



# “Computable Phenotypes”

Understanding Their Importance in Regulatory Submissions of RWE

Scott L. DuVall, PhD

VA Informatics and Computing Infrastructure (VINCI)

May 2024

# Disclosure and Disclaimers

---



- Within the last 36 months, I have received grant funding from Alnylam Pharmaceuticals, Inc., Astellas Pharma, Inc., AstraZeneca Pharmaceuticals LP, Biodesix, Inc, Celgene Corporation, Cerner Enviza, GSK PLC, IQVIA Inc., Janssen Pharmaceuticals, Inc., Moderna, Inc., Novartis International AG, Parexel International Corporation through the University of Utah or Western Institute for Veteran Research outside the submitted work.
- Views expressed are those of the presenter and do not necessarily represent the views or policy of the Department of Veterans Affairs or the United States Government.
- Mention of a commercial product should not be construed as actual or implied endorsement

# Acknowledgements



Julie A. Lynch, PhD, RN, MBA, FAAN



Scott L. DuVall, PhD



Michael E. Matheny, MD, MPH, MS



Catherine C. Chanfreau, PhD



Marc A. Suchard, MD, PhD



Kyung M. Lee, PhD, MSPH



Nai-Chung N. Chang, PhD, MS



Benjamin Viernes, PhD, MPH



Craig C. Teerlink, PhD, MS



Patrick R. Alba, MS



Jianlin Shi, MD, PhD, MS



Tia L. DiNatale, MPH



Jeffrey P. Ferraro, PhD, MS



Jacob T. Kean, PhD, MEd, MS



Olga V. Efimova, MD, PhD, MA



B. Kevin Malohi, BSc



Corinne Halls, MS

That which we call a rose ...









That which we call a CODef ...



## What do you need to consider when analyzing real world data?

### Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research

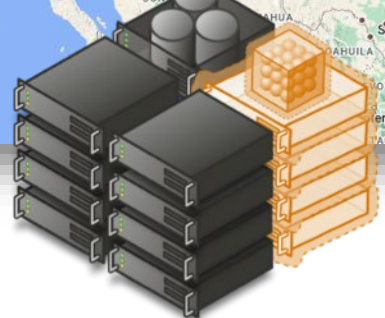
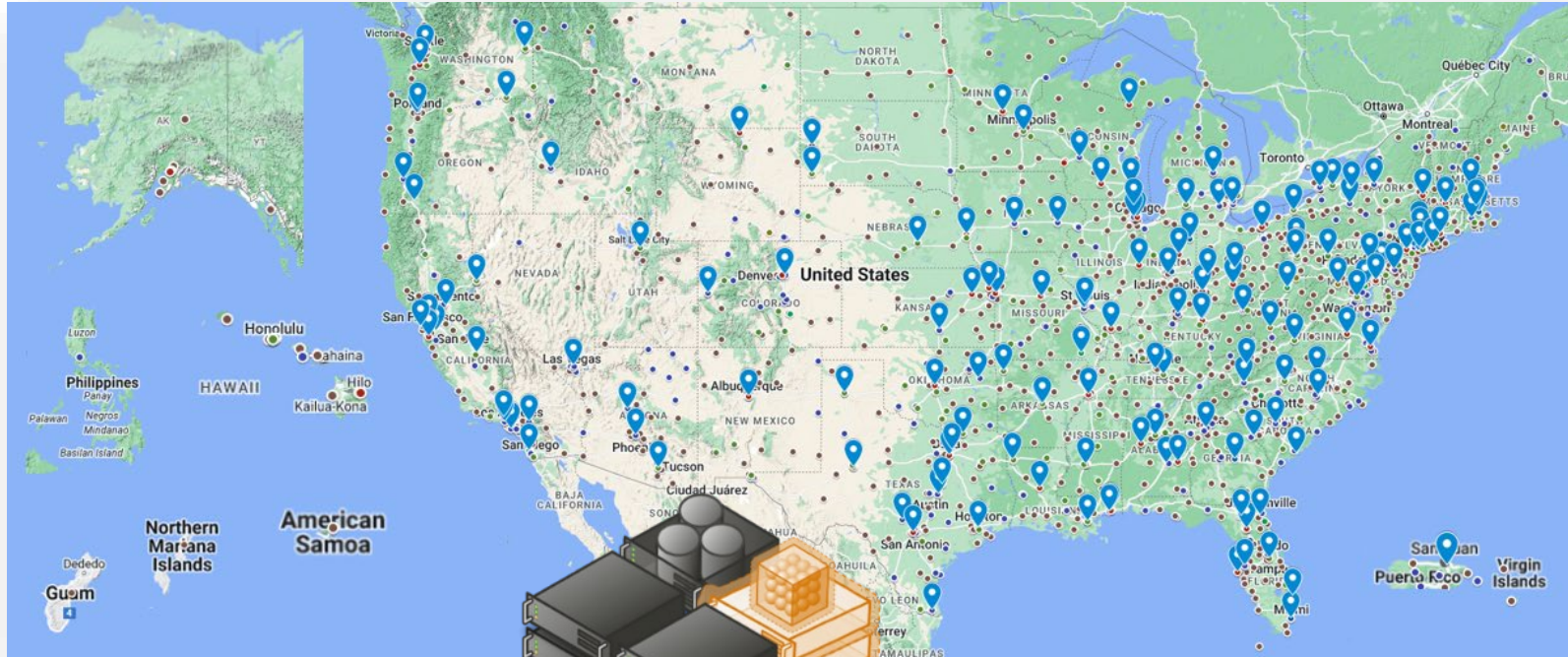
*William R. Hersh, MD,\* Mark G. Weiner, MD,† Peter J. Embi, MD, MS,‡ Judith R. Logan, MD, MS,\* Philip R.O. Payne, PhD,‡ Elmer V. Bernstam, MD, MSE,§ Harold P. Lehmann, MD, PhD,|| George Hripcsak, MD, MS,¶ Timothy H.artzog, MD, MS,# James J. Cimino, MD,\*\* and Joel H. Saltz, MD, PhD††*

Medical Care Volume 51, Number 8 Suppl 3, August 2013

- Inaccurate Data
- Not Complete Story
- Coded / Recorded for Other Purposes
- Information Locked in Notes
- Multiple Sources that may Conflict
- Data Granularity Mismatches
- Research Protocol ≠ Clinical Care



# Overview of VA



**163** medical centers

**985** outpatient clinics

**82** nursing homes

**128** state Veteran homes

**68** residential psychiatric centers

**111** domiciliary rehabilitation centers

**308** Veteran centers

national Veteran data centrally consolidated and standardized

# Overview of VA



**3.5** Diagnoses  
**billion**

**4.1** Outpatient  
Visits  
**billion**

**15.2** Inpatient  
Stays  
**million**

**258** Telehealth  
Visits  
**million**

**3.3** Procedures  
**billion**

**7.0** Medication  
Fills / Admins  
**billion**

**6.0** Clinical  
Notes  
**billion**

**180** Imaging  
Reports  
**million**

**11.9** Labs  
**billion**

**6.0** Vitals  
**billion**

**5.0** Genetic  
Tests  
**million**

**153** Vaccine  
Admins  
**million**



How many people  
are in the  
VA health care system?



# How many people are in the VA health care system?



- 26.9M persons in VA master list
- 17.8M persons with at least one visit in VA
- 13.9M persons with at least one lab in VA
- 8.7M persons with at least one lab in VA and no record of death

**8.7m → 26.9m**

3x difference in  
simple denominator for common question



How many people  
have experienced heart failure in the  
VA health care system?



# How many people have experienced heart failure in the VA health care system?



1,904,982

-5.5%

1,801,273

-18.4%

1,469,075

-83.6%

241,367

8x difference in HF denominator

**Algorithm Details**

≥1 ICD-9-CM diagnosis for HF

**Algorithm Validation**

**Algorithm Details**

≥1 ICD-9-CM diagnosis for HF and HF-specific treatment

**Algorithm Validation**

**Algorithm Details**

≥1 inpatient or ≥2 outpatient ICD-9-CM diagnosis for HF (≥30 days apart)

**Algorithm Validation**

**Algorithm Details**

≥2 ICD-9-CM diagnosis for HF (≥30 days apart) and HF-specific treatment and elevated NT-proBNP

**Algorithm Validation**

	Sensitivity	Specificity	PPV	NPV	Sample Size
<b>Gold Standard</b>	41.6	99.8	86.5	98.3	76254

**Geography** USA

**Data Source(s)** The provider-linked medical records from each institution Olmsted County, Minnesota

**Gold Standard** Manual medical record review

**Reference**

**Citation** Tison GH, et al. Identifying heart failure using EMR-based algorithms. Int J Med Inform. 2018 Dec;120:1-7. doi: 10.1016/j.ijmedinf.2018.09.016. [PMID: 30409334]

# Weighing the Benefits and Risks of Proliferating Observational Treatment Assessments

## Observational Cacophony, Randomized Harmony

---

**Robert M. Califf, MD**  
Verily Life Sciences  
(Alphabet), South San  
Francisco, California.

---

**Adrian F. Hernandez,  
MD, MHS**

Duke Clinical Research  
Institute, Durham,  
North Carolina; and  
Division of Cardiology,  
Department of  
Medicine, Duke  
University School of  
Medicine, Durham,  
North Carolina.

---

**Martin Landray,  
MBChB**

Nuffield Department of  
Population Health,  
University of Oxford,  
Headington, Oxford,  
United Kingdom.

“there is growing concern about whether attempts to infer causation about the benefits and risks of potential therapeutics from nonrandomized studies are providing insights that improve clinical knowledge and accelerate the search for needed answers, or whether these reports just add

**noise,  
confusion,  
and false confidence.”**

Califf RM, Hernandez AF, Landray M. Weighing the Benefits and Risks of Proliferating Observational Treatment Assessments: Observational Cacophony, Randomized Harmony. *JAMA*. 2020;10.1001/jama.2020.13319. doi:10.1001/jama.2020.13319


FOREWORD




## Transforming data into actionable insights

Carolyn M. Clancy

*Deputy Undersecretary for Health, Discovery, Education and Affiliate Networks (DEAN),  
Veterans Health Administration, US Department of Veterans Affairs,  
Washington, DC, USA*

 [Carolyn.clancy@va.gov](mailto:Carolyn.clancy@va.gov)

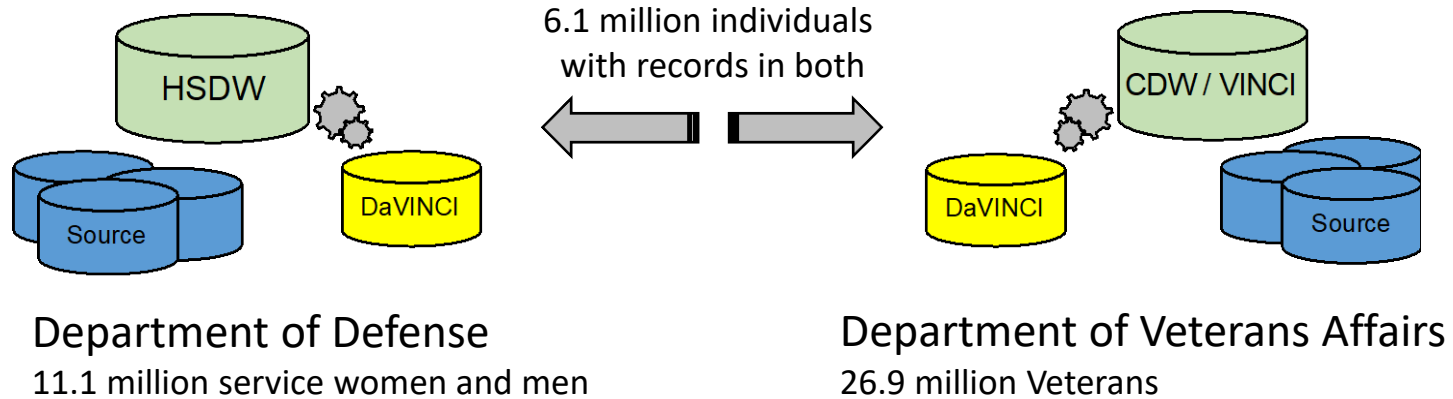
 <http://orcid.org/0000-0002-0308-5535>

### ***Biostatistics & Epidemiology Special Issue***

U.S. Department of Veterans Affairs Panel on Statistics and Analytics on Healthcare Datasets: Challenges and Recommended Strategies. [https://www.hsrd.research.va.gov/news/research\\_news/datasets-051420.cfm](https://www.hsrd.research.va.gov/news/research_news/datasets-051420.cfm)



# Increase completeness through linkages



Fully linked and incorporated:

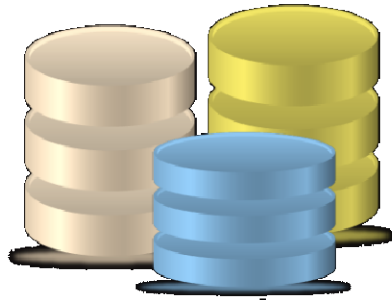
- Department of Defense (DoD/DHA) medical records
- Claims from Centers for Medicare and Medicaid Services (CMS)
- Centers for Disease Control and Prevention (CDC) National Death Index cause of death



network studies

- Active participant in OHDSI network studies

# Increase consistency through standardization



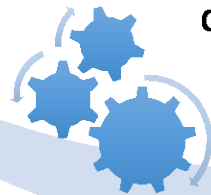
source data



quality assurance

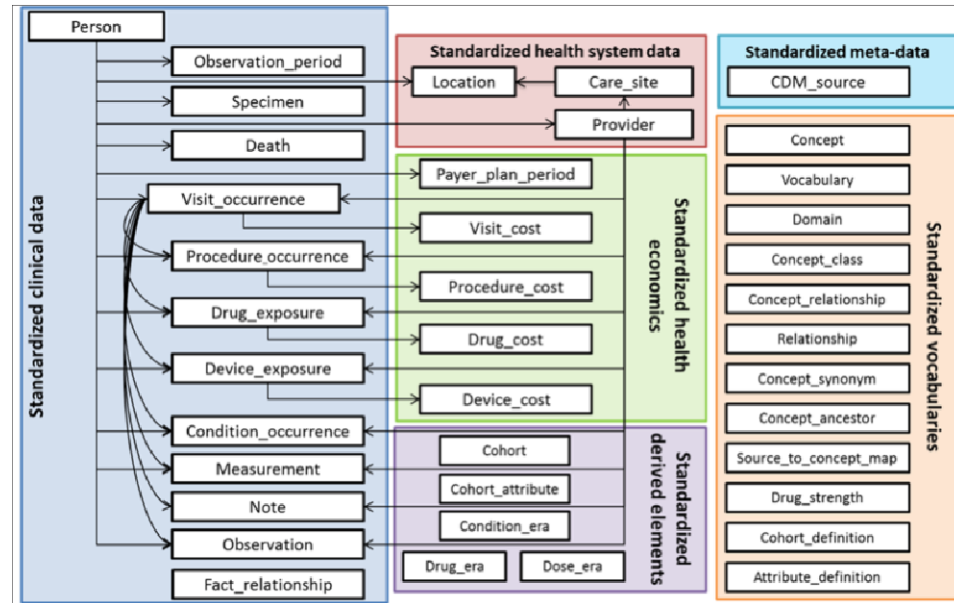
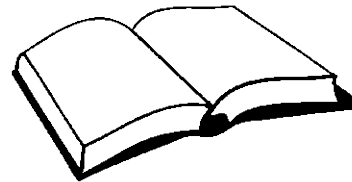
community use and contribution

extracted, transformed, derived data



national terminologies

best practices



# Increase granularity by including clinical text



# Increase accuracy by validating key variables



- *with* COVID  
VS  
*for* COVID
- Other reasons for codes:
  - Prophylaxis
  - History of
  - Acute event
  - Ongoing justification

The screenshot displays the ChartReview interface for a patient with Chronic Kidney Disease. The main window is divided into several panels:

- Annotation Panel:** Shows a task titled "Blood Transfusions Associated with Dialysis" for a "Chronic Kidney Disease Patient 0001". It lists three annotations: "Evidence of Transfusion" (highlighted in blue), "Indication for Transfusion" (highlighted in green), and "Relevant Laboratory Value" (highlighted in orange). Below this, a table shows a record for "transfused with 2 units RBC" on 08/23/2014, classified as "Evidence of Transfusion".
- CKD Diagnoses Panel:** Lists 39 diagnoses. A table shows the following data:

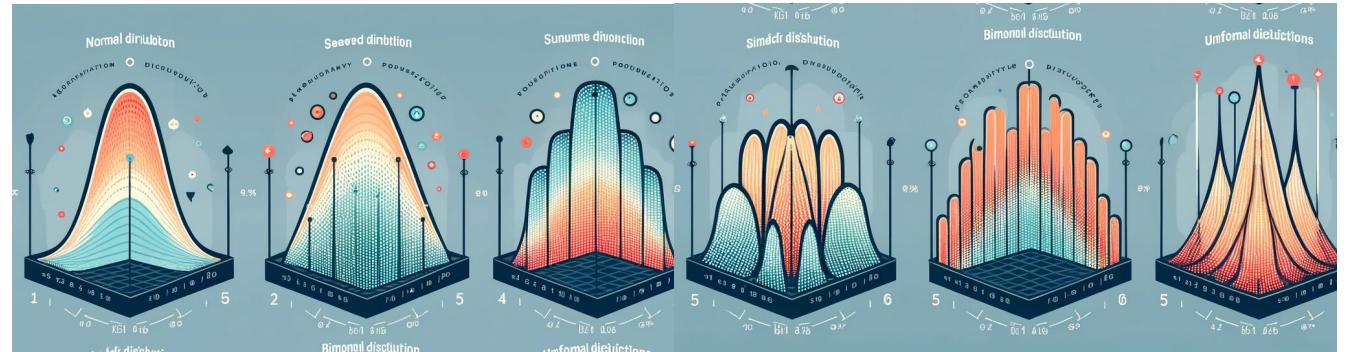
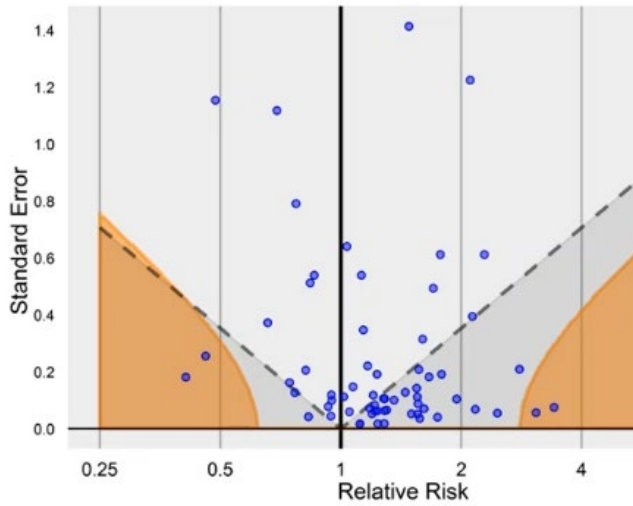
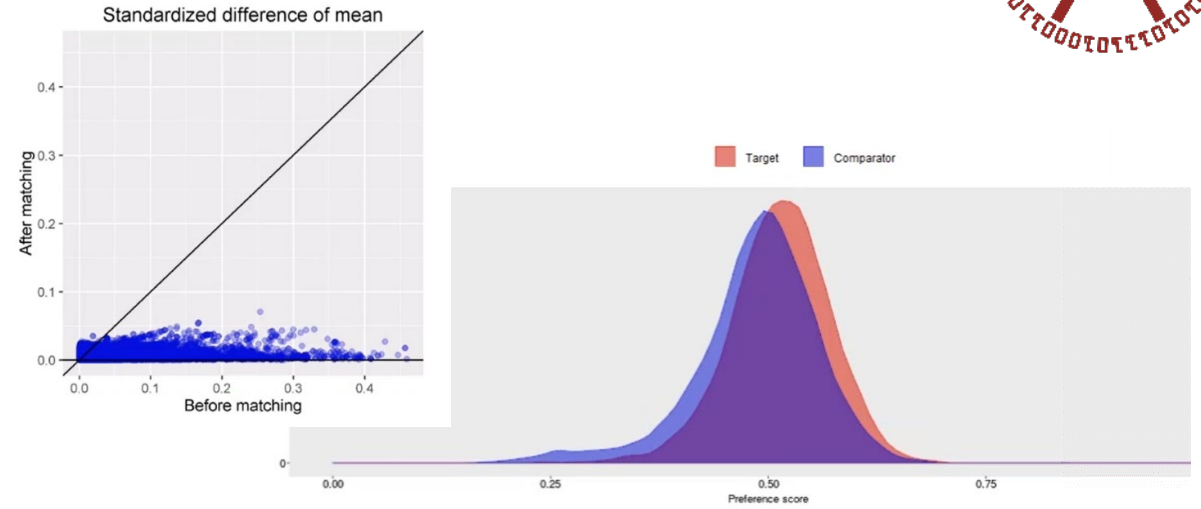
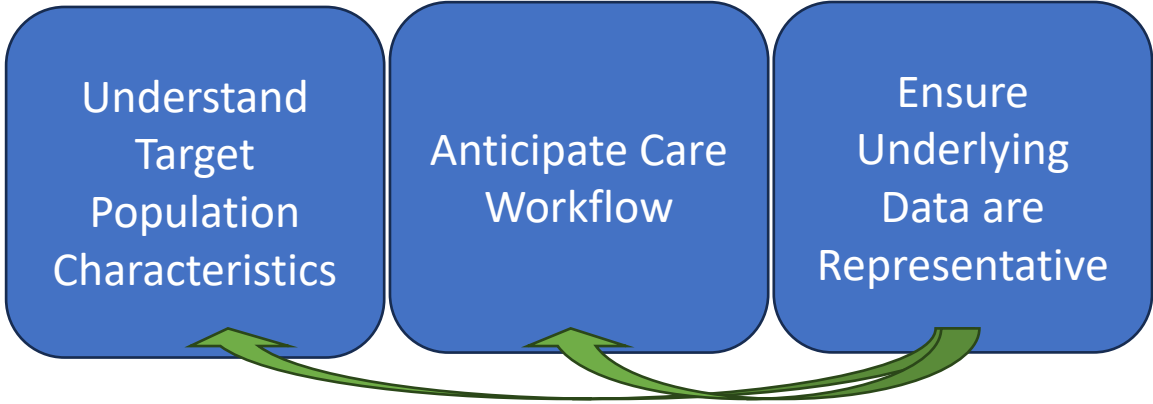
Diagnosis D	Code	Description	Type
01/12/2011	V65.40	Other Unspecified Counseling	Outpatient
01/14/2011	V65.40	Other Unspecified Counseling	Outpatient
01/20/2011	V45.11	Renal Dialysis Status	Outpatient
01/20/2011	250.42	Diabetes with renal manifestations, type II	Outpatient
01/22/2011	V45.11	Renal Dialysis Status	Outpatient
01/24/2011	280.0	Iron deficiency anemia secondary to blood loss	Outpatient
01/24/2011	250.42	Diabetes with renal manifestations, type II	Outpatient
- CKD Medications Panel:** Lists 432 medications. A table shows the following data:

Medication Date	Medication Name	Days Supply
12/14/2010	Metformin 500mg TAB	30
12/14/2010	Insulin (Injectable)	30
01/12/2011	Metformin 500mg TAB	30
01/12/2011	Insulin (Injectable)	30
01/24/2011	Packed Red Blood Cells	1
02/13/2011	Metformin 500mg TAB	30
02/13/2011	Insulin (Injectable)	30
- CKD Note Panel:** Lists 221 notes. A table shows the following data:

Reference Date	Title
01/24/2011	Nephrology Progress Note
02/13/2011	Nursing Note
02/13/2011	Admission Report
02/15/2011	Inpatient Progress Note
02/16/2011	Inpatient Progress Note
02/17/2011	Inpatient Progress Note
02/18/2011	Inpatient Progress Note
02/19/2011	Inpatient Progress Note
02/20/2011	Discharge Summary

synthetic medical record data

# Reduce bias, confounding, and other noise





Questions? Contact:

Scott DuVall

[scott.duvall@va.gov](mailto:scott.duvall@va.gov)