Closing The Real-World Evidence Gap: Pragmatic Clinical Trials & Observational Studies

December 4th, 2019

MassBio
MASSACHUSETTS BIOTECHNOLOGY COUNCIL

Register Today!

- **Building A Bulletproof Communications & Crisis Management Strategy**
  December 11, 2019 • 8:00 AM - 10:00 AM • MassBio Offices

- **MassBio Holiday Party**
  December 12, 2019 • 5:00 PM - 7:00 PM • Hyatt Regency, Cambridge

- **JP Morgan 2020 Recap**
  January 16th, 2020, 4:00PM – 6:00PM

For full forum schedule visit the MassBio website; go to Events, Forums.

BSDMCT Working Group

Co-Chairs:

Kevin Anderson, MBA, Director, Global Clinical Operations, Alexion Pharmaceuticals

Michelle Harrison, Associate Director, Clinical Data Management, Vertex Pharmaceuticals

Miganush Stepanians, PhD, President & CEO, PROMETRIKA, LLC

Ilker Yalcin, PhD, Vice President, Biostatistics, GSK

We are looking for additional Co-Chairs; if interested speak to us after the forum. Thank you!

Closing The Real-World Evidence Gap: Pragmatic Clinical Trials & Observational Studies

Our Distinguished Speakers:

- **Robert M. Califf, MD, MACC**, Former FDA Commissioner; Vice Chancellor for Clinical and Translational Research, Duke University; Head of Strategy and Policy for Verily Life Sciences and Google Health divisions
- **Jane Liang White, ScD**, Sr. Director, Statistical Group Lead for Oncology Hematology Franchise, Pfizer
- **Rebecca Miksad, MD**, Senior Medical Director, Flatiron Health
- **Miganush Stepanians, PhD**, President & CEO, PROMETRIKA, LLC **(Moderator)**

If you have a question, please raise your hand and wait for the microphone.
Thank you!

Innovative Clinical Development Solutions

Closing The Real-World Evidence Gap: Pragmatic Clinical Trials & Observational Studies

December 4, 2019

PROMETRIKA
A FULL-SERVICE CRO

Agenda

- Opening remarks
- Individual panelists presentations
- Moderated discussion with audience
  - Q&A dialogue
  - Audience to share experiences

Definitions

- **Real-World Data (RWD):** data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources.
- **Real-World Evidence (RWE):** clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD.
- **Pragmatic Randomized Clinical Trial:** a randomized clinical trial (RCT) embedded into real clinical practice with eligibility criteria designed to enroll a diverse/broad population of patients and capturing clinical data already collected as a part of routine care.

Definitions - Continued

- **Patient Registry:** an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined either by diagnosis of a disease (disease registry) or usage of a treatment (exposure registry).
- **Prospective Observational Study:** a non-interventional clinical study in which the population of interest is identified at the start of the study, and exposure/treatment and outcome data are collected from that point forward.
- **Retrospective Observational Study:** a clinical study that identifies the population and determines the exposure/treatment from historical data (i.e., data generated before the initiation of the study).

PROMETRIKA
A FULL-SERVICE CRO

Explanatory & Pragmatic RCTs & Observational Studies

	Interventional?	Randomized?	Prospective?	RWD Collection
Explanatory RCT	Yes	Yes	Yes	
Pragmatic RCT	Yes	Yes	Yes	
Single arm trial with external control	Yes	No	Yes	
Patient Registry	No	No	Yes	
Prospective Observational Study	No	No	Yes	
Retrospective Observational Study	No	No	No	

PROMETRIKA
A FULL-SERVICE CRO

Speakers

Robert M. Califf, MD, MACC

- Vice Chancellor for Clinical and Translational Research, Duke University

Jane Liang White, ScD

- Sr. Director, Statistical Group Lead for Oncology Hematology Franchise, Pfizer

Rebecca Miksad, MD

- Senior Medical Director, Flatiron Health

Miganush Stepanians, PhD

- President and CEO, PROMETRIKA, LLC (Moderator)

PROMETRIKA
A FULL-SERVICE CRO

Overview of Discussion Topics

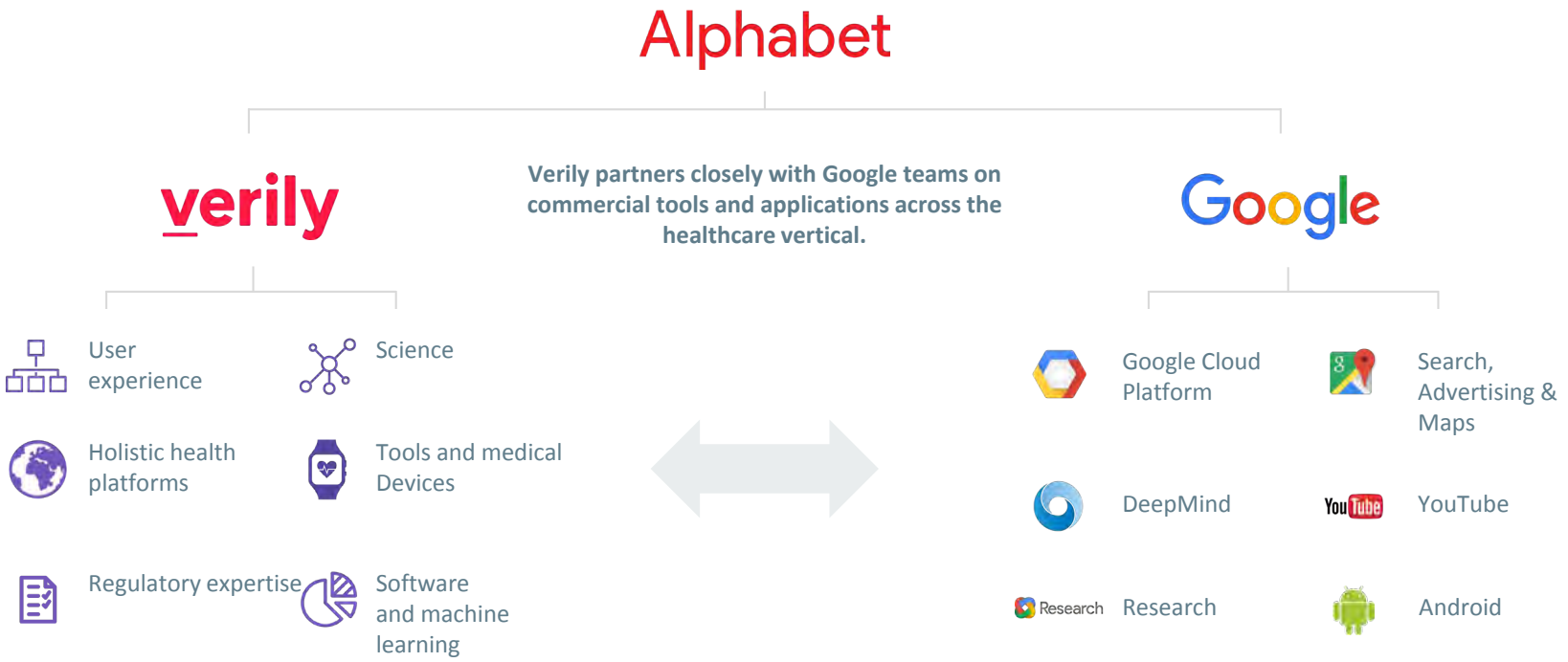
- Real-World Evidence in Drug Development
  - Regulatory Perspective
- Pragmatic Clinical Trials and Observational Studies
  - Present and Future
- Case Study: Single Arm Trial with Synthetic Control Arm
  - Statistical and Study Design Considerations
- Sources of Real-World Data
  - Available Databases

PROMETRIKA
A FULL-SERVICE CRO

Evolving Changes in Evidence Generation to Assess the Benefits, Risks and Value of Medical Products

Robert M Califf MD
Head of Medical Strategy and Policy
Verily Life Sciences and Google Health

The Alphabet family



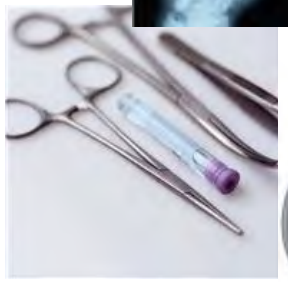
FDA Regulates a Spectrum of Health Products : 20-25 cents of every GDP dollar

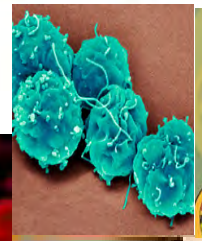
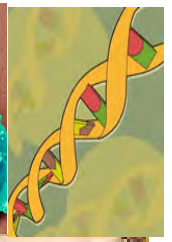
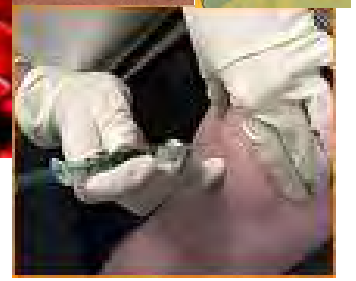








FDA Mission

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

FDA Mission

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

FDA Mission

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

FDA Mission

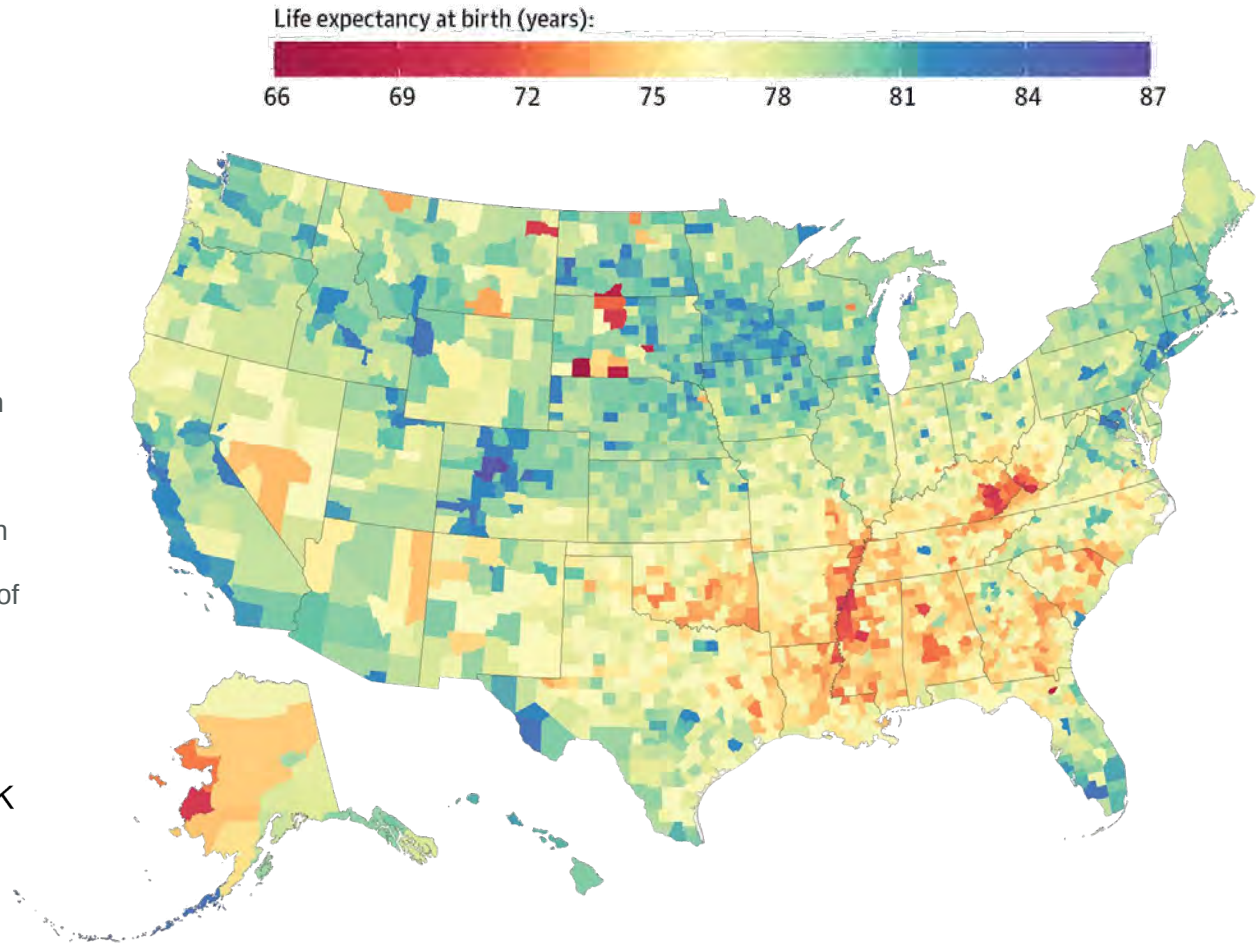
Finally, FDA plays a significant role in the Nation's counterterrorism capability. FDA fulfills this responsibility by ensuring the security of the food supply and by fostering development of medical products to respond to deliberate and naturally emerging public health threats.

The FDA: Big Picture

- Regulatory Agency
- Science Agency
- Public Health Agency
- Multiple disciplines always in play
  - Science/Medicine/Public Health
  - Policy
  - Law

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.

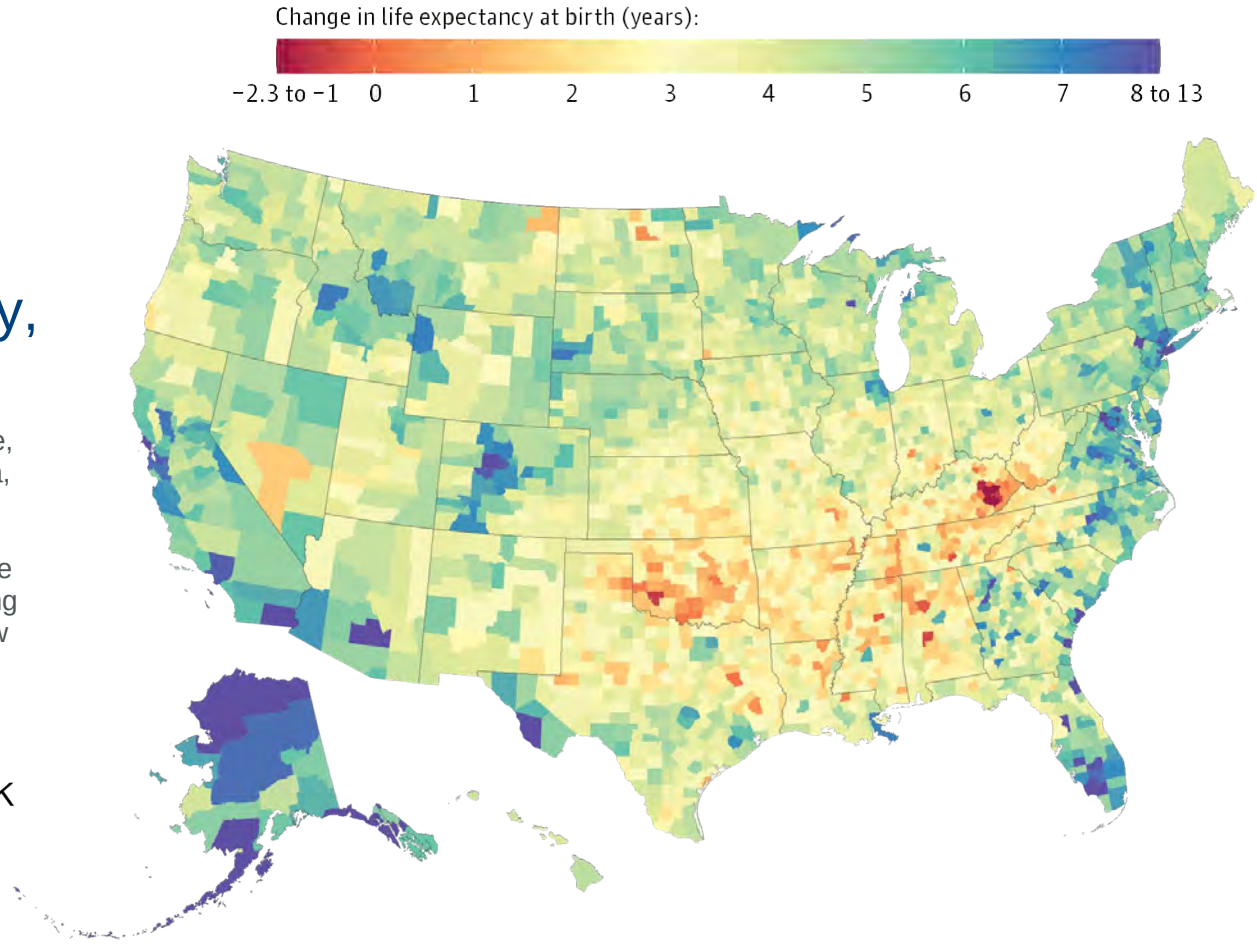
The JAMA Network

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

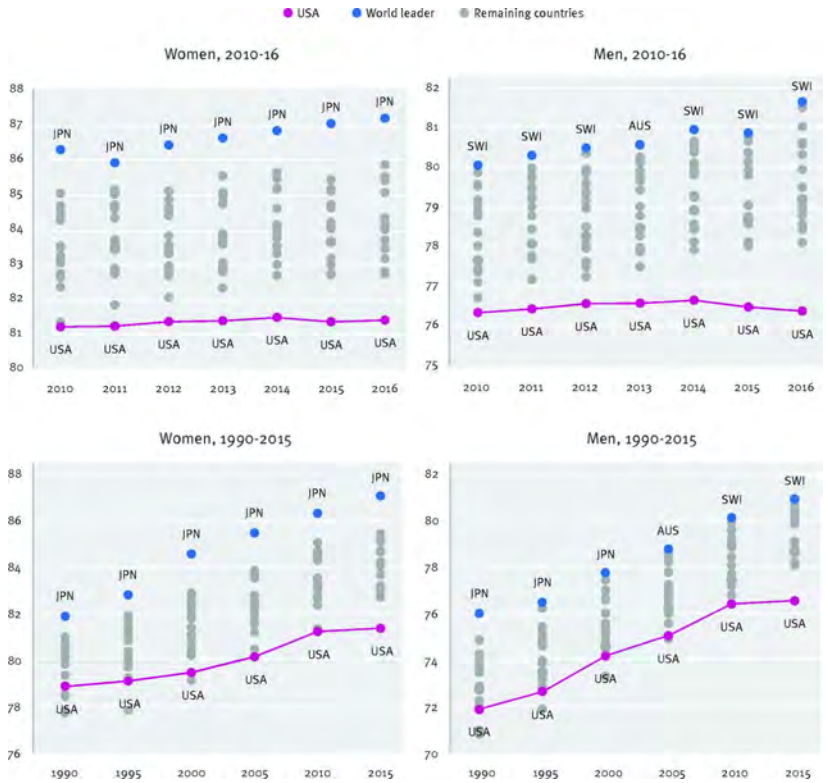
Variable	Summary Statistics, Mean (SD) [Range]	Bivariate Regression Results	
		Coefficient (SE)	R^2
Socioeconomic and race/Ethnicity factors			
Population below the poverty line, %	16.3 (6.4) [3.1-62.0]	−0.24 (0.005)	0.47
Median household income, log $	10.6 (0.2) [9.8-11.6]	6.06 (0.130)	0.41
Graduates, 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

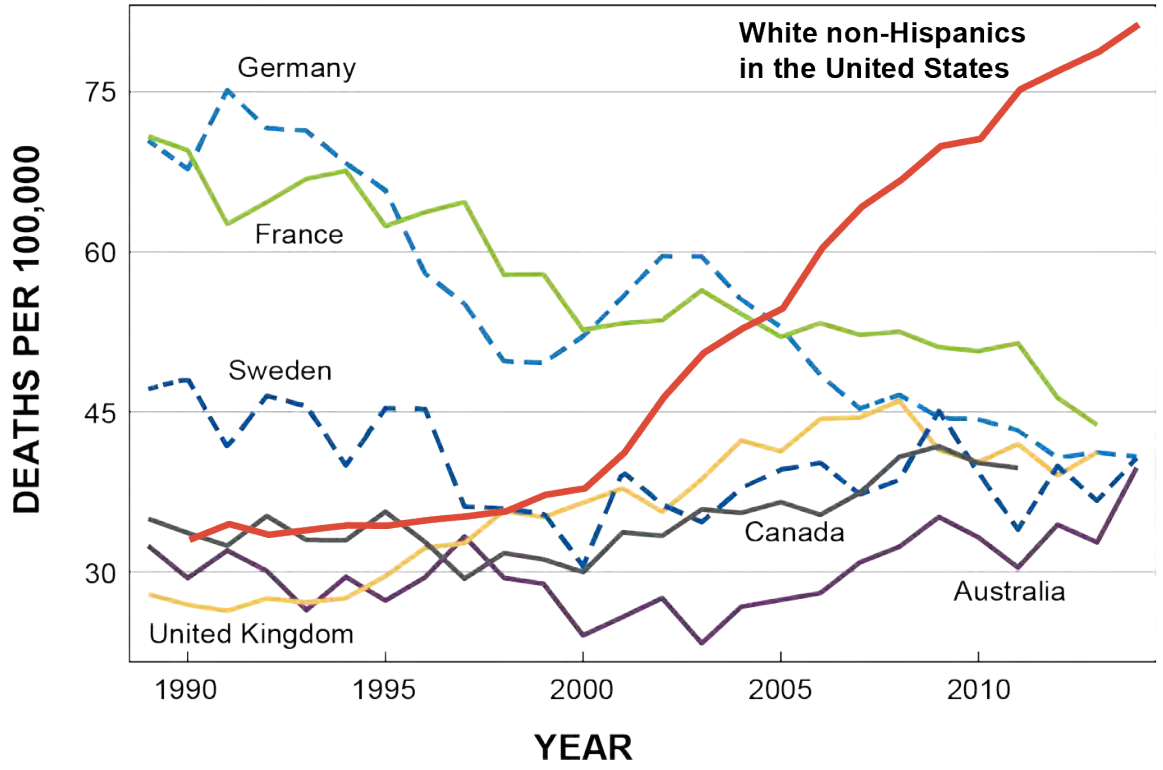
Life expectancy at birth (years) in 18 high income countries for women and men during 2010-16 and 1990-2015.



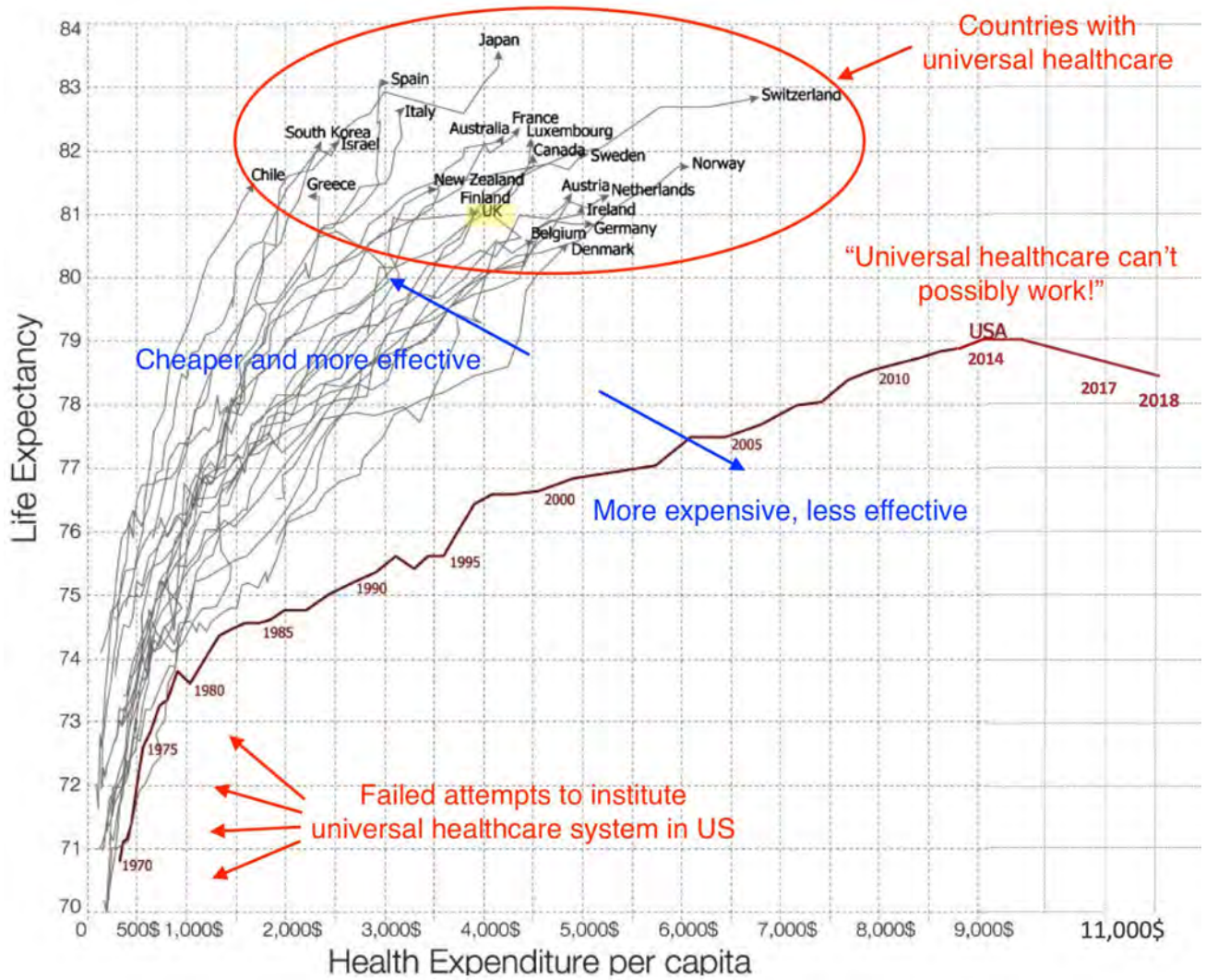
Jessica Y Ho, and Arun S Hendi BMJ 2018;362:bmj.k2562

Midlife mortality from “deaths of despair” across countries

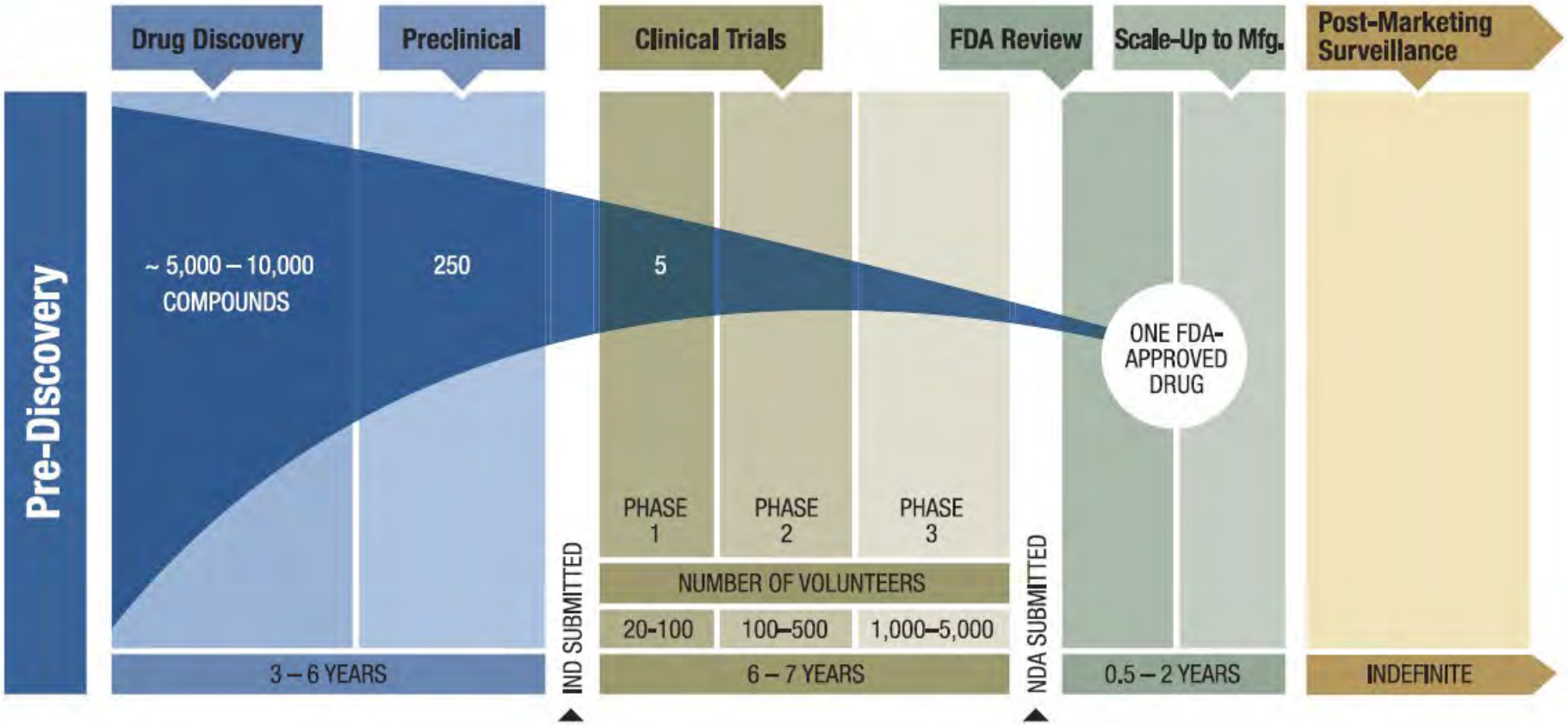
Men and women ages 50–54, deaths by drugs, alcohol, and suicide, 1989–2014



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


Countries with universal healthcare
Japan
Spain
Italy
Switzerland
France
Australia
Luxembourg
South Korea
Israel
Canada
Sweden
Norway
Chile
Greece
New Zealand
Austria
Netherlands
Finland
UK
Ireland
Belgium
Germany
Denmark
"Universal healthcare can't possibly work!"
Cheaper and more effective
USA
2014
2010
2017
2018
2005
2000
More expensive, less effective
1995
1990
1985
1980
1975
1970
Failed attempts to institute universal healthcare system in US
Life Expectancy
84
83
82
81
80
79
78
77
76
75
74
73
72
71
70
0
500$
1,000$
2,000$
3,000$
4,000$
5,000$
6,000$
7,000$
8,000$
9,000$
11,000$
Health Expenditure per capita

Drug Discovery and Development Timeline



American Association of Cancer Research 2011 Cancer Progress Report

Top 10 drugs in the United States: evidence for a massive structural shift in drug development
(courtesy of Clive Meanwell; Medicines Company)

	2000	2015	Change
Revenue	$34 billion	$84 billion	2.5-fold increase
Population addressed	413 million	54 million	7.5-fold decrease

2000: Celebrex, Claritin, Glucophage, Lipitor, Paxil, Prevacid, Prilosec, Prozac, Zocor, Zoloft

2015: Avastin, Embrel, Harvoni, Herceptin, Humira, Lantus, Remicade, Revlimid, Rituxan, Solvadi

Our National Clinical Research System is Well-intentioned But Flawed

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

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

Tricoci P et al. JAMA 2009;301:831-41

Which Treatment is Best for Whom?

High-Quality Evidence is Scarce

< 15% of Guideline Recommendations Supported by High Quality Evidence

ORIGINAL CONTRIBUTION

Scientific Evidence Underlying the ACC/AHA Clinical Practice Guidelines

Pierluigi Tricoci, MD, MHS, PhD
Joseph M. Allen, MA
Judith M. Kramer, MD, MS
Robert M. Califf, MD
Sidney C. Smith Jr, MD

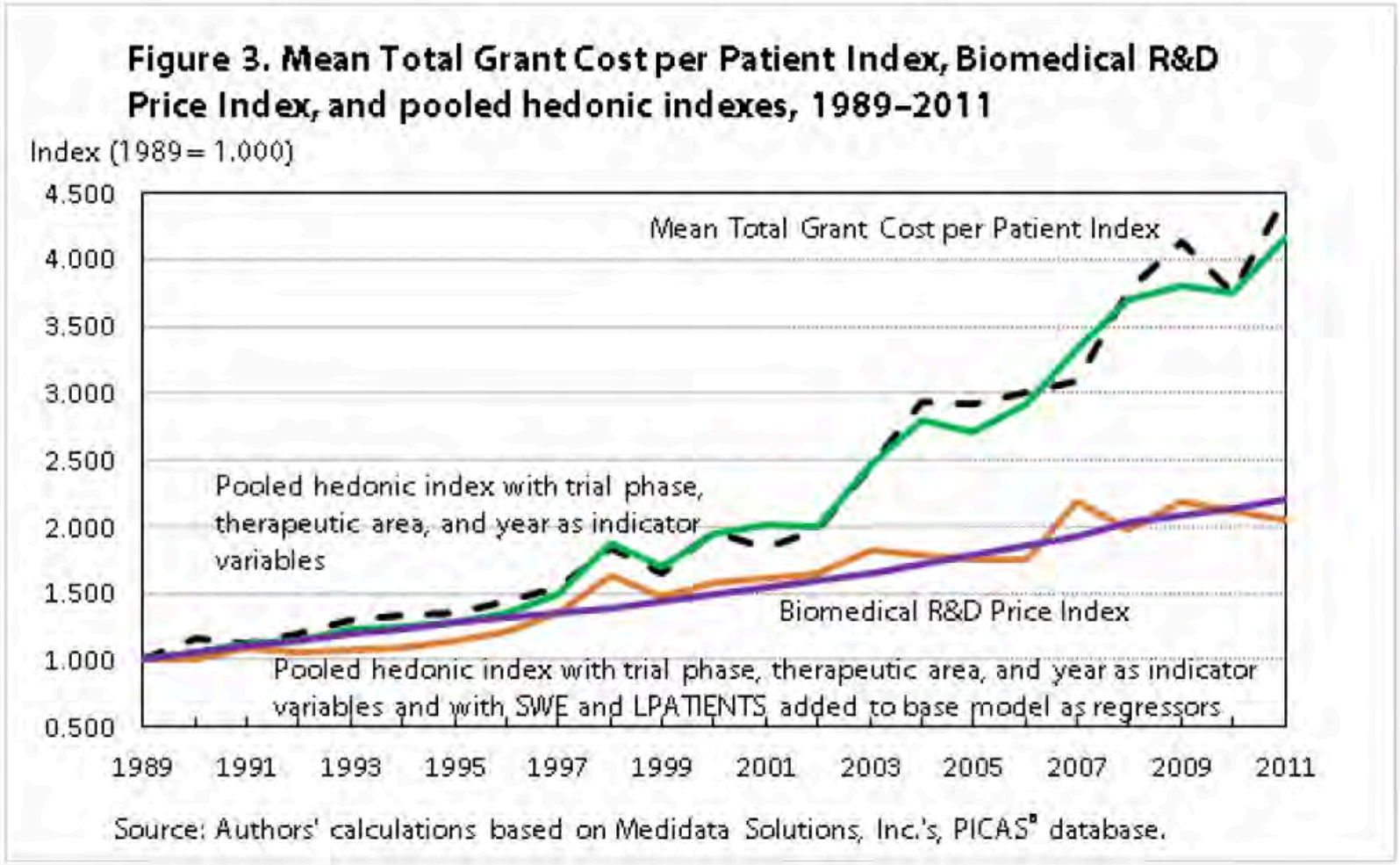
CLINICAL PRACTICE GUIDELINES are systematically developed statements to assist practitioners with decisions about appropriate health care for spe-

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

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

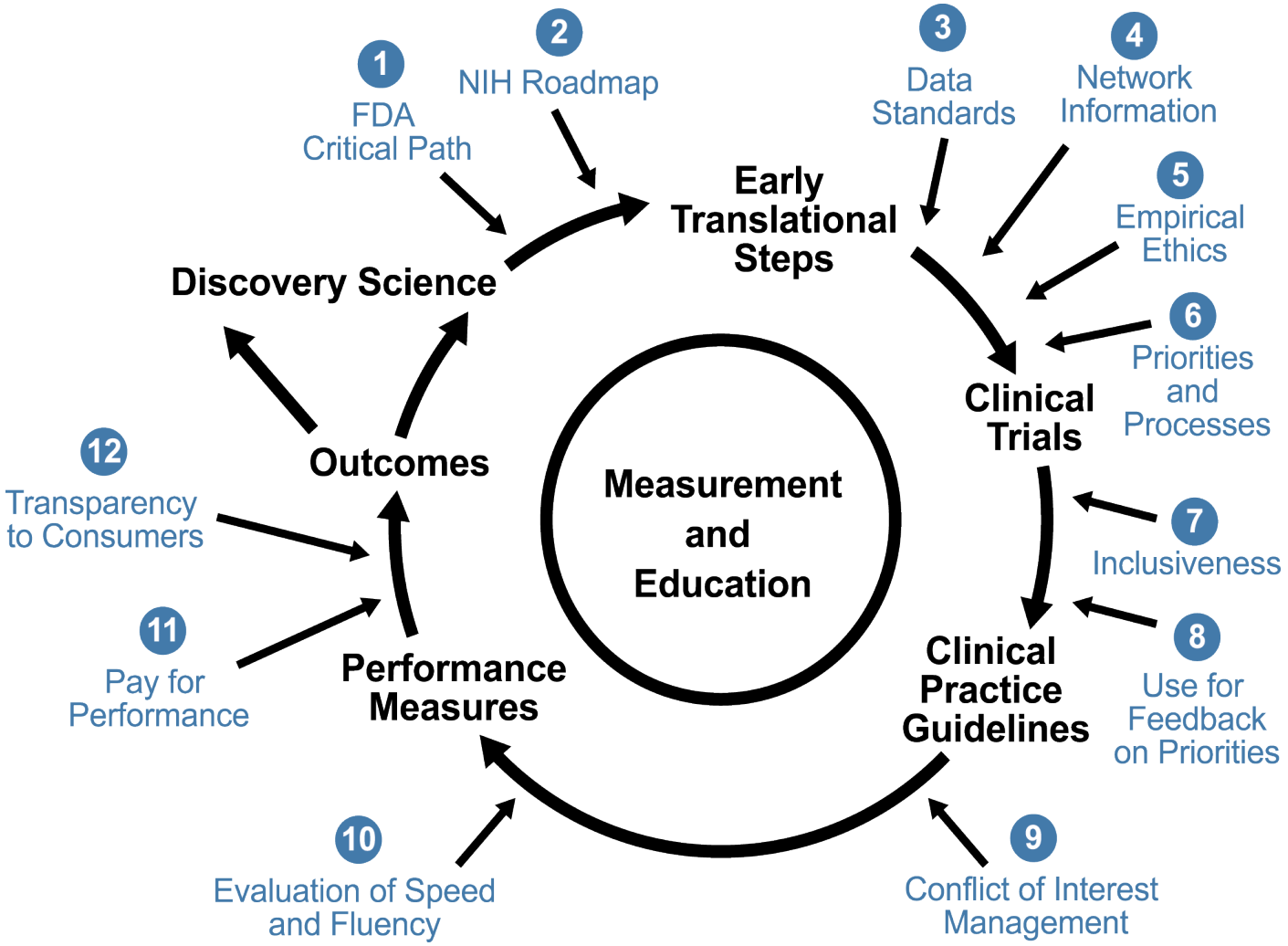
Data Sources and Study Selection Data from all ACC/AHA practice guidelines issued from 1984 to September 2008 were abstracted by personnel in the ACC Science and Quality Division. Fifty-three guidelines on 22 topics, including a total of 7196 recommendations, were abstracted.

Trial Hyperinflation



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

Generating Evidence to Inform Decisions



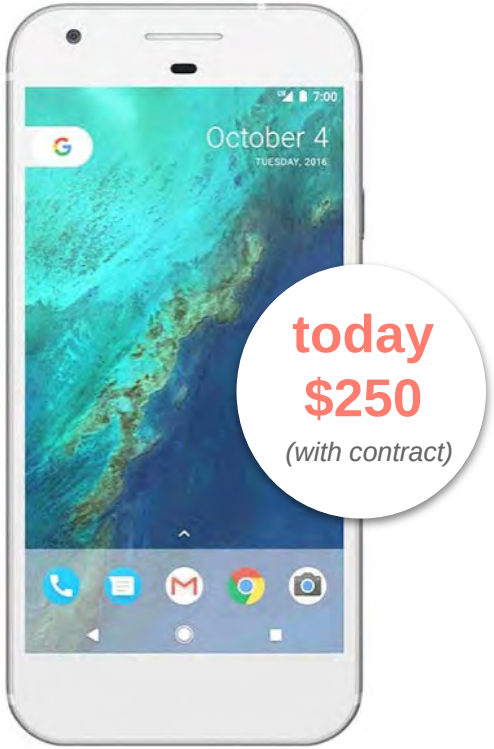
Duke | FORGE
HEALTH DATA IN ACTION


“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



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)

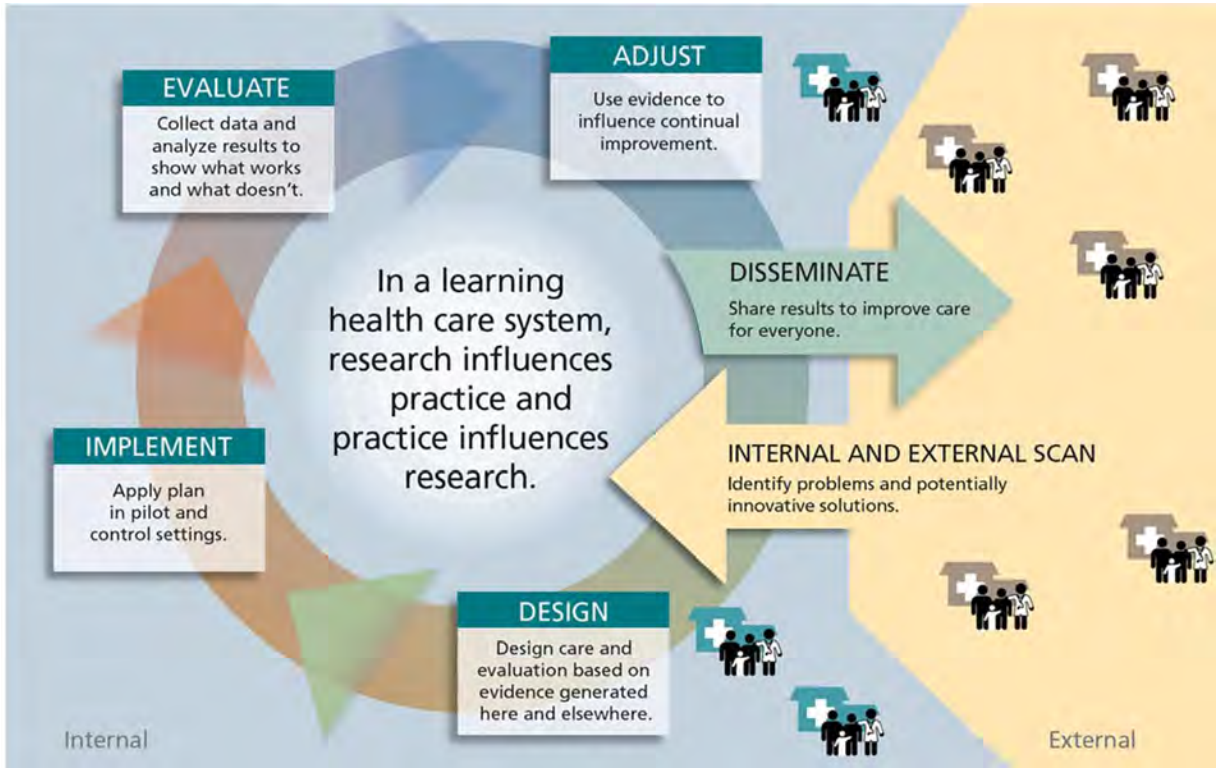
Encyclopedia
$2,200 (Encyclopedia)

Processor
$32M (Cray)

Portable TV
$665 (Casio Mini TV)

Video Conference
$110,520 (Future Sys)

Learning health care systems



www.fda.gov

Pragmatic Trials

- An intent to inform decision makers (patients, clinicians, administrators and policy makers) as opposed to elucidating a biological or social mechanism
- An intent to enroll a patient population relevant to the decision in practice and representative of the patients/populations and clinical setting for whom the decision is relevant
- Either an intent to:
  - Streamline procedures and data collection so that the trial can focus on adequate power for informing the clinical and policy decisions targeted by the trial or
  - Measure a broad range of outcomes
    - Califf and Sugarman; Clinical Trials 2015; 12: 436-441

The Core FDA Issue in Medical Products

- Do the benefits outweigh the risks for the condition of use for which the product is labeled?
  - Adequate and well controlled clinical studies
- Is the device safe and effective for its intended use?
  - Valid scientific evidence

Substantial Evidence

- “evidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, **on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use** prescribed, recommended, or suggested in the labeling or proposed labeling thereof.”

Adequate and Well Controlled

- To demonstrate that a trial supporting an effectiveness claim is adequate and well-controlled, extensive documentation of trial planning, protocols, conduct, and data handling is usually submitted to the Agency, and detailed patient records are made available at the clinical sites. **From a scientific standpoint, however, it is recognized that the extent of documentation necessary depends on the particular study, the types of data involved, and the other evidence available to support the claim.** Therefore, the Agency is able to accept different levels of documentation of data quality, as long as the adequacy of the scientific evidence can be assured.
    - Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drugs and Biological Products

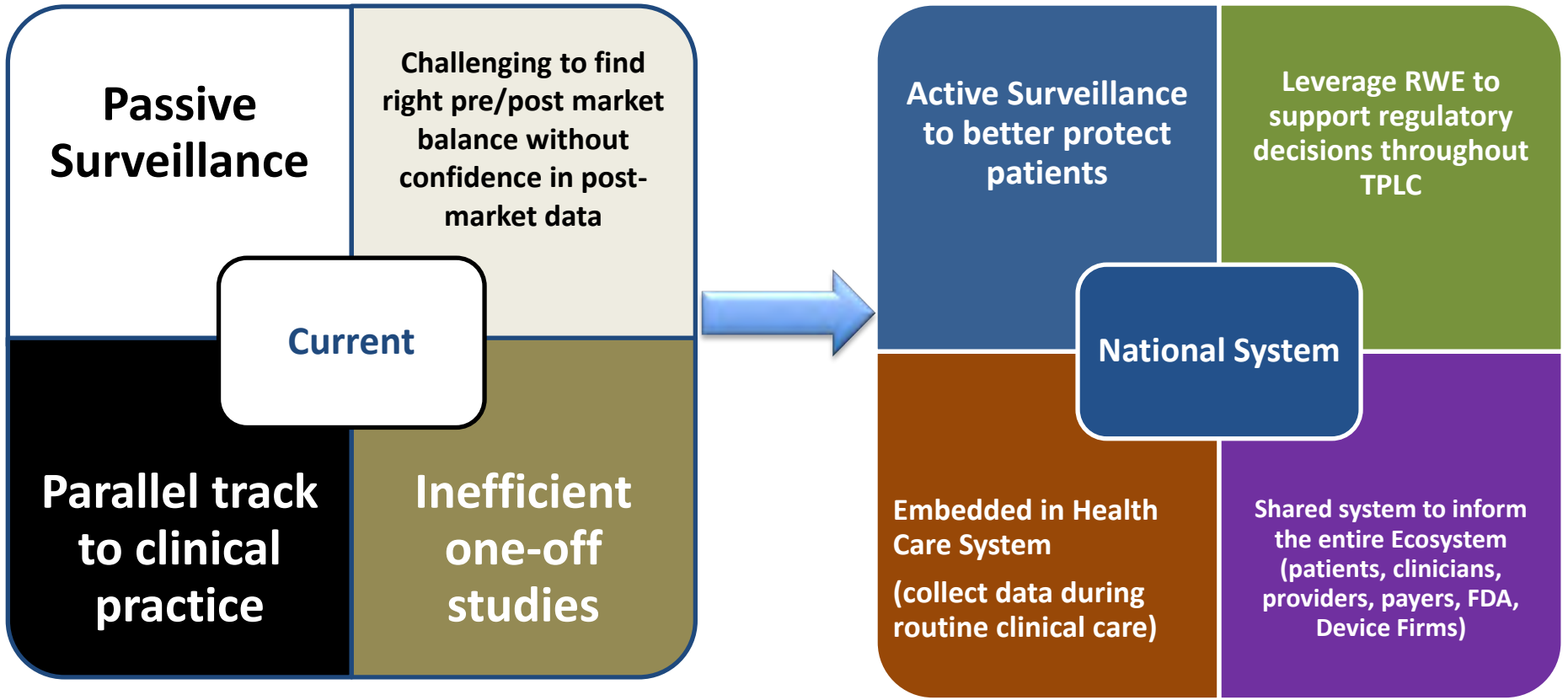
Good Clinical Practice

- An international ethical and scientific quality standard for the design, conduct, performance, monitoring, auditing, recording, analyses and reporting of clinical trials. It also serves to protect the rights, integrity and confidentiality of trial subjects.
- Instantiated in ICH documents
  - The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) is brings together the regulatory authorities and pharmaceutical industry to discuss scientific and technical aspects of drug registration.

The Evidence Continuum

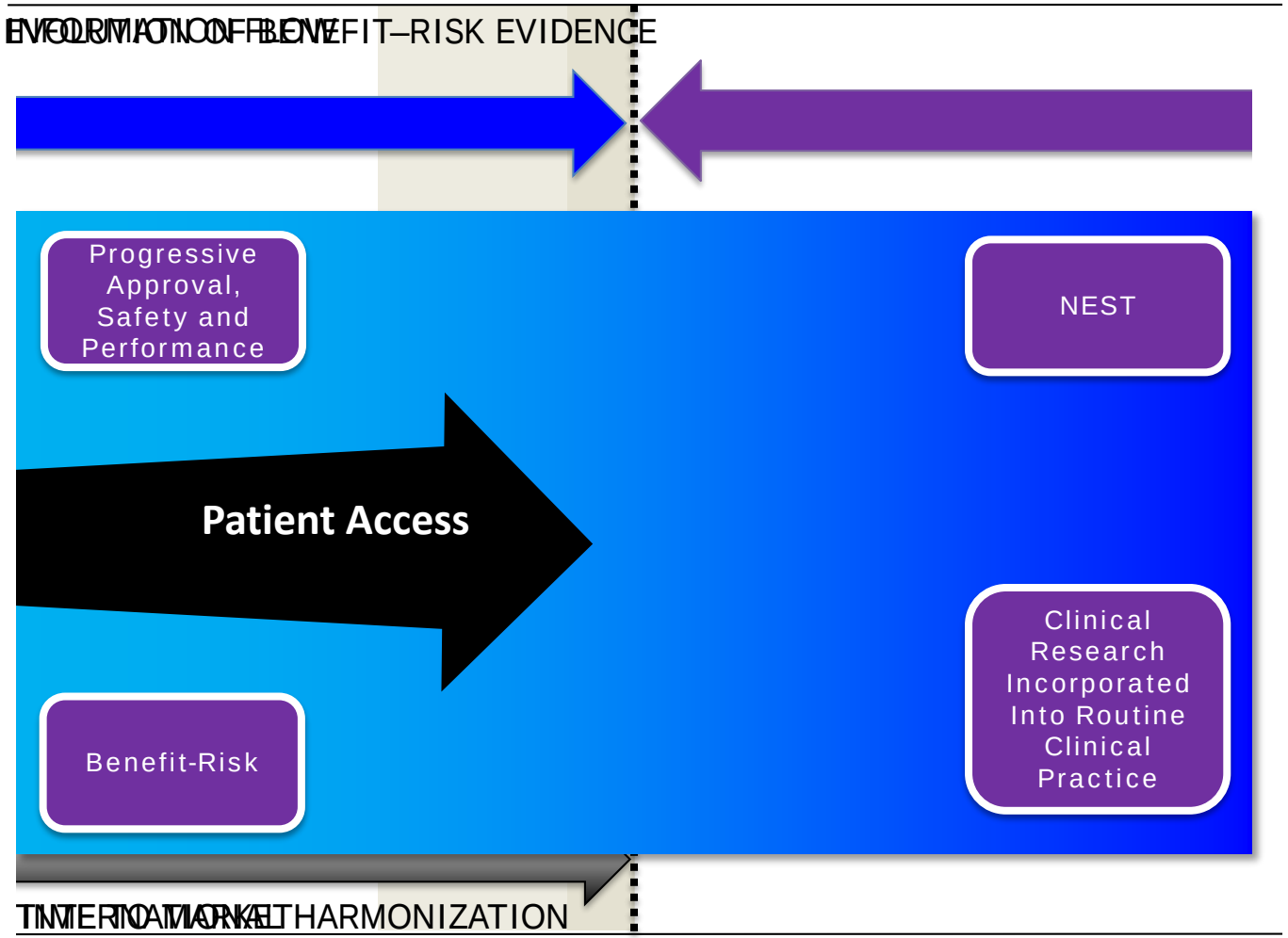
- Step 1: Regulatory approval for marketing
- Step 2: Health Technology Assessment
- Step 3: Payor Decisions
- Step 4: Individual provider/patient decisions
- **PREMISE: THIS SHOULD BE A CONTINUUM, NOT DISCRETE STEPS**

National System Paradigm Shift



Learning Medical Device Ecosystem

Total Product Life Cycle (TPLC) Framework



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

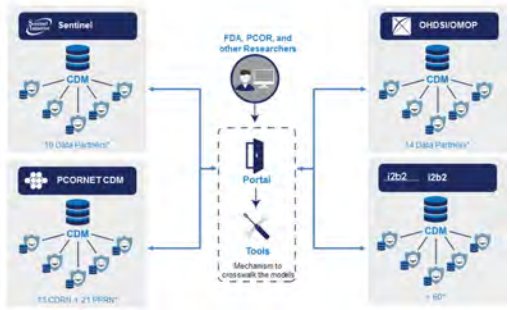
Reinforcing of a Learning Health Care System:

- 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

Data Standards

Stakeholder Engagement



Duke | MARGOLIS CENTER for Health Policy

CLINICAL TRIALS TRANSFORMATION INITIATIVE

Guidances

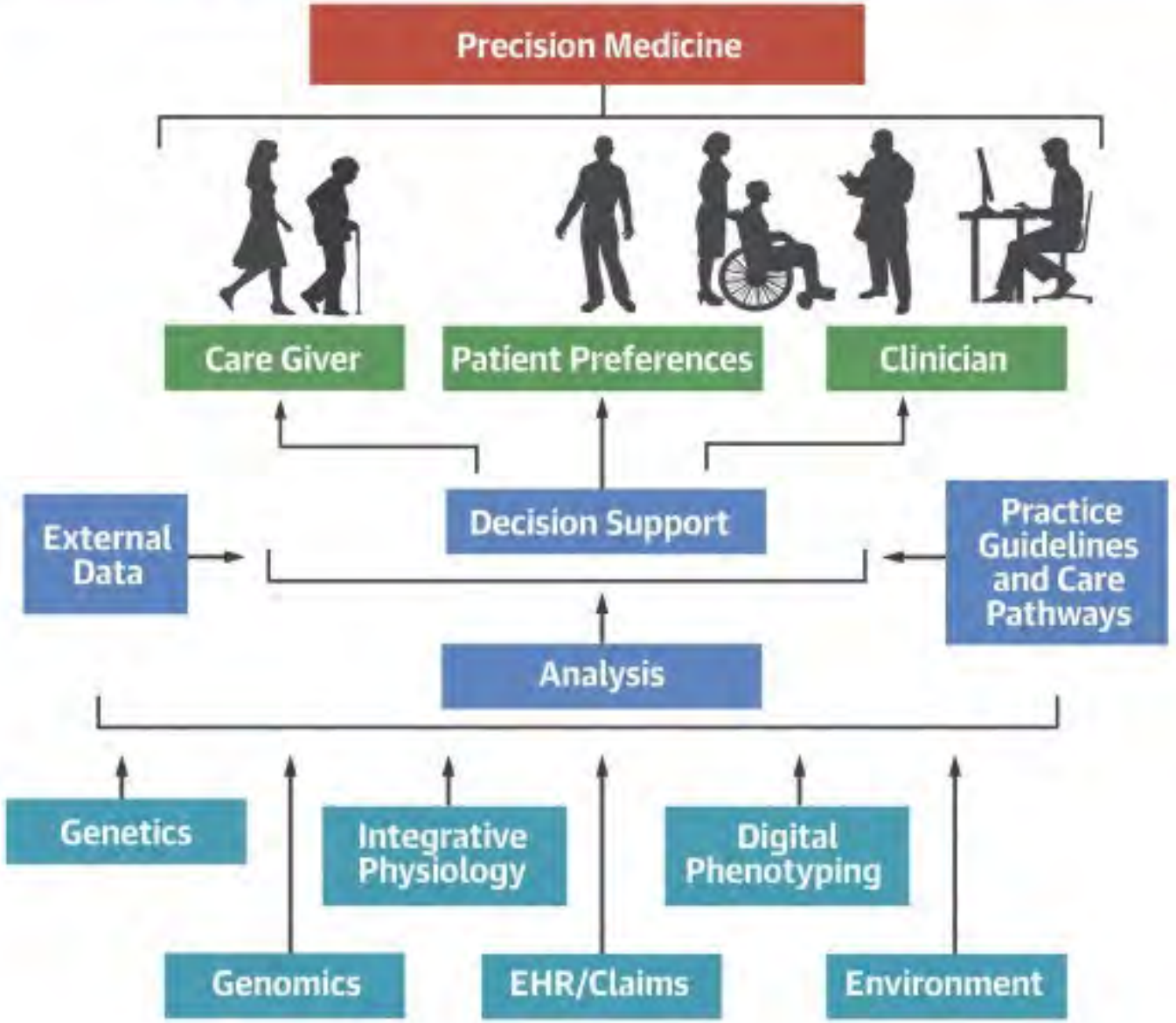
Draft: Use of Electronic Health Record Data in Clinical Investigations

Demonstratio

Electronic Source Data in Clinical Investigations

Use of Electronic Informed Consent

CENTRAL ILLUSTRATION: Precision Medicine



Califf, R.M. J Am Coll Cardiol. 2018;72(25):3301-9.

Designing a Registrational Study with Real World Data as a Synthetic Control Arm

Jane Liang White, ScD
Tao Wang, PhD

Pfizer GLOBAL PRODUCT DEVELOPMENT

TIME IS LIFE
Every Minute Matters

Recent Trend in Regulatory Approvals

- **Amgen was among the first to use historical RWD comparator to support accelerated approval of Blincyto in March 2018 for 1L treatment of Acute Lymphoblastic Leukemia (ALL) Minimal Residual Disease positive (MRD+) patients based on "high unmet need"**
  - ➢ Blincyto was previously approved by the FDA for the treatment of adult and pediatric patients with relapsed or refractory B-cell precursor ALL

- **Pfizer received the approval for Ibrance for male breast cancer in April 2019. Dr. Richard Pazdur stated in a press release:**

"Today we are expanding the indication for Ibrance to include male patients based upon data from post marketing reports and **electronic health records** showing that the safety profile for men treated with Ibrance is consistent with the safety profile in women treated with Ibrance,"

  - ➢ FDA approved Ibrance for use in combination with Faslodex in pretreated patients with HR-positive, HER2-negative metastatic breast cancer in 2016

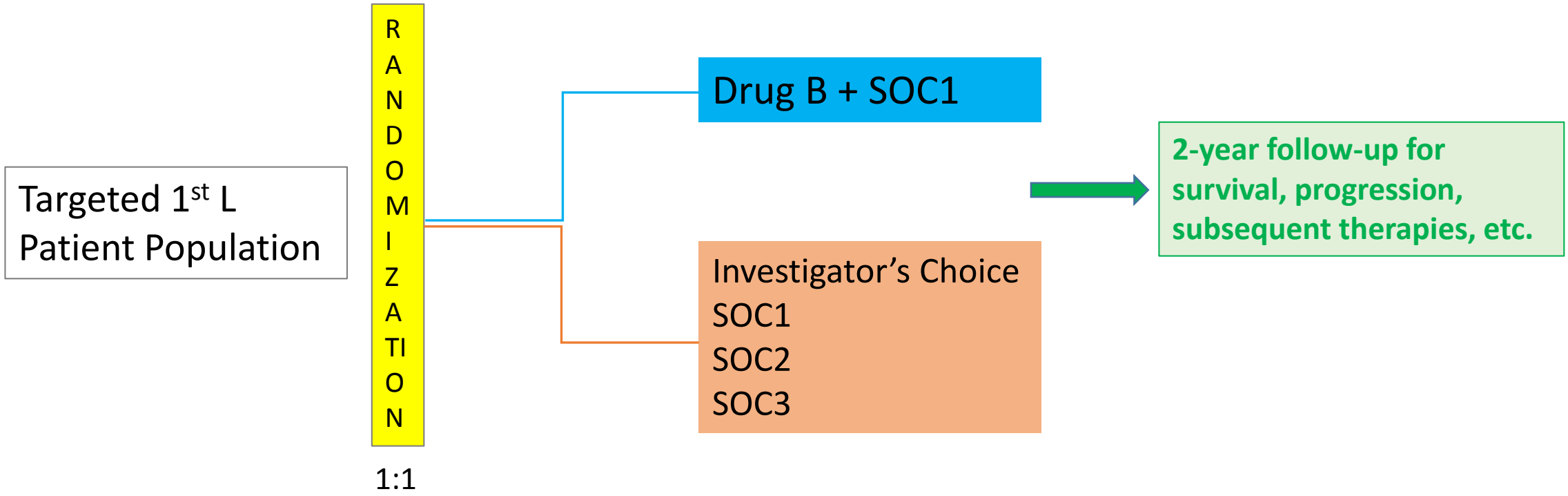
Evolving Regulatory Mindset and Pros of Using RWD

- FDA issued Framework for Real World Evidence (RWE) Program in Dec., 2018
  - Evaluates the potential use of RWE to help support the approval of a new indication for a drug already approved or to help support or satisfy drug post-approval study requirements
  - Evaluating RWE in the context of regulatory decision-making depends not only on the evaluation of the methodologies used to generate the evidence but also on the reliability and relevance of the underlying RWD; these constructs may raise different types of considerations.
- FDA held a Webinar for the Framework in March 2019
- Pros of using RWD - Faster to patients, significant cost reduction, and Extended long-term follow-up data, etc.

Our Situation for Drug B

- Previously approved by FDA and EMA for relapsed/refractory patient population with disease A
- Planning to expand the indication for the 1st Line treatment of disease A
  - Competitive landscape
  - Cost efficient
  - Fast to patients
  - Regular full approval as the optimistic goal and accelerated approval as the baseline goal

Traditional Design of a Registrational Study



Stratification factors:
- Factor 1
- Factor 2

Primary Endpoints: Response and/or Overall Survival (OS)
Required sample size N = 396
FSFV to Primary Completion Date (PCD) for response ~24 months
Total study duration (for OS) ~50+ months

Lessons Learned from Selinexor's ODAC

- At the Oncologic Drugs Advisory Committee (ODAC) meeting on February 26, 2019, Karyopharm Therapeutics, Inc sought approval of selinexor, an oral, first-in class, exportin 1 (XPO1) inhibitor, in combination with low-dose dexamethasone for the treatment of patients with relapsed refractory multiple myeloma (RRMM) who have received at least 3 prior therapies and whose disease is triple-refractory.

- The NDA is primarily based on Part 2 of the phase 2b trial, KCP-330-012 (STORM). STORM was a multicenter, open-label, single arm trial evaluating selinexor in combination with dexamethasone in patients with RRMM. Part 2 enrolled 123 patients. The primary endpoint was overall response rate (ORR), and key secondary endpoints included duration of response (DOR), progression-free survival (PFS), and OS.

- STORM study showed activity in primary endpoint ORR.

Lessons Learned from Selinexor's Recent ODAC

- Selinexor team included a retrospective observational study (KS-50039) in the submission that attempted to characterize the survival distribution of patients similar to those in STORM Part 2 using real-world data (RWD) and to compare it with the OS result from STORM Part 2.
- The FDA first stated that "Agency is committed to the use of Real-World Data (RWD) to support regulatory decision-making and recently published a Framework outlining considerations for RWD studies" However, the agency identified the following issues.
  - RWD analyses should be pre-specified and discussed with the Agency to ensure they are carefully designed to minimize bias. KS-50039 was not pre-specified or discussed with the Agency and has design issues that lead to bias and confounding.
  - Selection criteria were not aligned resulting in critical differences between the Flatiron (FHAD) population and the population evaluated in STORM

The agency concluded that **comparison of survival between FHAD and STORM is not appropriate**.

Key Learnings from the Recent Cases

Recent Cases

- Blincyto's ODAC on Mar. 7, 2018 with favorable votes and the subsequent approval by the FDA
- Selinexor's ODAC on Feb. 26, 2019 with ODAC's vote (8:5) to delay the decision until the readout of the ongoing Phase 3 study. Subsequently FDA made a final decision in sync with ODAC's recommendation on April 1, 2019
- Ibrance approval for male breast cancer patients on April 4, 2019

A few key points for future study design and analysis seeking approval using RWD

- Pre-specify study design and analysis using RWD
  - Comparable patient populations (same key selection criteria) between the investigational data and the RWD
  - Propensity Score methods to remove/reduce confounding effect, e.g., propensity score matching, inverse probability of treatment weighting (IPTW), etc.
- Consult with FDA on the plan before the initiation of the study

Alternative Study Design Utilizing the RWD – Part 1

A single-arm design with 200 patients enrolled and treated with drug B + SOC1

- Primary endpoint: Response
- Key secondary endpoint: OS
- 1-sided alpha=0.025
- >90% power
- H_0: Response rate<=60%
- Ha: Response rate>=80%
- FSFV to PCD: 14 mons
- Total study duration: 40 mons

Alternative Study Design Utilizing RWD Control – Part 2

Since it is unknown how many patients will be included in the RWD control arm, 3 scenarios for the potential number of patients in the RWD control arm (200, 100 and 67 patients [i.e., equivalent to 1:1, 2:1, and 3:1 randomization ratio]), provide a range of possible design characteristics for comparison of patients treated with drug B combo arm and the RWD control arm assuming the study were a 1:1, 2:1, or 3:1 randomized study with 1-sided alpha of 0.025.

Note that power of testing response rate between the single arm (80%) and the RWD control arm (60%) would still be adequate (>85%) even with the 3:1 ratio.

Ratio	Sample Size	OS HR (Δ%)	Total # OS Events (% of Total Patients)	Power
1:1	400 (200/200)	0.667 (50%)	260 (65%)	90%
2:1	300 (200/100)	0.667 (50%)	224 (75%)	80%
3:1	267 (200/67)	0.667 (50%)	181 (67%)	60%

Alternative Study Design – Testing Sequence

If response rate is tested positive in the single-arm study (i.e., 1-sided p-value <0.025), then proceed to comparisons below and gatekeeping testing strategies will be used to adjust for multiple statistical testing and to control the overall Type I Error rate at 0.025 (1-sided).

- Compare response rates between the single-arm and the synthetic control arm from RWD, if positive (i.e., 1-sided p-value<0.025), then
- Compare the key secondary endpoint OS between the single-arm and the synthetic control arm from RWD at 1-sided 0.025 significance level.

Considerations of the Protocol

- The alternative study design proposed essentially consists of two studies
  - Single-arm study
  - Comparisons of efficacy and safety between the single-arm study and the RWD control arm
- Given the regular full approval as our optimistic goal, one protocol with two studies included has been planned.

Question: How to mitigate bias and concerns for conducting comparative analyses for the data from two non-randomized arms?

Propensity Score

- The propensity score is a balancing score: conditional on the propensity score, the distribution of observed baseline covariates will be similar between the patients in the two arms.
- The propensity score can help mimic the effect of randomization by creating a balance between the two arms.
- A propensity score for each patient is often estimated using a logistic regression model, in which treatment status is regressed based on observed baseline characteristics.
  - Variables used in the logistic regression model will include factors known to be associated with clinical outcome for patients with the disease under study (e.g., age, ECOG PS, etc.).
  - Conditioning on the propensity score, the distribution of these baseline characteristics is expected to be similar between the two arms.

Methods for Propensity Score Analysis

- Matching on the propensity score - forming matched sets of patients between the two arms who share a similar value of the propensity score.
- Stratification on the propensity score – Patients are ranked according to their propensity score. A common approach is to divide patients into five equal-size groups using the quintiles of the estimated propensity score.
- Covariate adjustment using the propensity score - The outcome variable is regressed on an indicator variable denoting treatment arm (i.e., Investigational arm=1 vs RWD control arm=0) and the propensity score.
- Inverse probability of treatment weighting (IPTW) - uses weights based on the propensity score to create a synthetic sample. Let Z_i be an indicator variable denoting treatment arm and e_i denote the propensity score for the ith patient. Weights can be defined as $w_i = \frac{Z_i}{e_i} + \frac{(1-Z_i)}{1-e_i}$, which is $w_i=\frac{1}{e_i}$ when $Z_i=1$ and $w_i=\frac{1}{1-e_i}$ when $Z_i=0$.

Pros and Cons for the Four Methods

- Matching – Can balance the known covariates and reduce selection bias. But it can also result in significant loss of observations of patients, particularly if the RWD per inclusion/exclusion criteria is small.
- Stratification - The overall treatment effect may not be interpretable when the treatment effects of strata are very different in scale especially in direction. In addition, patients in different strata may not separate into distinguishable groups that are meaningful to clinicians.
- Covariate adjustment assumes the nature of the relationship between the propensity score and the outcome has been correctly modeled, i.e., can perform poorly if the sample linear discriminant based on covariates is not a monotone function of propensity score.
- IPTW – Use IPTW weighted estimators to obtain treatment effects adjusting for known confounders and produce one overall estimate of treatment effect. Allows to include all data available according to the inclusion/exclusion criteria

IPTW and sIPTW

- In the pseudo data using IPTW assuming a total of N patients from both arms, the number of observations is the sum of weights, which is always greater than the original sample size of the data N. $$N_w = \sum_{i=1}^{N} w_i$$
- An improvement to the IPTW is the use of stabilized IPTW which is shown that it reduces the type I error by preserving the original sample size of the data. The sIPTW for the ith patient (sw) is $sw_i = \frac{p}{e_i}$ when $Z_i = 1$ and $sw_i = \frac{1-p}{1-e_i}$ when $Z_i = 0$, where p is the probability of being in the investigational arm without covariates.

Therefore, the propensity score analysis using sIPTW will be used to adjust for a patient's propensity score in the analyses of efficacy endpoints (i.e., response, OS, EFS, etc.).

Analyses of Time-to-Event and Response Endpoints

- Time-to-event endpoints such as OS and EFS (i.e., hazard ratio and its 95% CI, etc.) will be estimated from the weighted Cox proportional hazards model using the sIPTW. The p-value will be estimated based on the weighted log-rank test using sIPTW.

- Response type of endpoints will be analyzed based on weighted Chi-Square test using sIPTW.

Closing Remarks

- Make sure to explore as many vendors as possible to maximize the amount of patients available for the RWD control arm.
- Include all the key inclusion/exclusion criteria if possible when selecting patients into the RWD control arm so the comparisons of endpoints between arms are appropriate.
- Different definitions and assessment schedules for disease assessments could impact the evaluation of efficacy endpoints such as PFS, EFS, DFS, response, etc. among the patients between the investigational arm and the RWD control arm.
- Other methods could be considered, i.e., marginal structural models, instrumental variable analysis, etc.
- **Consult with FDA in the planning stage to get their buy-in on the study design and the possible registrational path.**

Acknowledgement

Thanks to the following colleagues to either provide helpful input or allow us to refer to their work for the presentation.

- Enayet Talukder
- Ray Lu
- Jack Mardekian
- Xin Huang
- Hui Zhang

References

- Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research, 46:399–424, 2011
- Austin PC. A Tutorial and Case Study in Propensity Score Analysis: An Application to Estimating the Effect of In-Hospital Smoking Cessation Counseling on Mortality. Multivariate Behavioral Research, 46:119–151, 2011.
- Khozin S, Blumenthal GM, Pazdur R. Real-world Data for Clinical Evidence Generation in Oncology. JNCI J Natl Cancer Inst (2017) 109(11): djx187
- Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, and Smith D. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value Health*. 2010 ; 13(2): 273–277.
- Framework for FDA's Real-World Evidence Program. Dec. 2018.
- Blincyto FDA Slides for ODAC. Mar. 2018.
- Blincyto FDA Briefing Document for ODAC. Mar. 2018.
- Blincyto Amgen Briefing Document for ODAC. Mar. 2018.
- Selinexor FDA Slides for ODAC. Feb. 2019.
- Selinexor FDA Briefing Document for ODAC. Feb. 2019.


thank you
danke
謝謝
ngiyabonga
рахмат
teşekkür ederim
gracias
merci
obrigado
bedankt
dziękuje
спасибо
Баярлалаа
go raibh maith agat
sukriya
kop khun krap
terima kasih
mochchakkeram
tapadh leat
dank je
grazie
arigatô
dakujem
takk
мерси
감사합니다
ありがとう
ευχαριστώ
hvala
mauruuru
sağolun
köszönöm
faafetai lava
kiitos
dankie
dhanyavad
nanni
nandri
enkosi
bayarlalaa
gràcie
sobodi
děkuji
mèsi
didi madloba
najis tuke
kam sah hamnida
rahmet
tanemirt
rahmet
diolch
dhanyavadagalu
shukriya
merce
trugarez
mamnun
chokrane
murakoze
tenki
obrigada
asante
manana
хвала
mahalo
paldies
grazzi
matondo
misaotra
welalin
tack
spas
barka
kia ora
mersi
vinaka
спасибі
blagodaram
akun dankon
aciú
djiere dieuf
tau
дякую
sulpáy
taiku
gràcies
gratias ago
chnorakaloutioun
xièxie

December 4, 2019

Learning from the real world: Electronic health records and real world evidence

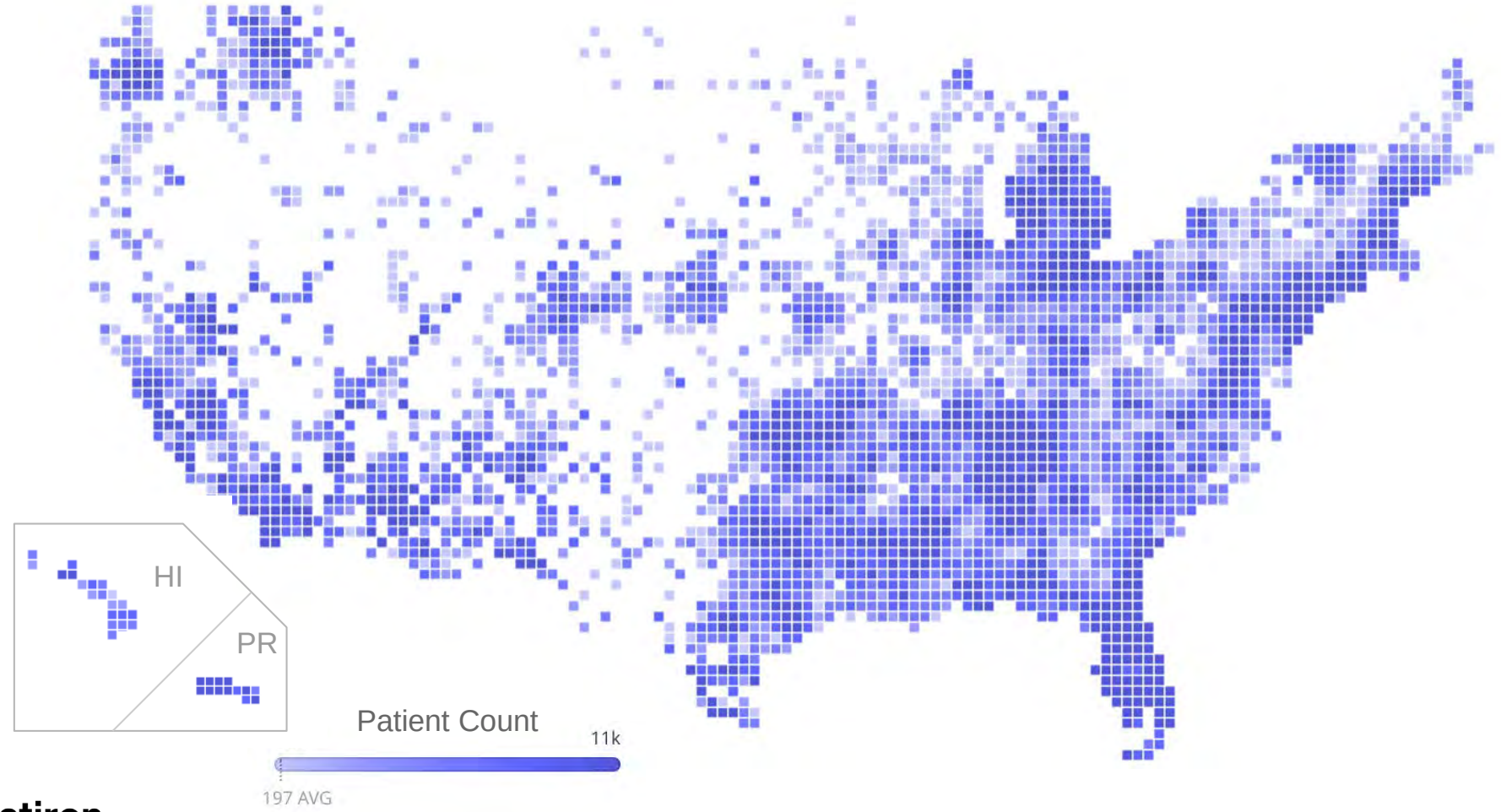
Rebecca Miksad, MD, MPH
Senior Medical Director
Flatiron Health

Rebecca Miksad, MD, MPH

Senior Medical Director, Flatiron

flatiron

The Flatiron Network



2.2M+
Patients

2,600+
Clinicians

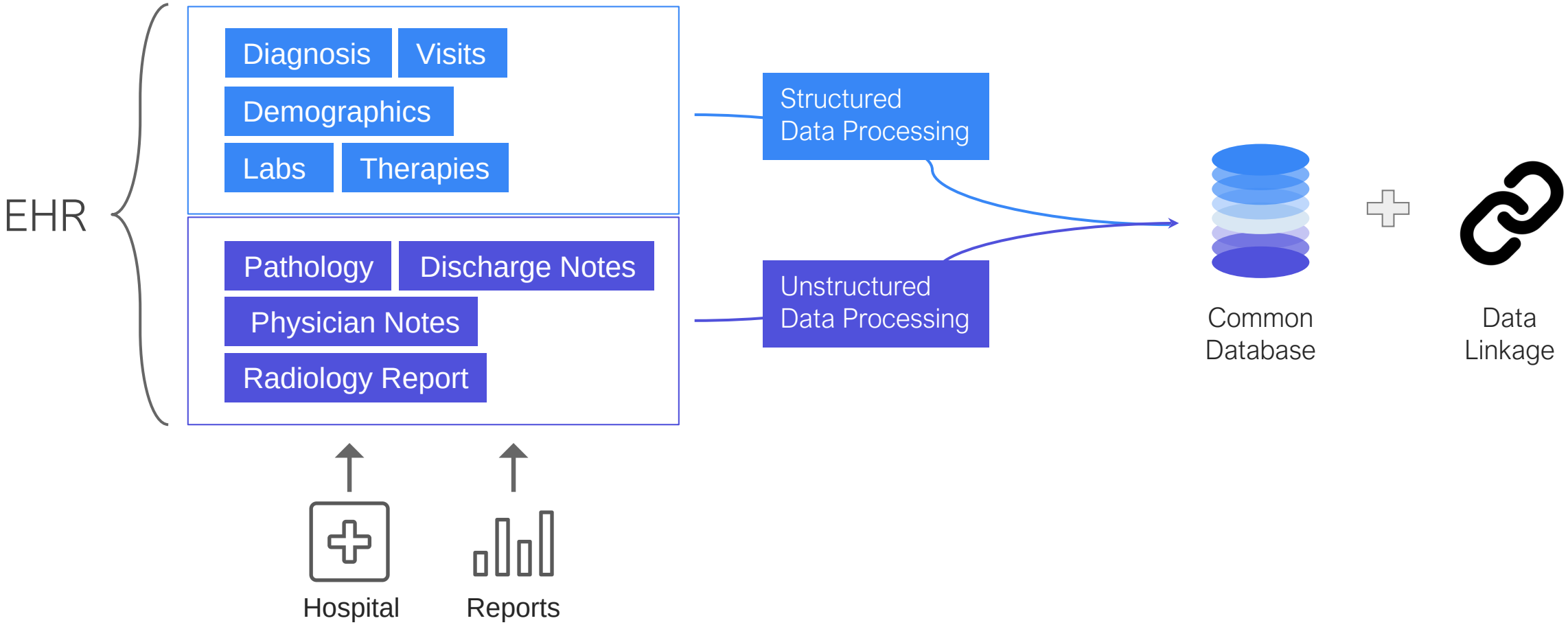
280
Cancer Clinics

7
Academic Medical Centers

800+
Unique Sites of Care

flatiron

Pathway to meaningful data: source and curation



Transforming structured data to a common data model

2220	Blood Serum Albumin	g/dL
QD25001600	ALBUMIN/GLOBULIN RATIO QD	(calc)
QD25001400	ALBUMIN QD	g/dL
QD50058600	ALBUMIN	%
QD50055700	ALBUMIN	g/dL
CL3215104	Albumin % (EPR)	%
LC001081	ALBUMIN, SERUM (001081)	g/dL
LC003718	Albumin, U	%
LC001488	Albumin	g/dL
LC133751	Albumin, U	%
CL3215162	Albumin%, Urine	%
CL3215160	Albumin, Urine	mg/24hr
3234	ALBUMIN SS	g/dL
LC133686	Albumin, U	%
QD50060710	MICROALBUMIN	mg/dL
QD50061100	MICROALBUMIN/CREATININE RATIO, RANDOM URINE	mcg/mg creat
QD85991610	ALBUMIN	relative %
50058600	ALBUMIN UPEP RAND	%
CL3210074	ALBUMIN LEVEL	g/dL
QD86008211	ALBUMIN/GLOBULIN RATIO	(calc)
LC149520	Albumin	g/dL
QD45069600	PREALBUMIN	mg/dL
QD900415245	ALBUMIN, SERUM	mg/dl
QD900429745	ALBUMIN	g/dL
CL3215124	Albumin Electrophoresis	g/dL
LC016931	Prealbumin	mg/dL
QD50060800	MICROALBUMIN, 24 HOUR UR	mg/24 h
QD50060900	MICROALBUMIN, 24 HOUR UR	mcg/min
QD85994821	ALBUMIN,SERUM	g/dL
CL3213320	PREALBUMIN	mg/dL
QD85995225	PROTEIN ELECTROPHORESIS ALBUMIN	g/dL

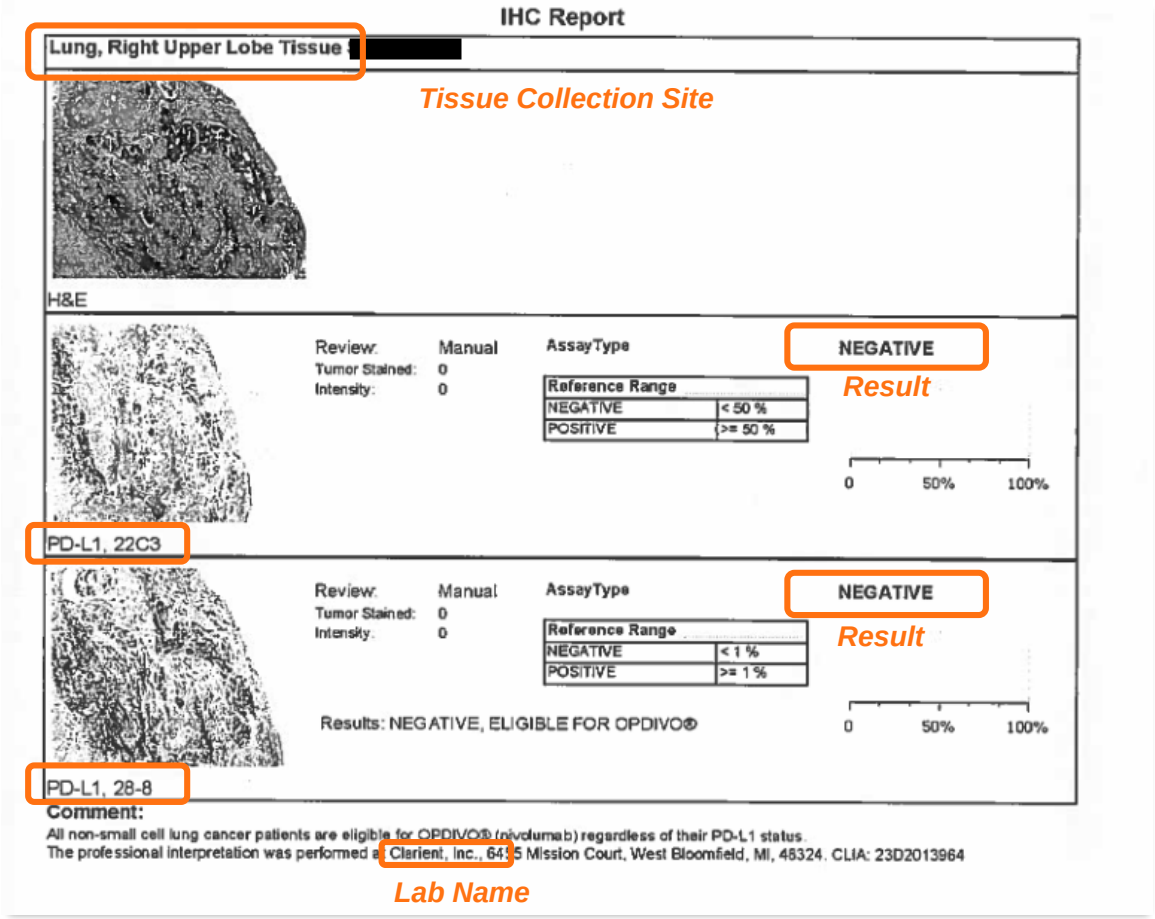
- Many structured data elements are coded and collected in multiple ways
- Flatiron combines structured data across sites, and maps all data elements to a single set of definitions (“data model”)

1751-7	**Albumin [Mass/volume] in Serum or Plasma**	**g/dL**

flatiron

Liberating critical oncology information from unstructured data

Section of PD-L1 Report



For every PD-1/PD-L1 test a patient receives, Flatiron biomarker data model captures:

- Test result
- Date biopsy collected
- Date biopsy received by laboratory
- Date result received by provider
- Lab name
- Sample type
- Tissue collection site
- Type of test (e.g., IHC)
- Assay / kit (e.g., Ventana 142)
- Percent staining & staining intensity

Technology Enabled Human “Abstraction” (unstructured data cleaning)

Expert Abstractors

1000+ expert data abstractors (oncology nurses and cancer registry professionals) follow precise policies to review unstructured documents and enter data in a structured format

Flatiron Technology

Flatiron Health software organizes EHR documents, manages data entry, controls access and monitors quality for efficient and reliable unstructured data processing

Real-world clinico-genomic data: Pathway to precision medicine

flatiron

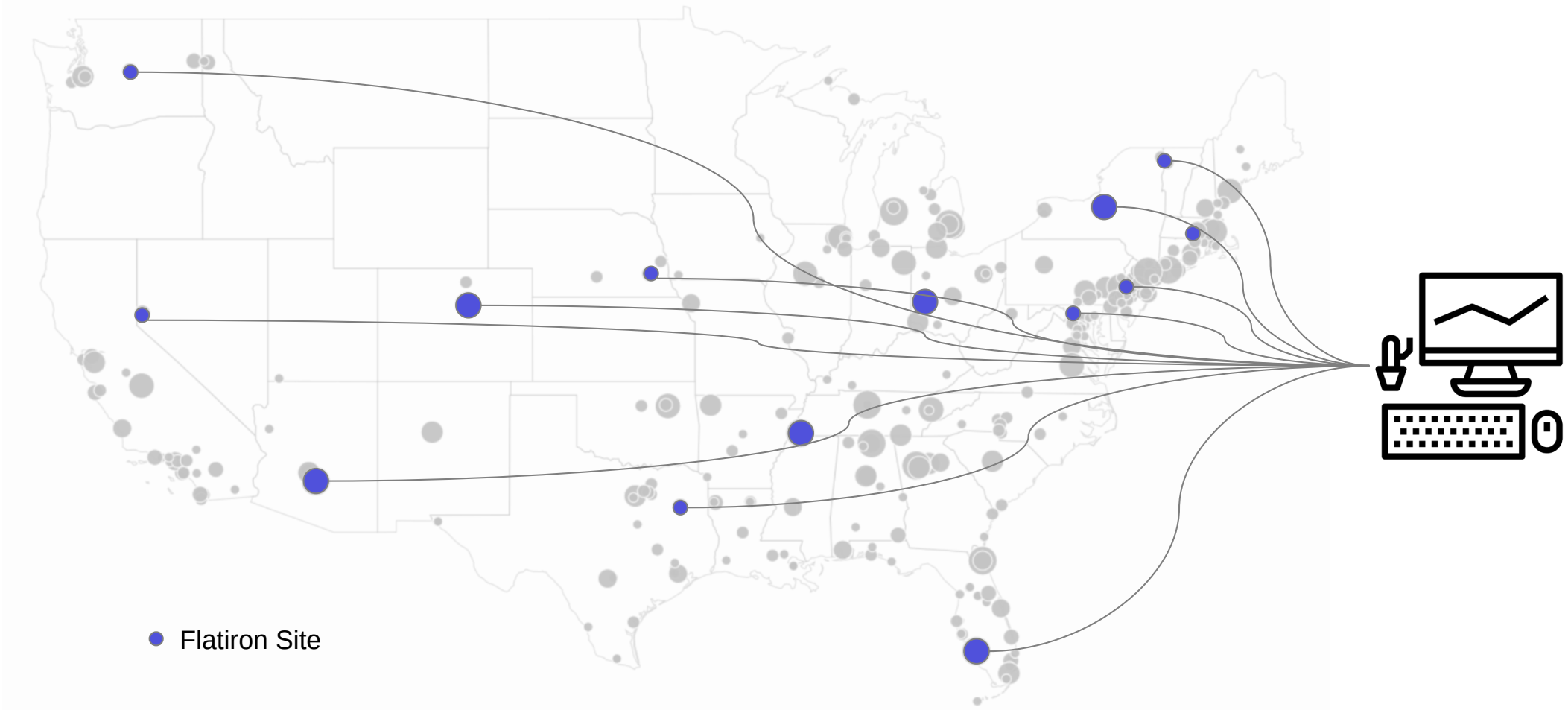
+

FOUNDATION MEDICINE®

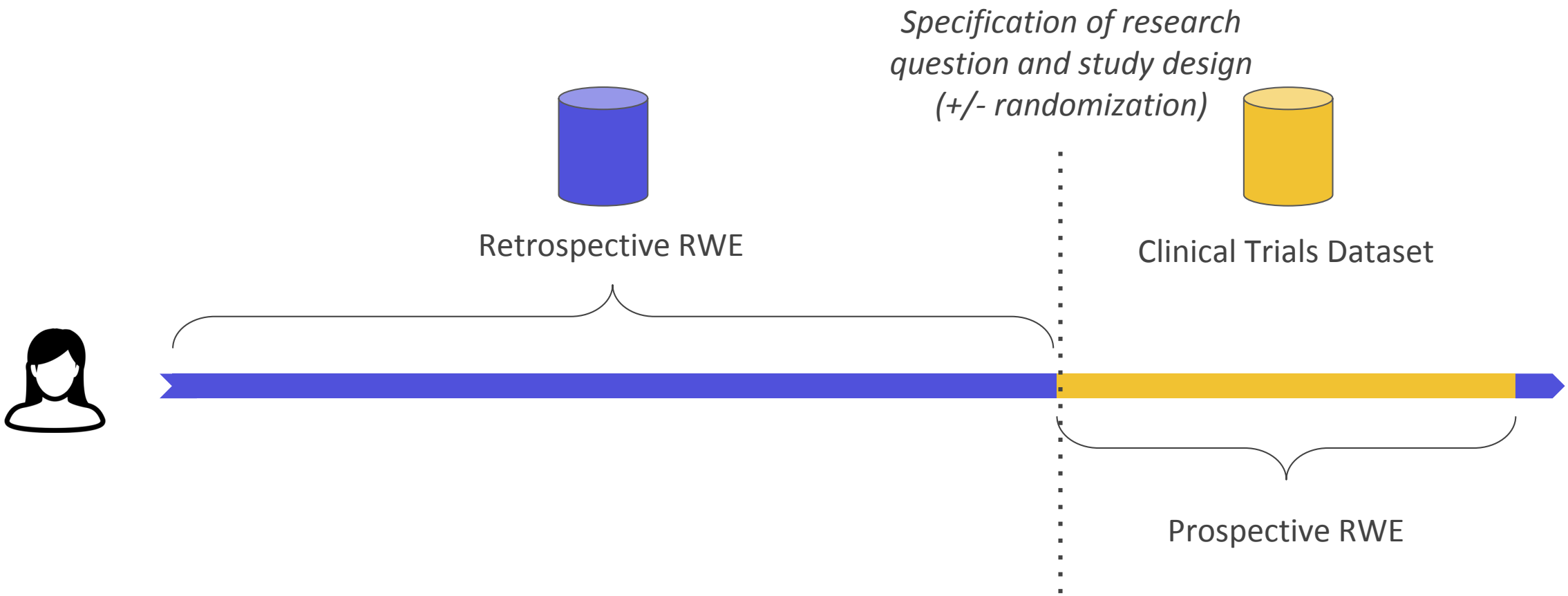
Real-World Clinico-Genomic Database

Patient population: >48,000 patients

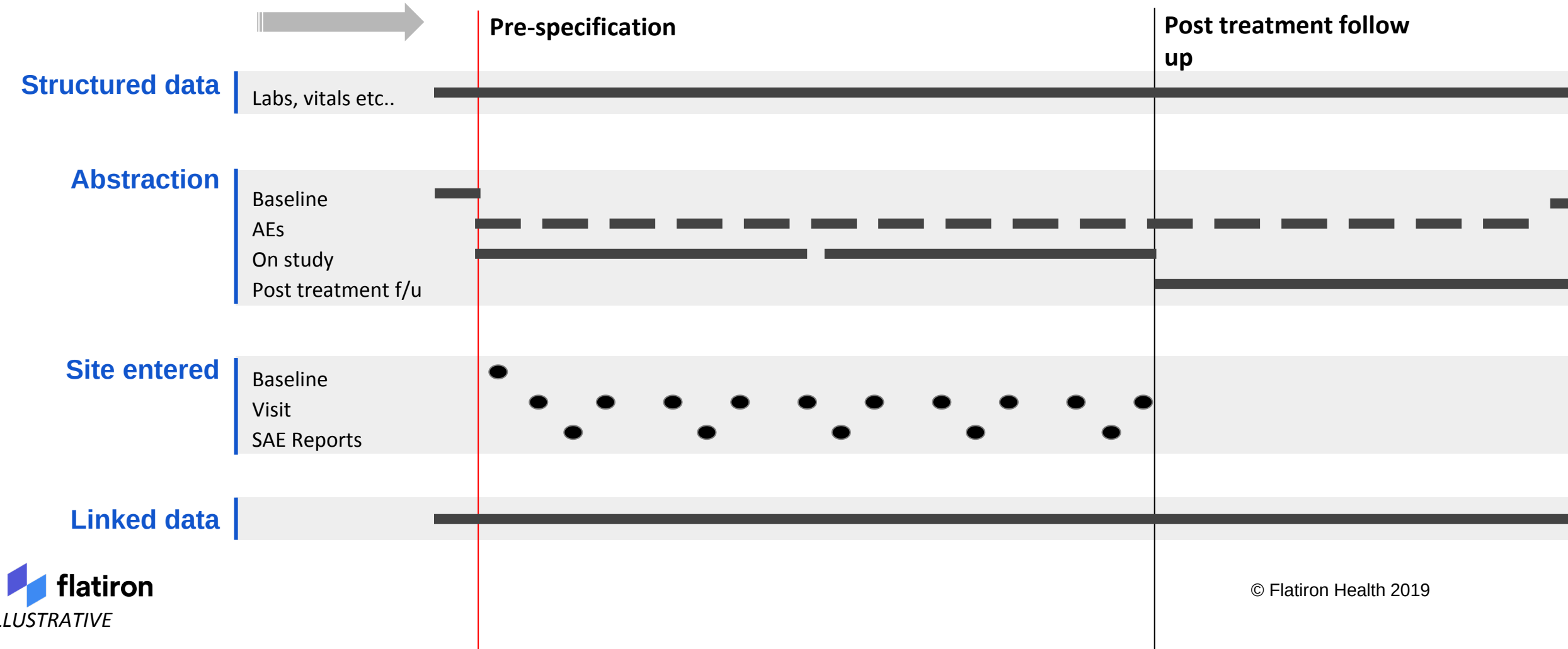
EHR data facilitates practical real world trials



Real world evidence is on a continuum with traditional clinical trials



Generate the real world trial dataset from retrospective and prospective sources

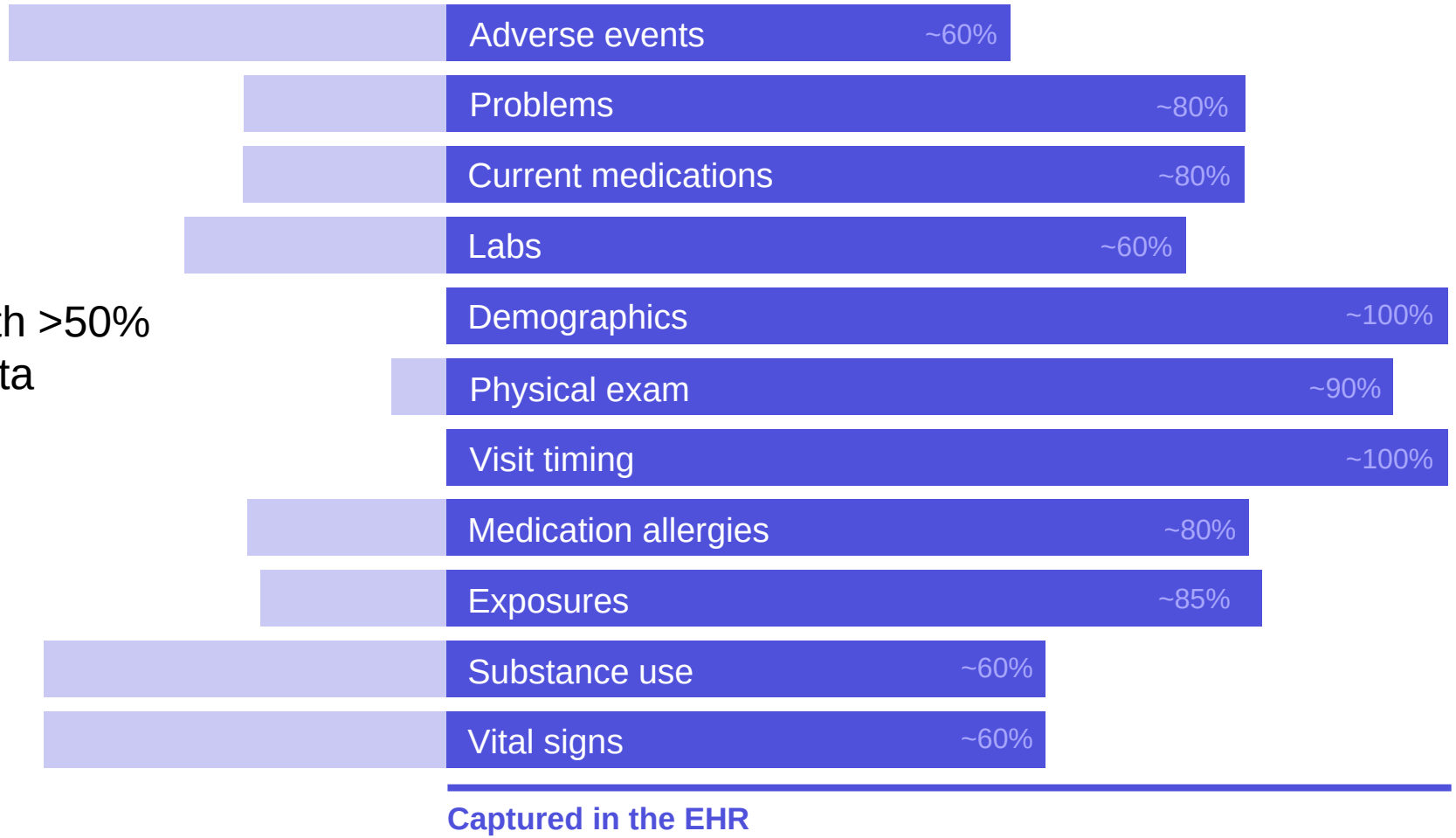


flatiron

ILLUSTRATIVE

This is possible because much of the clinical data collected in trials is captured in the EHR

CDISC E2C domains with >50% overlap in EHR data



Source: CDISC “EMR to CDASH” working group field mappings compared to OncoEMR; preliminary analysis does not consider important factors such as completeness and standardization

EHR based systems reduce the need for on-site activities

Traditional monitoring

- Site training
- Protocol compliance
- Deviations
- Consent
- Eligibility
- SDV
- ALCOA
- Missing data
- Safety reconciliation
- Facilities
- Discrepancies
- Other
- ISF / TMF review

Legend

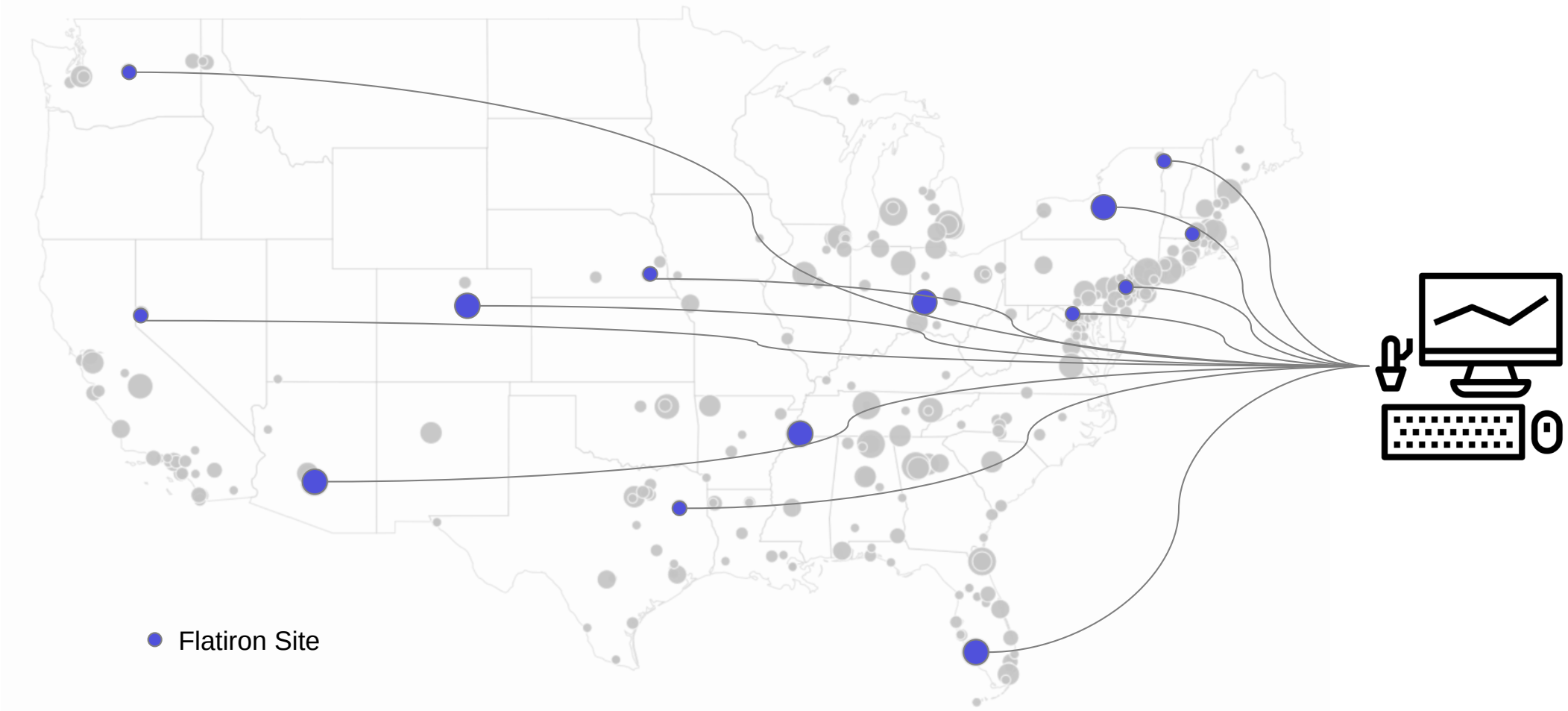
- Performed on-site
- Performed centrally

EHR-enabled monitoring

- Site training
- Protocol compliance
- Deviations
- Consent
- Eligibility
- SDV
- ALCOA
- Missing data
- Safety reconciliation
- Facilities
- Discrepancies
- Other
- ISF / TMF review

SDV: Source Document Verification
ALCOA: Attributable, Legible, Contemporaneous, Original, Accurate

Study follow-up AND traceability of source data enabled through EHR



Fit for use data quality is a core principle underlying fit-for-use RWE

While the technology of RWD has evolved, quality and reliability remain paramount

World War II bomber

21st century

Thank you

Questions? Comments?
rmiksad@flatiron.com