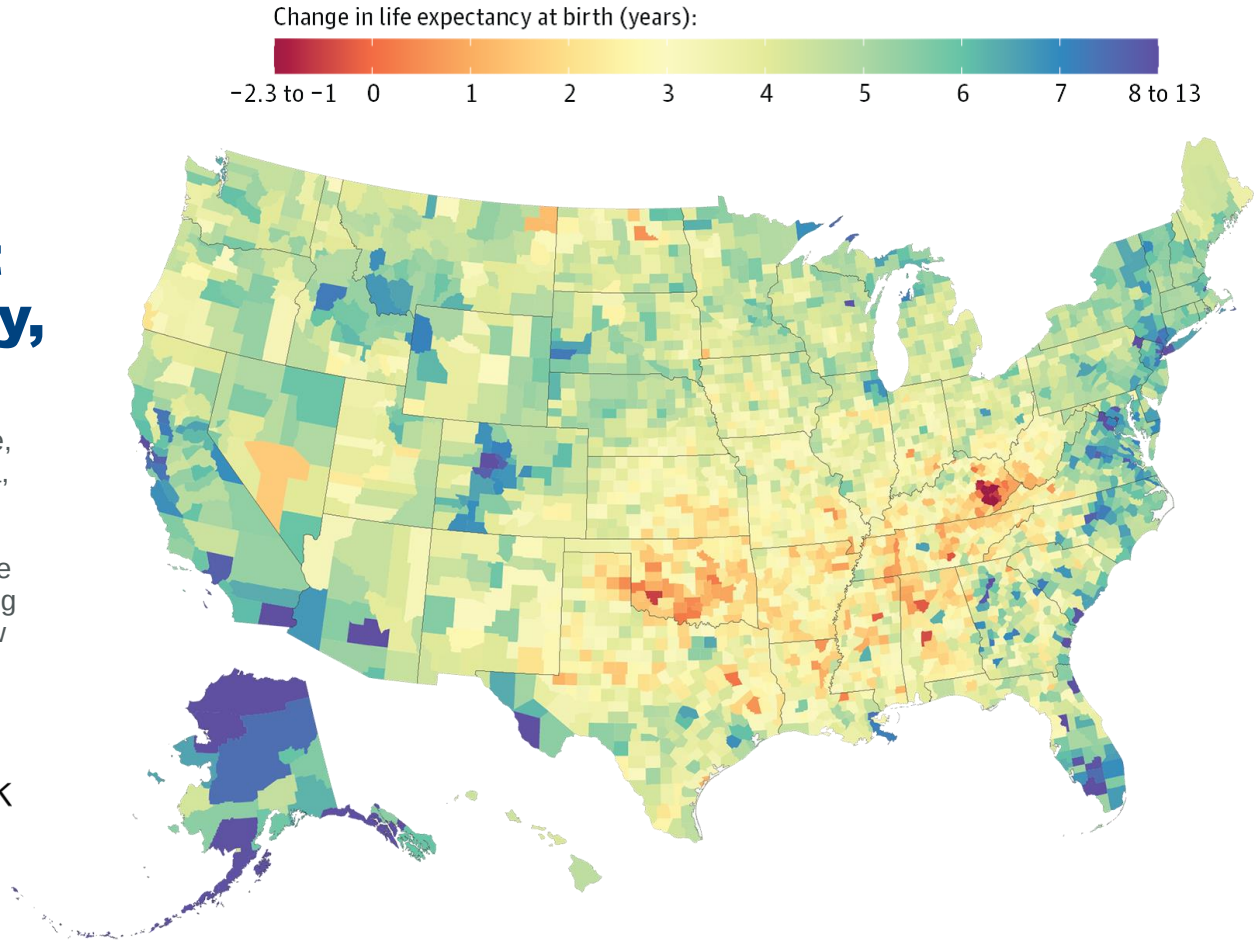
The JAMA Network

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.

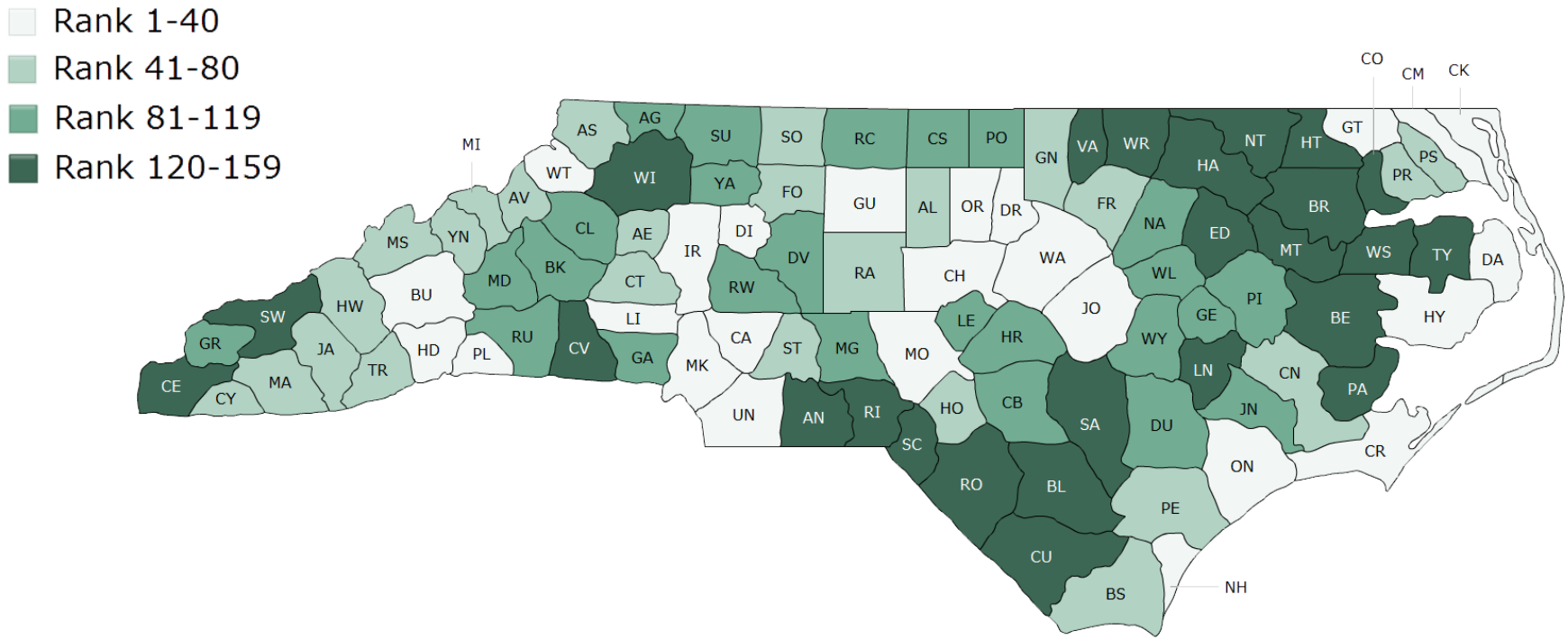


The JAMA Network

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

Duke | FORGE
HEALTH DATA IN ACTION

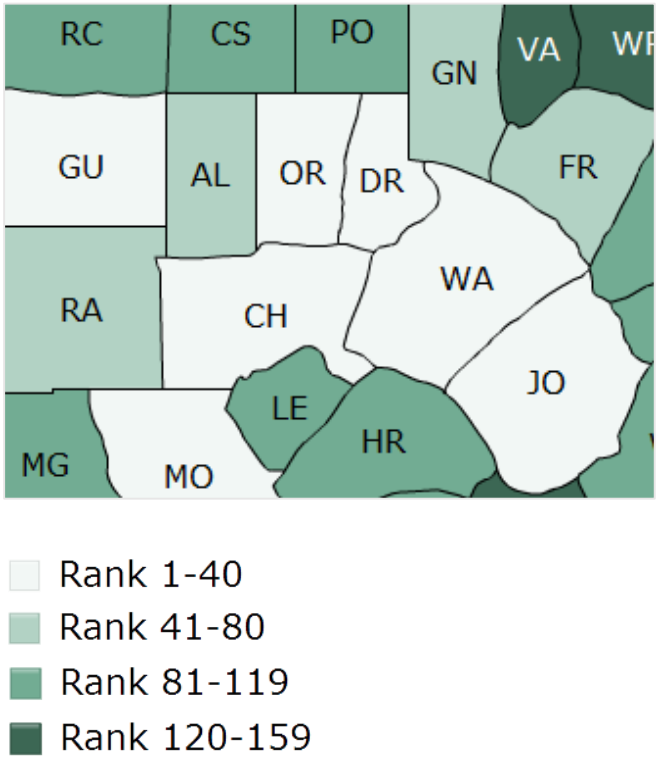
This same dynamic can be seen in North Carolina…



http://www.countyhealthrankings.org/app/north-carolina/2018/overview
https://www.rwjf.org/en/library/interactives/whereyouliveaffectshowlongyoulive.html?cid=xtw_rwjf_unpd_ini:reply_dte:2018030 1

Duke | FORGE
HEALTH DATA IN ACTION

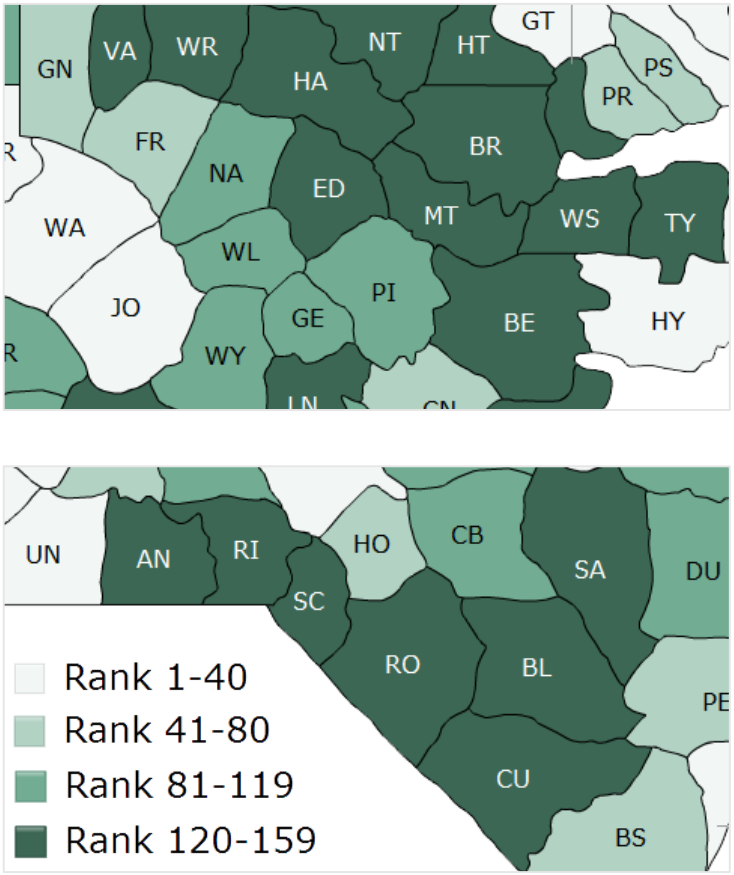
...where better health outcomes in and around the Research Triangle area...



http://www.countyhealthrankings.org/app/north-carolina/2018/overview
https://www.rwjf.org/en/library/interactives/whereyouliveaffectshowlongyoulive.html?cid=xtw_rwjf_unpd_ini:reply_dte:2018030 1

Duke | FORGE
HEALTH DATA IN ACTION

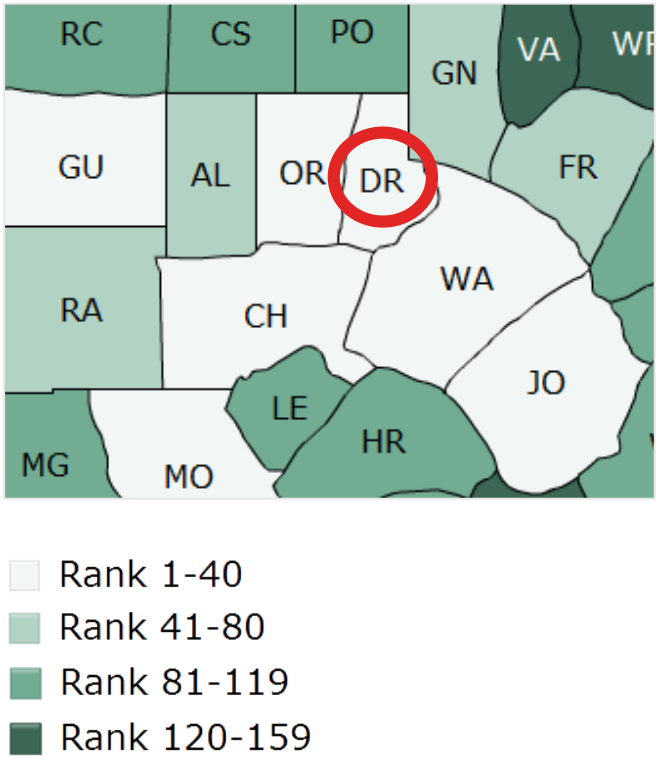
...contrast with poorer outcomes in rural counties in southern and eastern NC.



http://www.countyhealthrankings.org/app/north-carolina/2018/overview
https://www.rwjf.org/en/library/interactives/whereyouliveaffectshowlongyoulive.html?cid=xtw_rwjf_unpd_ini:reply_dte:2018030 1

Of note

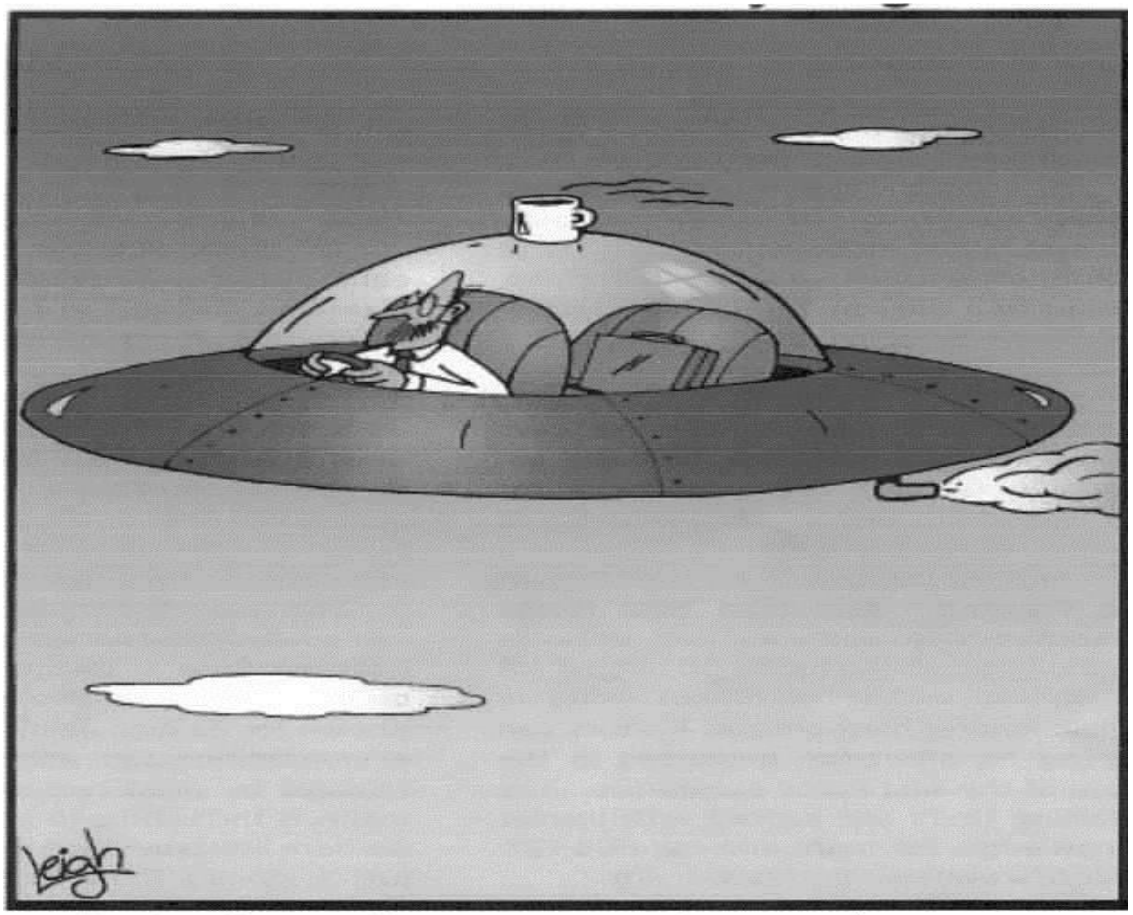
Although ranked 11th out of 100 for health outcomes and 7th out of 100 for life expectancy, Durham County life expectancy remains slightly below the national average.



http://www.countyhealthrankings.org/app/north-carolina/2018/overview
https://www.rwjf.org/en/library/interactives/whereyouliveaffectshowlongyoulive.html?cid=xtw_rwjf_unpd_ini:reply_dte:2018030 1

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



Technology advances; people stay the same.

Duke | FORGE
HEALTH DATA IN ACTION

Facebook Fiasco

The Economist

Topics Current edition More

What Zuckerberg should do

Facebook faces a reputational meltdown

This is how it, and the wider industry, should respond

Print edition | Leaders ›
Mar 22nd 2018

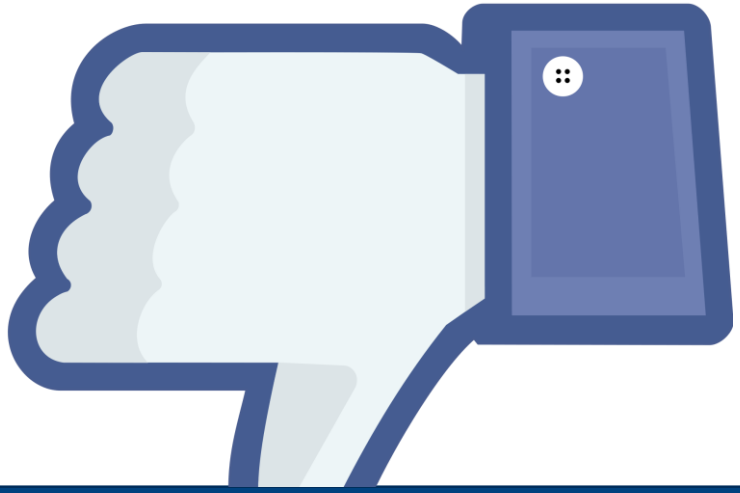
LAST year the idea took hold that Mark Zuckerberg might run for president in 2020 and seek to lead the world's most powerful country. Today, Facebook's founder is fighting to show that he is capable of leading the world's eighth-biggest listed company or that any of its 2.1bn users should trust it

https://www.economist.com/news/leaders/21739151-how-it-and-wider-industry-should-respond-facebook-faces-reputational-meltdown

The Guardian

Facebook: is it time we all deleted our accounts?

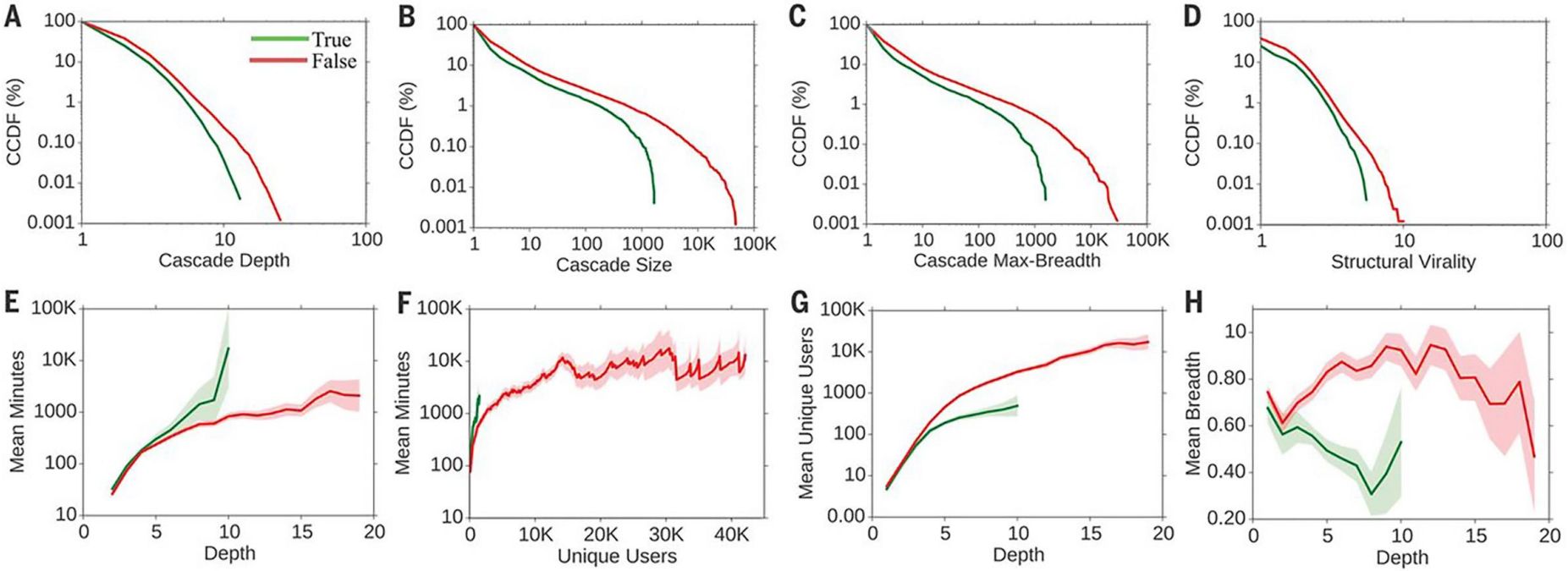
The Cambridge Analytica revelations may be the final nudge we need to turn away from the social network. And it's only the tip of the iceberg when it comes to big tech harvesting private information



https://www.theguardian.com/technology/2018/mar/20/facebook-is-it-time-we-all-deleted-our-accounts

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

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

Print this Page for Your Records Close Window

Control/Tracking Number: 2018-SW-2357-ISPE
Activity: Symposium/Workshop
Current Date/Time: 2/14/2018 9:52:36 PM

Long Live the "Medical Data Janitors": International Data Quality Assurance Practices in Distributed Data Networks

Author Block: Judith C. Maro[1], Christian G. Reich[2], Keith Marsolo[3], Yoshiaki Uyama[4], Kristian B. Filion[5], Miriam C. J. M. Sturkenboom[6]. [1]Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA; [2]IQVIA, Cambridge, MA; [3]Cincinnati Children's Hospital Medical Center, Cincinnatti, OH; [4]Pharmaceuticals and Medical Devices Agency, Tokyo, Japan; [5]McGill University, Montreal, QC, Canada; [6]University Medical Center Utrecht, Utrecht, Netherlands

Proposal / Abstract:

Background: Ensuring data quality for distributed data networks is challenging.

Objectives: We will examine international practices in five distributed data networks that house a mixture of administrative claims data and electronic health record data including: the U.S. Food and Drug Administration's (FDA's) Sentinel Initiative (Sentinel), the FDA's Biologics Effectiveness and Safety (BEST) Initiative, the U.S. National Patient Centered-Clinical Research Network (PCORnet), Japan's Medical Information Database Network (MID-NET), and the Canadian Network for Observational Drug Effect Studies (CNODES).

Duke | FORGE
HEALTH DATA IN ACTION

How it works: analytics and data science

Ask the right questions

- Tap knowledge to specify the question that defines the challenge

Find the right data

- Examine all data sources (access, feasibility, content, quality)

Optimize methods

- Tailor methods to address the question
- Develop new tools to improve analytics

Decide wisely

- Improve health
- Improve care
- Reduce cost
- Improve clinical practice

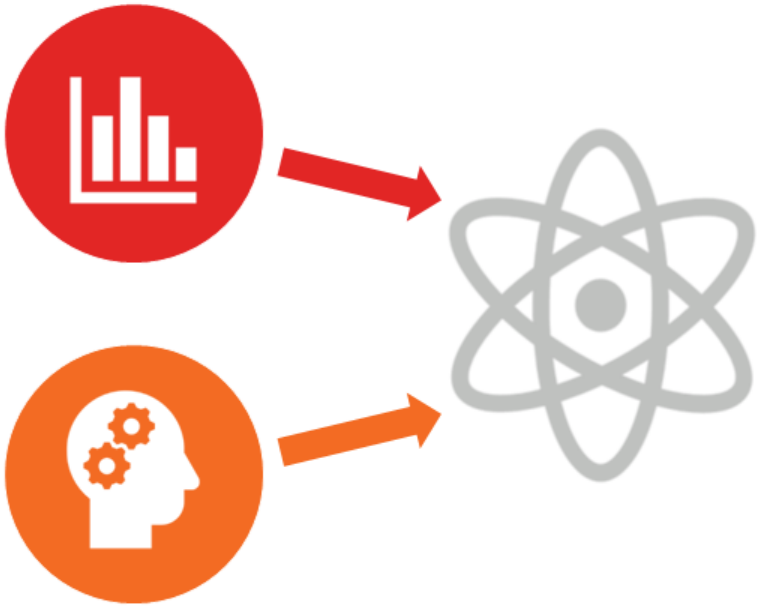
Duke | FORGE
HEALTH DATA IN ACTION

The grand fusion

Melding strengths across disciplines and between professionals

Fostering the **comprehensive toolbox across the spectrum** including frequentist statistics, Bayesian statistics, machine learning, and deep learning

Developing the **right framework for teams** including clinicians and quantitative expertise



- Biostatistics and bioinformatics
- Population health
- Clinical research
- Research training and support
- Oversight and quality assurance

- Basic science departments
- Clinical departments
- Clinical research units

- Engineering
- Computer science
- Statistical science
- Big data analytics

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

Duke | FORGE
HEALTH DATA IN ACTION

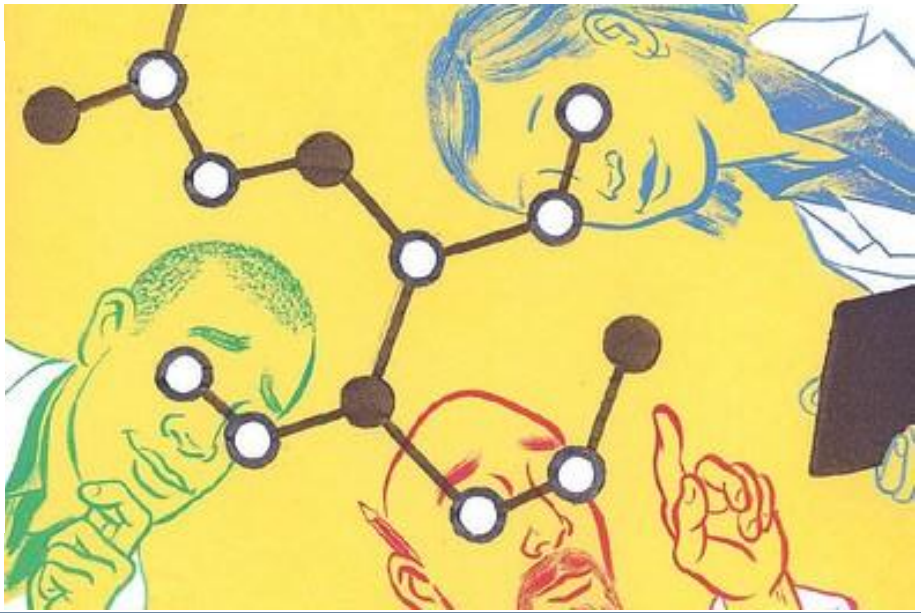
The New Einsteins Will Be Scientists Who Share

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.

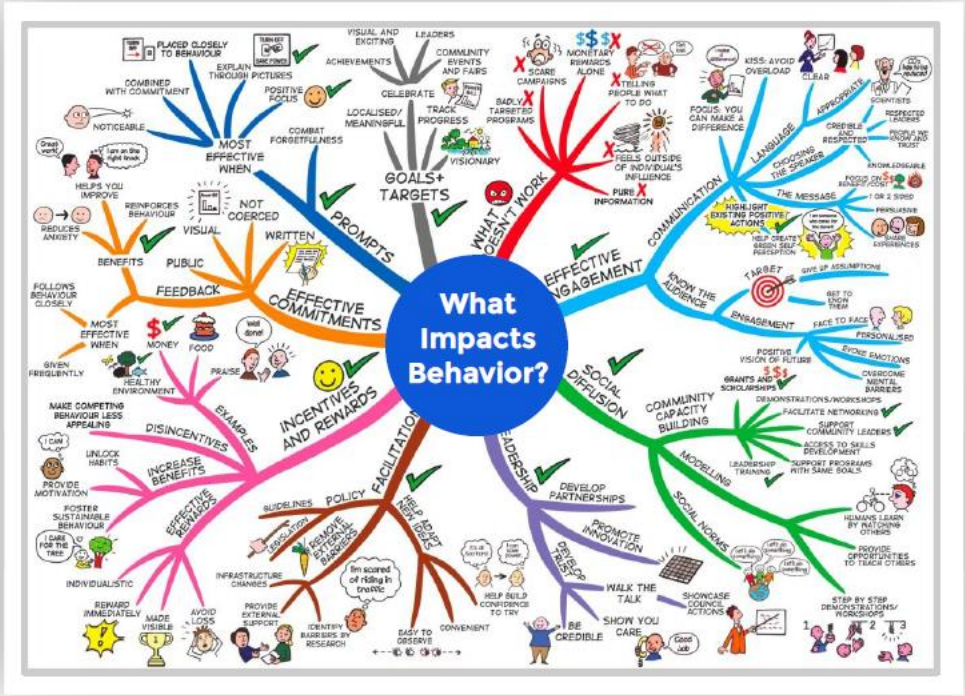
Several hours after Mr. Gowers opened up his blog for discussion, a Canadian-Hungarian mathematician posted a comment. Fifteen minutes later, an Arizona high-school math teacher chimed in. Three minutes after that, the UCLA mathematician Terence Tao commented. The discussion ignited, and in just six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

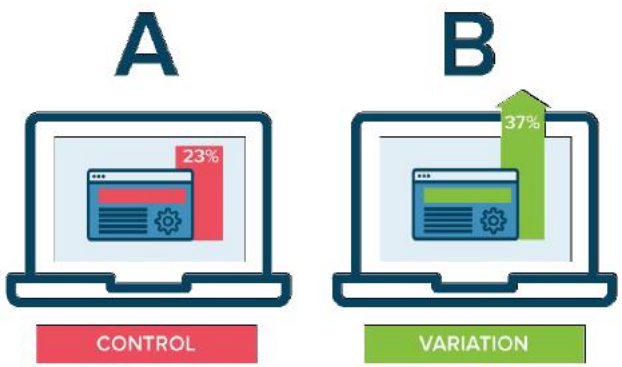


THE WALL STREET JOURNAL.

Duke | FORGE
HEALTH DATA IN ACTION

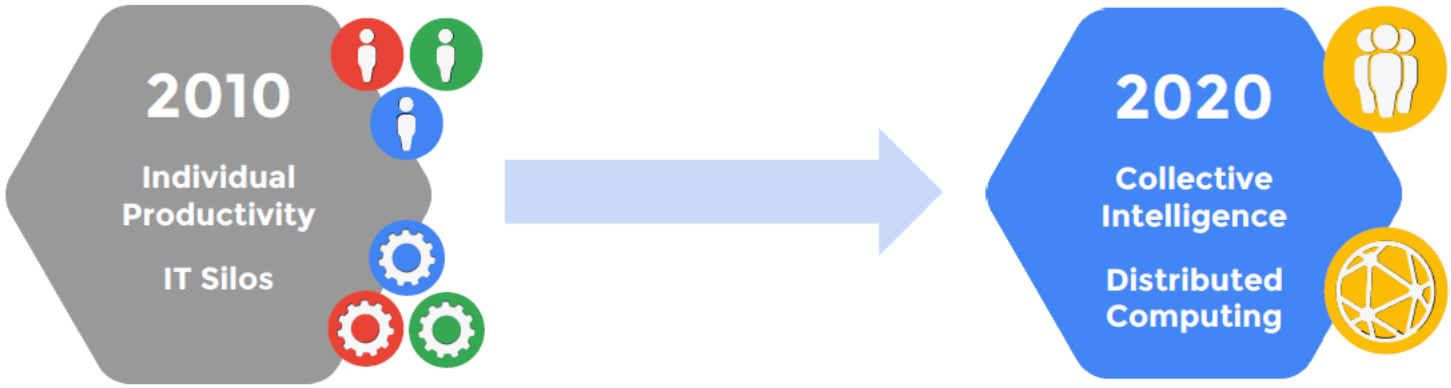
Data Activation and Testing Outcomes





Duke | FORGE
HEALTH DATA IN ACTION

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

Duke | FORGE
HEALTH DATA IN ACTION

CRITICAL ISSUES IN NEED OF PROGRESS

- Continued focus on curating the data (”data janitorship”)
- Connecting clinicians and people (consumers, patients, carers) in functional networks to advance knowledge
- Forming the right teams to convert data into knowledge
- Culture and ethics
  - Privacy and confidentiality
  - Monitoring the function of the system
  - Data sharing
  - Priority setting

Duke | FORGE
HEALTH DATA IN ACTION

USE OF BIG DATA: THE NEW VIEW

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."

Harvard Business Review Oct 2012

The shortage of data scientists is becoming a serious constraint in some sectors.