Thoughts on Public Health Data

Dale Sanders

Feb 27, 2023

Today's Storyline

- My background and how that influences my approach to healthcare data strategies
- The psychology of data strategies
- Healthcare data engineering
- What can vendors do?
- Data related observations from COVID

My Professional Genetics

And how it relates to healthcare data

I’ll let you in on a secret…

I’m a patient and clinician activist, disguised as a vendor and investor

The accidental career that prepared me for today’s data strategies…

- Healthcare terminology vendor– 90% of US EHRs 2021
- Healthcare IT investor– biotech, clinical trials, virtual care 2019
- Healthcare software & analytics vendor-- ~180M patients 2011
- Healthcare delivery– EHRs, data warehousing, analytics 1997
- Quantitative nuclear risk modeling, communication, and mitigation
- Nuclear warfare command centers
- Satellite and space operations 1984

At 24 years old, this was my introduction to data and decision making

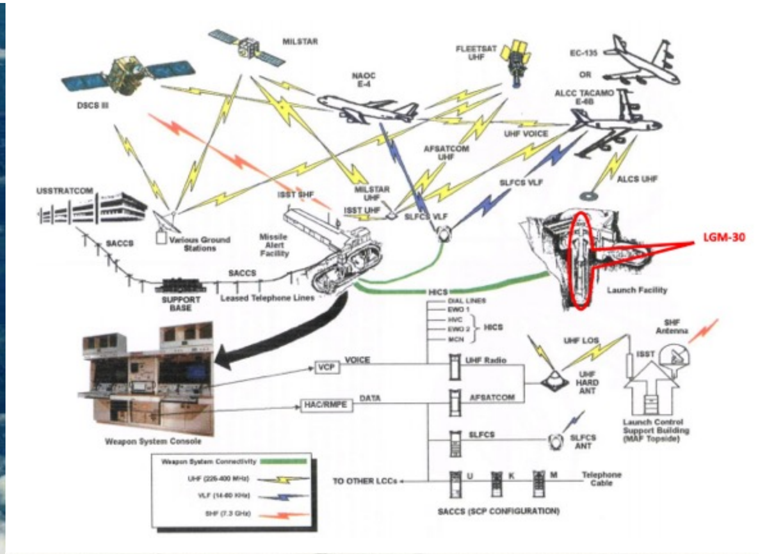
Underground
Command Center,
for US Nuclear Forces, Strategic Air Command (SAC)

“Peace is Our Profession”

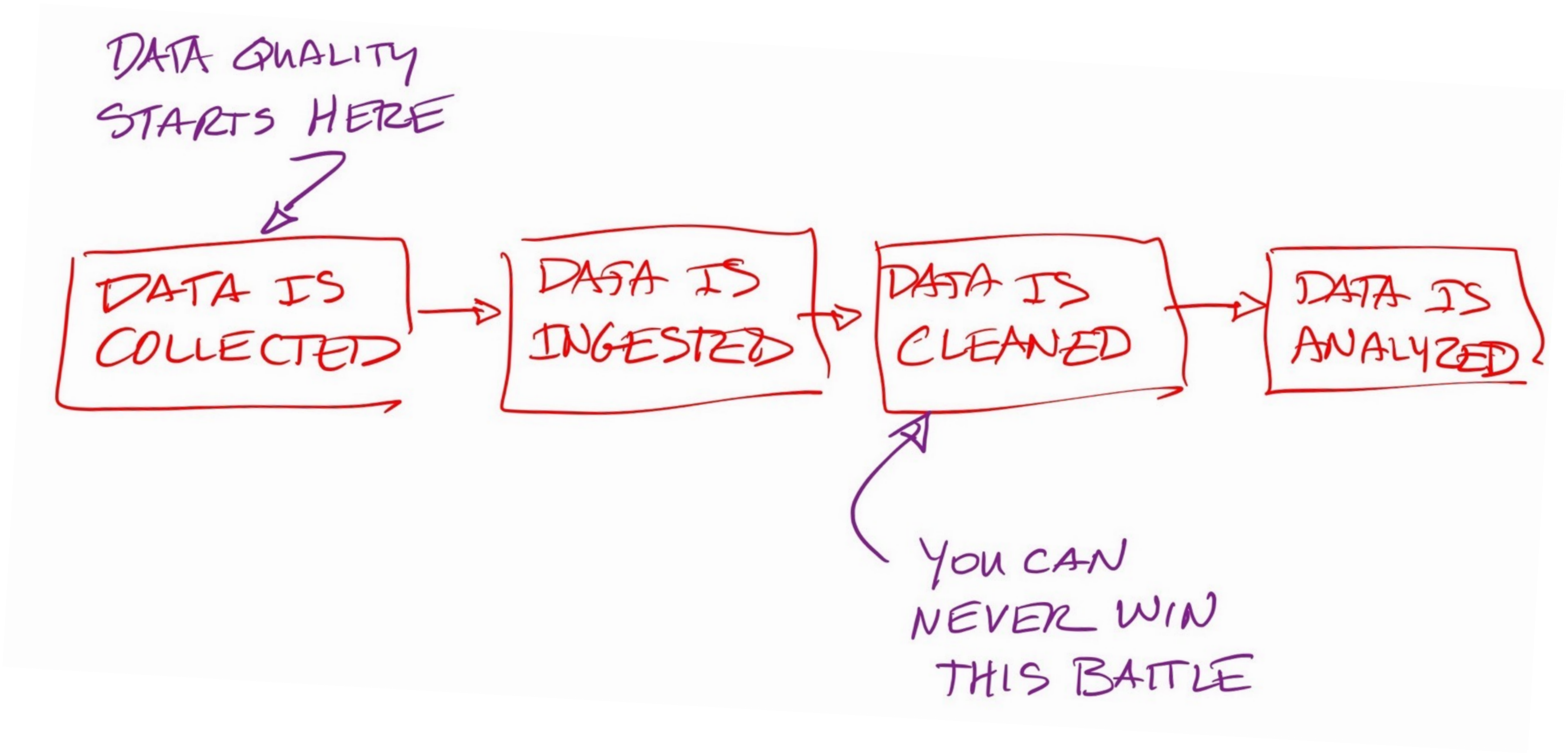
“To Err is Human, to Forgive is Not SAC Policy.”

Emergency Airborne Command Post System

“Doomsday Planes”



I’ve been on the right side of this diagram for 27 years in healthcare. We must move upstream.



The Psychology of Data Strategies

More difficult now, than the technology

My Three “– ilities”

Successful data strategies require good leadership, and good leaders are characterized by…

- **Believability**: Is this person honest, sincere, trustworthy, and transparent? If I don’t know them well enough, is there evidence of such in their background?
- **Relatability**: Can this person empathize directly with my situation and/or role? Can I relate to this person along other dimensions of empathy, e.g., upbringing, religion, age, gender, race, ethnicity, experiences, hobbies, education, etc.?
- **Credibility**: Even though we’re similar, does this person bring expertise or knowledge that I don’t have; and that I value, respect, and need?

Mindset, Skillset, Toolset

Develop all three, in this order…

Putting data technology ahead of mindset and skillset is a recipe for trouble

Mastery, Autonomy, Purpose

Successful data strategies feed
these fundamental human needs.

For example, if CDC is not giving STLTs the data the STLTs need to satisfy all three fundamental needs, it's game over. Same applies to OPHDST and the Programs and Centers.

Thank you for the inspiration, Daniel Pink

The four psychological barriers to data sharing

If you tread on any of these, it's game over

- **Punishment**: You're going to use my data to measure and punish me
- **Embarrassment**: You're going to embarrass me in front of my peers or stakeholders, with my data
- **Demotion**: You will lower my professional value if I share my data
- **Liability**: You will misuse my data, and it's going to come back on me

Three lines of effort in a holistic data strategy…

- **Optimize** the value of the data you have
  - Quality, accessibility, usability, literacy
- **Acquire** the data you need, but don't have
  - The data exists somewhere, but we don't have it
  - The data doesn't exist, so we need to build the technology or human capability to collect it
- **Disseminate**: Create a communication capability to address each of the stakeholders who you want to reach with the data
  - Including uncertainty in the data, and a misinformation mitigation plan

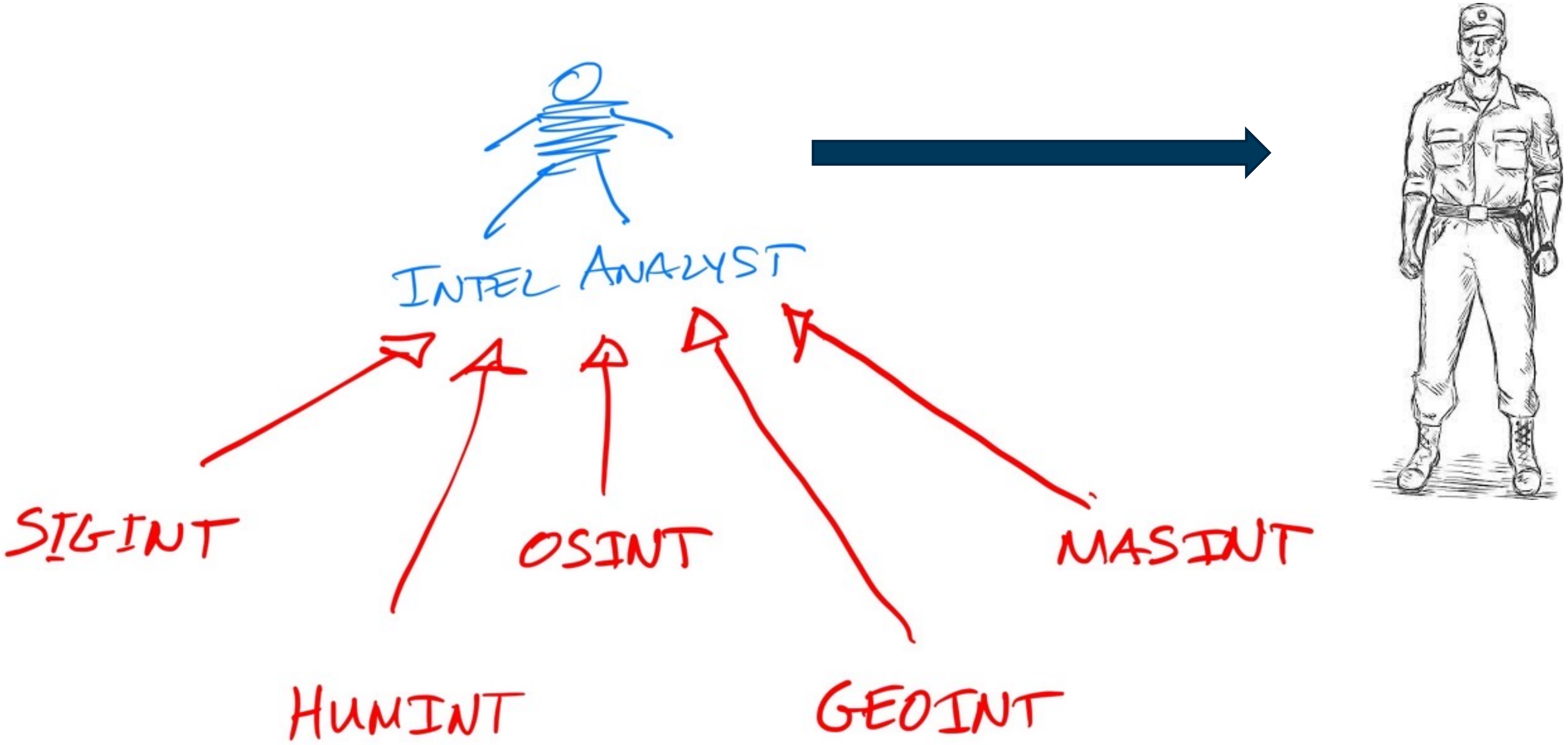
Public Health is, essentially, a warfighting mission and the enemies are contagions

Our passion and **insistence** on high-quality, real-time data must match that of battlefield commanders and frontline troops

How is US national intelligence organized around data?

Five fundamental sources… are there analogies in Public Health?

- **OSINT**: Open Source Intelligence
- **HUMINT**: Human Intelligence
  - Contact tracing, EHR data, case data
- **SIGINT**: Signals Intelligence
  - Lab results, Rx orders and retail OTC, ED visits, bio-integrated sensors
- **GEOINT**: Geospatial Intelligence
  - Heat maps, contact maps
- **MASINT**: Measurement and Signature Intelligence
  - Wastewater, air monitoring



Warfighting Commanders:
"What does this mean relative to our Force Status?"

Some of the best Intelligence Analysts have a background in "soft sciences"… sociology, psychology, ethnography, anthropology, economics, library science, et al

Public health is, essentially, in the risk management business.

Therefore, formal risk management methods should be the basis for the public health data strategy

Communicating Risk: Meeting people where they are…

Risk = Probability x Consequence

"What's the probability of me being infected AND what's the consequence?"

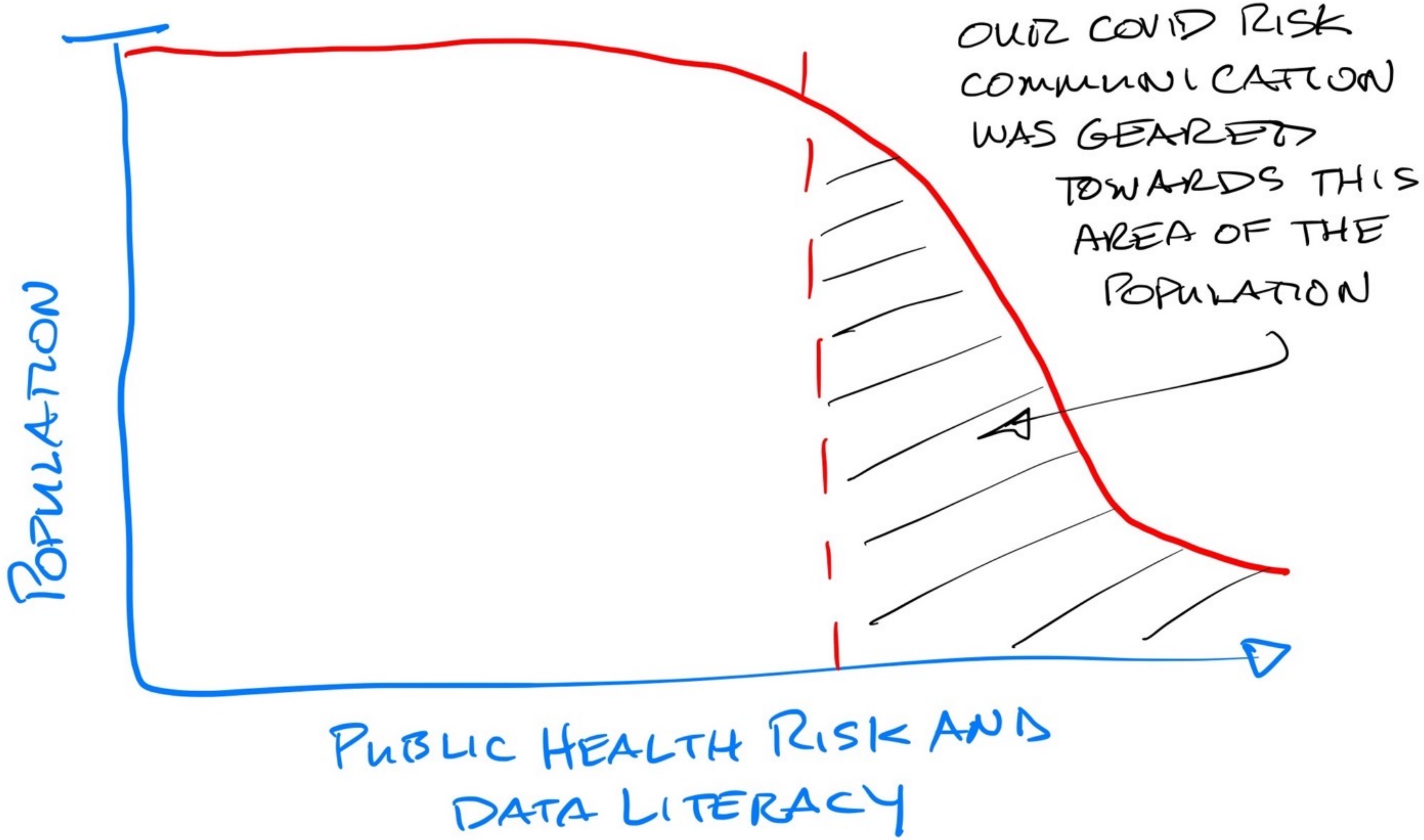
Healthcare tends to assume worst case for either or both, but…

“Over-mitigating” wastes money, wrecks credibility, stifles agility, and creates its own cascading risks that are usually worse than the initial scenario.

The goal is to walk right up to the edge of risk without going over.

But you need to know where the edge resides. That’s what a holistic data strategy enables.

Our COVID risk communication strategy met a subset of the population



So we end up with this, from misplaced fear and misunderstanding of risk

Fewer dashboards, more narratives

As data people, we tend to put more faith in dashboards than dashboards deserve

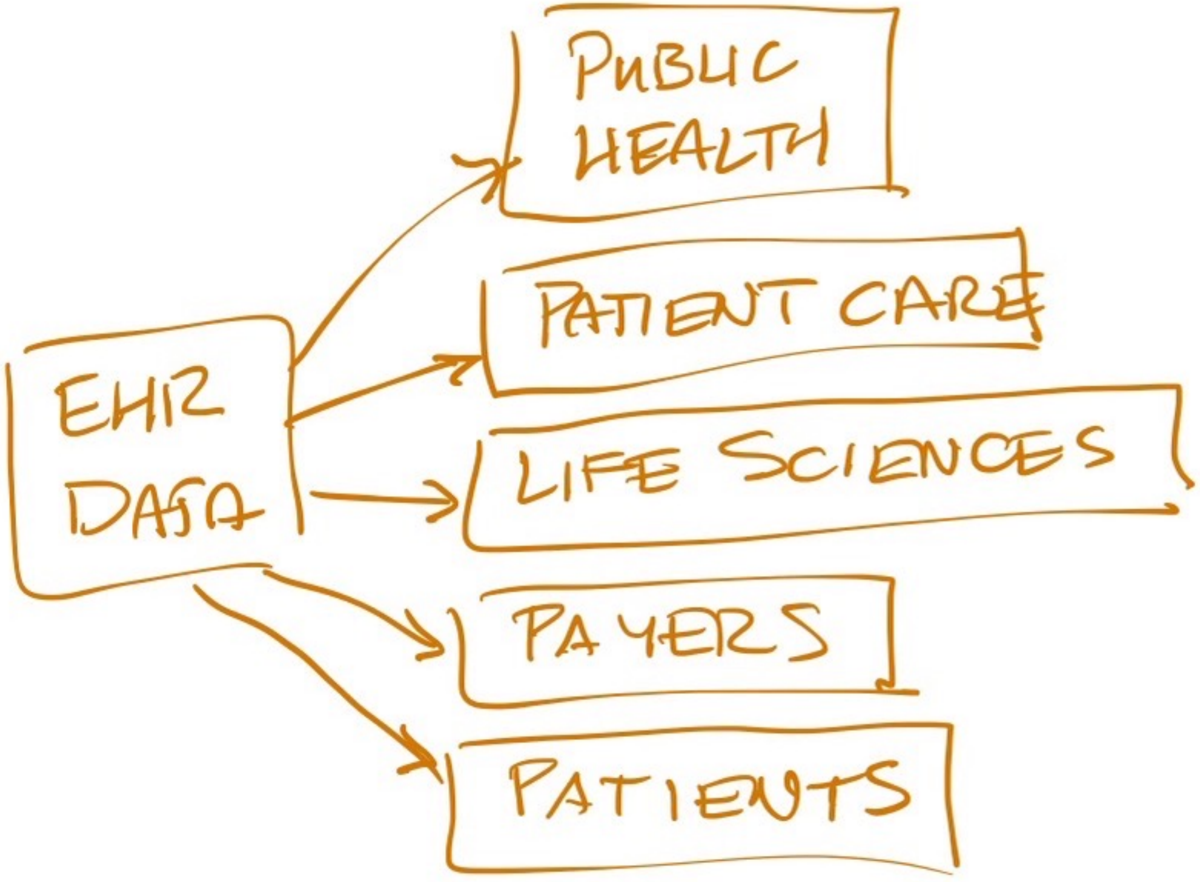
A BIG part of the population needs and wants a narrative translation between the data in the dashboard, and what it means to their daily lives

Healthcare Data Engineering

Subtleties of success that are often missed

Downstream dependence on EHR data quality

EHR data quality cascades throughout the healthcare industry



Healthcare data quality is poor… but not useless

Opportunities in Machine Learning for Healthcare

Marzyeh Ghassemi
University of Toronto, Vector Institute
Toronto, Canada
marzyeh@cs.toronto.edu

Tristan Naumann
Microsoft Research
Redmond, WA 98052
tristan@microsoft.com

Peter Schulam
Johns Hopkins University
Baltimore, MD 21218
pschulam@cs.jhu.edu

Andrew L. Beam
Harvard School of Public Health
Boston, MA 02115
andrew_beam@hms.harvard.edu

Irene Y. Chen
Massachusetts Institute of Technology
Cambridge, MA 02139
iychen@mit.edu

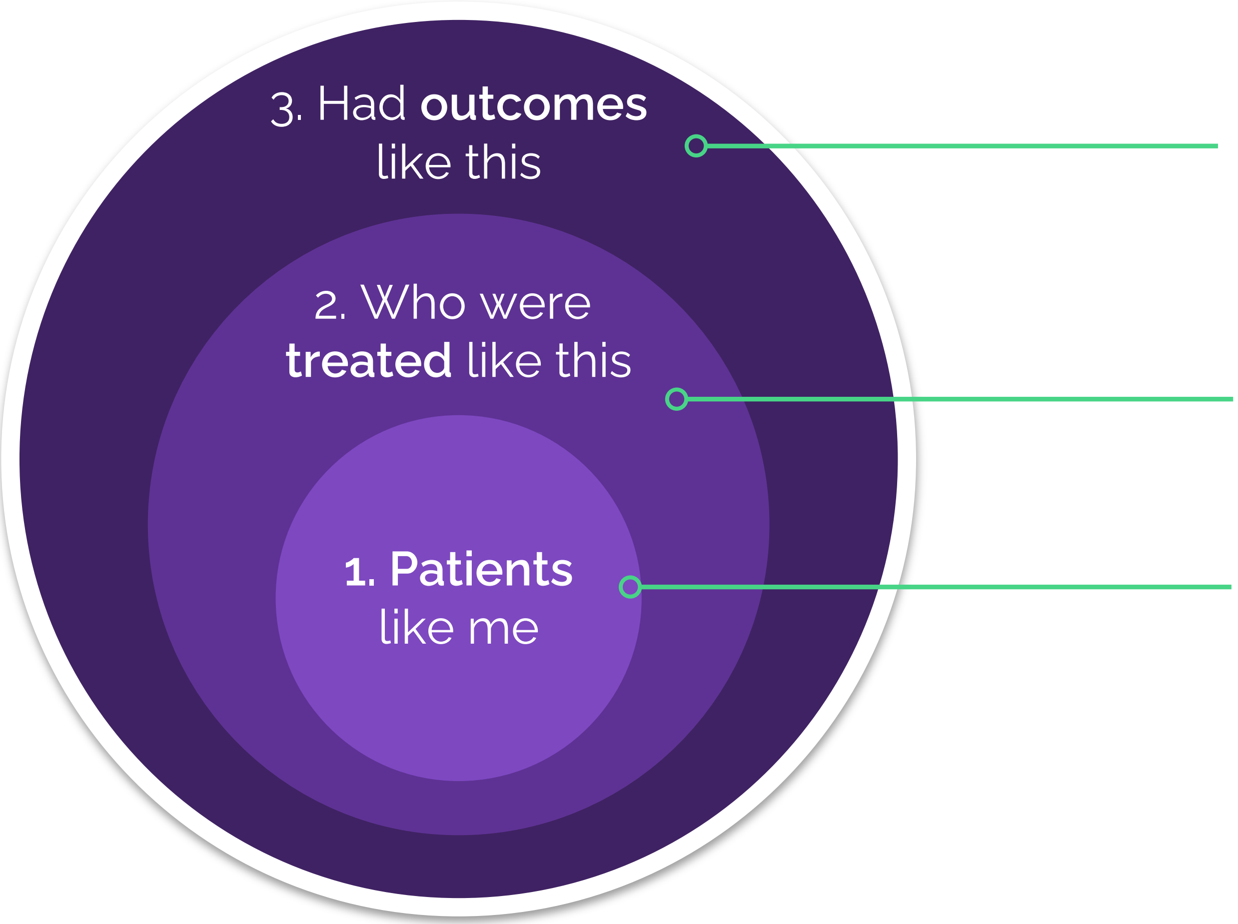
Rajesh Ranganath
New York University
New York, NY 10011
rajeshr@cims.nyu.edu

*"…**diseases in EHRs are poorly labeled, conditions can encompass multiple underlying endotypes**, and **healthy individuals are underrepresented**. This article serves as a primer to illuminate these challenges and highlights opportunities for members of the machine learning community to contribute to healthcare."*

July 2019, U Toronto, Microsoft, Johns Hopkins, Harvard, MIT, NUY

This is the patient data we need…

It can be flexed to support endless secondary use cases



REQUIRED TERMINOLOGY AND DATA

Outcomes
- Patient Reported Outcomes
- Quality of life
- Length of life
- Cause of death
- Cost of care

Orders & Interventions
- Medications
- Procedures
- Patient initiated

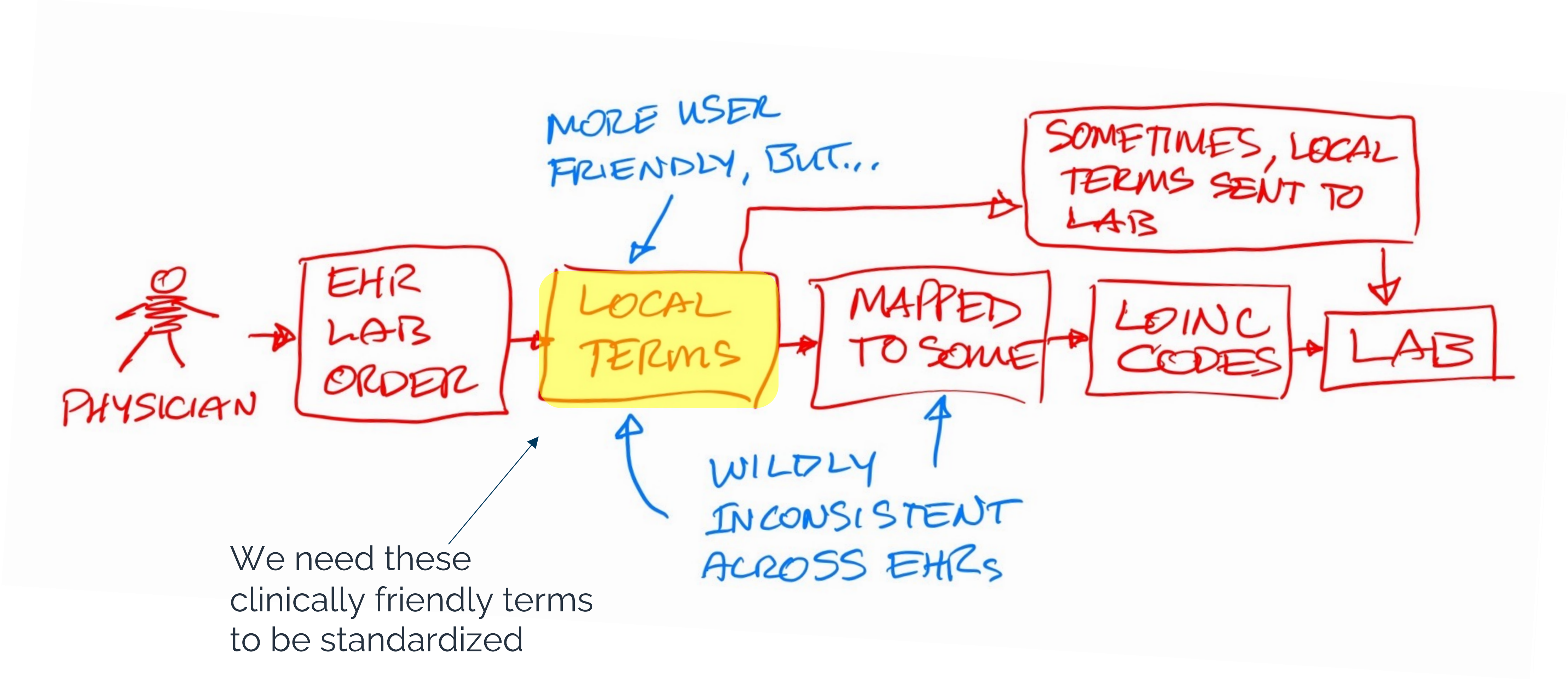
Patient cohorts
- Diagnoses
- Demographics
- SDOH
- Genomics
- Vitals
- Allergies
- Lab results

- Temporality
- Clinician and Patient Recorded Notes

Terminology challenges

- **ICD**: Widely used, but confusing to use for clinicians and understand; weird mix of coarse and fine-grained
- **LOINC**: Designed for lab techs, not ordering clinicians; our biggest terminology problem right now
- **SDOH**: Inconsistent interpretation, non-standard terms
- **SNOMED**: Academically flexible, practically unusable at the point of care

One of the problems with LOINC

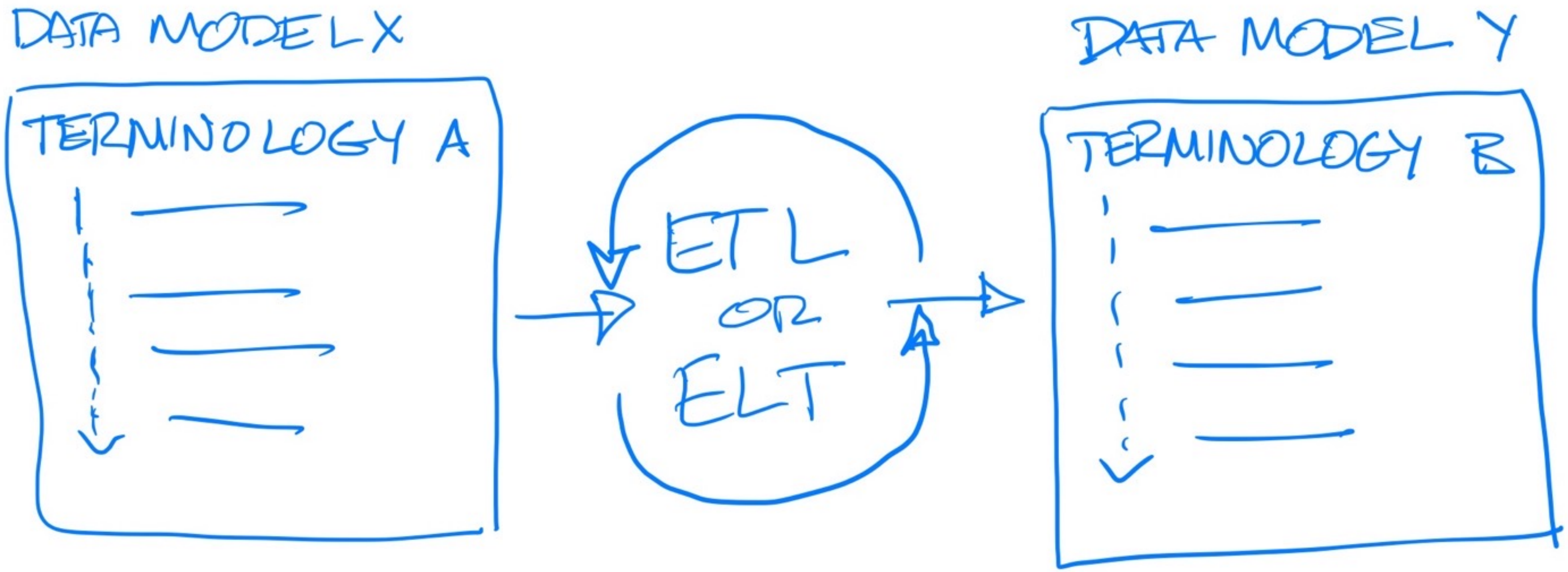


We need these clinically friendly terms to be standardized

Terminology challenges

- **Race/Ethnicity:** Lack of comprehensive adoption and persistent agreement
- **Outcomes:** Lack of comprehensive adoption and persistent agreement
- **Cause of death**: Every jurisdiction has its own process
- **Quality of life/functional status**: Lack of comprehensive adoption and persistent agreement
- **Temporality of disease**: No standards, very important

Moving data between data models and terminologies does not affordably scale… and it's very lossy



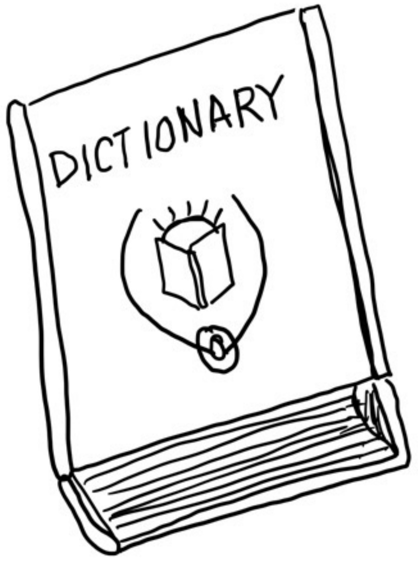
FHIR is an improvement, but...

Data models are almost useless for interoperability and analytics if you fill them with inconsistently used, random terminologies

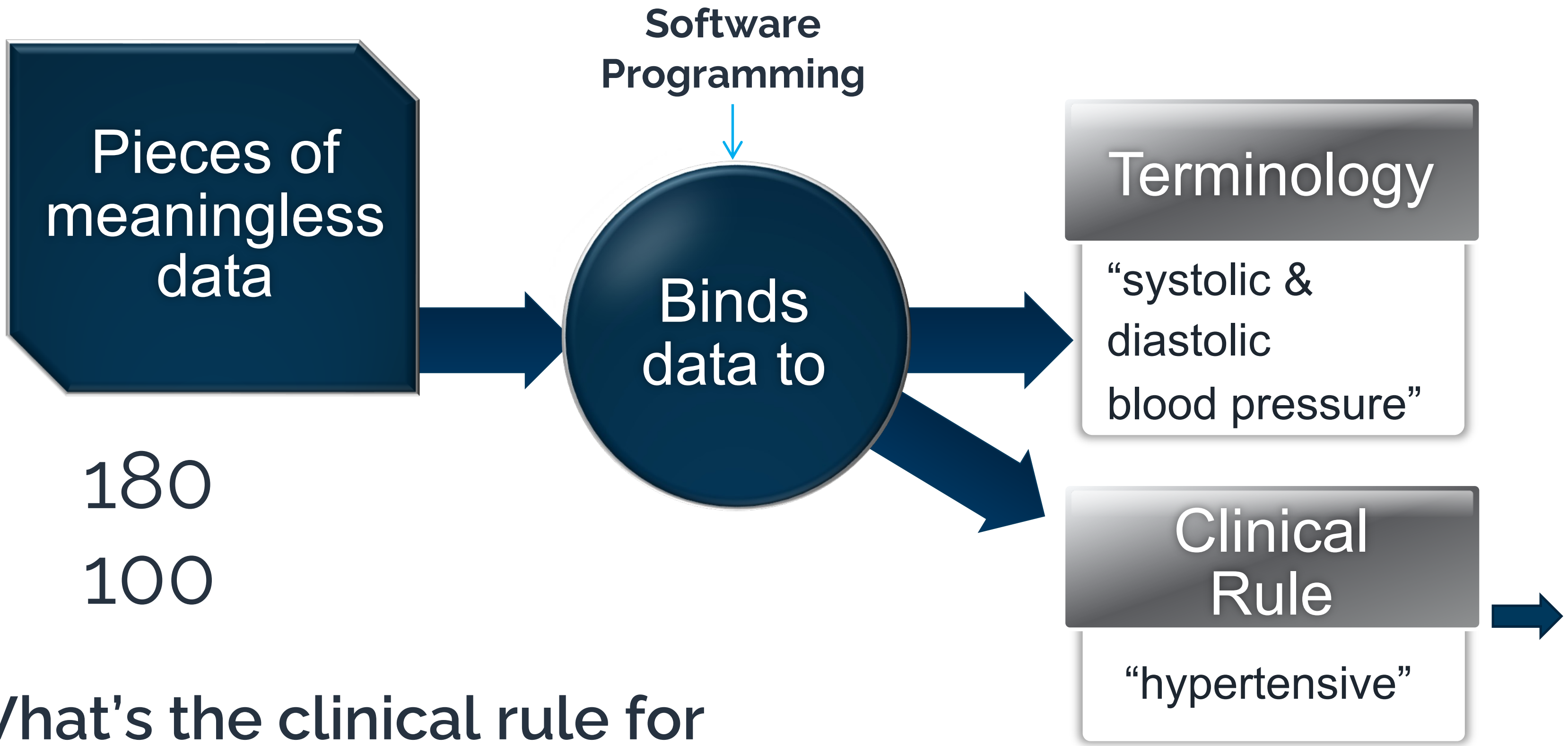
TEFCA and **USCDI** are a step in the right direction for the adoption of standard clinical terminology

Data binding in healthcare data engineering

- Atomic data must be "bound" to business rules about that data and to vocabularies related to that data in order to create information
- **The words:** Data binding at the terminology and master data layer
  - Unique patient and provider identifiers
  - Standard facility, department, and revenue center codes
  - Standard definitions for gender, race, ethnicity
  - ICD, CPT, SNOMED, LOINC, RxNorm, RADLEX, etc.
- **The sentences:** Data binding at the data logic, value sets layer
  - Length of stay
  - Patient relationship attribution to a provider
  - Revenue (or expense) allocation and projections to a department
  - Revenue (or expense) allocation and projections to a physician
  - Data definitions of general disease states and patient registries
  - Patient exclusion criteria from disease/population management
  - Patient admission/discharge/transfer rules



For example, binding data to “hypertension”



What’s the ICD code for hypertension? I10

What’s the clinical rule for declaring a “hypertensive patient”?

Why is this concept important?

Knowing when to bind data, and how tightly, to vocabularies and rules is THE KEY to analytic success and agility

Comprehensive Agreement?

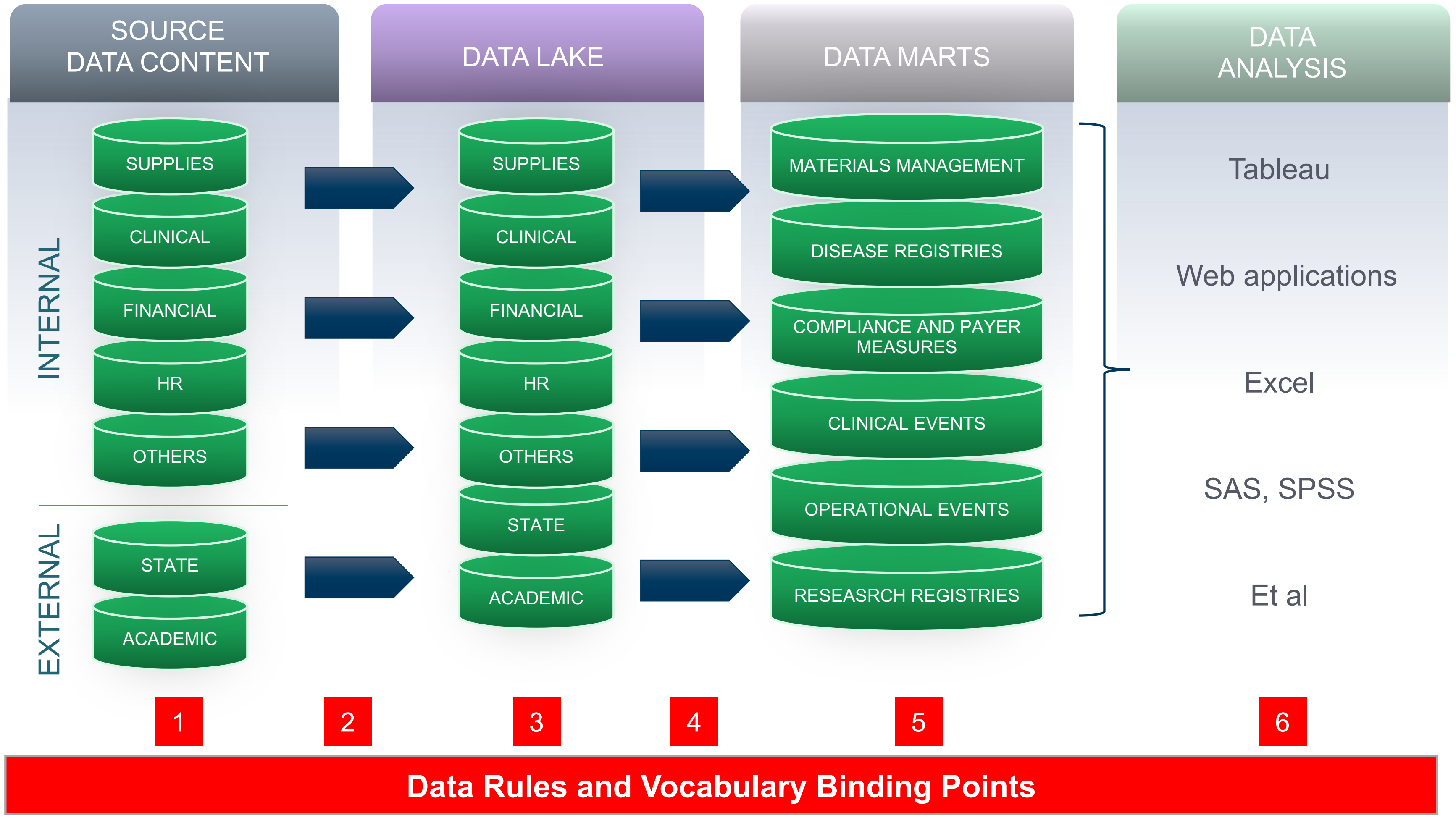
Is the rule or vocabulary widely accepted as appropriate in the organization or industry?

Persistent Agreement?

Is the rule or vocabulary stable and rarely change?

Two tests for tight, early binding

Six binding points in a typical data lifecycle



Widespread agreement and persistent use of vocabulary & clinical rules? = Early binding

Minimal agreement and inconsistent use of vocabulary and clinical rules = Late binding

What About Public Health and Vendors ?

"When times are tight, goodwill takes flight."

Vendors and investors won’t stay long, where there’s no money

CDC and the STLTs need to pool the limited money they have, reduce the number of overlapping data systems, and work together towards common platforms and applications

We need a national Terminology as a Service (TaaS) solution

Could be provided by multiple vendor solutions

Must be certified by ONC

Should also include other ”master reference data” such as National Provider ID and mortality data

The lack of multi-tenant architectures in many of today's healthcare clinical technology platforms, makes rapid, agile data response to a public health emergency nearly impossible

Cloud x Multitenant = Agility + Scalability

The public health mission needs more data, on more patients, more often.

Public health depends on EHR data, but a patient only visits their clinician an average of three times per year.

What about all those patients who didn't or couldn't seek treatment?

RESEARCH ARTICLE | STATISTICS

Can auxiliary indicators improve COVID-19 forecasting and hotspot prediction?

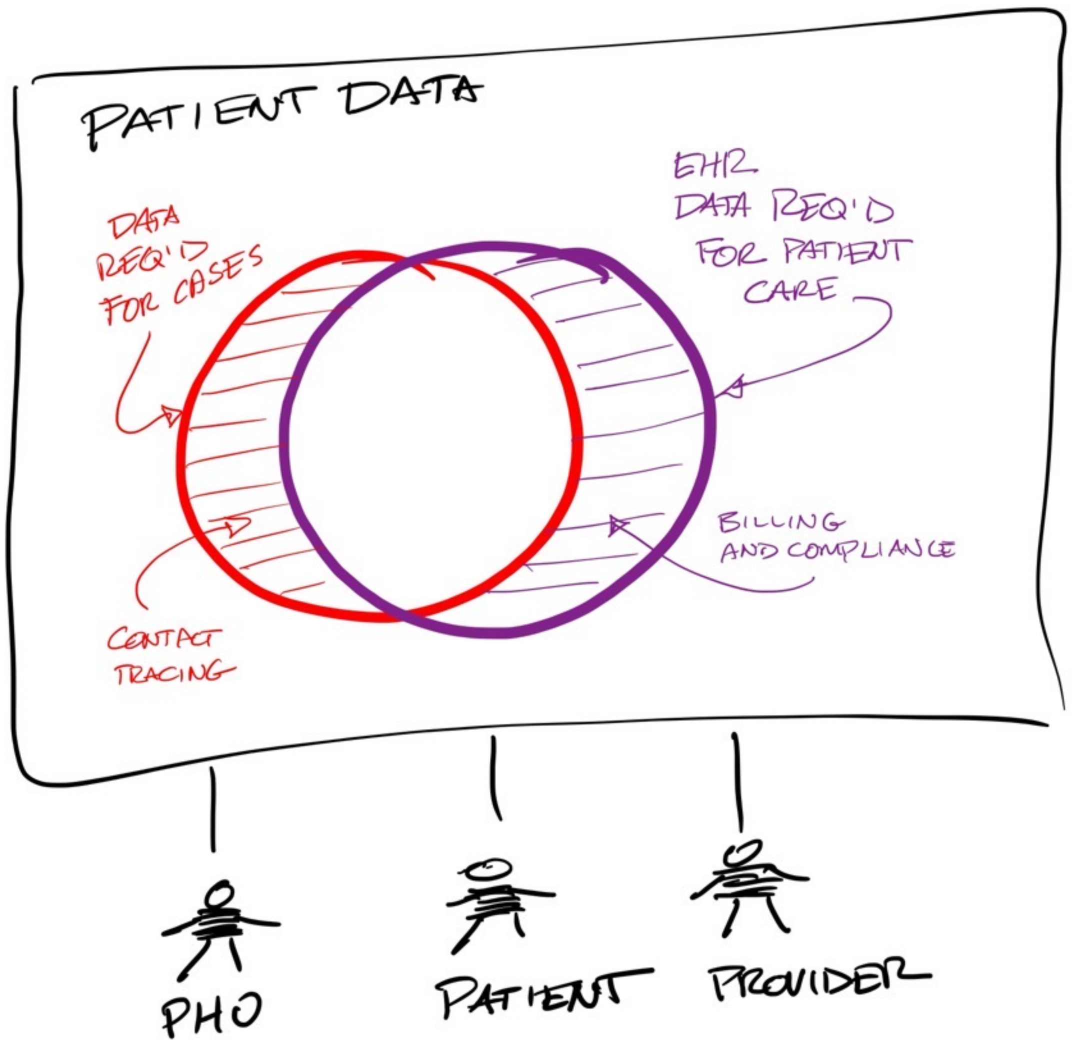
Daniel J. McDonald, Jacob Bien, Alden Green, +9, and Ryan J. Tibshirani Authors Info & Affiliations

Edited by Roni Rosenfeld, Carnegie Mellon University, Pittsburgh, PA; received June 23, 2021; accepted November 2, 2021 by Editorial Board Member William H. Press

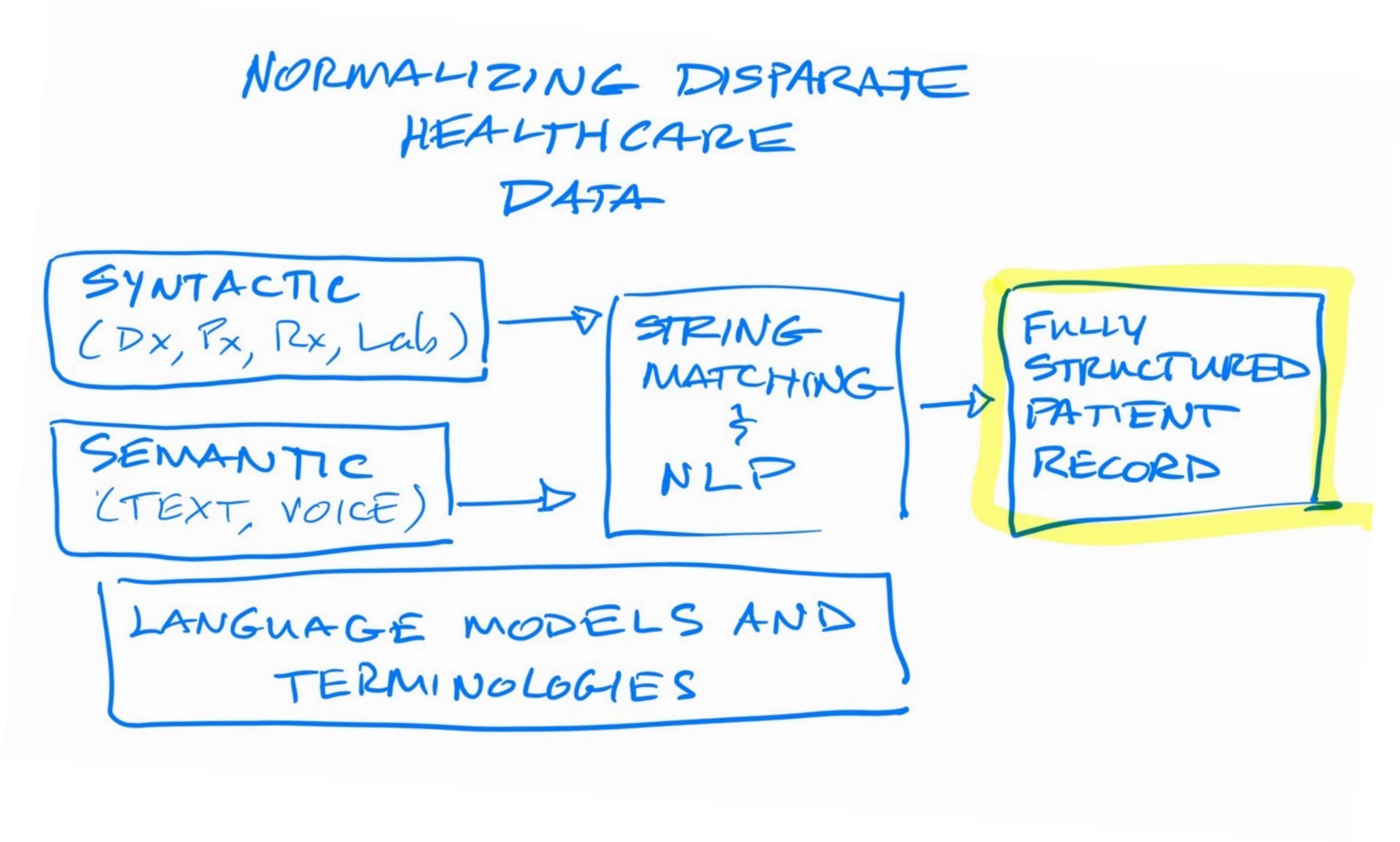
December 13, 2021 | 118 (51) e2111453118 | https://doi.org/10.1073/pnas.2111453118

There is significant overlap in data required for clinical care, and public health

We need to make these two data worlds more seamless with enabling technology



A fully structured patient record is the goal… and now it's possible



Remember how bad our COVID predictive models were?

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Model-based ensembles: Lessons learned from retrospective analysis of COVID-19 infection forecasts across 10 countries

Martin Drews [a,*], Pavan Kumar [b], Ram Kumar Singh [c], Manuel De La Sen [d], Sati Shankar Singh [b], Ajai Kumar Pandey [b], Manoj Kumar [e], Meenu Rani [f], Prashant Kumar Srivastava [g]

[a] Department of Technology, Management and Economics, Technical University of Denmark, Kgs. Lyngby 2800, Denmark
[b] Rani Lakshmi Bai Central Agricultural University, Jhansi 284003, India
[c] Department of Natural Resources, TERI School of Advanced Studies, New Delhi 110070, India
[d] Institute of Research and Development of Processes IIDP, Department of Electricity and Electronics, University of the Basque Country, PO Box 48940, Leioa, Spain
[e] Forest Research Institute, Dehradun, Uttarakhand 248006, India
[f] Department of Geography, Kumaun University, Nainital, Uttarakhand 263001, India
[g] Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi 221005, India

INTRODUCTION | COMPUTER SCIENCES

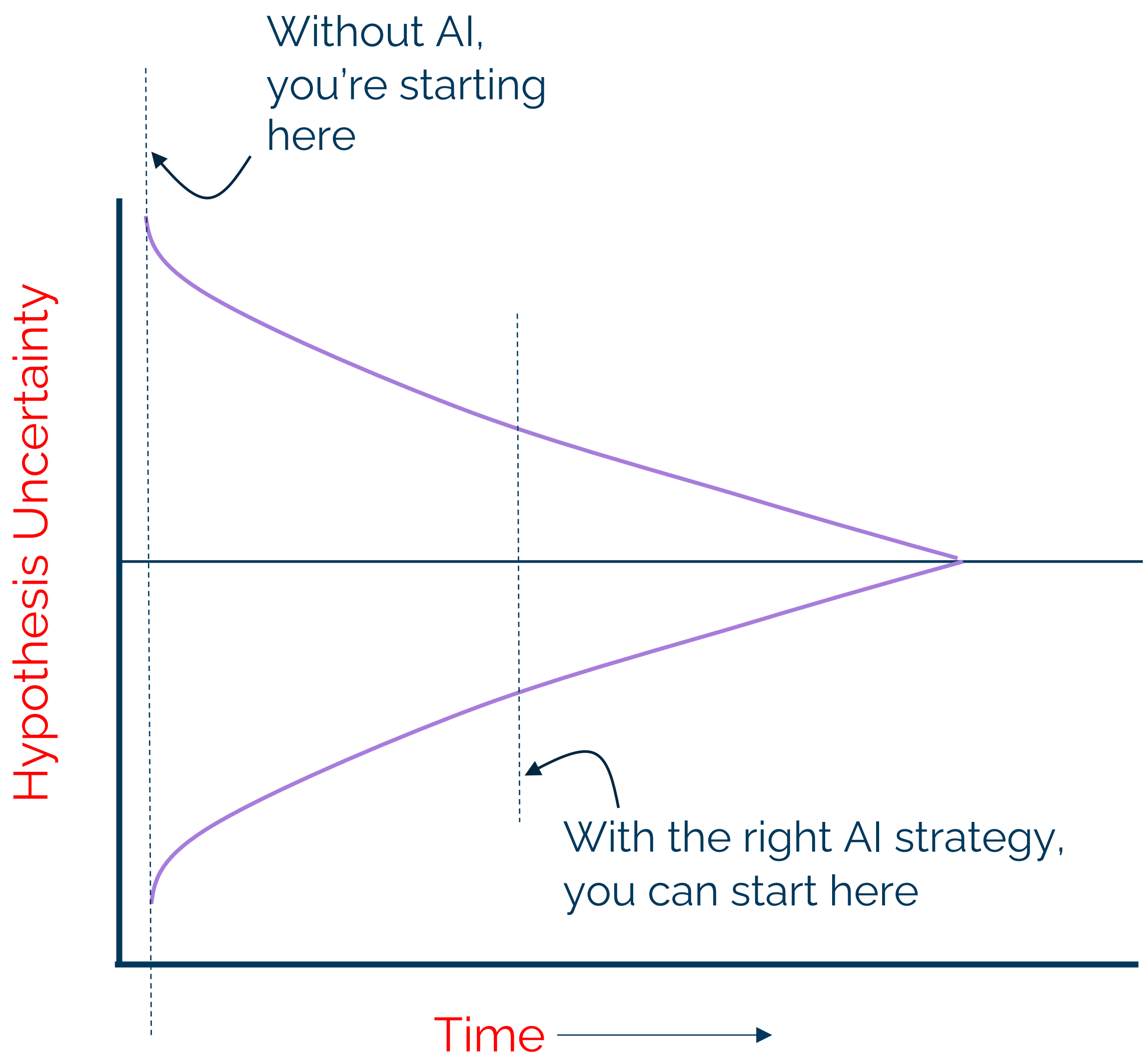
Epidemic tracking and forecasting: Lessons learned from a tumultuous year

Roni Rosenfeld and Ryan J. Tibshirani Authors Info & Affiliations

December 13, 2021 | 118 (51) e2111456118 | https://doi.org/10.1073/pnas.2111456118

Adjust our national AI strategy according to data quality

Given that our EHR and patient data quality is relatively poor and incomplete right now, we should adjust our national AI/ML strategy to support rapid hypothesis generation, not specific predictions



In AI, let the data speak for itself…

From a mentor in the Space, Defense, and National Intelligence Sector

“Let the model fluctuate around the data.”

When data is messy, go with supervised and unsupervised clustering

nature > scientific reports > articles > article

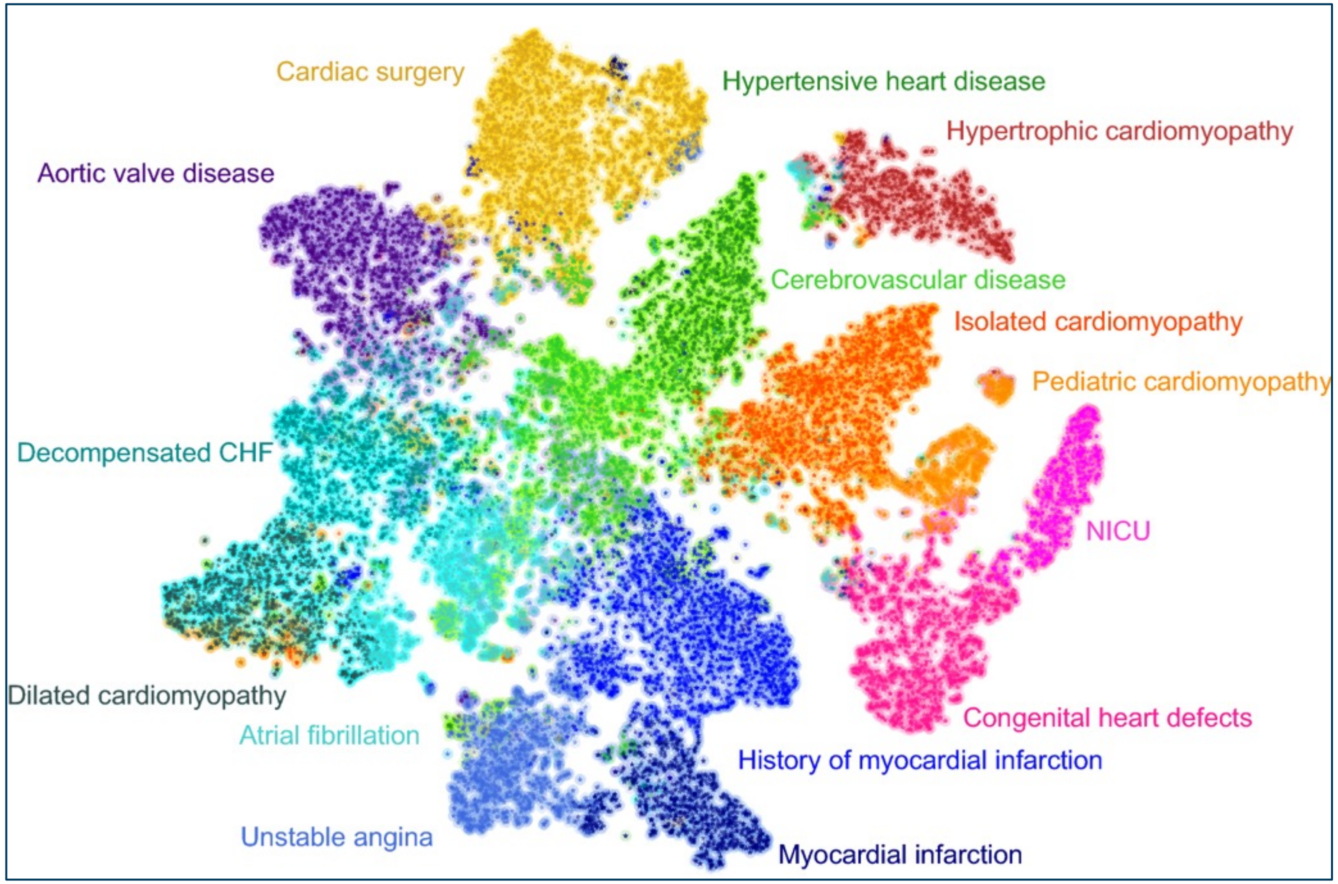
Article | Open Access | Published: 07 December 2020

Multiscale classification of heart failure phenotypes by unsupervised clustering of unstructured electronic medical record data

Tasha Nagamine, Brian Gillette, Alexey Pakhomov, John Kahoun, Hannah Mayer, Rolf Burghaus, Jörg Lippert & Mayur Saxena

Scientific Reports **10**, Article number: 21340 (2020) | Cite this article

2399 Accesses | **3** Altmetric | Metrics



Data Related Observations from COVID

As an informed member of the community

I won't mention the obvious challenges, e.g., data use agreements.

Some of the challenges I mention were a symptom of an unfortunate political environment

Technical, human, and administrative agility is required throughout the lifecycle of the battle

Responding with data agility and uncertainty to an outbreak is important, but so is ongoing adaptability as the outbreak evolves

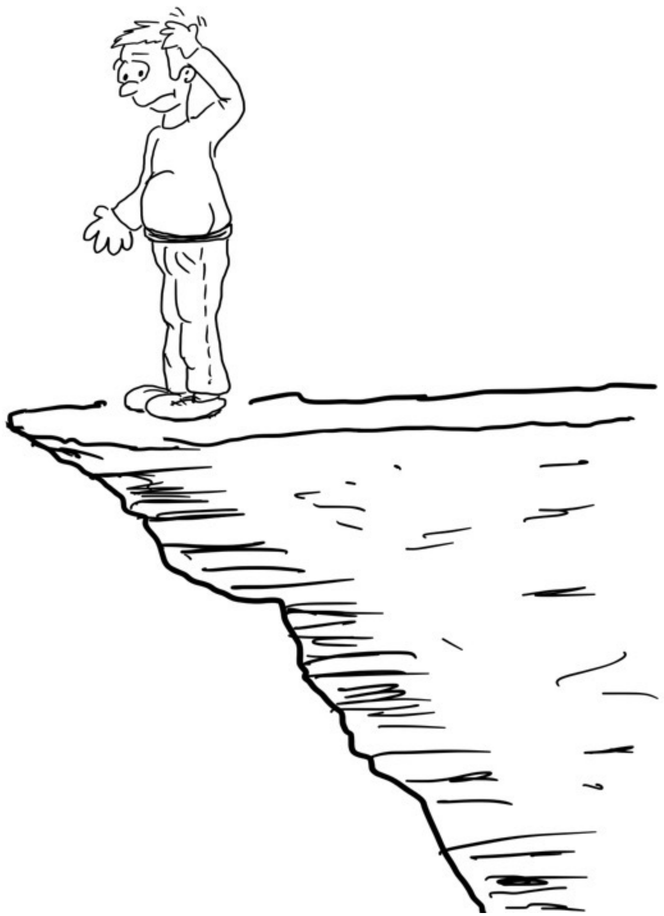
Thank goodness for Johns Hopkins and the New York Times

As a nation, we suspended quality measures for providers during the pandemic.

CMS, NCQA, and commercial payers need to incorporate public health data requirements in the overall quality measures space, by reducing and making room for public health, not adding yet another data entry requirement on physicians.

There was wide variation in communicating COVID risk conditions, nationally and locally

It took too long to develop a risk and threat communication framework, and the final products weren't easily actionable at the personal and family level



We need to be better prepared and creative in quickly estimating disease prevalence in the population

RESEARCH ARTICLE | PHYSICAL SCIENCES |

Estimating SARS-CoV-2 infections from deaths, confirmed cases, tests, and random surveys

Nicholas J. Irons and Adrian E. Raftery Authors Info & Affiliations

Contributed by Adrian E. Raftery, June 16, 2021 (sent for review February 17, 2021); reviewed by Mary Bassett and Constantin T. Yiannoutsos

July 26, 2021 | 118 (31) e2103272118 | https://doi.org/10.1073/pnas.2103272118

Risk management is basically guesswork without understanding the denominator

While lab tests were being developed, other data case definitions (i.e., the data logic) for "This is a COVID patient" took too long to publish and implement in EHRs and data warehouses

WHO COVID-19: Case Definitions

Updated in Public health surveillance for COVID-19, 22 July 2022

World Health Organization | **Case Definitions**

Suspected case of SARS-CoV-2 infection (3 options)

A A person who meets the clinical **OR** epidemiological criteria:

Clinical criteria:

- acute onset of fever AND cough (ILI)

OR

- acute onset of **ANY THREE OR MORE** of the following signs or symptoms: fever, cough, general weakness/fatigue[1], headache, myalgia, sore throat, coryza, dyspnoea, nausea/diarrhoea/anorexia

OR

Epidemiological criteria [2] **:**

- contact of a probable or confirmed case, or linked to a **COVID-19 cluster.**[3]

B A patient with **severe acute respiratory illness** (SARI: acute respiratory infection with history of fever or measured fever of ≥38 °C; and cough; with onset within the last 10 days; and requires hospitalization)

C A person with no clinical signs or symptoms **OR** meeting epidemiologic criteria with a **positive professional-use or self-test** SARS-CoV-2 Antigen-RDT.[4]

[1] Signs separated with slash (/) are to be counted as one sign.

[2] In light of the heightened transmissibility of emerging variants and the high likelihood that any close contact could be infected, epidemiological criteria alone are included in order to qualify asymptomatic contacts for testing, when possible, for the countries with the capacity to adapt more sensitive testing strategies; this is particularly relevant in high-risk populations and settings.

[3] A group of symptomatic individuals linked by time, geographic location and common exposures, containing at least **one NAAT-confirmed** case or at least **two** epidemiologically linked, symptomatic (meeting clinical criteria of Suspect case definition A or B) persons with **positive professional use OR self-test Ag-RDT** (based on ≥97% specificity of test and desired >99.9% probability of at least one positive result being a true positive)

Probable case of SARS-CoV-2 infection (2 options)

A A patient who meets **clinical criteria AND** is a **contact of a probable or confirmed case**, or linked to a **COVID-19 cluster**[3]

B **Death**, not otherwise explained, in an adult with **respiratory distress** preceding death AND who **was a contact of a probable or confirmed case** or linked to a **COVID-19 cluster**[3]

Confirmed case of SARS-CoV-2 infection (2 options)

A A person with a positive **Nucleic Acid Amplification Test (NAAT), regardless of** clinical criteria **OR** epidemiological criteria

B A person meeting clinical criteria **AND/OR** epidemiological criteria (suspect case A) with a **positive professional-use or self- test** SARS-CoV-2 Antigen-RDT.[4]

[4] <u>**Ag RDT antigen-detection rapid diagnostic tests (Ag-RDT) are available for use by trained professionals or for self-testing by individuals:**</u>

- **Professional-use SARS-CoV-2 antigen-RDT :** WHO EUL-approved Ag-RDT, in which sample collection, test performance and result interpretation are done by a trained operator

- **Self-test SARS-CoV-2 antigen-RDT :** WHO EUL-approved Ag-RDT in which sample collection, test performance and result interpretation are done by individuals by themselves.

Note: Clinical and public health judgment should be used to determine the need for further investigation in patients who do not strictly meet the clinical or epidemiological criteria. Surveillance case definitions should not be used as the sole basis for guiding clinical management.

WHO reference number: **WHO/2019-nCoV/Surveillance_Case_Definition/2022.1**

World Health Organization

There was a propensity to lean on randomized trials and peer reviewed evidence before making a decision

Battlefield commanders are trained and must make high-risk decisions with limited objective data

In Closing…

- HITECH funded EHR adoption which gave US healthcare a toehold on computable patient data for care delivery
  - Public health was an afterthought
- COVID put a spotlight on, and tailwind behind, the shortcomings of data for public health
- The ONC, CDC, and STLTs have an opportunity to impact the quality and quantity of patient data for all stakeholders