Acknowledgments

Welcome

Housekeeping

- User-testing of CURE ID during breaks

- Reception in Chesapeake Room – top floor, see Jennifer Stephens if you haven't RSVP'd

- If you are scheduled for an Interview please make your way to the Roosevelt room next door at your scheduled time

Agenda

CDRC Annual Meeting and Workshop Day 2: Wednesday, April 19 Morning

Development of open-sourced tools to automate the collection of repurposed drugs from electronic medical records to generate real-world data.

Meeting kickoff on automating data extraction from EHRs

1. Welcome and Overview: Marco Schito (C-Path) and Heather Stone (FDA)   8:00-8:15
2. Keynote: Jacqueline Corrigan -Curay (FDA)   8:15-8:45

Cochairs: Jagdeep Podichetty (C-Path) & Matthew Robinson (JHU)

EHRs Data Extraction
- Danielle Boyce (JHU) – Pulling structured data and the Edge tool   8:45-9:00
- Wes Anderson (C-Path) Unstructured data using AWS NLP   9:00-9:15
- Ruth Kurtycz (CDC) – Propensity score modeling   9:15-9:30
- Kerry Howard (Clemson) – Lasso regression   9:30-9:45
- BREAK   9:45-10:00

Agenda

CDRC Annual Meeting and Workshop Day 2: Wednesday, April 19 Morning

Moderators: Ewy Mathe (NCATS) & Raghav Tirupathi (C-Path)

1. **Smitty Heavner** (C-Path) –   10:00-10:15
   - *COVID-19 in the Real World: Preliminary Results from the Edge Tool Suite*
2. Roundtable: Discussion of results of Edge tool extraction and analysis   10:15-11:45
   - **Laura Evans** (SCCM)
   - **Nathalie Strub-Wourgaft** (DNDi)
   - **Matt Robinson** (JHU)
   - **Anup Challa** (AstraZeneca)
3. Wrap up and next steps   11:45-12:00
4. **LUNCH**   **12:00 – 1:00**

Introduction and Welcome Address

Keynote

Jacqueline Corrigan-Curay

Jacqueline Corrigan-Curay, JD, MD

Principal Deputy Center Director
FDA's Center for Drug Evaluation and Research (CDER)

Jacqueline Corrigan-Curay provides executive leadership on strategic initiatives that advance CDER's mission to deliver safe, effective and high-quality medications to the public.

Outline

- FDA's RWD/RWE Efforts

- Why is EHR automated extraction important?

- Challenges with automating EHR extraction

- Opportunities regarding EHR extraction

  - FDA/NCATS/C-Path effort to automate EHR extraction with support from HHS OS-PCORTF

- What is the potential utility to FDA?

U.S. 21st Century Cures Act of 2016 – *current status*

- **FDA** *established* a program to evaluate the potential use of real- world evidence (RWE) to:
    - support new indication for a drug approved under section 505(c)
    - satisfy post-approval study requirements

- **Draft framework** *issued* December 2018
    - describes sources of RWE, challenges, pilot opportunities, etc.

- **Draft guidance** for industry *issued* Sep, Oct, Nov, & Dec 2021
    - other guidance in development

- *Commitments met* under Prescription Drug User Fee Act (PDUFA);
    - standard for substantial evidence *remains unchanged*

RWE for Effectiveness: Overview of FDA Approach
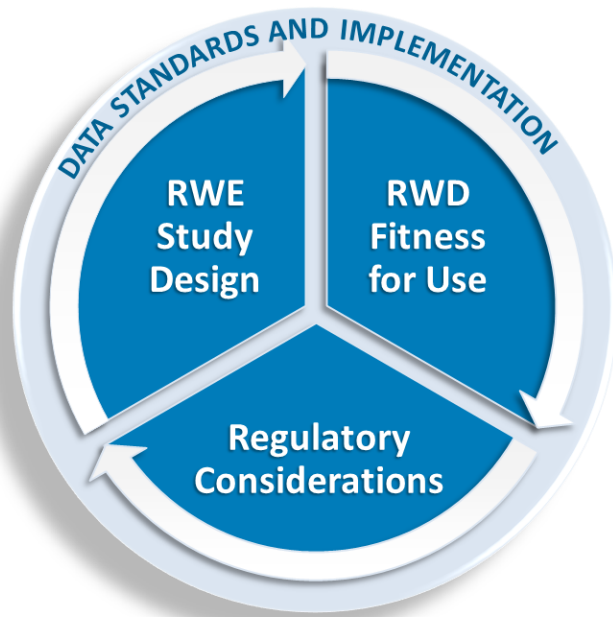
Key considerations (from 2018 Framework):

- Whether the RWD are fit for use

- Whether the trial or study design used to generate RWE can provide adequate scientific evidence to answer or help answer the regulatory question

- Whether the study conduct meets FDA regulatory requirements

Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products

Draft Guidance for Industry

SEPTEMBER 2021

Download the Draft Guidance Document   Read the Federal Register Notice

Draft   Level 1 Guidance

GUIDANCE DOCUMENT

Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products

FEBRUARY 2023

Download the Draft Guidance Document   Read the Federal Register Notice

GUIDANCE DOCUMENT

Considerations for the Use of Real-World Data and Real-World Evidence To Support Regulatory Decision-Making for Drug and Biological Products
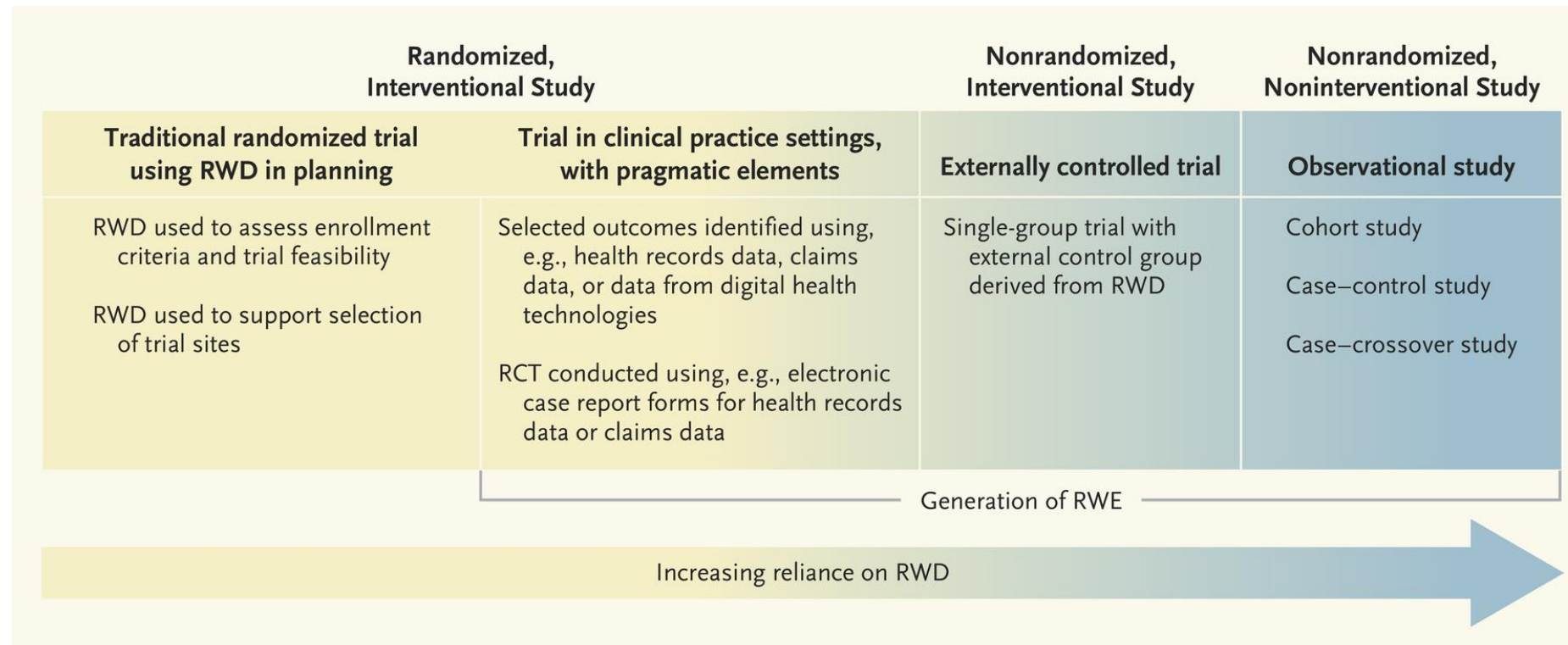
Draft Guidance for Industry

DECEMBER 2021

Download the Draft Guidance Document   Read the Federal Register Notice

Real-World Evidence — Where Are We Now?

John Concato, M.D., M.P.H., and Jacqueline Corrigan-Curay, J.D., M.D.

Randomized, Interventional Study		Nonrandomized, Interventional Study	Nonrandomized, Noninterventional Study
Traditional randomized trial using RWD in planning	**Trial in clinical practice settings, with pragmatic elements**	**Externally controlled trial**	**Observational study**
RWD used to assess enrollment criteria and trial feasibility RWD used to support selection of trial sites	Selected outcomes identified using, e.g., health records data, claims data, or data from digital health technologies RCT conducted using, e.g., electronic case report forms for health records data or claims data	Single-group trial with external control group derived from RWD	Cohort study Case–control study Case–crossover study

Generation of RWE

Increasing reliance on RWD →

Reliance on RWD in Representative Types of Study Design.
RCT denotes randomized, controlled trial; RWD real-world data; and RWE real-world evidence.

RWD and Clinical Endpoints

Biomarker

Clinical endpoint

Type of endpoint	Studies %	Examples of endpoints measured
Chemistry	21%	HBA1c, pregnancy test, GFR
Hematology	4%	Severe neutropenia Apheresis yield > 5 million CD34+ cells/kg
Pathology	1%	Increase/decrease of parabasal cells; biopsy proven acute rejection, clearing of anterior chamber cells
Microbiology	9%	Sustained virological response, plasma viral load, conversion to negative sputum
Imaging +/- (survival, clinical signs)	10%	Bone mineral density; vertebral fractures, spleen volume, progression free survival
Physiological/ functional measurement	10%	6 minute walk, normal sinus rhythm, FEV1, sleep studies
Clinical event /clinical sign	13%	Death, hospitalization, MACE, MS relapse, Lice free head
CRO/PRO	31%	Toronto western spasmodic torticollis rating scale, Hamilton depression rating scale, Rheumatology scale ankylosing spondylitis scale, psoriasis severity index, seizures, sleep, prostate symptom score

Real World Data
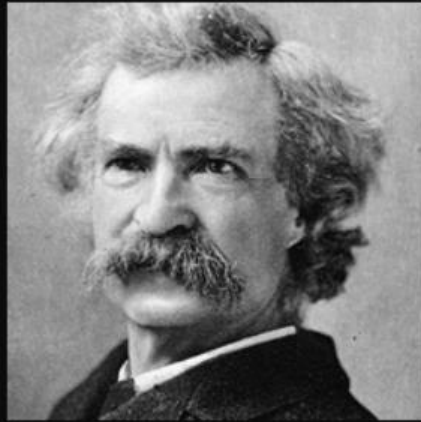
In the real world, nothing happens at the right place at the right time . . .

~ Mark Twain

Why is EHR automated extraction important?

- Enables the ability to access and analyze large quantities of real-world data, while removing the burden on clinical care providers to enter reams of data

- EHR data can have enormous value and can provide information on:
  - Patient characteristics/demographics
  - Co-morbidities
  - Hospitalizations
  - Treatments
  - Outcomes
  - And more...

Automating the Extraction of EHR data is just the first step

- Different EHR systems collect and store information differently
    - Lack of standardization and common data models in some settings
- The information you are looking for may not be captured in an EHR or may not be captured in a usable format
- Incorrect information (e.g., race and gender are often inaccurately reported and do not align with how the patient identifies)
- Missing data is not uncommon but unclear if missing at random

Automating the Extraction of EHR data is just the first step

- It is estimated that up to 80% of data is unstructured and will be difficult to automatically extract

- Therefore, while extraction of structured data is incredibly useful, need to identify the use case for which these data will be most informative

CURE ID - Explore
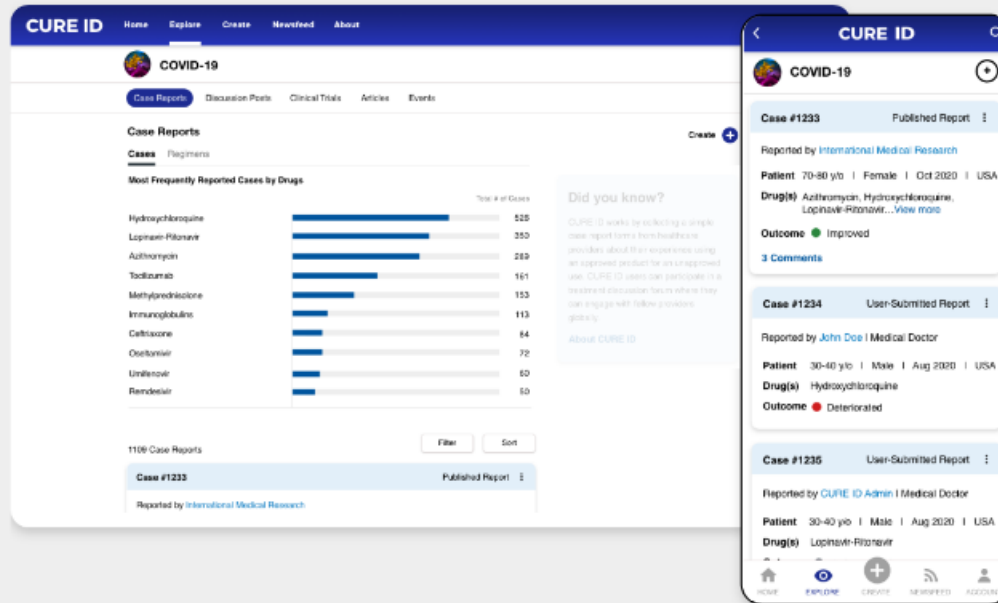
Explore

Explore platform that brings together peer reviewed literature, case studies, insights into practice patterns and selected outcomes, as well as patient-generated data

Getting Cases from EHRs and Registries

- $8.3 million grant to build and deploy an automated EHR extraction tool, the "Edge Tool"
    - Grant from the U.S. Department of Health and Human Services Assistant Secretary for Planning and Evaluation's Office of the Secretary's Patient-Centered Outcomes Research Trust Fund (HHS ASPE OS-PCORtf)
    - Award to FDA and their fellow government health agency partner, the National Center for Advancing Translational Sciences, a part of the National Institutes of Health (NCATS/NIH)
    - In collaboration with the partners below

- CURE Drug Repurposing Collaboratory (CDRC) – a public-private partnership (PPP) convened by The Critical Path Institute (C-Path)

- Institutional partners-
    - Johns Hopkins School of Medicine,
    - Emory School of Medicine,
    - The Society of Critical Care Medicine (SCCM), and
    - The Infectious Diseases Data Observatory (IDDO) at the University of Oxford

What has been Accomplished so Far

- 10,000+ COVID cases automatically extracted from an electronic health record (EHR) after mapping the data to a common data model through the "Edge Tool" – dataset available in the Observational Medical Outcomes Partnership (OMOP) common data model

- Partnership with Infectious Diseases Data Observatory at Oxford to host large dataset in Clinical Data Interchange Standards Consortium (CDISC)

- CURE ID to host a smaller dataset that is displayed as case reports on open-source CURE ID platform

Importance of the Tools being Developed

- Conversion of EHR into a common data model
  - Opens up participation to hospitals with EHRs that have not been designed for supporting research
  - Prior to being able use the tool, participating academic institution took about 2000 hours to convert EHR to OMOP
  - First conversion using Edge tool to OMOP – extract, translate and load (ETL)- took 200-250 hours;
    - Next three conversion down to approx. 100 hours
    - Goal is to bring it down to about 50 hours

- Conversion from OMOP to CDISC completed by IDDO
  - While more limited number of variables has practical considerations as FDA used CDISC
  - Could enable integration with data from RCTs submitted to FDA

FDA recognizes that a range of approaches may be used to apply currently supported data standards (e.g., Clinical Data Interchange Standards Consortium's (CDISC's) Study Data Tabulation Model (SDTM)) to RWD sources such as EHR or claims data.

GUIDANCE DOCUMENT

Data Standards for Drug and Biological Product Submissions Containing Real-World Data

Draft Guidance for Industry

OCTOBER 2021

Download the Draft Guidance Document   Read the Federal Register Notice

Draft   Level 1 Guidance

Importance of the Tools being Developed

- The ability to use clinical practice data alone to answer regulatory questions regarding effectiveness is extremely challenging because there are often a number of factors that influence a clinician's decision to select a particular drug for a particular patient

- Being able to extract the data from an EHR, however, is an important step in bringing clinical practice data together quickly and at a reasonable cost

- Assembly of these data may lead to new hypotheses that drive additional research

- Is an important step for a learning health care system

- For very rare conditions, when results are extremely compelling, may be an important part of the evidence package assuming one can isolate the impact of a single intervention

What utility does automatically extracted-EHR data hold for FDA?

- FDA currently uses EHR and claims data to help analyze safety signals

- FDA's real-world data/real-world evidence (RWD/RWE) efforts are exploring in what circumstances data extracted from EHRs might be used to support regulatory actions (e.g., approvals of supplemental indications, post-market surveillance)
  - For example these efforts also have the potential to enable randomized controlled trials (RCTs) to be integrated in routine clinical practice

Importance of the Tools being Developed

Tools that can facilitate RCTs that use RWD to establish efficacy is important

Looking Forward

- FDA remains committed to exploring tools and methods that can enhance our use of RWE in support of regulatory decisions

- Tools are not just developed to enhance evidence generation for regulatory purposes- they are important if we want to turn each individual encounter into a point for potential learning

- FDA can be a catalyst but we need to promote collective efforts of academics, health care systems and drug developers to realize the potential

CDERMedicalPolicy-RealWorldEvidence@fda.hhs.gov

EHRs Data Extraction

Cochairs: Jagdeep Podichetty (C-Path) & Matthew Robinson (JHU)

EHRs Data Extraction

- **Danielle Boyce** (JHU) – Pulling structured data and the Edge tool

- **Wes Anderson** (C-Path) Unstructured data using AWS NLP

- **Kerry Howard** (Clemson) – Lasso regression

- **Ruth Kurtycz** (CDC) – Propensity score modeling

Jagdeep Podichetty, PhD

Senior Director of Predictive Analytics in
the Quantitative Medicine Program, Critical Path Institute

Dr. Jagdeep Podichetty is the Senior Director of Predictive Analytics in the Quantitative Medicine Program at the Critical Path Institute where he is developing quantitative solutions such as disease progression models, survival models, clinical trial simulation, artificial intelligence models to advance the field of drug discovery and development.

Matthew Robinson, MD

Assistant Professor in the Division of Infectious Diseases, Johns Hopkins University School of Medicine

Matthew Robinson, MD, is Assistant Professor in the Division of Infectious Diseases at Johns Hopkins University School of Medicine. He is interested in leveraging diagnostic innovation and precision medicine to reduce diagnostic and prognostic uncertainty for infectious diseases.

CURE ID with the OHDSI Edge Tool Suite: Automating Data Extraction from Eletronic Medical Records

Danielle Boyce

Danielle Boyce

Faculty, Johns Hopkins University School of Medicine

Danielle Boyce is a data scientist, instructor, patient advocate, and researcher. She has served on several patient and caregiver advisory panels for the Patient-Centered Outcomes Research Institute, the U.S. Food and Drug Administration as well as academic institutions, pharmaceutical companies, and nonprofits.

Introducing OHDSI

- The Observational Health Data Sciences and Informatics (OHDSI) program is a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics

- OHDSI has established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University
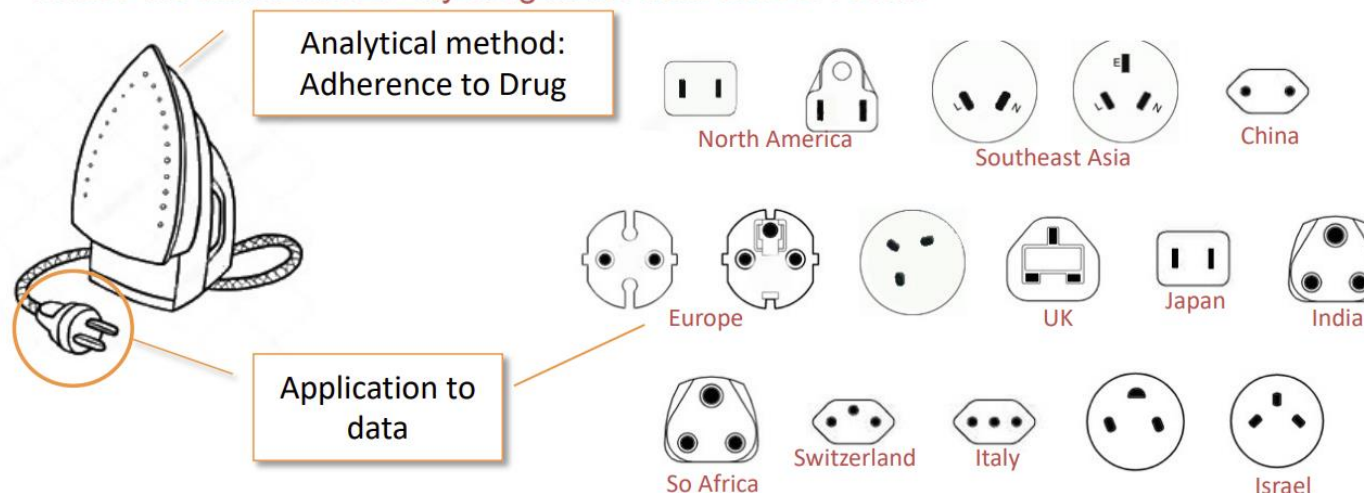
http://ohdsi.org

OHDSI's Mission

To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care
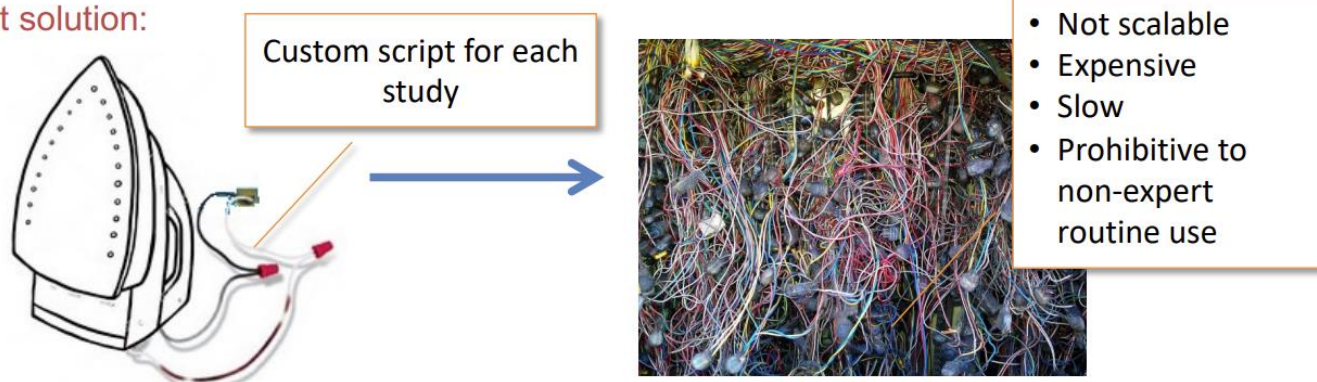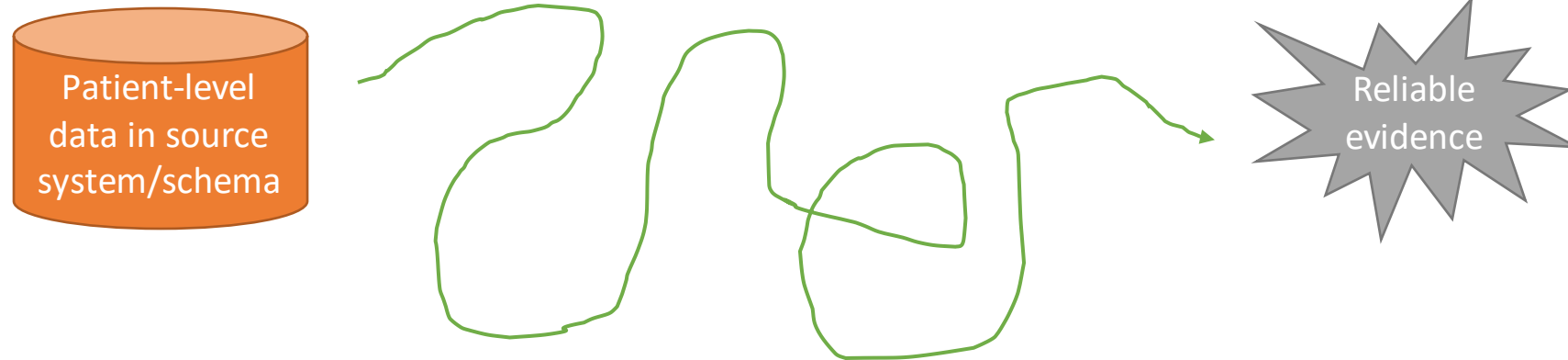
Current Approach: "One Study – One Script"

"What's the adherence to my drug in the data assets I own?"

Analytical method: Adherence to Drug

Application to data

North America

Southeast Asia

China

Europe

UK

Japan

India

So Africa

Switzerland

Italy

Israel

Current solution:

Custom script for each study

- Not scalable
- Expensive
- Slow
- Prohibitive to non-expert routine use

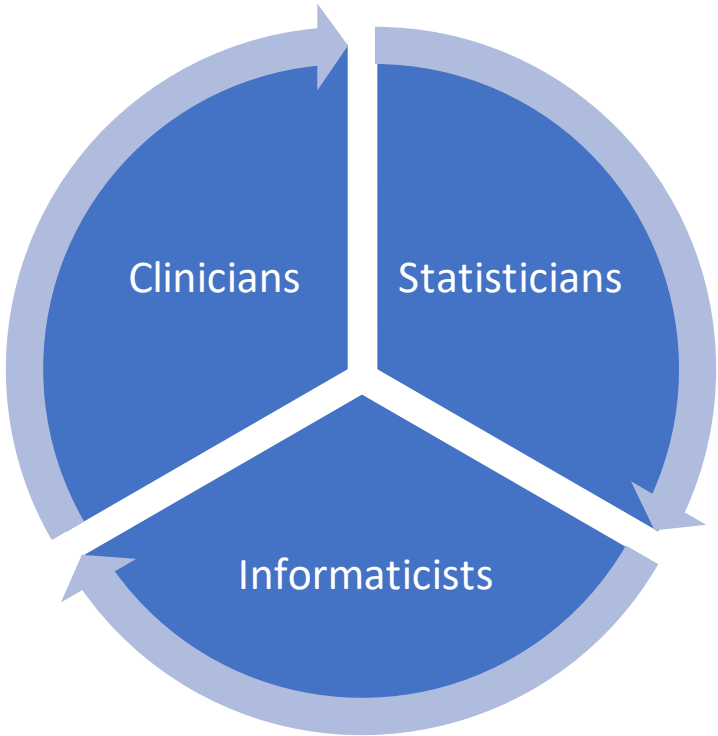# The Journey to Real-World Evidence

OHDSI Community

We're all in this journey together…

# Multi-disciplinary Innovation with Open Science at Scale

Translate EMR and claims data into a common data model tied to standard terminologies.

Clinicians

Statisticians

Informaticists

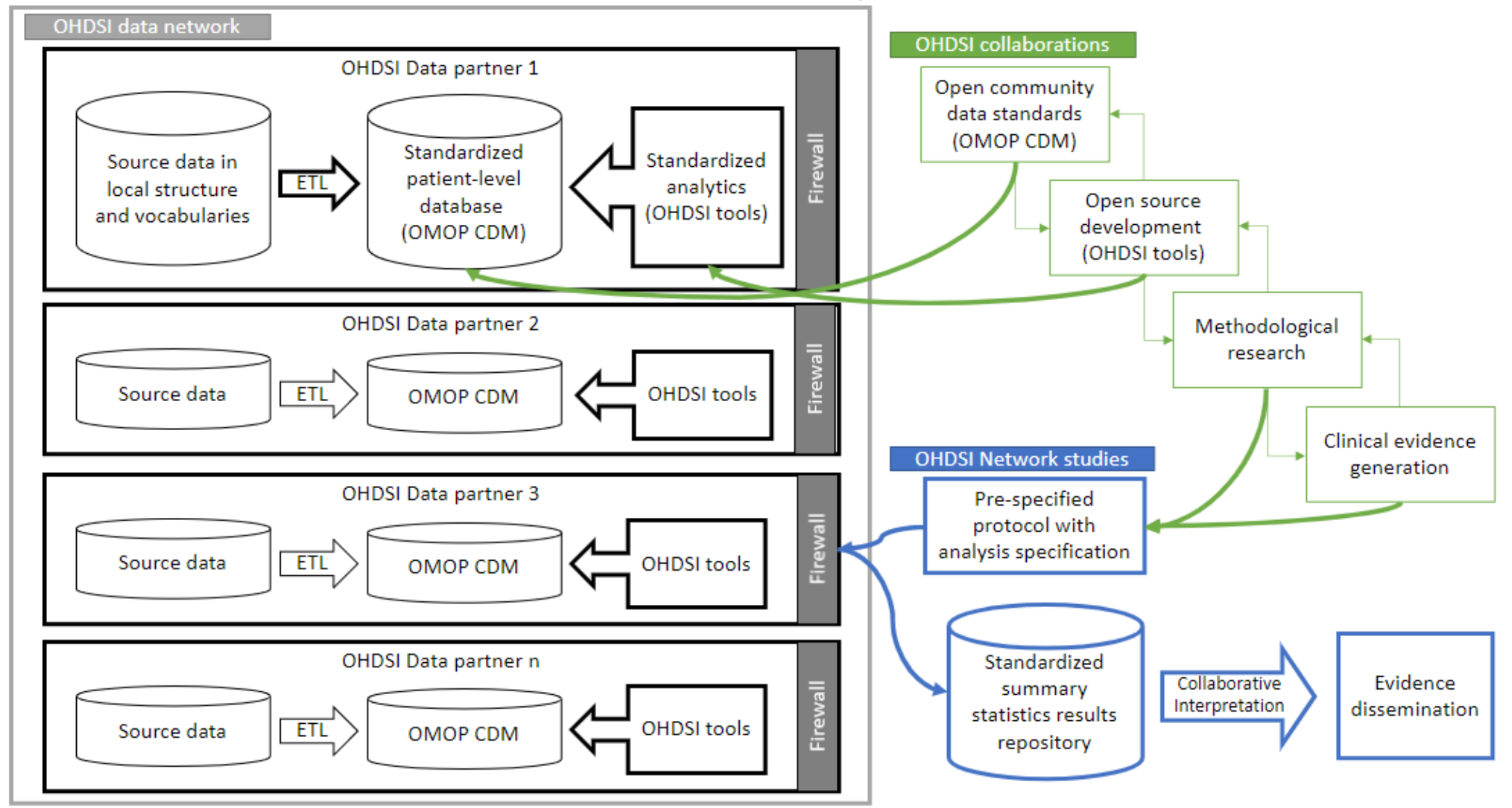Open-source ecosystem of validated statistical tools on observational health data

Open-source community and research network with 100's millions health records
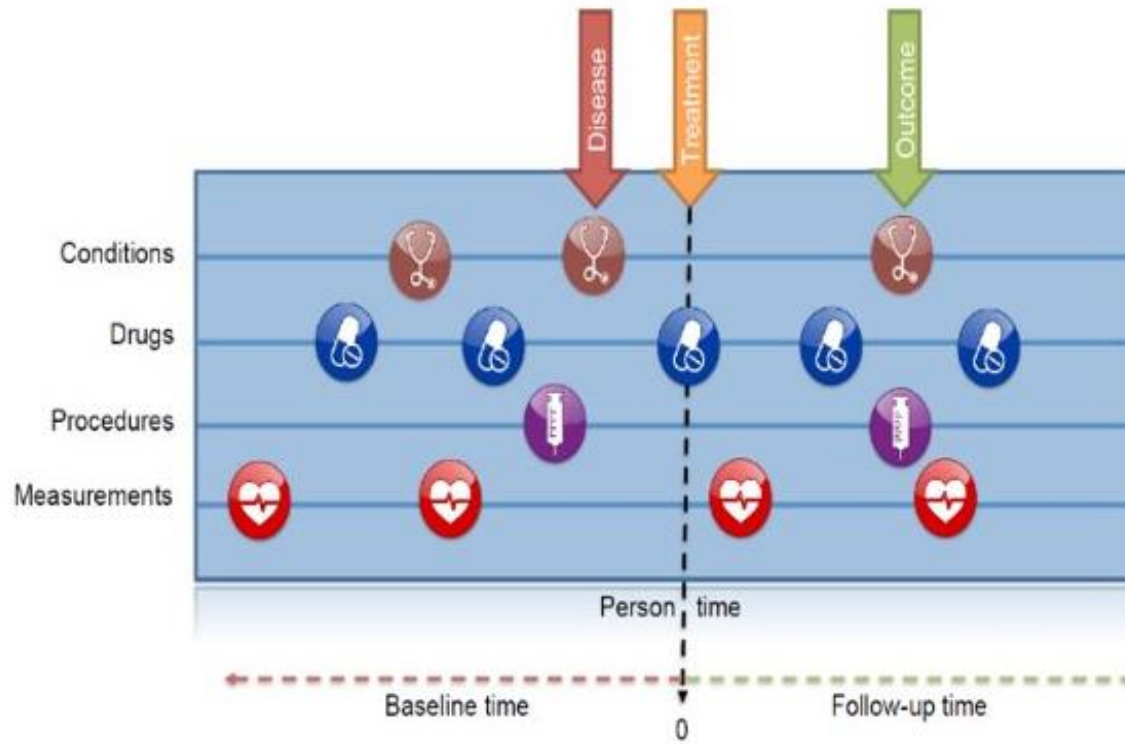
OMOP CDM V5.3.1
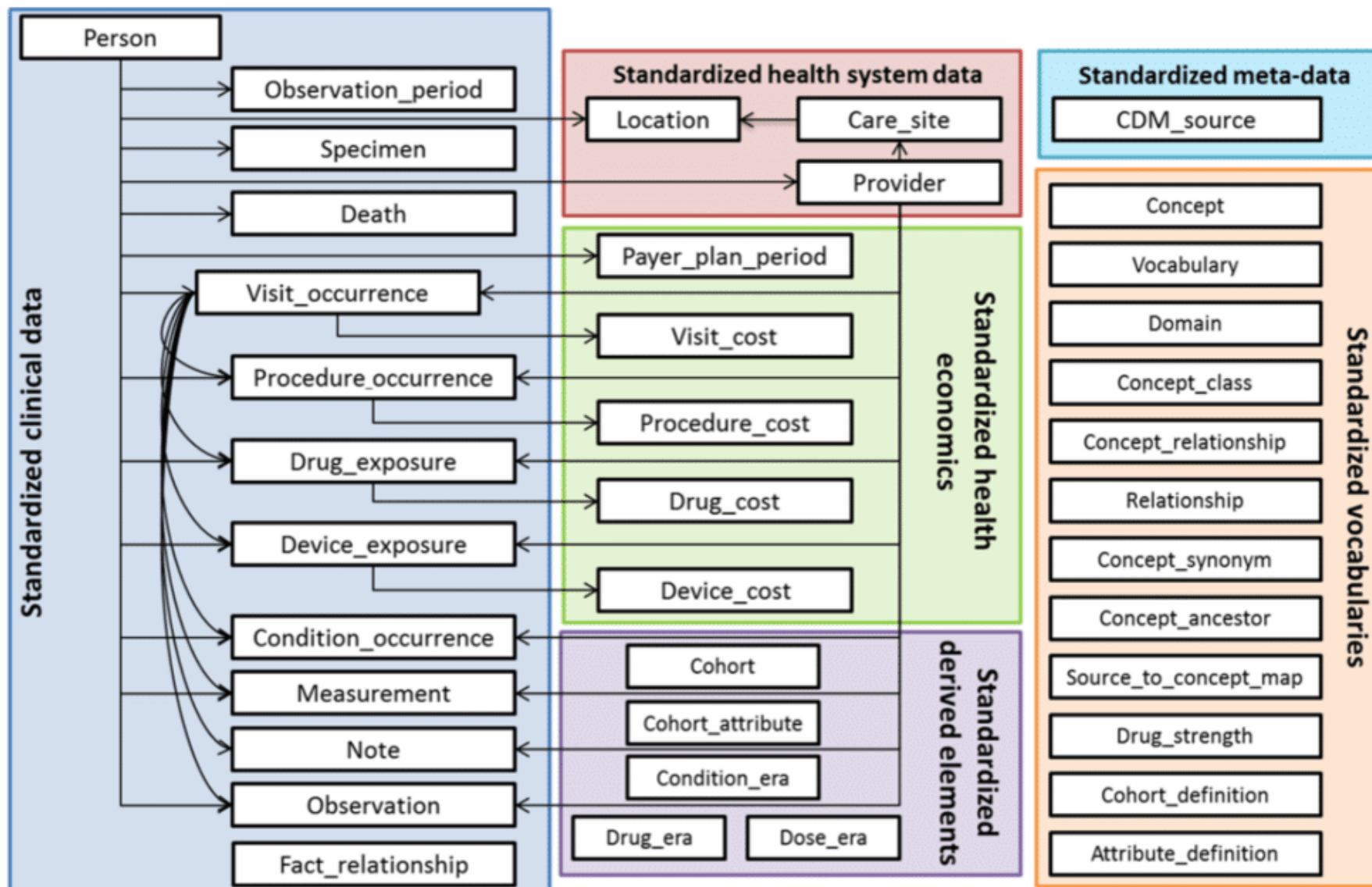
OHDSI Community

Standardized Concept Vocabularies

Literature

Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases

Patricia Biedermann, Rose Ong, Alexander Davydov, Alexandra Orlova, Philip Solovyev, Hong Sun, Graham Wetherill, Monika Brand & Eva-Maria Didden ✉

Mapping registry data to the OMOP CDM facilitates more efficient collaborations between researchers and establishment of federated data networks

Trauma registry methodology: A survey of trauma registry custodians to determine current approaches

Gerard M. O'Reilly [a,b,*], Belinda Gabbe [a], Peter A. Cameron [a,b,c]

[a] Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Commercial Rd, Melbourne, 3004, Australia
[b] Emergency and Trauma Centre, Alfred Health, Commercial Rd, Melbourne, Victoria 3004, Australia
[c] Emergency Services, Hamad Medical Corporation, Doha, Qatar

Table 3
Human resources–Single hospital registries ($n = 40$).

Staffing type	Number of persons			
	0	1	2	>2
All (total) staff	0	2	4	34
Director/Head	14	25	1	0
Manager	18	21	1	0
Data manager	17	21	0	2
Database programmer	33	2	4	1
Database analyst	31	7	1	1
Trauma nurse coordinator	17	15	6	2
Data collector	12	2	4	22
Data entry clerk	23	9	2	6
ICD coder	26	4	0	10
AIS coder	21	4	2	13
Data analyst	28	9	2	1
Office administrator	29	11	0	0

0.5 FTE for every 200-300 patients.

Courtesy – Jon Duke, MD, Georgia Tech Research Institute

> "The vision of creating accessible, reliable clinical evidence by accessing the clinical experience of hundreds of millions of patients across the globe is a reality."

George Hripcsak, MD, MS
Vivian Beaumont Allen Professor & Chair of
Biomedical Informatics
Columbia University Medical Center

Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers

George Hripcsak[a], Jon D Duke[b], Nigam H Shah[c], Christian G Reich[d], Vojtech Huser[e], Martijn J Schuemie[f,g], Marc A Suchard[h], Rae Woong Park[i], Ian Chi Kei Wong[f], Peter R Rijnbeek[j], Johan van der Lei[j], Nicole Pratt[k], G Niklas Norén[l], Yu-Chuan Li[m], Paul E Stang[g], David Madigan[n], and Patrick B Ryan[g]

Goals of Edge Tool Suite

- Assist up to 60 health systems to develop automated ETL processes from their EMR into the OHDSI OMOP CDM format.

- Support OHDSI Broadsea open source project

- Validate and deploy at sites

- Docker Container

Components of the Edge Tool Suite

1. Web-based decision support for concept mapping

2. Base configuration settings for major EMR Vendors.

3. Configuration management documentation tool

4. Inspection Report of DevOps on ETL processes

5. Data Quality Dashboard framework of 3,000+ data quality tests

6. Virus Registry cohort subset definition

7. Perform de-identification and submission to the Virus Registry

ETL Mapping Decision Support Tools

ATLAS

- Home
- Data Sources
- Search
- Concept Sets
- Cohort Definitions
- Characterizations
- Cohort Pathways
- Incidence Rates
- Profiles
- Estimation
- Prediction
- Jobs
- Configuration
- Feedback

Apache 2.0

Cohort Entry Events

Events having any of the following criteria:

a condition occurrence of [Type 2 Diabetes Mellitus Diagn...]

a drug exposure of [Type 2 Diabetes Mellitus Prescri...]

with continuous observation of at least [0▼] days before and [0▼] days after event index date

Limit initial events to: [all events▼] per person.

Restrict initial events to:

having [all▼] of the following criteria:

with [at most▼] [0▼] using all occurrences of:
a condition occurrence of [Type 1 Diabetes Mellitus Diagn...]
where [event starts] between [All▼] days [Before▼] and [All▼] days [After▼] [index start date] add additional constraint
☐ restrict to the same visit occurrence
☐ allow events from outside observation period

And having [any▼] of the following criteria:

having [all▼] of the following criteria:

with [at most▼] [0▼] using all occurrences of:
a condition occurrence of [Type 2 Diabetes Mellitus Diagn...]
where [event starts] between [All▼] days [Before▼] and [All▼] days [After▼] [index start date] add additional constraint
☐ restrict to the same visit occurrence
☐ allow events from outside observation period

and with [at least▼] [1▼] using all occurrences of:
a drug exposure of [Type 2 Diabetes Mellitus Prescri...]
where [event starts] between [All▼] days [Before▼] and [All▼] days [After▼] [index start date] add additional constraint
☐ restrict to the same visit occurrence
☐ allow events from outside observation period

REVIEW — Open Access

Extracting research-quality phenotypes from electronic health records to support precision medicine

Wei-Qi Wei[1] and Joshua C Denny[1,2*]

Figure decision-tree nodes:
- T1DM ICD-9 code? — NO → T2DM ICD-9 code? — NO → Rx T2DM medication? — YES → ...
- T2DM ICD-9 code? — YES → Rx Insulin? — YES → Rx T2DM medication? — NO → ≥2 dates when clinician entered T2DM Dx
- Rx T2DM medication? — YES → Rx T2DM medication before insulin Rx? — YES → Rx T2DM medication?
- Rx Insulin? — NO → Rx T2DM medication? — YES → Rx T2DM medication?; NO → Abnormal glucose or HbA1c? — YES → Rx T2DM medication?
- Rx T2DM medication? — YES

Figure 1 Algorithm for the identification of subjects with type 2 diabetes. Normal glucose values are random glucose >200 mg/dl, fasting glucose >125 mg/dl. Normal HbA1c ≥6.5%. Dx, diagnosis; HbA1c, hemoglobin A1c; ICD-9, International Classification of Diseases, Ninth Revision; Rx, treatment; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus. Figure reprinted with permission from Kho *et al.* [57].

Data Quality Dashboard

CDRC CRITICAL PATH INSTITUTE | **CURE ID** Challenging cases... New approaches

DATA QUALITY ASSESSMENT

JOHNS HOPKINS MEDICINE ENTERPRISE

DataQualityDashboard Version: 1.0.0
Results generated at 2021-08-28 09:20:06 in 7 hours

Examples of DQ Checks from Kahn et al (2016)

Atemporal Plausibility	• 48% of labs outside of normal range
Temporal Plausibility	• Unexpected change in number of records from month to month
Completeness	• 62% of *route_concept_id* is missing
Value Conformance	• ICD9 codes in *condition_concept_id*
Relational	• *visit_date* and *visit_datetime* inconsistency in

	Pass
Plausibility	1938
Conformance	64
Completeness	35
Total	293

STATUS	CONTEXT	CATEGORY	SUBCATEGORY	LEVEL	DESCRIPTION	% RECORDS
FAIL ▾	Verification ▾	Completeness ▾	▾	▾		▼
FAIL	Verification	Completeness	None	FIELD	The number and percent of records with a value of 0 in the standard concept field UNIT_CONCEPT_ID in the SPECIMEN table. (Threshold=5%).	0%
FAIL	Verification	Completeness	None	FIELD	The number and percent of records with a value of 0 in the standard concept field UNIT_CONCEPT_ID in the DOSE_ERA table. (Threshold=0%).	0%

Showing 6 to 7 of 7 entries (filtered from 3,424 total entries)

Previous 1 2 Next

Concepts for VIRUS Registry

- 111 Concepts
- Many are standard to OMOP
- Some come from forms within EMR
  - Respiratory Devices

Concept ID	Concept Code	Concept Name	Domain	Vocabulary
38003563	Hispanic	Hispanic or Latino	Ethnicity	Ethnicity
38003564	Not Hispanic	Not Hispanic or Latino	Ethnicity	Ethnicity
8570	A	AMBIGUOUS	Gender	Gender
8532	F	FEMALE	Gender	Gender
8507	M	MALE	Gender	Gender
8521	O	OTHER	Gender	Gender
8551	U	UNKNOWN	Gender	Gender
8657	1	American Indian or Alaska	Race	Race
8515	2	Asian	Race	Race
8516	3	Black or African American	Race	Race
8557	4	Native Hawaiian or Other	Race	Race
8527	5	White	Race	Race
3655896	870386000	Vasopressor therapy	Procedure	SNOMED
4230167	40617009	Artificial respiration	Procedure	SNOMED
4052536	233573008	Extracorporeal membrane	Procedure	SNOMED
3655950	870533002	Heated and humidified hig	Procedure	SNOMED
4177224	428311008	Noninvasive ventilation	Procedure	SNOMED
4201025	315041000	High concentration oxyge	Procedure	SNOMED
4162736	371908008	Oxygen administration by	Procedure	SNOMED
4119964	304577004	Humidified oxygen therap	Procedure	SNOMED
4155151	371907003	Oxygen administration by	Procedure	SNOMED
4216695	71786000	Intranasal oxygen therapy	Procedure	SNOMED

Deidentification

- Reassignment of Person IDs
- Date Shifting
- Date Truncation
- Removal of other identifiers

Preserving temporal relations in clinical data while maintaining privacy

George Hripcsak,[1] Parsa Mirhaji,[2] Alexander FH Low,[3] and Bradley A Malin[4,5]

ABSTRACT

Leveraging AWS Resources and Natural Language Processing to Develop an End-to-End Data and Analytics
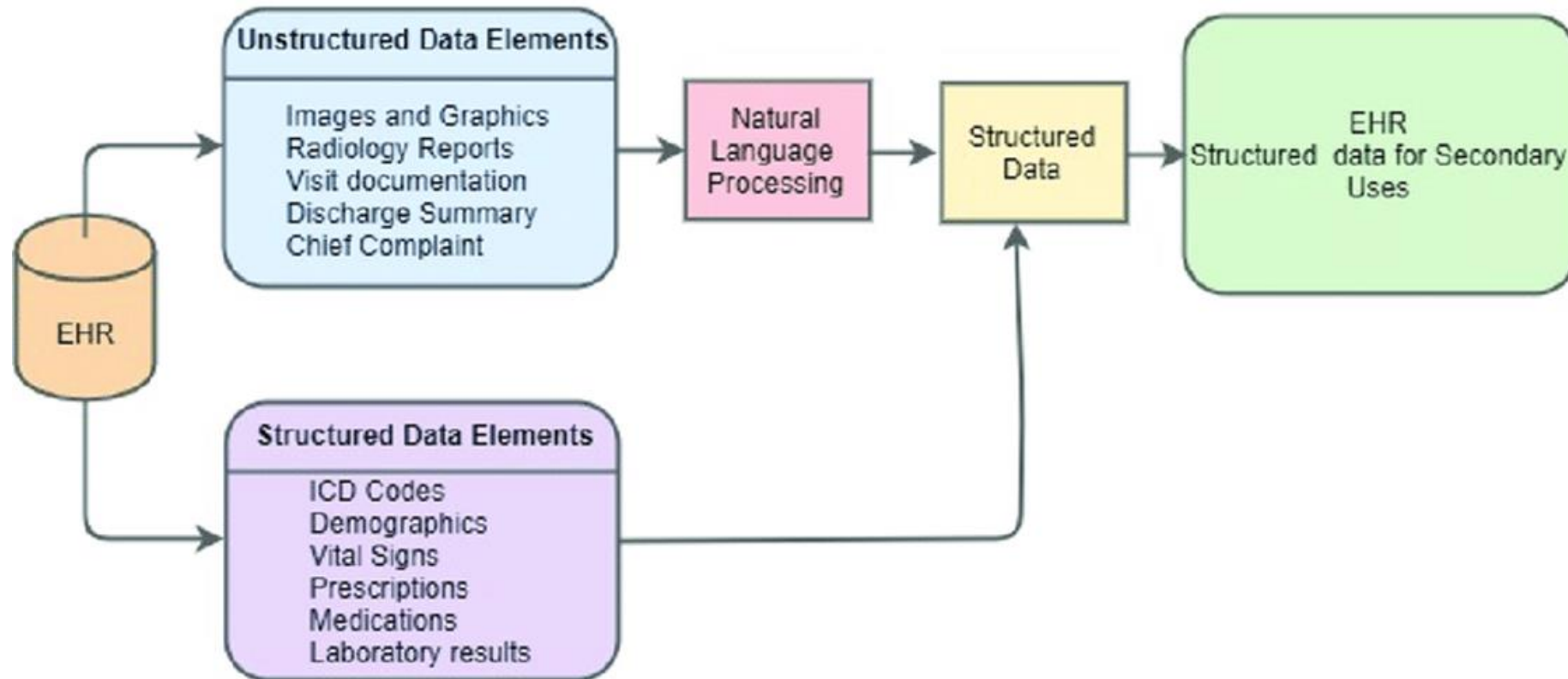
Wes Anderson

Wes Anderson, PhD

Postdoctoral Fellow in the Quantitative Medicine
Program at the Critical Path Institute

Dr. Wes Anderson is a Postdoctoral Fellow
in the Quantitative Medicine Program at
the Critical Path Institute, where he is
developing quantitative solutions such
through machine learning and artificial
intelligence to advance the field of drug
discovery and development.

Electronic Health Records (EHRs)

Shah et. al. "Secondary Use of Electronic Health Record: Opportunities and Challenges," *IEEE Access, vol. 8,* 2020

Implementation Directly to the EHR

CURE ID
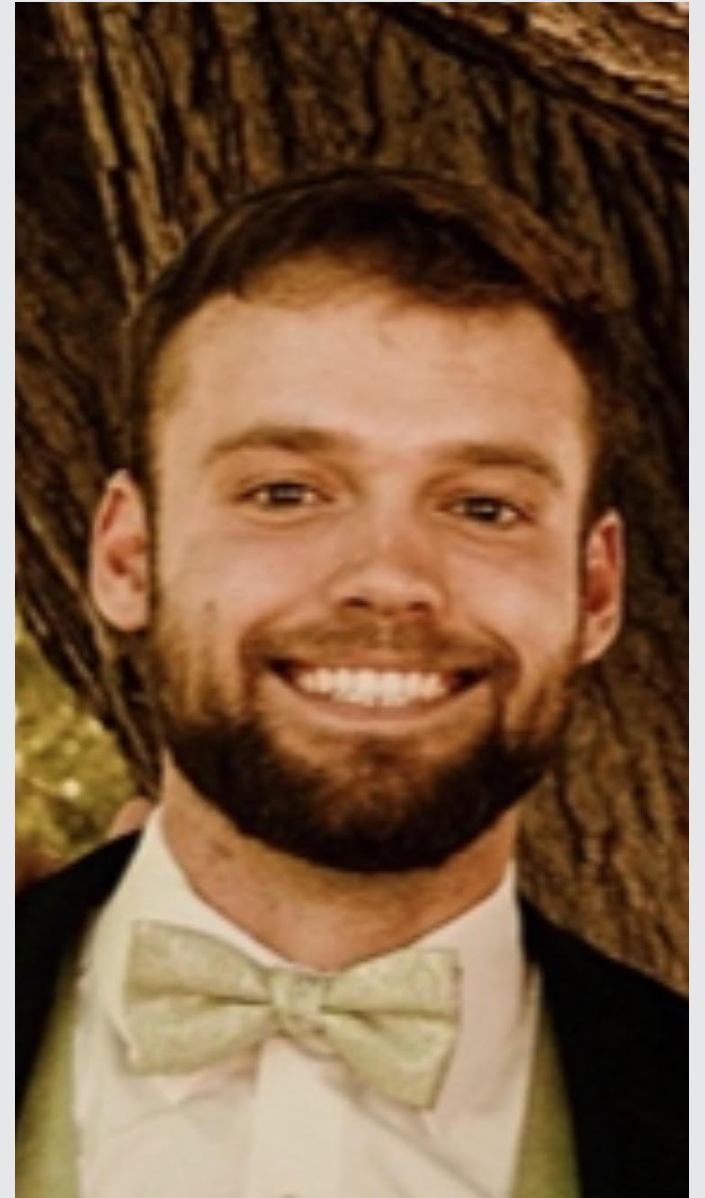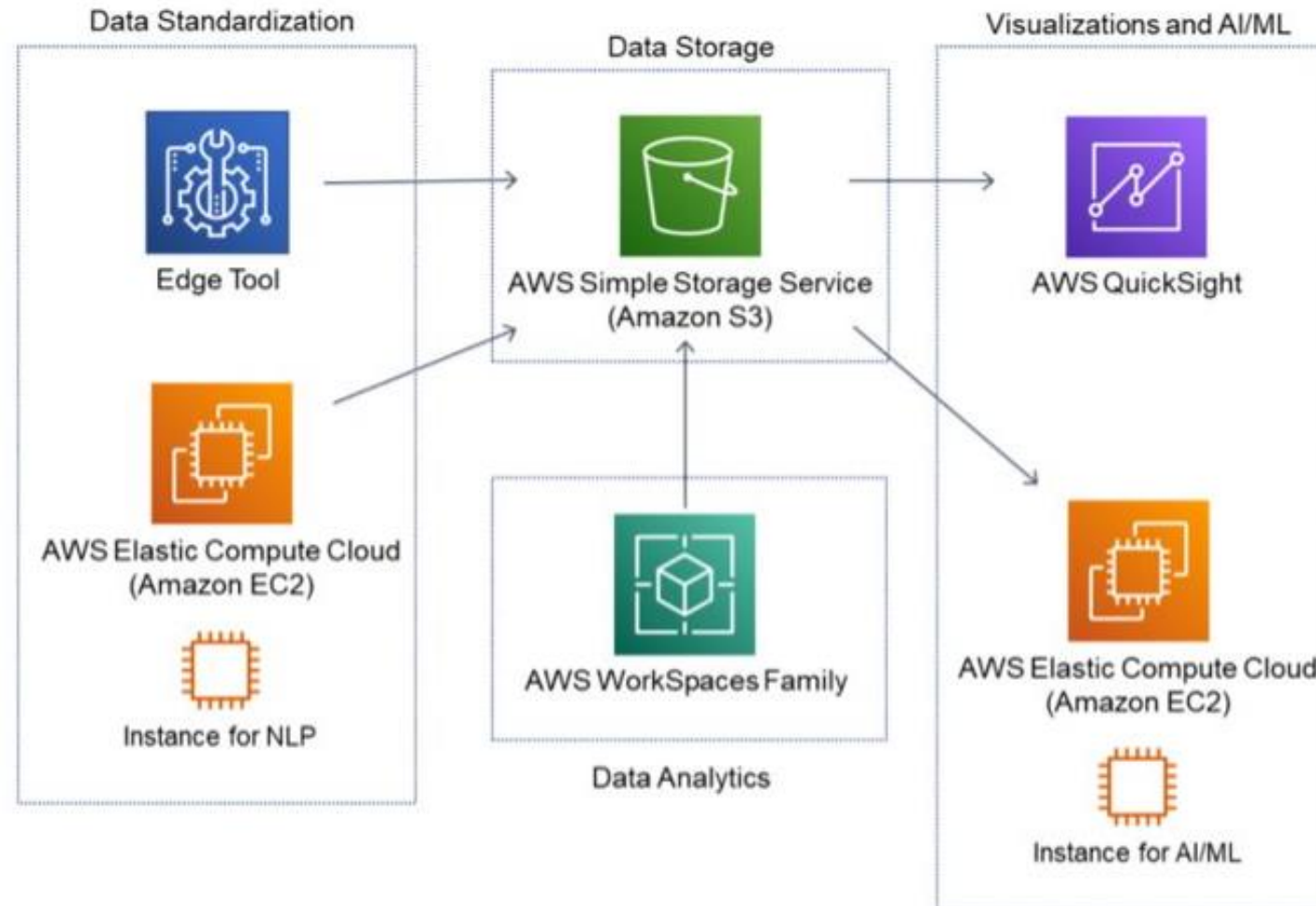
- A platform to capture novel uses of existing drugs

- Web-based tool
  - Computer, smartphone or mobile device

- Capture and share real-world experience treating patient through a simple online case report form

- All data collected is HIPAA compliant and contains no PII

- **Multiple fields of free-text information**

https://cure.ncats.io/explore

CURE ID: Structured Information
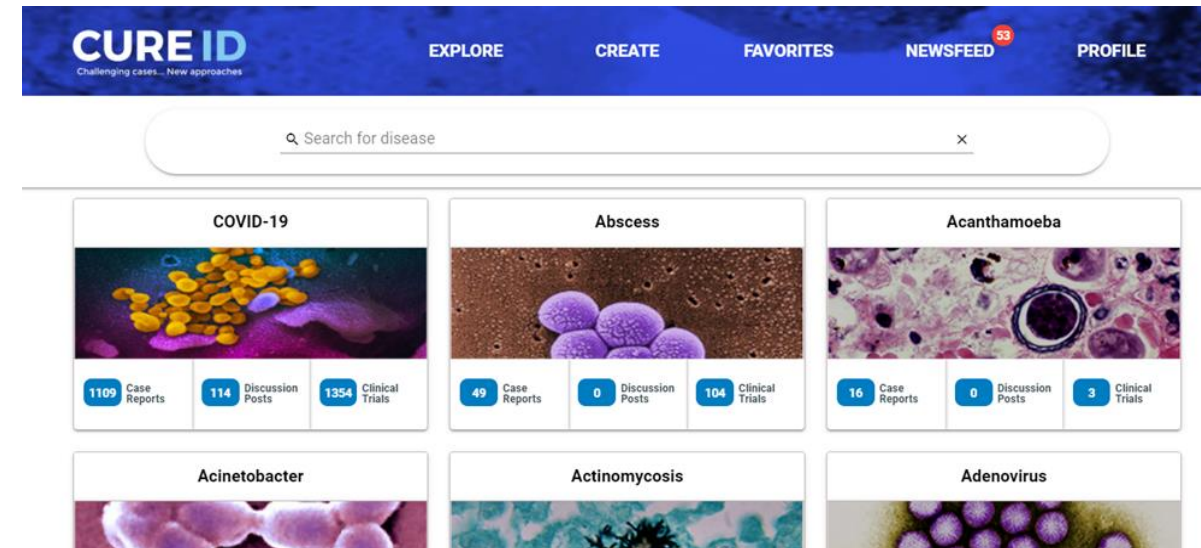
- Case characteristics
  - Age range, race(s), ethnicity, sex assigned at birth, pregnancy status

- Disease Name

- Medications
  - Drug name, dosage, frequency, route, duration

- Outcomes
  - Improved, deteriorated

CURE ID Free-Text Information

Examples of fields with free-text information

- -"Additional information"
  - "Vaccinated. Received Hydroxyurea per protocol with improvement."
- -"Clinical syndrome"
  - "History of flu like symptoms which worsened along with pregnancy at gestational age of 34 weeks."
- -"Adverse events"
  - "Intolerable gastrointestinal adverse effects from lopinavir/ritonavir."
- -"Site of disease"
  - "Multisystems: severe muscle pain, headache; moderate cough, fever, congestion, dizziness, brain fog"
- -"Other coinfections"
  - "Acute renal failure, hepatic transaminitis"

This information can be extracted and standardized with **Natural Language Processing (NLP)**

NLP Tasks

Lipizzi et. al. "Meshing Capability and Threat-based Science and Technology (S&T) Resource Allocation," *Research Gate*, July 2020.
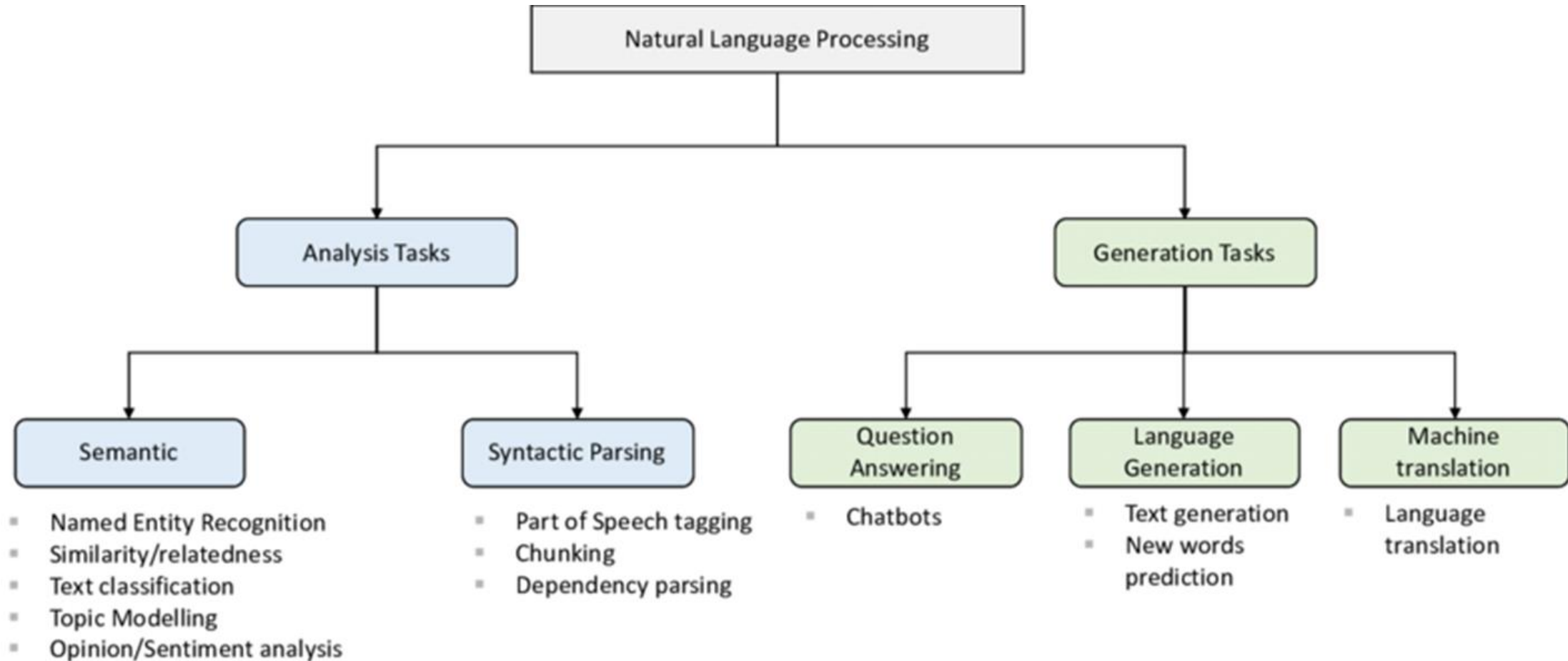
CDRC AWS Data and Analytics Architecture
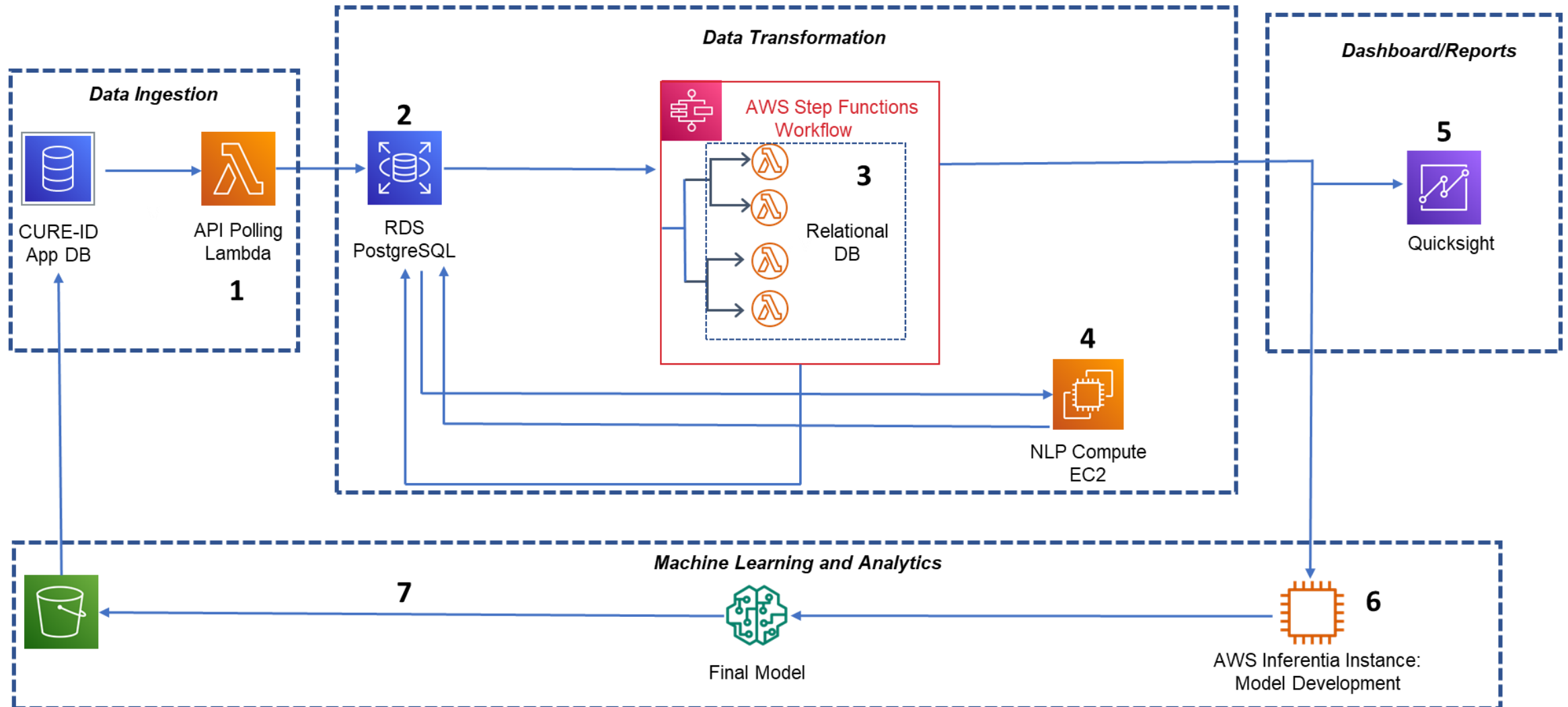
Investigated Commercial Services for NLP

	Pricing Model	Cost
AWS Comprehend Medical https://awsvideocatalog.com/machine_learning/amazon_comprehend/	Pay-as-you-go	https://aws.amazon.com/comprehend/medical/pricing/
Spark NLP https://www.nobleprog.com/spark-nlp-training	License	$42K/year
Microsoft Azure https://en.wikipedia.org/wiki/Microsoft_Azure	Pay-as-you-go	Sentiment analysis (available in containers) Key phrase extraction (available in containers) Language detection (available in containers) Named entity recognition, including PII 0M-0.5M text records - $1 per 1,000 text records 0.5M-2.5M text records - $0.75 per 1,000 text records 2.5M-10.0M text records - $0.30 per 1,000 text records 10.0M+ text records - $0.25 per 1,000 text records Text analytics for health (available in containers) 0M-0.005M text records - Included 0.005M-0.5M text records - $25 per 1,000 text records 0.5M-2.5M text records - $15 per 1,000 text records 2.5M+ text records⁴ - $10 per 1,000 text records https://azure.microsoft.com/en-us/pricing/details/cognitive-services/language-service/

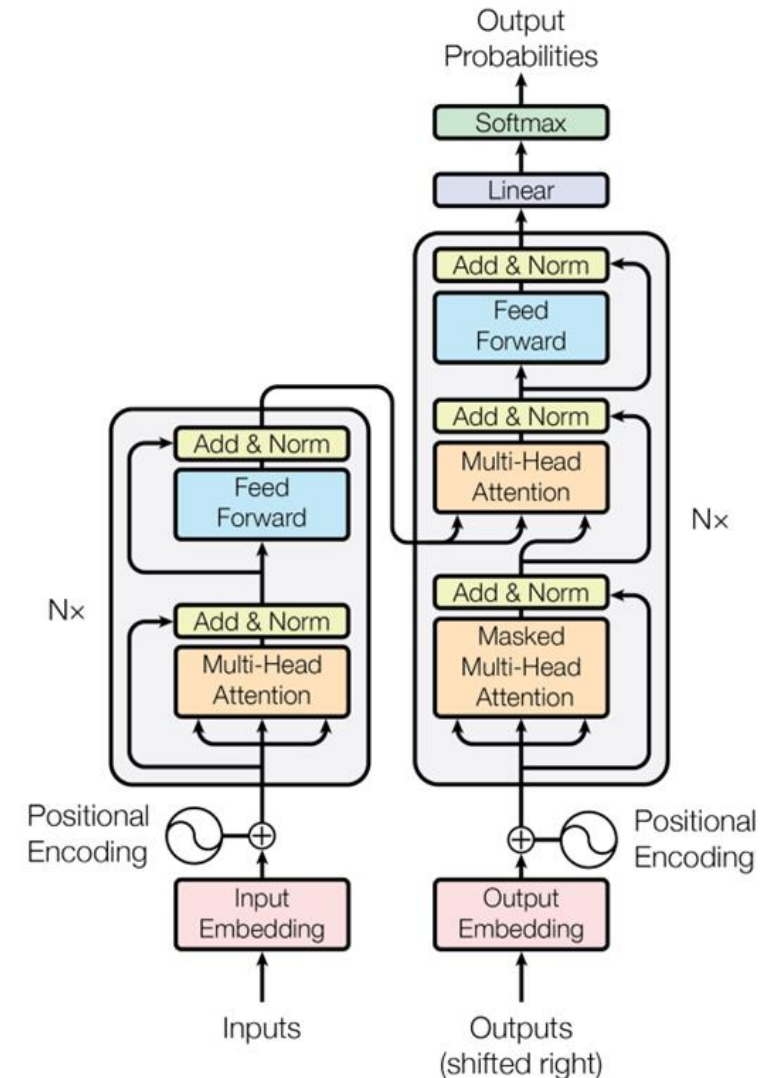# Investigated NLP Tools

- ChatGPT and GPT-4

  - Ongoing research into these models as zero and one-shot learners

  - Privacy concerns outside of healthcare center's firewall

  - Investigation is ongoing

Pre-Trained Language Models (PLM's)

- Large-scale, unlabeled corpora relatively easy to access and construct
  - Wikipedia, Brown Corpus
  - PubMed, MIMIC III

- Advantages:
  - Universal language representation
  - Better generalization performance
  - Opportunity for transfer learning and fine-tuning

- Transformer models
  - Bi-directional Encoder Representations from Transformers (BERT)

Vaswani et. al. Attention is All You Need, *Advances in Neural Information Processing Systems (NIPS)*, 5998-6008, 2017.

CDRC Goal

- Gather RWD (EHR, case reports)

- **Utilize open-source, widely-accessible tools** to process information through extraction pipelines from medical documents (such as the EHR) into a structured format for downstream analytical purposes

- Analyze repurposed drugs with structured data

Open-source NLP Libraries: HuggingFace

Open-source NLP Libraries: spaCy and ScispaCy

https://spacy.io/

https://github.com/allenai/scispacy

NLP Implementation in AWS
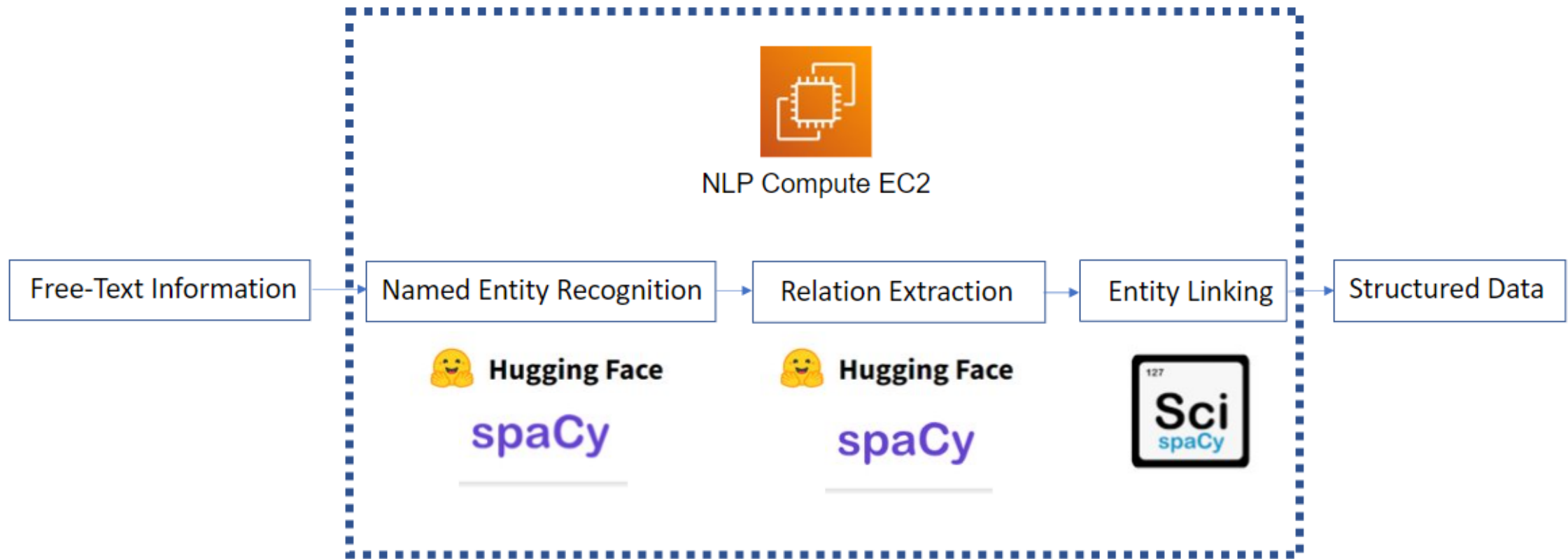
Methods

- Annotate data: Label Studio (labelstud.io)
  - Entity types: Anatomy, medical condition, medication, TTP
  - Relations: Medication: dosage, duration, frequency; condition duration; adverse event of treatment; test value and unit
- Utilize labeled data for supervised fine-tuning using spaCy pipeline and a BERT model architecture from HuggingFace
  - Bio-Clinical-BERT [1]
  - 3588 "documents"
  - 860 relations
  - 70/30 train/test split
- Evaluate performance metrics: Precision, recall, F1-score

https://medium.com/data-reply-it-datatech/training-and-deployment-using-amazon-sagemaker-of-a-custom-spacy-ner-model-by-integrating-e355c8e90240
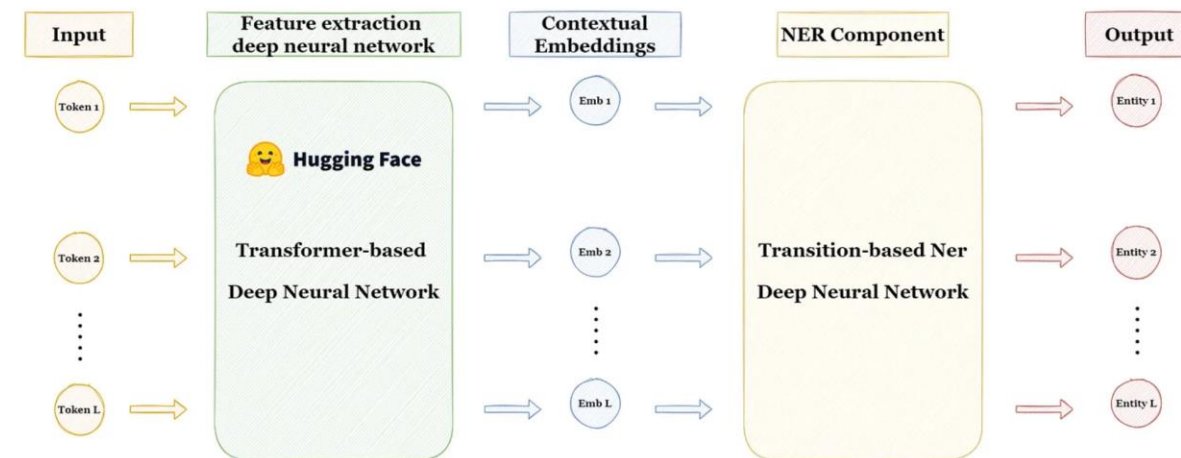
[1] Bio_ClinicalBERT at main. https://huggingface.co/emilyalsentzer/Bio_ClinicalBERT/tree/main.

Methods: Pipeline Architecture

Adverse drug event: Intolerable gastrointestinal adverse effects from lopinavir/ritonavir

Named Entity Recognition:

DX_NAME 2   GENERIC_NAME y

Intolerable gastrointestinal adverse effects from lopinavir/ritonavir

Relation Extraction:

DX_NAME 2   adverse_treatment_effect   GENERIC_NAME y

Intolerable gastrointestinal adverse effects from lopinavir/ritonavir

Entity Linking:

DX_NAME 2   adverse_treatment_effect   GENERIC_NAME y

Intolerable gastrointestinal adverse effects from lopinavir/ritonavir

ICD-10 = T47.8X5A   RxNorm = 847745

Results
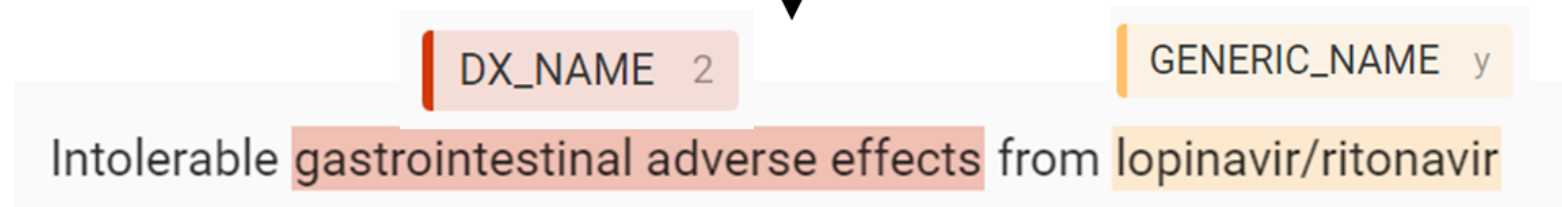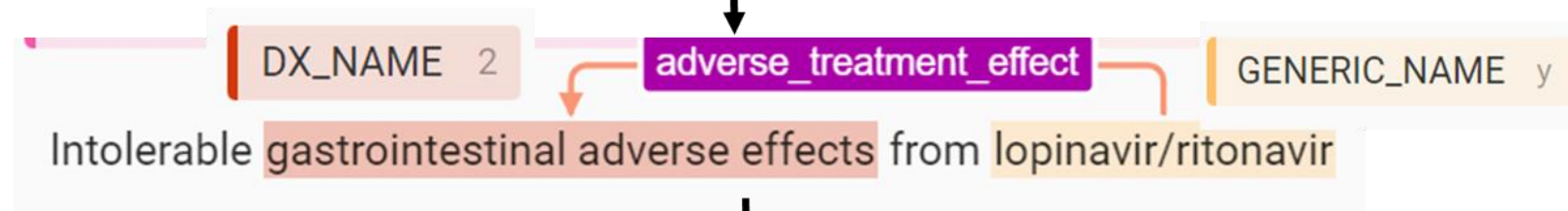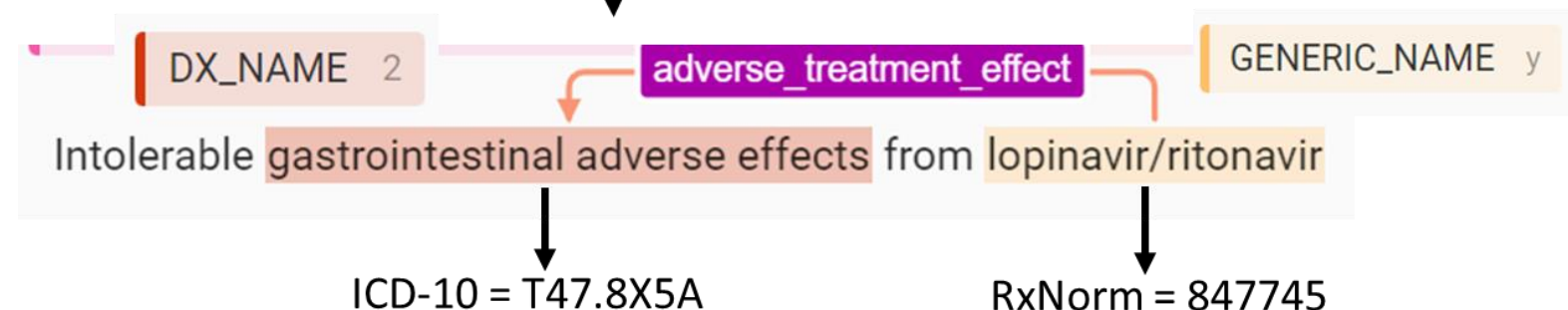
Performance Metrics for NER and Relation Extraction*

	Precision	Recall	F1-Score
NER	91	89.1	90
Relation Extraction	96.74	61.59	75.26

*Using the Bio-Clinical-BERT Transformer model

Implementation Directly to the EHR
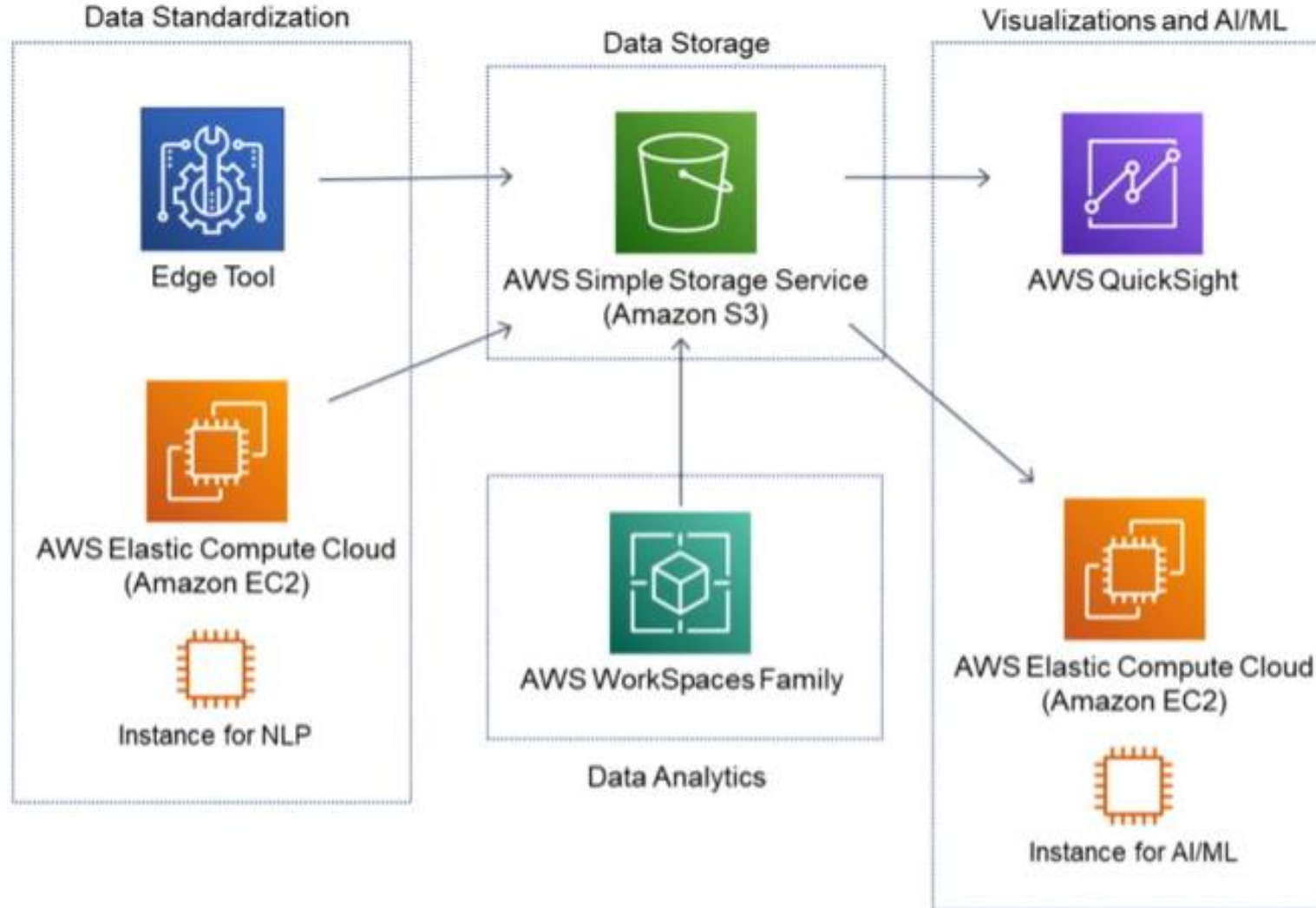
Future Steps

- Test the architecture with data from all diseases

- Perform predictive analytics

- Expand implementation to the EHR
  - Combine with Edge Tool

- Experiment with HIPAA-compliant ChatGPT/GPT-4 as a tool for clinical information extraction from unstructured data

Methodology for Validating a Minimal Dataset

Kerry Howard

Dr. Kerry Howard

Research Manager for the Center for Public Health Modeling and Response, Clemson University

Dr. Kerry Howard (she/her) is the Research Manager for the Center for Public Health Modeling and Response. She has served as a research assistant in the fields of public health and psychology since 2014, accumulating extensive experience with research design, data maintenance and analysis, and scientific writing, as well as an understanding of how these areas can be applied to improve real-world outcomes and disseminate findings to the community.

Utilizing EHR Data

- Importance of clinical research and data-driven approaches

- Focus on EHR data as a promising source of systematic and collaborative work in drug repurposing

- Avoid data quality issues, including unreliable data sources and inefficient data sharing

- Enhance utility and ensure reproducibility

LASSO Regression: Overview

- Least Absolute Shrinkage and Selection Operator (LASSO)

- Factor selection tool
  - Variables with the least influence or redundancy are excluded
  - Process the dataset and ensure workability

- Accurate identification of prognostic factors for COVID-19 outcomes

Variables	Estimate
Age	1.4
Years of Education	.
Socioeconomic Status	2.5

LASSO Regression: Methodological Benefits

- Avoid data quality issues, enhance utility, and ensure reproducibility

- Plausible problems:
  - Large number of potential predictors
  - No a-priori knowledge to inform variable use

- Necessity of enhancing research and data processes with an effective, minimal dataset

Laboratory Values
Serum creatinine
Leukocyte count
Total bilirubin
ALT
AST
Respiratory rate
Oxygen saturation

LASSO Regression: Methodological Benefits

- K-fold cross-validation
  - Splits the data into training subsamples and testing subsamples
  - Allows for internal validation
  - Reduces overfitting

- Result:
  - A validated model consisting of selected variables for best prediction of the outcome without being appropriate only for the dataset
  - Validate a minimal dataset with confidence in the incorporated variables and which can be used with additional data

LASSO Regression: Preliminary Insight

- Outcome: 28-day all-cause mortality

- Predictors:

Measures of Central Tendency
Minimum
Maximum
Median
Mean

Laboratory Values
Serum creatinine
Leukocyte count
Total bilirubin
ALT
AST
Respiratory rate
Oxygen saturation

LASSO Regression: Preliminary Insight

Laboratory Variable	Most Predictive Measure
Serum creatinine	Minimum, mean was excluded
Leukocyte count	Mean
Total bilirubin	Minimum, mean was negligible
ALT	Minimum, mean was excluded
AST	Minimum
Respiratory rate	Mean
Oxygen saturation	Mean

Implications

- Aids in variable selection and verification

- Provides evidence for the utility of the dataset with validation of its use, the selected variables, and ability to incorporate more data as it becomes available

- Provides departure point for relevant prognostic variables in future research and streamlines the data's utility for real-world applications

Exploring the potential of causal inference modeling in CURE ID to replicate clinical trial findings

Ruth Kurtycz

Ruth Kurtycz, PhD

Public Health Informatics Fellow, CDC

Ruth Kurtycz is a Public Health Information Fellow at CDC. Ruth is leading the statistical analysis to identify promising treatments for diseases with inadequate therapy from the data in CURE ID and contributing to the publications of treatments that may merit further study, as well as inform planned clinical trials.
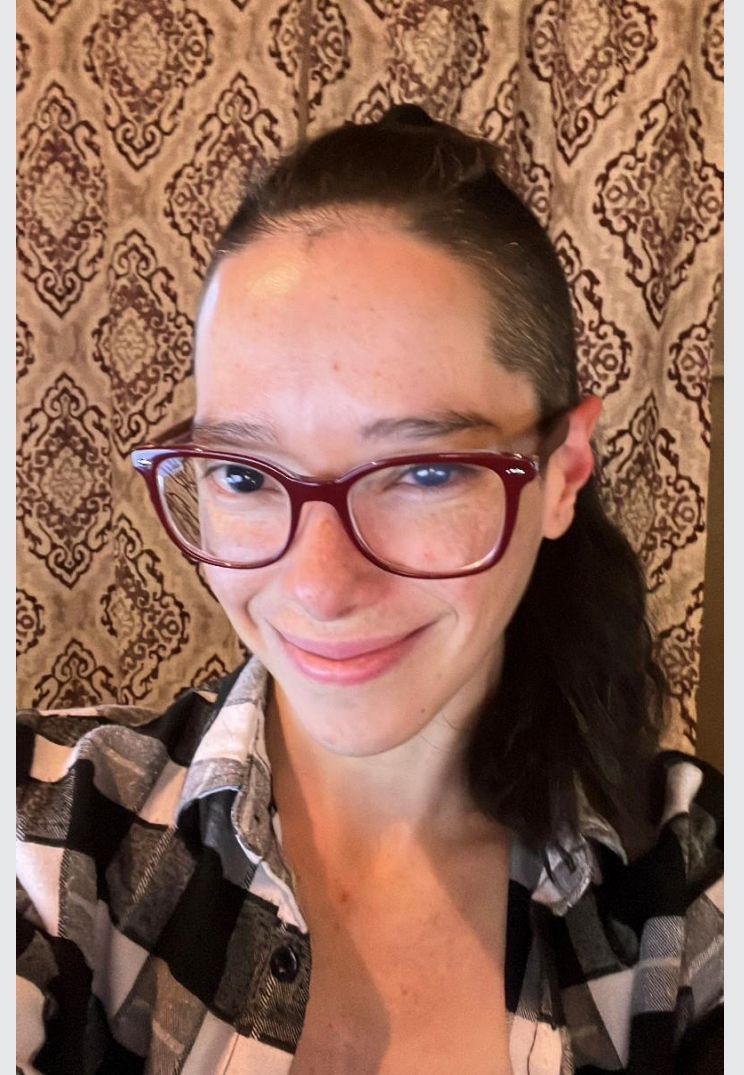
What is Causal Inference?

- **Causal inference** is the process of inferring causes from data.
  - It seeks to determine the effect of a particular phenomenon in the context of a larger system and analyze the response when the effect variable is changed

- Randomized controlled trial (RCT) vs causal inference methods
  - Observational data
  - Widely used
  - Strong assumptions
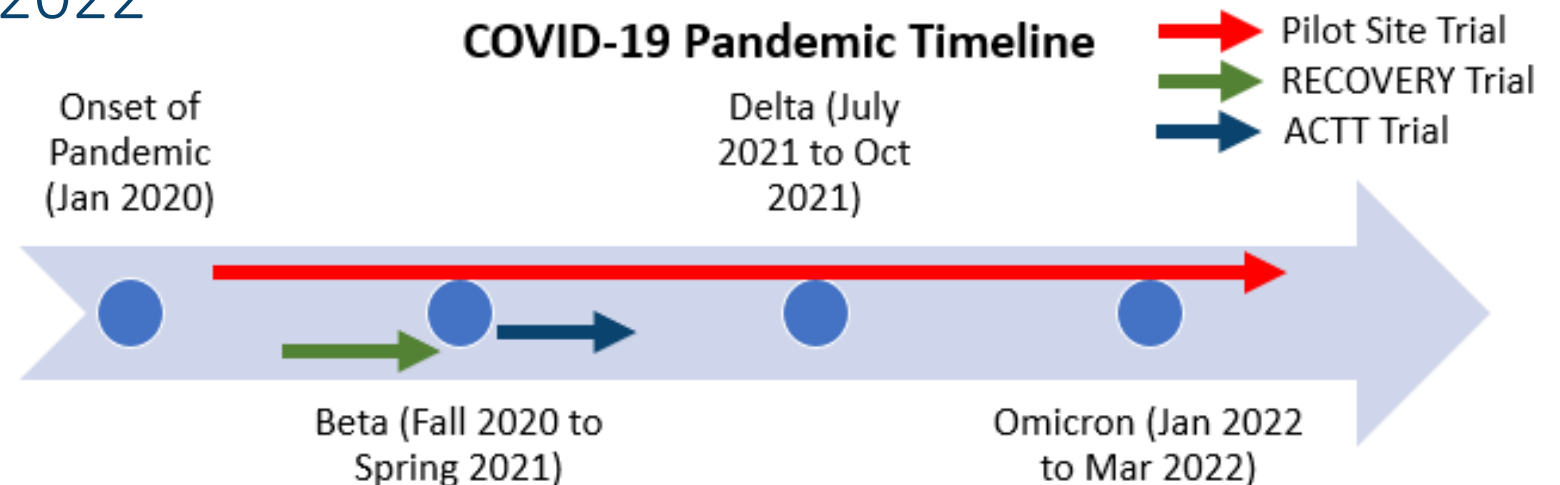
Causal Inference Methodology

- **Propensity Score Matching** (PSM) is a causal inference method that attempts to estimate the effect of a treatment by accounting for the covariates within the data that predict an individual's likelihood of receiving the treatment

- PSM Steps
  - Build logistic regression model
  - Generate individual scores
  - Match on scores

CURE ID Causal Inference Methodology

- Goal: Generation of real-world data (RWD) with the Edge Tool, using COVID-19 as a use case

- Demonstrate the initial utility of Edge Tool by attempting to replicate findings of clinical trials in COVID-19
  - Methodology: PSM in conjunction with clinical trial data analysis (logistic regression, survival analysis, etc.)

CURE ID Causal Inference Methodology

- Evaluate key agents: dexamethasone, remdesivir, ivermectin
    - Dexamethasone – RECOVERY[1] trial (July 2020)
    - Remdesivir – ACTT[2] trial (November 2020)

- Pilot site data
    - March 2020–March 2022
    - N = 12,129

COVID-19 Pandemic Timeline

→ Pilot Site Trial
→ RECOVERY Trial
→ ACTT Trial

Onset of Pandemic (Jan 2020)

Delta (July 2021 to Oct 2021)

Beta (Fall 2020 to Spring 2021)

Omicron (Jan 2022 to Mar 2022)

Preliminary RECOVERY Replication Findings

- RECOVERY Trial
  - N = 6,424 (2,104 treatment, 4,321 control)
  - Methods: Logistic regression, survival analysis
  - Outcomes: 28-day mortality, time to discharge

- Edge Tool Replication
  - N = 11,973 (3,911 treatment, 7,982 controls)
  - PSM covariates: age, race, sex, level of oxygen support, diabetes, cardiovascular disease, chronic kidney disease, chronic lung disease, HIV

Preliminary RECOVERY Replication Findings

Results	Overall	Oxygen Support	
RECOVERY	**Lower** 28-day mortality in treatment group	**Invasive Mechanical Ventilation: Lower**	
		Oxygen Alone: Lower	
		No Oxygen: Not significantly different	
Edge Tool Replication	**Higher** 28-day mortality in treatment group	**Invasive Mechanical Ventilation vs No Oxygen: Lower**	**Oxygen Alone vs No Oxygen: Lower**

Limitations of Causal Inference

- Association is not causation

- Randomization

- Unmeasured covariates

- Missing data problem

- Effects of Spillover/Interference

- Rule of Large Numbers

Additional Limitations of Trial Replication

- Limited covariates for match
- Pilot site data only
  - Matching with replication of controls
- Timing limitations
- Context of treatment administration

Implications and Future Directions

- Edge tool data has potential to be used to support and supplement the generation of RWD for treatment of emerging disease such as COVID-19

- Plan to refresh existing replication analysis with the addition of more sites' data
    - Refining methods to include dose/timing

- Explore replication with other treatments and diseases

Dr. Ewy Mathé

Director of Informatics in the Division of Preclinical Innovation, NCATS

Dr. Ewy Mathé is the Director of Informatics in the Division of Preclinical Innovation at NCATS. She leads a diverse team of experts in bioinformatics, cheminformatics, data science and software development that empower translational scientists to make meaningful data-driven decisions in their research.

Dr. Raghave Tirupathi

Medical Director, Keystone infectious Diseases/HIV; Keystone community health services a Chambersburg PA

Dr. Raghave Tirupathi is the Medical Director of Keystone Infectious Diseases/HIV and Keystone Community Health Services. He also serves as the chair of the infection prevention committees of WellSpan Chambersburg and Waynesboro Hospital. Dr. Tirupathi has a teaching appointment at Penn State College of Medicine, as clinical assistant professor of medicine and also serves as a clinical consultant for Cure ID/CDRC, a FDA and NIH joint initiative.

COVID-19 in the Real World: Progress Report on the Edge Tool Suite

Smith Heavner

Smith Heavner, PhD, RN

Senior Scientific Director of the CURE Drug Repurposing Collaboratory

Dr. "Smitty" Heavner serves as Senior Scientific Director of the CURE Drug Repurposing Collaboratory. He earned his PhD in Applied Health Research and Evaluation from Clemson University in 2021 and completed a one-year post-graduate training in clinical trial design and analysis at Harvard Medical School in 2023.

Real-World Data

- Explore clinical practice
- Develop and refine hypotheses
- Provide external controls
- Observational research
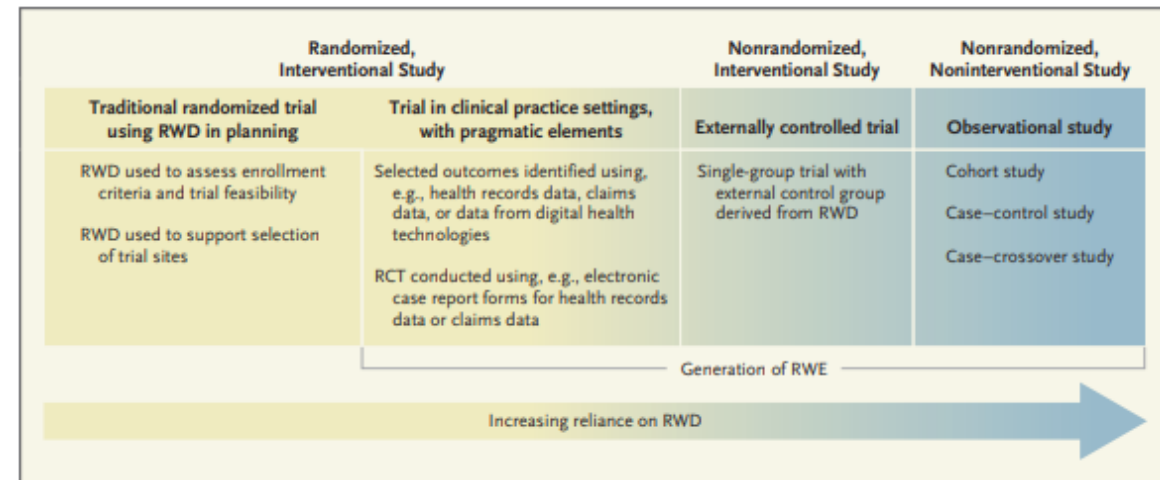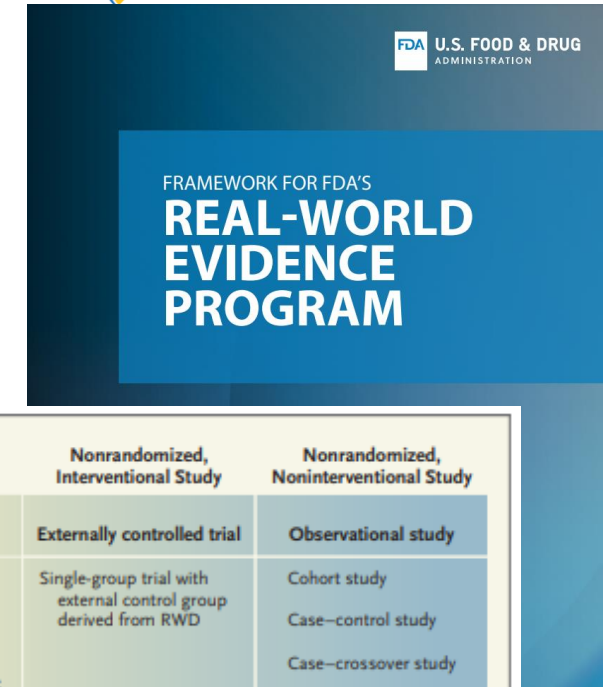
Reliance on RWD in Representative Types of Study Design.
RCT denotes randomized, controlled trial; RWD real-world data; and RWE real-world evidence.

Concato J, Corrigan-Curay J. Real-World Evidence — Where Are We Now? N Engl J Med. 2022;386(18):1680-1682. doi:10.1056/NEJMp2200089

Real-World Data: Challenges

Data harmonization

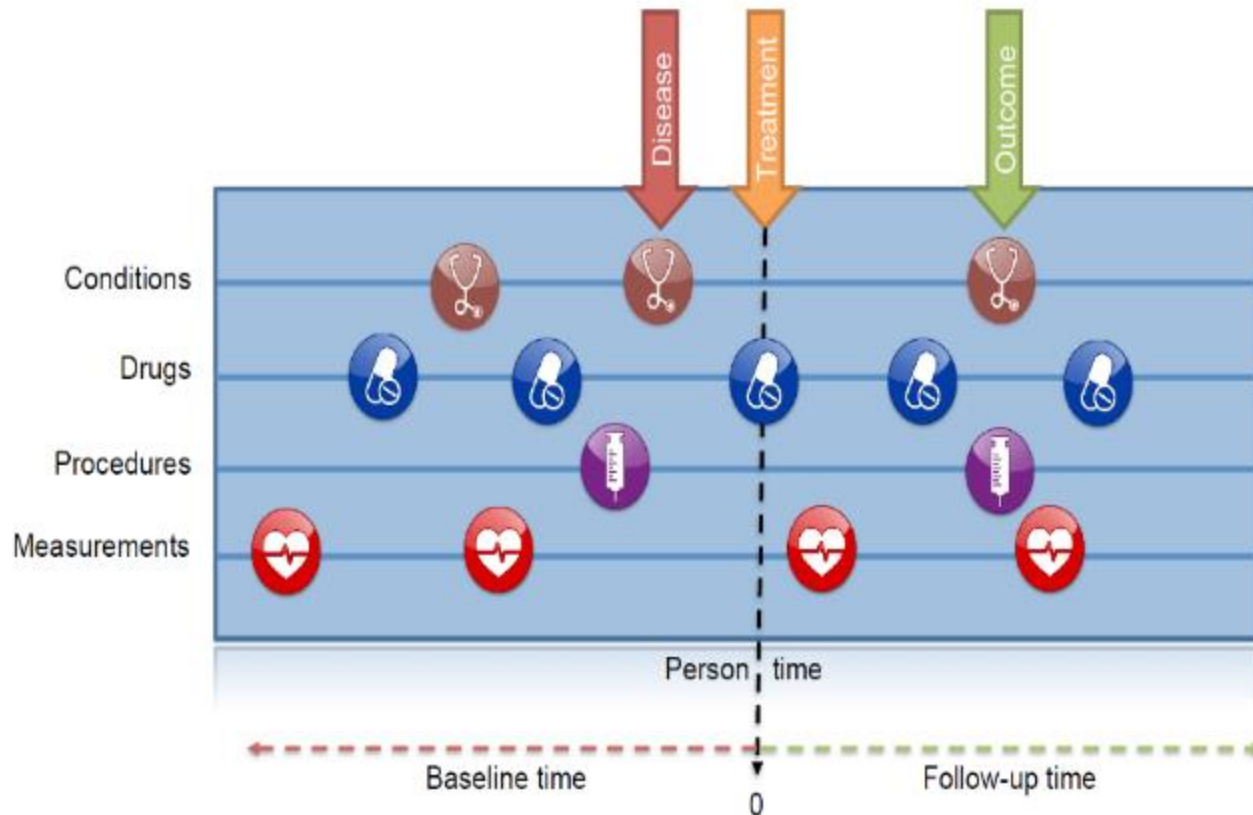Unstructured data

Privacy concerns

Probabilistic linkage

Defining outcomes

OMOP

153 Controlled Vocabularies
9 Million concepts

Real-World Data: Data Harmonization

- De-centralized harmonization
- Sites not required to have common data model experience
- International
- Disease agnostic

National COVID Cohort Collaborative (N3C)

The N3C is a partnership among the NCATS-supported **Clinical and Translational Science Awards (CTSA) Program** hubs, the **National Center for Data to Health (CD2H)**, and NIGMS-supported **Institutional Development Award Networks for Clinical and Translational Research (IDeA-CTR)**, with overall stewardship by NCATS. Collaborators will contribute and use COVID-19 clinical data to answer critical research questions to address the pandemic.

Scientists Use N3C Data to Identify Common Features of Long COVID

NIH-supported researchers used electronic health record data from the National COVID Cohort Collaborative (N3C) Data Enclave to identify people with long COVID and those likely to have it.
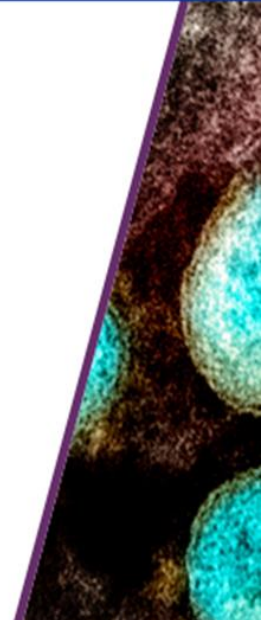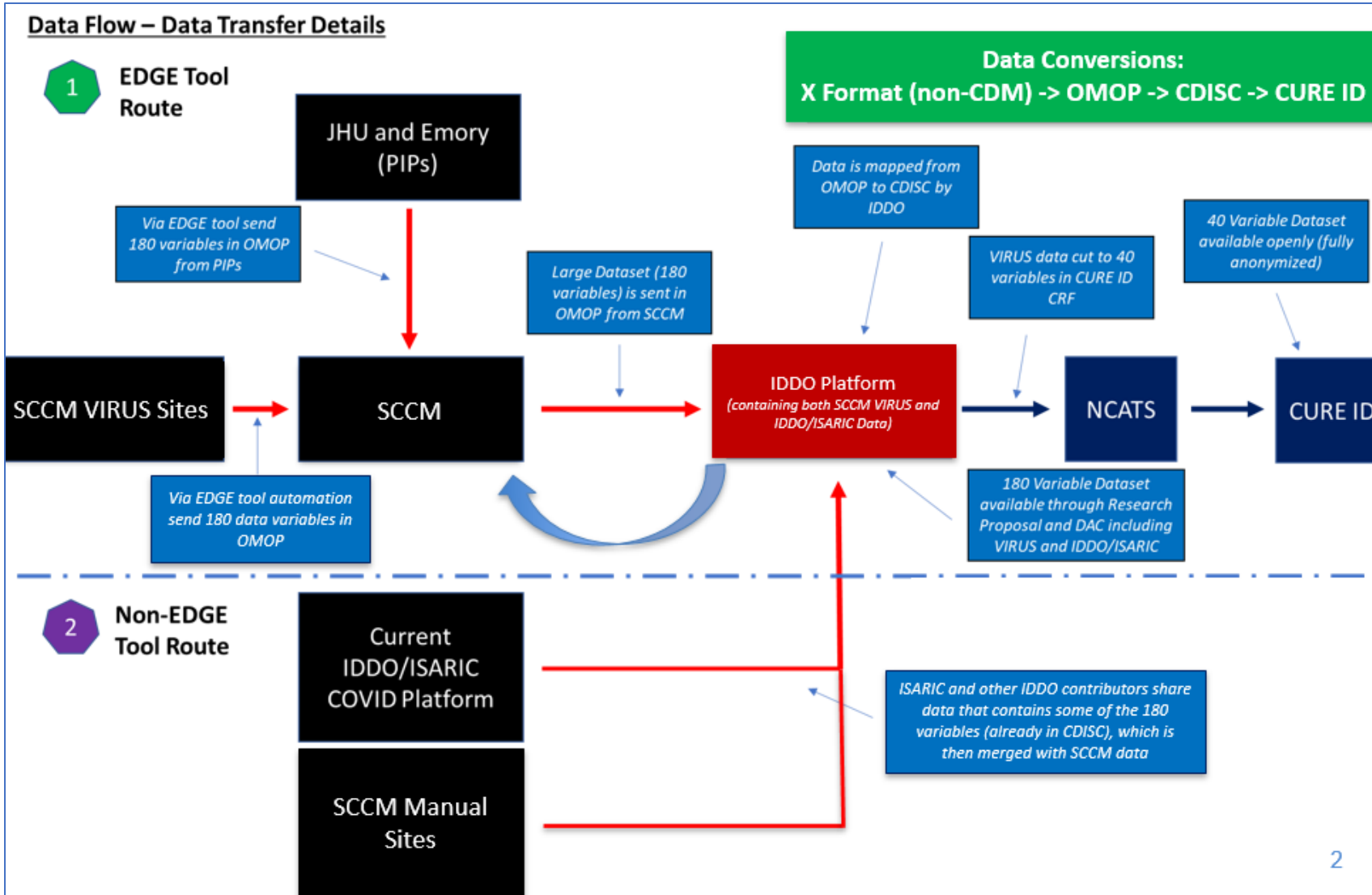
COVID-19 as Use Case

- Drug Repurposing
- Broad Impact
- Identifiable Cases
- Discreet Data
- Acute Disease
- Definitive Outcomes

Data Flow and Partners

Progress

- 26 sites actively engaged
- 4 sites finalizing data transfer
  - Data quality activities
  - ~40,000 patients
- 1 site complete
  - ~9,000 patients in CURE ID (dev)

LASSO Regression: Preliminary Insight

Laboratory Variable	Most Predictive Measure
Serum creatinine	Minimum, mean was excluded
Leukocyte count	Mean
Total bilirubin	Minimum, mean was negligible
ALT	Minimum, mean was excluded
AST	Minimum
Respiratory rate	Mean
Oxygen saturation	Mean

Preliminary RECOVERY Replication Findings

Results	Overall	Oxygen Support	
RECOVERY	<u>Lower</u> 28-day mortality in treatment group	**Invasive Mechanical Ventilation: <u>Lower</u>**	
		Oxygen Alone: <u>Lower</u>	
		No Oxygen: Not significantly different	
Edge Tool Replication	**Higher** 28-day mortality in treatment group	**Invasive Mechanical Ventilation vs No Oxygen: <u>Lower</u>**	**Oxygen Alone vs No Oxygen: <u>Lower</u>**

Progress

- 3 Webinars

- 4 Conference presentations

- 1 publication

Heavner, et al. A Path to Real-World Evidence in Critical Care Using Open-Source Data Harmonization Tools. Critical Care Explorations 5(4):p e0893, April 2023.

A Path to Real-World Evidence in Critical Care Using Open-Source Data Harmonization Tools

ABSTRACT: COVID-19 highlighted the need for use of real-world data (RWD) in critical care as a near real-time resource for clinical, research, and policy efforts. Analysis of RWD is gaining momentum and can generate important evidence for policy makers and regulators. Extracting high quality RWD from electronic health records (EHRs) requires sophisticated infrastructure and dedicated resources. We sought to customize freely available public tools, supporting all phases of data harmonization, from data quality assessments to de-identification procedures, and generation of robust, data science ready RWD from EHRs. These data are made available to clinicians and researchers through CURE ID, a free platform which facilitates access to case reports of challenging clinical cases and repurposed treatments hosted by the National Center for Advancing Translational Sciences/ National Institutes of Health in partnership with the Food and Drug Administration. This commentary describes the partnership, rationale, process, use case, impact in critical care, and future directions for this collaborative effort.

KEY WORDS: critical care; data harmonization; drug repurposing; electronic health record; real-world data; Observational Medical Outcomes Partnership (OMOP)

Smith F. Heavner, PhD, RN[1,2]
Wesley Anderson, PhD[3]
Rahul Kashyap, MBBS, MBA[4,5]
Pamela Dasher, BA[1]
Ewy A. Mathé, PhD[6]
Laura Merson[7]
Philippe J. Guerin, MD, PhD[8,9]
Jeff Weaver, MBA[10]
Matthew Robinson, MD[11]
Marco Schito, PhD[1]
Vishakha K. Kumar, MD, MBA[12]
Paul Nagy, PhD[13]

Coming Soon...

- Webinar:
  - OHDSI community call

- Manuscripts:
  - Lessons learned: standardizing ventilator data
  - COVID-19 in immunocompromised patients
  - Immunomodulation in Hospitalized Patients with COVID-19 Infection
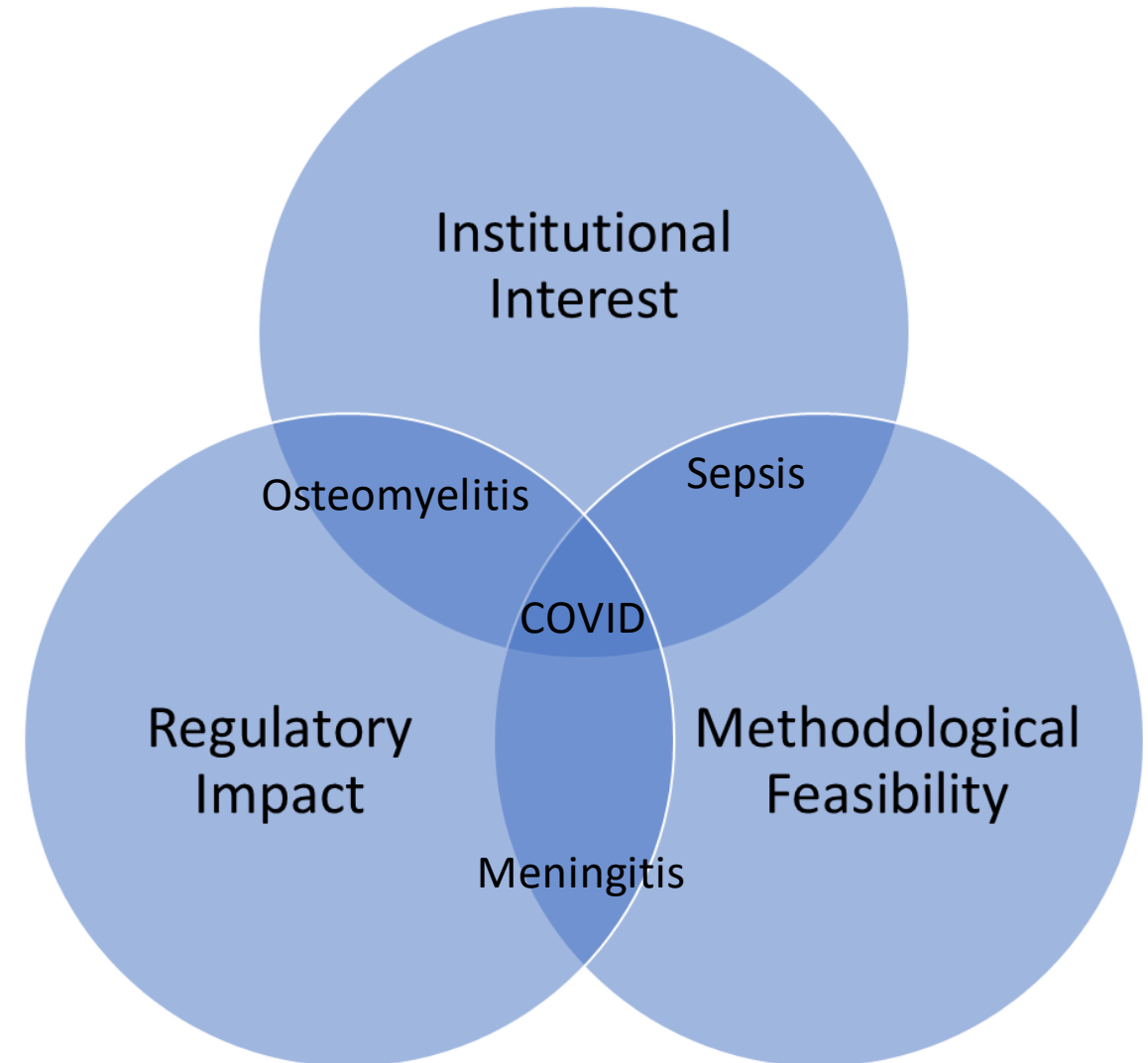
RWD: Towards Sustainability

What motivates institutions to build infrastructure to extract and share RWD?

https://episframework.com/
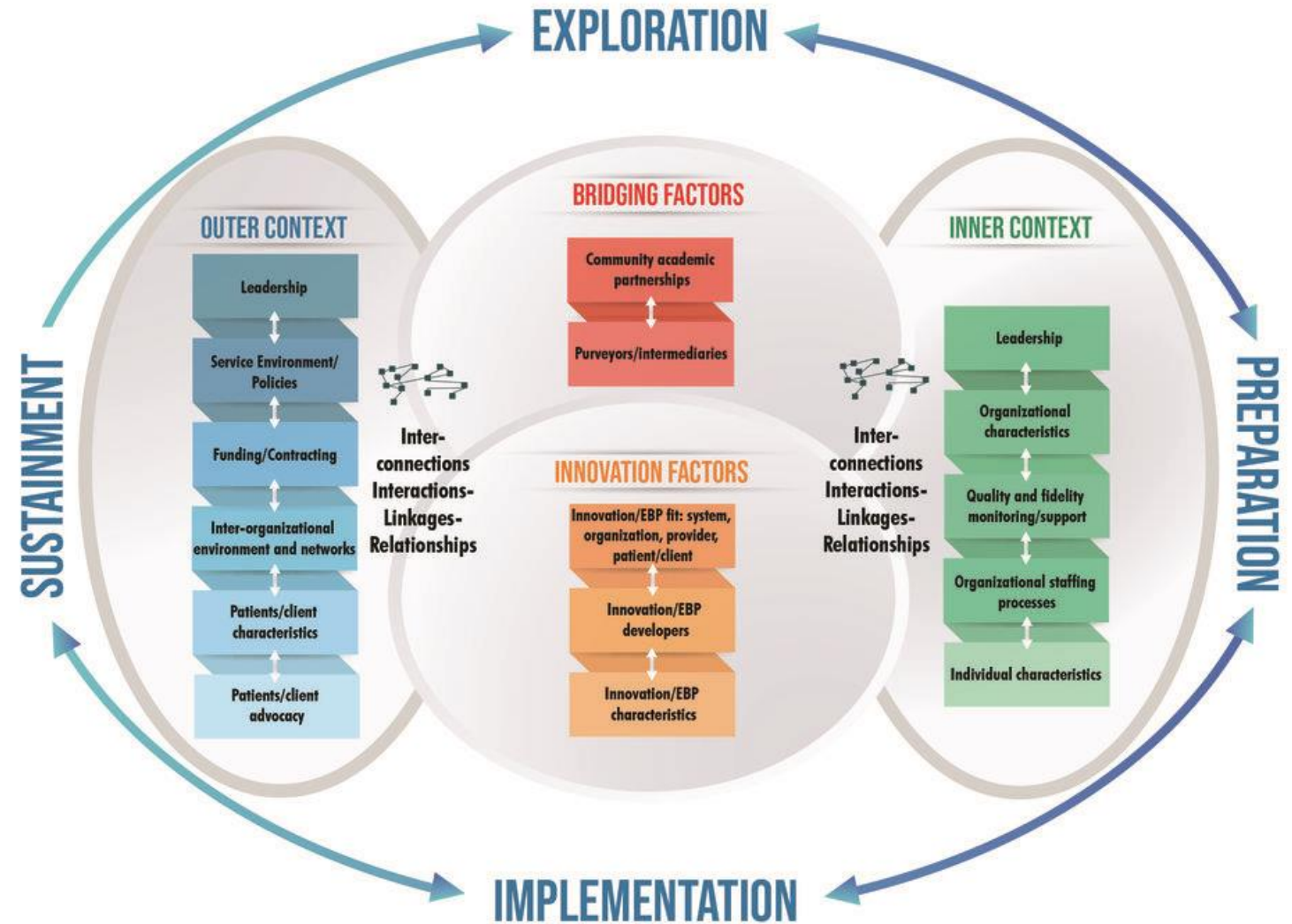
RWD: Beyond COVID

- Sepsis
  - Isolating organisms
  - Case definition
  - Sites highly motivated

- Meningitis
  - Isolating organisms
  - Rarer cases of more interest

- Osteomyelitis
  - Isolating organisms
  - Linking encounters
  - Lost to follow up

Roundtable

Discussion of results of Edge tool extraction and analysis

Co-Chairs: Ewy Mathe and Raghav Tirupathi

- **Laura Evans** (SCCM)

- **Nathalie Strub-Wourgaft** (DNDi)

- **Matt Robinson** (JHU)

- **Anup Challa** (AstraZeneca)

Dr. Ewy Mathé

Director of Informatics in the Division of Preclinical Innovation, NCATS

Dr. Ewy Mathé is the Director of Informatics in the Division of Preclinical Innovation at NCATS. She leads a diverse team of experts in bioinformatics, cheminformatics, data science and software development that empower translational scientists to make meaningful data-driven decisions in their research.

Dr. Raghave Tirupathi

Medical Director, Keystone infectious Diseases/HIV; Keystone community health services a Chambersburg PA

Dr. Raghave Tirupathi is the Medical Director of Keystone Infectious Diseases/HIV and Keystone Community Health Services. He also serves as the chair of the infection prevention committees of WellSpan Chambersburg and Waynesboro Hospital. Dr. Tirupathi has a teaching appointment at Penn State College of Medicine, as clinical assistant professor of medicine and also serves as a clinical consultant for Cure ID/CDRC, a FDA and NIH joint initiative.

Dr. Laura Evans

Professor and Medical Director of Critical Care, University of Washington

Dr. Laura Evans is a professor of medicine at the University of Washington and the Medical Director of Critical Care at the University of Washington Medical Center. Her interests focus on sepsis, severe acute respiratory infection as well as preparedness for high consequence infectious diseases.
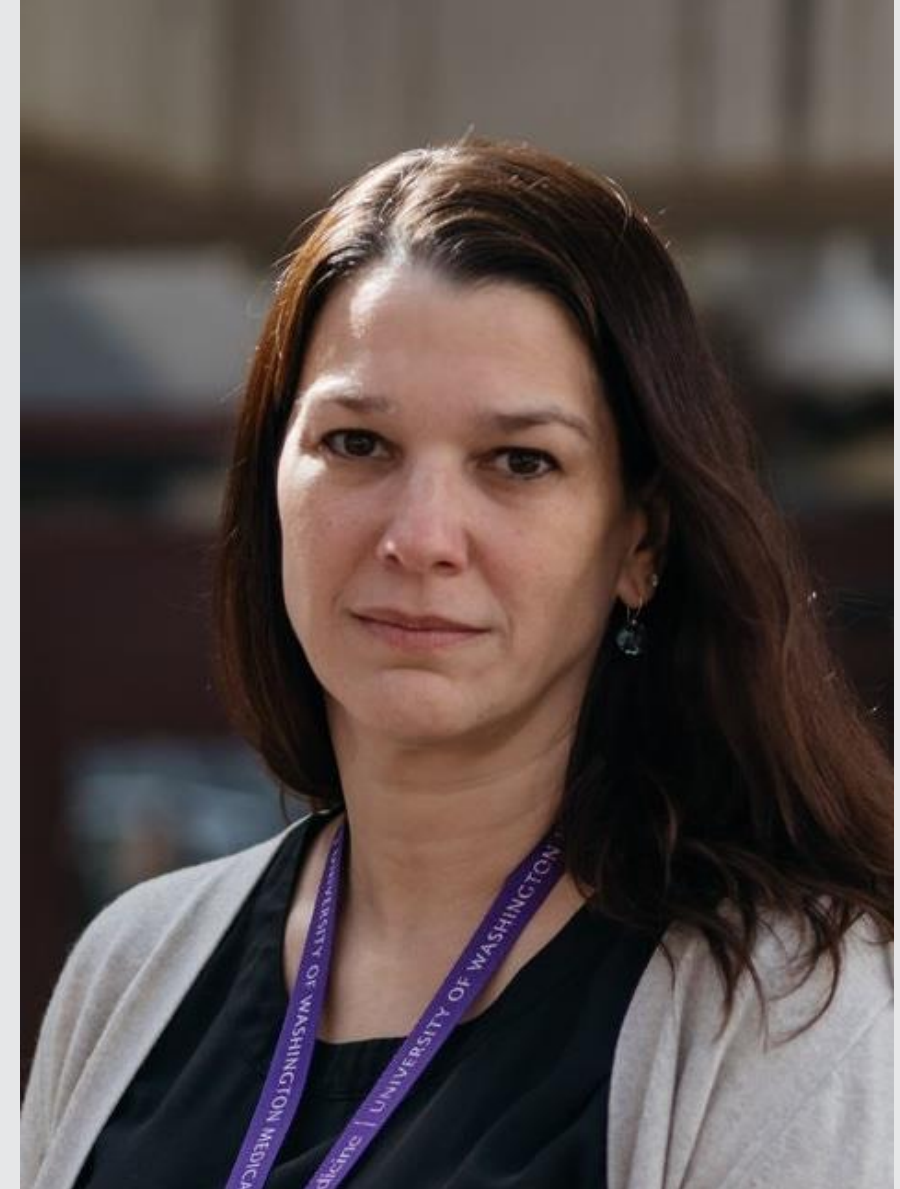
Dr. Nathalie Strub-Wourgaft

DNDi

Dr. Nathalie Strub-Wourgaft joined Drugs for Neglected Diseases Initiative (DNDi) in 2009 and is member of the Global Executive Team as COVID 19 and preparedness Director. As Director of NTDs since 2018, Dr. Strub-Wourgaft provided strategic and technical oversight to a wide portfolio of R&D and access plans for therapeutic areas covering Sleeping Sickness, Chagas disease, Cutaneous and Visceral Leishmaniasis, Filaria and Mycetoma.

Matthew Robinson, MD

Assistant Professor in the Division of Infectious Diseases, Johns Hopkins University School of Medicine

Matthew Robinson, MD, is Assistant Professor in the Division of Infectious Diseases at Johns Hopkins University School of Medicine. He is interested in leveraging diagnostic innovation and precision medicine to reduce diagnostic and prognostic uncertainty for infectious diseases.

Anup Challa

Associate Director, AstraZeneca

Anup Challa is an Associate Director at AstraZeneca, supporting global pediatric vaccines and immunotherapies development as a pharmacovigilance scientist. He is a content expert in the management and evaluation of RWE, with particular interest in its use for benefit/risk contextualization of pharmaceuticals and associated regulatory interactions.

Wrap up and next steps

Agenda

CDRC Annual Meeting and Workshop Day 2: Wednesday April 19 Afternoon

Challenges and opportunities for confirming hypothesis generated by real-world data for repurposed drugs through randomized clinical trials.

Pragmatic adaptive platform trials

Cochairs: Marco Schito and David Simon

1. Welcome and Overview: **Marco Schito** (C-Path) and Heather Stone (FDA)   1:00-1:15
2. Keynote: **David Fajgenbaum** (Every Cure and CDRC AC Co-Chair)   1:15-1:45

Cochairs: Marco Schito (C-Path) and Leonard Sacks (FDA)

Undiagnosed Diseases Network (UDN): Discovering Rare Disease Therapies Through Team Science   1:45-2:00

Moderators: David Fajgenbaum (U Penn) and Heather Stone (FDA)

1. Critical considerations for clinical trials in DR   2:00-2:15
2. Roundtable Discussions on decentralized trials and embedding trials in clinical practice.
   - Inpatient trials and innovations in embedding trials in practice   2:15-3:15

- Panelists:
  - **Trevan Locke** (Duke Margolis)
  - **Jon Sevransky** (Emory)
  - **Stacey Coe** (C-Path)
  - **Clare Thibodeaux** (Cures within Reach)
  - **Cynthia Adinig** (Patient)

Break   3:15-3:30

Agenda

CDRC Annual Meeting and Workshop Day 2: Wednesday April 19 Afternoon

Moderators: David Fajgenbaum (U Penn) and Heather Stone (FDA)

1. Outpatient/decentralized/patient-centric trials   3:30-4:30
   - Chris Lindsell (Duke)
   - Nathalie Strub -Wourgraft (DNDi)
   - Suanna Buinooge (ASCO)
   - Oved Amitay (Solve CFS)
   - Amy Morris (Emory)
   - Vidula Sukhatme (GlobalCures)
   - Michael Sieverts (Long-COVID Patient)
   - Ingrid Oakley-Girvan (Medable)

2. Wrap up and next steps   4:30-4:45
3. Evening: Reception   6:00-8:00
   - Keynote: Patient advocate
   - Hors d'oeuvres
   - Open bar
   - Taco stand

Leonard Sacks, MD

FDA

Leonard Sacks, MD, joined the FDA in 1998 as medical reviewer in the Office of New Drugs. Subsequent positions included acting director of the Office of Critical Path Programs and associate director for clinical methodology in the Office of Medical Policy in the Center for Drug Evaluation and Research. In this capacity he has led efforts to support novel approaches to clinical trials including the use of electronic technology.

Marco Schito, PhD

Executive Director, CDRC

Dr. Marco Schito is Executive Director of C-Path's CURE Drug Repurposing Collaboratory, the Scientific Director for the Inflammatory Bowel Disease Group, and an Adjunct Professor at the University of Arizona, James E. Rogers College of Law. His work aims to discover potentially safe and effective repurposed therapies for diseases with high unmet medical need by capturing and sharing global, real-world clinical data.

Keynote

David Fajgenbaum

Every Cure and CDRC AC Co-Chair

David Fajgenbaum, MD, MBA, MSc

University of Pennsylvania, CDCN, Every Cure

Dr. David Fajgenbaum is a physician-scientist at the University of Pennsylvania, co-Founder & President of the Castleman Disease Collaborative Network, co-founder of Every Cure, and national bestselling author of 'Chasing My Cure: A Doctor's Race to Turn Hope Into Action'.

F. Sessions Cole, MD

Professor, Washington University School of Medicine

Dr. F. Sessions Cole is a neonatologist and the Park J. White, M.D., Professor of Pediatrics at Washington University School of Medicine. His longstanding engagement in the undiagnosed and rare disease space has provided him with opportunities to see the impact of leveraging team science to end diagnostic odysseys and the importance of discovering therapeutic strategies for rare disease patients, families, biomedical discovery, and society.
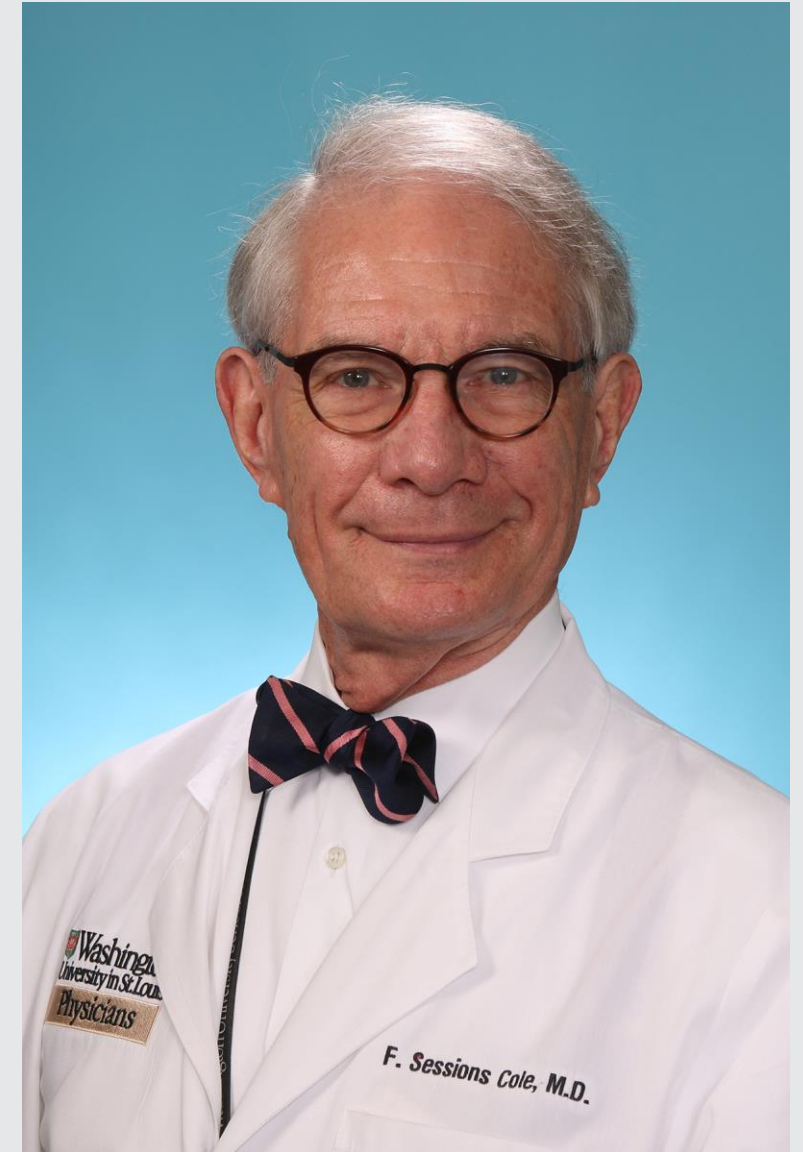
Andrew Crouse, PhD

Director of Research and Operations, UAB

Experienced in Genomic and Precision Medicine with a demonstrated history of working in the biotechnology industry. Skilled in DNA Sequencing, go-to-market Strategy, Immune System, Molecular Biology, and Biotechnology. Strong community and social services professional with a Ph.D. focused in Molecular Genetics from University of Alabama at Birmingham.

The Undiagnosed Diseases Network: Discovering Rare Disease Therapies Through Team Science

Andrew Crouse, Ph.D.

University of Alabama Hugh Kaul Precision Medicine Institute/Undiagnosed Diseases Network

F. Sessions Cole, M.D.

Washington University in St. Louis Department of Pediatrics/Undiagnosed Diseases Network

Pragmatic Adaptive Platform Trials

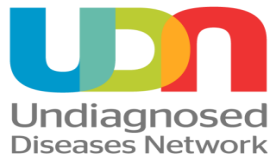Cure Drug Repurposing Collaboratory Annual Meeting

19 April 2023
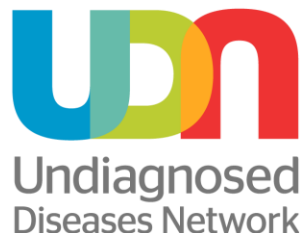
Disclosures

- Neither Dr. Crouse nor Dr. Cole has relevant conflicts of interest to disclose.

Background: The Undiagnosed Diseases Network

Goal of UDN: Solve medical mysteries through team science

- 2008 – Founded as an intramural NIH program (Undiagnosed Diseases Program) by Dr. William Gahl
- 2014 – National Undiagnosed Diseases Network funded by NIH Common Fund
- 2018 – Phase 2 of the Undiagnosed Diseases Network

- 6,376 applications received; 2,535 participants accepted; 2,165 participants evaluated; 627 participants diagnosed (who include 53 newly described conditions); 210 manuscripts published

UDN Components:

- Central Institutional Review Board (NIH)

- 3 UDN Cores: Sequencing (Baylor College of Medicine), Metabolomics Center (Mayo Clinic), Model Organism Screening Core (Baylor College of Medicine, University of Oregon, and WUSM)
- Central biorepository (Vanderbilt University School of Medicine)
- Coordinating Center (Harvard Medical School Department of Biomedical Informatics)
- 12 Clinical Sites
- Patient Engagement and Empowerment Resource (PEER)
- Undiagnosed Diseases Network Foundation (UDNF)

UDN Current Sites
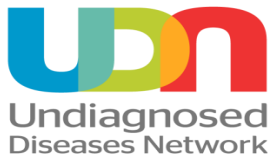
Background: UDN Therapeutic Matching Committee

- TMC Leadership Group: Matthew Might, Aleksandra Foksinska, Andrew Crouse, Hugo Bellen, David Adams, David Eckstein, Kimberly LeBlanc, Sarah Marshall, Meghan Halley, Rachel Mahoney

- UDN mission: solve medical mysteries through team science
  - Extend UDN mission to discover rare disease therapies through team science

- Therapeutic road map development process benefits patients/families, third party payers, investigators, and industry partners
  - Diagnosed UDN participants have genomic and functional characterization of mechanism of genetic disruption

Therapeutics Matching Committee Process

- UDN participants identified by TMC Leadership Group based on biology of encoded pathogenic mechanism, phenotype, and likelihood of therapeutic option
  - Identification of quantitative therapeutic metrics that can be used to evaluate effectiveness in a clinical trial
- UDN internal experts, external experts, pharmaceutical representatives, individuals with FDA experience, and patients/families review therapeutic options
  - Repurposed FDA-approved drugs, gene therapy, genome editing, ASOs, tRNA regulation, small molecules

WARNING:
I am a Biologist not a Computer Scientist

mediKanren

imagine a person that knows all science

Translator and mediKanren

IN THE LAB

NIH-funded project aims to build a 'Google' for biomedical data

By Ruth Hailu — July 31, 2019   Reprints

A view of the NIH campus
NIH

Every year, the National Institutes of Health spends billions of dollars for biomedical research, ranging from basic science investigations into cell processes to clinical

Not Google

Better…
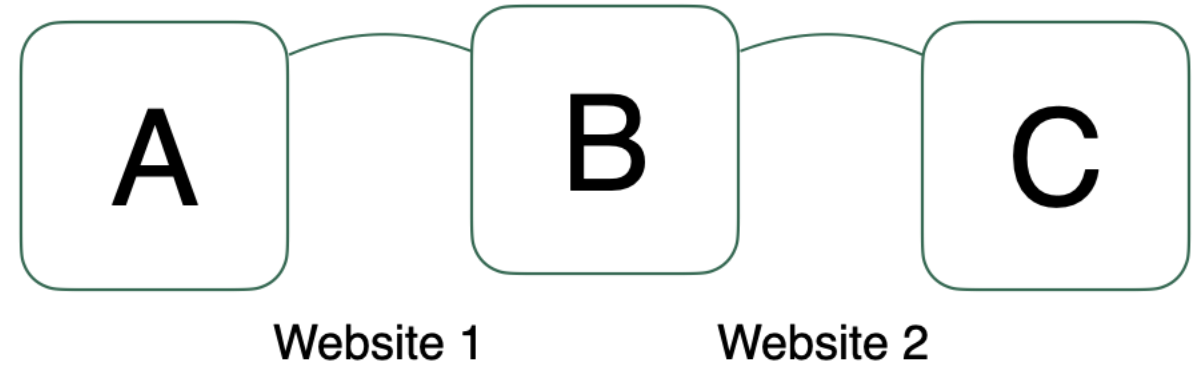
Translator and mediKanren

Search for <u>relationships</u> not just terms
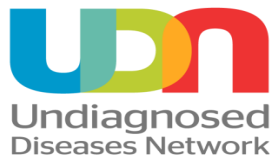
Relationship Search

Relationships can be chained together:

Disease — **TREATED BY** → Drugs — **REGULATES** → Genes — **REGULATED BY** → NEW Drugs
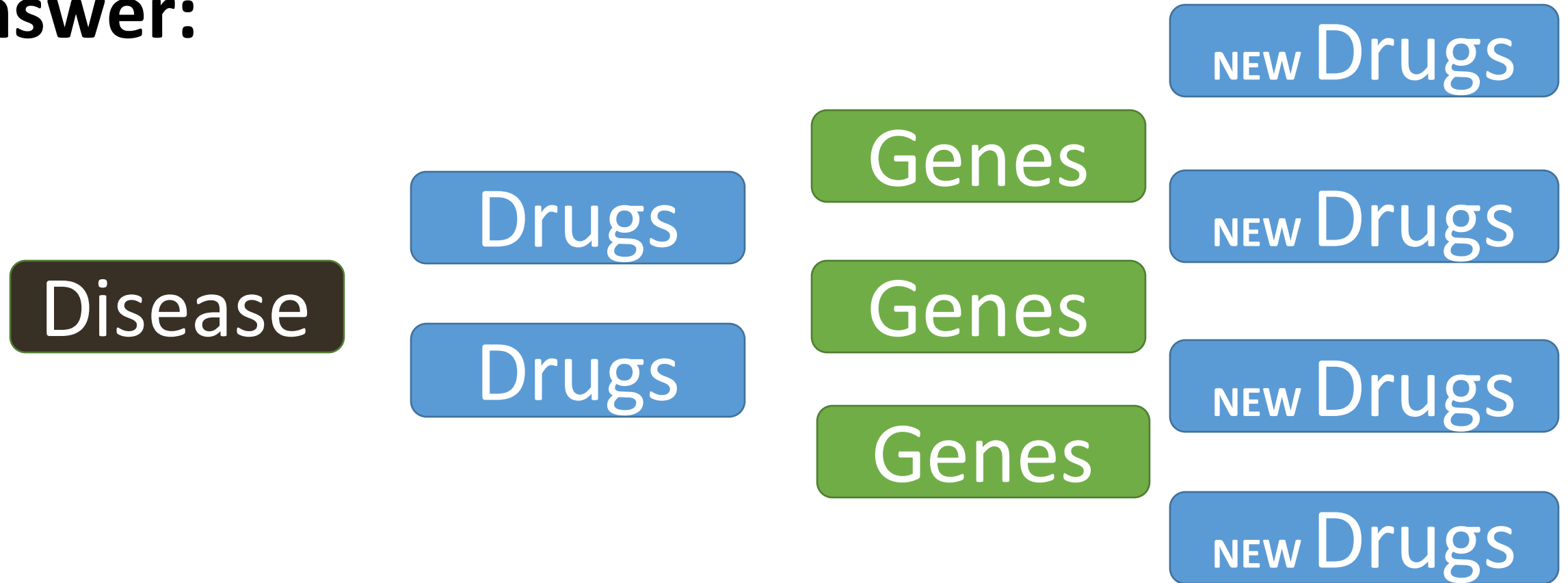
Translator and mediKanren

Success with simpler examples:

Translator and mediKanren
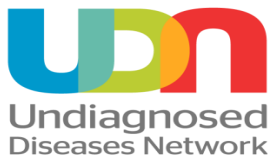
Cyclic Vomiting Syndrome

The *unknown known:*

10s of thousands of papers on nausea

Thousands of papers on treatment

Many treatments have >50 publications

"Everything"

"Everything else"

Few have 3-5 clinical trials

What about nasally inhaled isopropyl alcohol?

The *unknown known:*
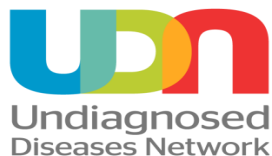
Translator and mediKanren

AI Guided Variant-centric Therapeutics

MAK8IP3

Age and Gender
- 6 year old / male

Description of symptoms
- Global DD
- Congenital hypotonia
- Autism spectrum
- Unable to walk or stand

AI Guided Variant-centric

mediKanren predicted Vitamin A as a potent up-regulator of MAPK8IP3 to harness the wild-type allele for increasing activity in MAPK8IP3

- Physician prescribed Vitamin A based on our research report

- Child began standing without support at approximately 3 months on treatment

- Child began taking 3-4 steps at a time unaided at approximately 6 months on treatment

User Stories!

Help us make this scalable for everyone!

Contact acrouse@uab.edu if you are interested in working with Translator

Discovering Rare Disease Therapies Through Team Science

Critical success factors for discovering rare disease therapies

- Engagement of patients and families
- Understanding the biology of the variant encoded disruption
- Careful phenotyping to identify standardized biomarkers and patient reported outcomes for evaluation of safety and efficacy
- Scalability of strategies for rare disease therapy discovery
- Sustainability through discovery research, advocacy, and partnerships with third party payers (private payers, Medicare, Medicaid, and the Veterans Administration), pharmaceutical companies, regulators, and philanthropy

Heather Stone, MPH

FDA

Heather Stone, MPH, is a Health Science Policy Analyst, Office of Medical Policy at U.S. Food and Drug Administration (FDA). She is an infectious disease epidemiologist and policy expert focused on global health and, in particular, the cycle between extreme poverty and infectious diseases.

Innovations in the Design and Conduct of Randomized Controlled Trials of Repurposed Drugs

Heather A. Stone, MPH
CURE ID Lead – an FDA-NCATS/NIH Collaboration
Office of Medical Policy
Center for Drug Evaluation and Research
U.S. FDA

April 19th, 2023

Outline

- Why do we still need RCTs if we have RWD?

  - Confirming/refuting hypotheses generated from RWD

- Where does this fit in FDA's RWD/RWE Efforts

- Challenges with RCTs of Repurposed Drugs

- Opportunities for Trial Innovation

Why do we still need RCTs?

- While causal inference techniques and experience have improved a lot in recent years, it is not always clear that these models can accurately replicate the findings of RCTs

- RCTs enable us to control for both known and unknown confounders

- Currently, they still remain the gold-standard for evidence generation in most diseases

How RWD and RCTs can compliment one another

- People often view it as a question of there either being value of RCTs or value of RWD, but not both

- They can in fact be complimentary

- RCTs can provide us controlled data that most clearly demonstrates a treatment's efficacy in a specific population

- RWD may provide more generalizable experience, but the increased heterogeneity may make it more difficult to isolate the treatment effect

Combining RWD and RCTs

- Another possibility is to use RWD within RCTs
  - Sometimes called "pragmatic trials", "embedded trials", or "point-of-care trials"
- Many different variations, but many of these designs utilize data that is captured during the course of routine clinical care in electronic health records (EHRs)
- May also use broader inclusion/exclusion criteria

Where does this fit in FDA's RWD/RWE efforts?

Clinical practice data and sites in RCTs

Real-World Evidence — Where Are We Now?

John Concato, M.D., M.P.H., and Jacqueline Corrigan-Curay, J.D., M.D.

Randomized, Interventional Study		Nonrandomized, Interventional Study	Nonrandomized, Noninterventional Study
Traditional randomized trial using RWD in planning	**Trial in clinical practice settings, with pragmatic elements**	**Externally controlled trial**	**Observational study**
RWD used to assess enrollment criteria and trial feasibility RWD used to support selection of trial sites	Selected outcomes identified using, e.g., health records data, claims data, or data from digital health technologies RCT conducted using, e.g., electronic case report forms for health records data or claims data	Single-group trial with external control group derived from RWD	Cohort study Case–control study Case–crossover study

Generation of RWE

Increasing reliance on RWD

Reliance on RWD in Representative Types of Study Design.

RCT denotes randomized, controlled trial; RWD real-world data; and [RWE] real-world evidence.

N ENGL J MED 386;18   NEJM.ORG   MAY 5, 2022

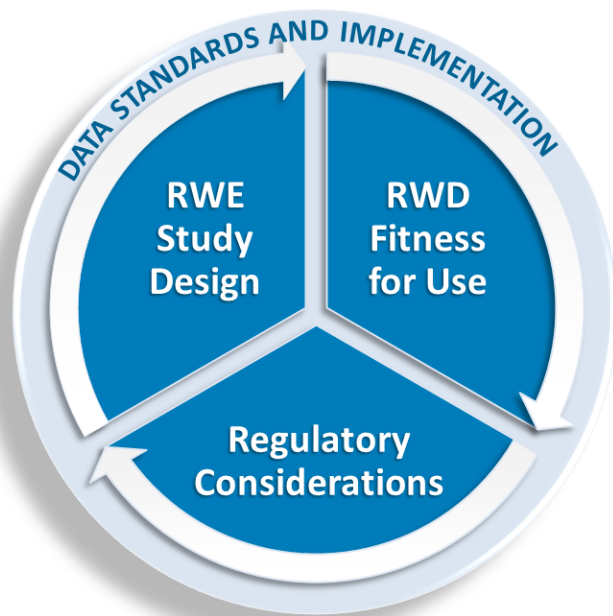# RWE for Effectiveness: Overview of FDA Approach

Key considerations (from 2018 Framework):

- Whether the RWD are fit for use
- Whether the trial or study design used to generate RWE can provide adequate scientific evidence to answer or help answer the regulatory question
- Whether the study conduct meets FDA regulatory requirements

Potential Advantages of Combining RWD and RCTs

- In some cases may provide a sort of "best of both worlds"
  - May provide more generalizable evidence than a traditional RCT with very strict inclusion/exclusion criteria
  - But still addresses both known and unknown confounders as a result of being randomized
  - Increases the potential for participation by non-traditional providers and patients by decreasing the burden of participation (e.g., using routinely collected data and routine practice sites)
    - May enable participants to see their routine care providers, who they know and trust
    - May decrease the cost of the trial
    - May be well-suited for the study of drugs that are approved and used in clinical practice (e.g., repurposed drugs)

Challenges to Combining RWD and RCTs

- EHRs were not designed for research, but primarily for billing
  - The data that is needed for the trial may not be routinely collected or may not be collected in a standardized format

- The healthcare infrastructure in the US is ill-suited for large coordinated efforts, given the lack of a national health system and high level of variation across sites
  - Compared to UK NHS or single-payor systems in Europe (e.g., Scandinavian countries)

- Challenges of how providers in routine practice actually provide the information? How will it be implemented? Is it a tab or pop-up within the EHR? Is it automatically extracted data? How do you get clinicians to agree to do this work?

Other Types of RCT Innovations to Study Repurposed Drugs

- ## Decentralized Trials

  - Decentralized Clinical Trials (DCT) hold promise to reduce patient and sponsor burden and increase accrual and retention of a more diverse trial population

    (https://www.fda.gov/about-fda/oncology-center-excellence/advancing-oncology-decentralized-trials)

    – Long COVID may be a particularly good use case for the use of Decentralized trials which can be conducted virtually from patient's homes

- ## Use of Digital Health Technologies

Trials That Have Informed the Discussion We are Having Today

- **PREDICT:** Platform for Randomized trials Evaluating Drugs for Inpatient Comparative Treatments

  - *Building a Sustainable Infrastructure for an Embedded, Adaptive, Platform Trial Comparing Multiple Repurposed Drug Treatment Arms in Sepsis and Other Serious Infections/Critical Care Conditions*

- **TLC:** Treatments of Long COVID

  - *Building a Sustainable Infrastructure for a predominantly Virtual, Adaptive, Platform Randomized Controlled Trial Comparing Multiple Drug Treatment Arms Against Symptoms of Long COVID/Post-Acute Sequelae of SARS-CoV-2 (PASC)*

Roundtable

Roundtable Discussions on decentralized trials and embedding trials in clinical practice.

- Inpatient trials and innovations in embedding trials in practice

Panelists:

- **Trevan Locke** (Duke Margolis)

- **Jon Sevransky** (Emory)

- **Raghave Tirupathi** (WellSpan Health)

- **Stacey Coe** (C-Path)

- **Barbara Goodman** (Cures within Reach)

- **Cynthia Adinig** (Patient)

Trevan Locke

Research Director, Duke-Margolis

Trevan Locke is an Assistant Research Director at Duke-Margolis working on issues related to biomedical innovation. He oversees Duke-Margolis' involvement as a founding member of the Coalition for Advancing Clinical Trials at the Point of Care as well as workstreams on evidence generation for Duke-Margolis' Real-World Evidence Collaborative.

Jonathan Sevransky, MD, MHS, FCCM

Emory University

Dr. Jonathan Sevransky is Professor of Medicine and Associate Division Director for Critical Care in the division of Pulmonary, Allergy, Critical Care and Sleep Medicine at Emory University. He also serves as the Director of Critical Care for Emory University Hospital. His research, clinical, mentoring and administrative responsibilities are centered on improving the care of patients with life threatening medical illness, including sepsis.

Stacey Coe, MS, CCRP

Senior Clinical Research Coordinator at the Critical Path Institute, CURE Drug Repurposing Collaboratory (CDRC)

Stacey Coe is a Senior Clinical Research Coordinator at the Critical Path Institute, CURE Drug Repurposing Collaboratory. She has over 10 years of experience conducting complex inpatient and outpatient drug trials and post-market device trials in the therapeutic areas of trauma and critical care, neurology, orthopedics, nephrology and podiatric wound care.

Clare Thibodeaux, PhD

Vice President, Scientific Affairs with Cures Within Reach

Dr. Clare Thibodeaux is the Vice President, Scientific Affairs with Cures Within Reach and has over 20 years of scientific research and philanthropic experience. Clare joined the Cures Within Reach team in November 2015, and collaborates with key opinion leaders, research institution partners, industry representatives and patient advocates from any disease area globally to identify, centralize and vet clinical repurposing research for funding.

Cynthia Adinig

Patient, co-founder of BIPOC Equity Agency

Cynthia Adinig is from Vienna, Virginia and developed Long Covid in March 2020. Cynthia has had a long career in community service and marketing, including work with nonprofits, small businesses, and tech companies. She is also the co-founder of BIPOC Equity Agency, a multidisciplinary consulting agency specializing in promoting racial equity in healthcare and scientific research.

Break

Dr. Chris Lindsell

Director of Data Science and Biostatistics,
Duke Clinical Research Institute

Dr. Chris Lindsell is director of data science and biostatistics at the Duke Clinical Research Institute, Professor of Biostatistics and Bioinformatics and Duke University, and Professor of Biomedical Informatics at Vanderbilt University Medical Center. His research portfolio includes clinical trials, health systems and services, and biomarker discovery and validation.

Dr. Nathalie Strub-Wourgaft

DNDi

Dr. Nathalie Strub-Wourgaft joined Drugs for Neglected Diseases Initiative (DNDi) in 2009 and is member of the Global Executive Team as COVID 19 and preparedness Director. As Director of NTDs since 2018, Dr. Strub-Wourgaft provided strategic and technical oversight to a wide portfolio of R&D and access plans for therapeutic areas covering Sleeping Sickness, Chagas disease, Cutaneous and Visceral Leishmaniasis, Filaria and Mycetoma.

Suanna Bruinooge, MPH

Division Director of Research Strategy and Operations, ASCO's Center for Research and Analytics (CENTRA)

Suanna Bruinooge, MPH, is the Division Director of Research Strategy and Operations in ASCO's Center for Research and Analytics. CENTRA generates, integrates, analyzes, and shares oncology data to foster innovation in research and patient care and help develop and evaluate ASCO's policy positions.

Oved Amitay

President and CEO, Solve ME

Oved Amitay is a pharmacologist by training, a drug-developer by trade, and a patient advocate by choice. He serves as President and Chief Executive Officer at the Solve ME/CFS Initiative - a national organization devoted to making Myalgic Encephalomyelitis (also known as chronic fatigue syndrome or ME/CFS) and other infection-associated diseases such as Long-Covid, understood, diagnosed and treatable.

Amy Morris

CEO, Consultant, IND 2 Results, LLC

Amy Morris has 40+ years' experience in the clinical drug development field. She has held senior management positions within the biopharma industry and nonprofit setting and spent most of her career working in chronic infectious diseases, immunology disorders and rare diseases.
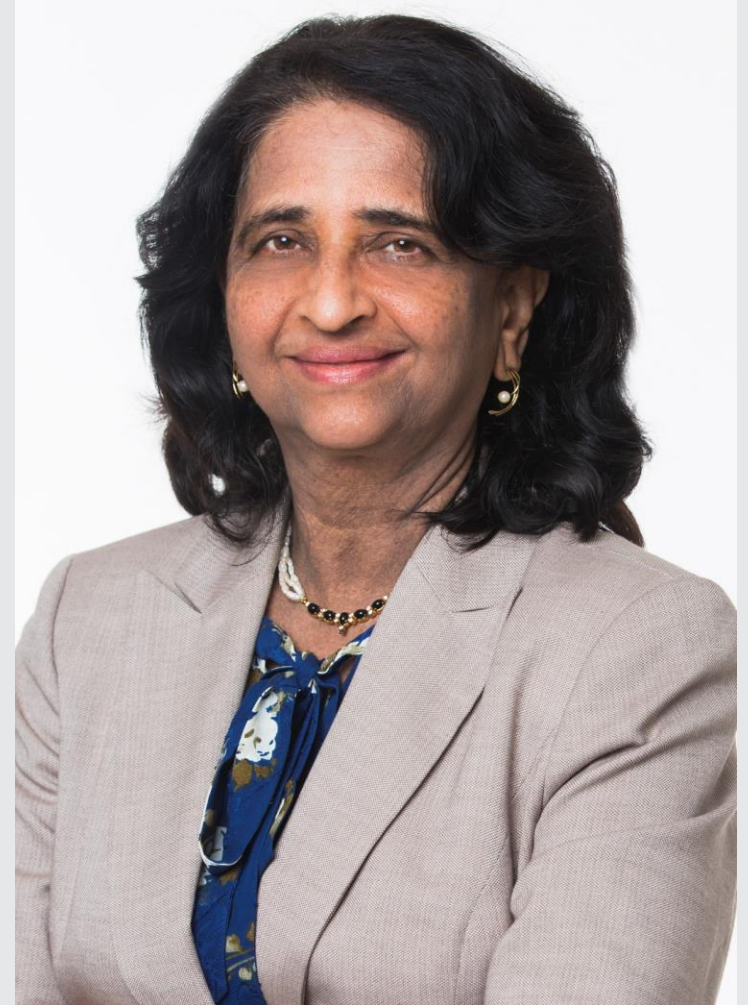
Vidula Sukhatme

CEO, GlobalCures

Ms. Vidula Sukhatme is the Co-founder and CEO of GlobalCures, Inc. Additionally, she played a role in founding the Morningside Center for Innovative and Affordable Medicine at Emory University.

Michael Sieverts

Patient-Advocate

Michael's career in federal science policy spans over 30 years with the National Science Foundation, including as Budget Director and Deputy Office Head in the Office of Budget, Finance, and Award Management. To his work as a patient-advocate, he brings budgeting, planning, and grants management experience, and continues to serve as an NSF Expert.

Dr. Ingrid Oakley-Girvan

SVP of Research and Strategy, Medable

Dr. Oakley-Girvan received her Ph.D. from the Stanford University School of Medicine and her M.P.H. from Tulane University School of Public Health and Tropical Medicine. She is affiliated with the Stanford Cancer Institute, the Canary Center at Stanford for Cancer Early Detection, and is SVP of Research and Academics at Medable Inc.

Wrap up and next steps

Reception