



## Database & Methods Cyberseminar Series

---

### ***Session #9: Diagnoses, Labs, and Meds, Oh My! Using CDW to Define a Study Cohort by Multiple Criteria***

June 3, 2019

Peter A. Richardson, PhD & Kathryn E. Royse, PhD  
Center for Innovations in Quality, Effectiveness and Safety (IQuEST)  
Houston, Texas

## Poll #1: Your role as a data user

***What is your role in research and/or quality improvement?***

- Research investigator
- Methodologist
- Data manager, analyst, or programmer
- Project coordinator
- Other – please describe via the Q&A function



## Poll #2: Your experience with VA data

*How many years have you worked with VA data?*

- One year or less
- More than 1, less than 3 years
- At least 3, less than 7 years
- At least 7, less than 10 years
- 10 years or more



*The objectives of this cyberseminar are to:*

- Gain some familiarity with using three distinct CDW data sources in a convergent way.
- Display some ways to structure data from all these sources for use in cohort definition.

# Session roadmap

- The mission
- Before CDW
- Finding first diagnoses
- Finding first medication exposure
- Finding first positive lab result
- Putting these together for further use
- Would this go better with OMOP?

# Session roadmap

- **The mission**
- Before CDW
- Finding first diagnoses
- Finding first medication exposure
- Finding first positive lab result
- Putting these together for further use
- Would this go better with OMOP?

# The Mission

We want to assist in constructing a disease-based patient cohort relying on:

**Diagnosis histories**

**Medication histories**

**Lab result histories**

This information will supply components for a suitable cohort definition.

# Our Approach

## What we're looking for

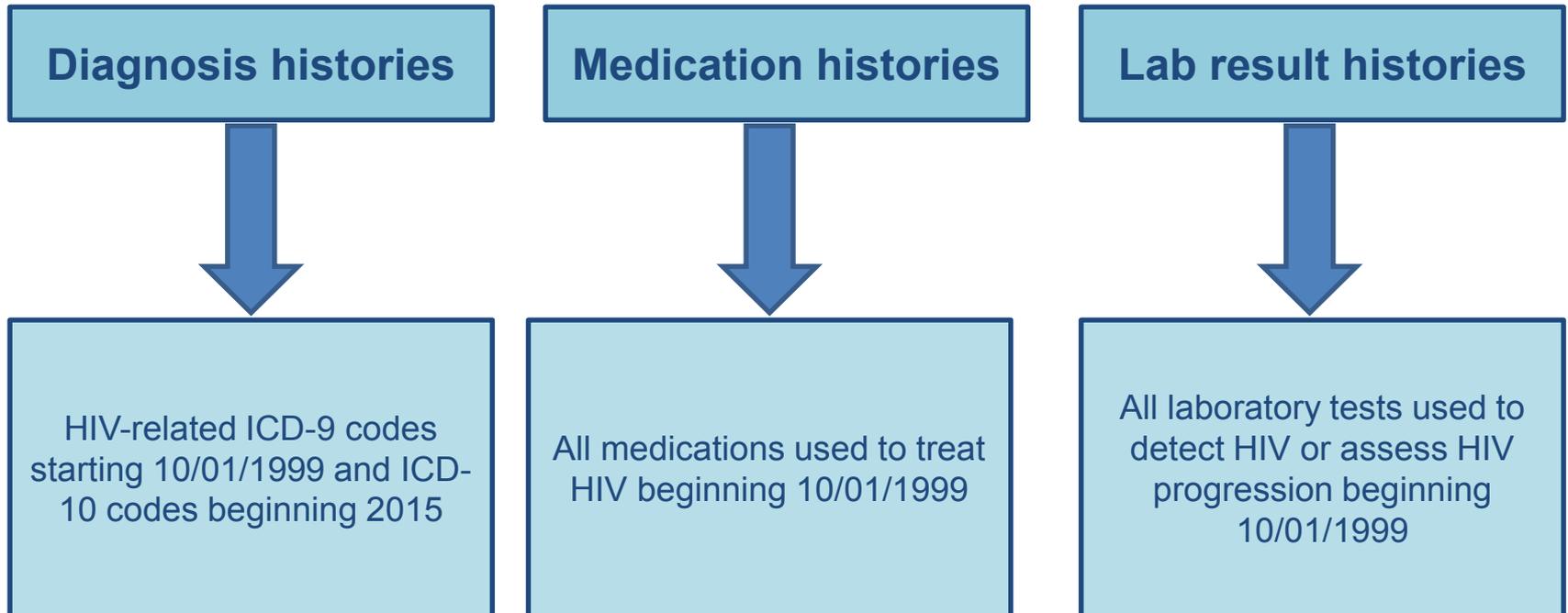
- ICD codes, pharmacy fills, and lab tests
- Dates of those codes, fills, and labs
- Results of the lab tests

## Where we're looking

- **CDW fact tables:** Include patient records for codes, fills and labs but need to be deciphered first using supporting tables (with patient ids)
- **CDW dim tables:** Are smaller, supporting tables containing deciphering tools to select patient characteristics and linking keys (no patient ids)

# Our HIV+ Cohort

- Identify all Veterans who received treatment for or diagnosis of HIV between 10/01/1999 and 12/31/2016
- **HOW?**



# Session roadmap

- The mission
- **Before CDW**
- Finding first diagnoses
- Finding first medication exposure
- Finding first positive lab result
- Putting these together for further use
- Would this go better with OMOP?

# Pre CDW

- looking for dx codes: ICD9 was found in MedSAS
- looking for medication fills: rx records were in the DSS (MCA) PHA table
- looking for lab records: selected lab results were in the DSS (MCA) LAR table.
- *All those legacy sources are still available with updating records*

# Session roadmap

- The mission
- Before CDW
- **Finding first diagnoses**
- Finding first medication exposure
- Finding first positive lab result
- Putting these together for further use
- Would this go better with OMOP?

## Poll #3: Your experience with Diagnosis Data

*How would you rate your overall knowledge of diagnosis data in CDW tables?*

- 1 Never Used
- 2
- 3
- 4
- 5 Frequently Use



# Diagnosis Codes-Using Dim tables

- looking for codes: ICD9: 042, V08 and ICD10: B20, Z21
- looking for dates of those codes
- fact tables
  - hospital stay diagnoses
  - outpatient visit diagnoses
  - diagnoses from fee basis
- dim tables

dim table	field	to link with fact table
ICD9	ICD9Code	ICD9SID
ICD10	ICD10Code	ICD10SID

# Diagnosis Codes-Linking to fact tables

- looking for codes: ICD9: 042, V08 and ICD10: B20, Z21
- looking for dates of those codes

fact table	DxDate
InpatientDiagnosis	CONVERT(Date, DischargeDateTime)
VDiagnosis	CONVERT(Date, VisitDateTime)
FeeInpatInvoiceICDDiagnosis	CONVERT(Date, TreatmentToDateTime)
InpatientFeeDiagnosis	CONVERT(Date, DischargeDateTime)

# Diagnosis Codes-Our approach

- The four codes 042, V08, B20 and Z21 were grouped together as a single diagnosis group.
- Date of first code at the PatientSID level from each source table was retained.
- Once we had a study subject crosswalk, these by-source first dates were aggregated at the PatientICN level and a single first ICD dx date was obtained by taking the MIN of these.

# Session roadmap

- The mission
- Before CDW
- Finding first diagnoses
- **Finding first medication exposure**
- Finding first positive lab result
- Putting these together for further use
- Would this go better with OMOP?

## Poll #4: Your experience with Pharmacy Data

*How would you rate your overall knowledge of pharmacy data in CDW tables?*

- 1 Never Used
- 2
- 3
- 4
- 5 Frequently Use



# Medication Fills: Many sources & many dates!

- looking for HIV medications: ABACAVIR, ATAZANAVIR, COBICISTAT, DARUNAVIR, DOLUTEGRAVIR, EFAVIRENZ, ELVITEGRAVIR, EMTRICITABINE, ENFUVIRITIDE, ETRAVIRINE, LAMIVUDINE, LOPINAVIR, MARAVIROC, NEVIRAPINE, RALTEGRAVIR, RILPIVIRINE, RITONAVIR, TENOFOVIR, TIPRANAVIR, ZIDOVUDINE
- looking for dates of those fills
  - Fill date vs. dispense date vs. release date?
- fact tables
  - Outpat RX
  - BCMA
  - fee basis pharmacy records
  - Non-VA drugs

# Medication Fills: Making the best of Dim tables

- looking for medications:
  - Additive versus reductive approach when there are MANY similar meds to pull
  - VA drug classes (IN150, AM800, IN160)

dim table	field	to link with fact table
LocalDrug	DrugClass (VA Class)	LocalDrugSID

# Medication Fills: Linking to fact tables

- fact tables
  - outpatient RX
  - BCMA
  - fee basis pharmacy records
  - non-VA drugs
- dim tables

dim table	field	to link with fact table
LocalDrug	LocalDrugNameWithDose	LocalDrugSID
LocalDrug	DrugNameWithoutDose	LocalDrugSID

# Medication Fills: Date selection & conversions

- looking for dates of those fills

fact table	MedicationDate
RXOutpatFill	CONVERT(Date, DispensedDateTime)
BCMADispensedDrug	CONVERT(Date, ActionDateTime)
FeePrescription	CONVERT(Date, FillDateTime)
NonVAMed	CONVERT(Date, StartDateTime)

# Medications Fills: Our approach

- The medications in the list were treated as a single medication group.
  - Union of all 4 medication tables
- Date of first fill at the PatientSID level was retained.
  - “HIVMedDate”
  - All HIV medication dates, names, and localdrugSIDs were saved to create time-varying exposures later
- Once we had a study subject crosswalk, these first dates were aggregated at the PatientICN level and a single first Rx date was obtained by taking the MIN of these.

# Session roadmap

- The mission
- Before CDW
- Finding first diagnoses
- Finding first medication exposure
- **Finding first positive lab result**
- Putting these together for further use
- Would this go better with OMOP?

## Poll #5: Your experience with Lab Data

*How would you rate your overall knowledge of lab data in CDW tables?*

- 1 Never Used
- 2
- 3
- 4
- 5 Frequently Use



# Lab Tests: An overview

- looking for tests: CD4, HIV viral load, Western Blot, HIV Qual test, HIV ELISA test
- Dim tables: LabChem & LOINC
- Fact table: LabChem
- looking for tests with positive results
  - some tests are categorical and easily decoded
  - some tests are numeric and require thresholding
  - units for numeric tests are often unclear
- dates for positive tests with a test type indicator are retained for further use

# Lab Tests: Identification is the hard part!

- Identification of tests can be tricky and requires lots of linking by dim tables.
- This also can involve good old fashion word search when LOINCs are missing (1999, 2000 ...)

dim table	field	to link to fact table
LabChem	LabChemTestName	LabChemTestSID
LabChem	LabChemPrintTestName	LabChemTestSID
LabChem	LabTestType	LabChemTestSID
LOINC	LOINC	LOINCSID
LOINC	Component	LOINCSID
LOINC	Units AS Units2	LOINCSID

# Lab Tests: Date selection & conversion

- Lab dates are the easiest part.
- Specimen dates are most biologically relevant, but when missing, complete dates *may* be substituted.

fact table	LabDate
LabChem	CONVERT(Date, LabChemSpecimenDateTime)
LabChem	CONVERT(Date, LabChemCompleteDateTime)

# Lab Tests: Our approach

- Date of first positive code at the PatientSID level from each lab type was retained.
- Once we had a study subject crosswalk, these by-type first dates were aggregated at the PatientICN level and a single first positive lab date was obtained by taking the MIN of these.

# Session roadmap

- The mission
- Before CDW
- Finding first diagnoses
- Finding first medication exposure
- Finding first positive lab result
- **Putting these together for further use**
- Would this go better with OMOP?

# What We Have So Far

- diagnosis code first dates
- medication fill first dates
- positive lab first dates

# What Do We Do With Them?

- These will be aggregated, depending on what overall criterion is chosen.
- That criterion will depend on comparison with a “gold standard”
- The process in our case would be the theme of a full presentation.

## SHORT COMMUNICATION

## Validation of HIV-infected cohort identification using automated clinical data in the Department of Veterans Affairs

JR Kramer <sup>1,2,3</sup>, C Hartman,<sup>1</sup> DL White,<sup>1,2,3,4</sup> K Royle,<sup>1,2</sup> P Richardson,<sup>1,2</sup> AP Thrift,<sup>2,3</sup> S Raychaudhury,<sup>1</sup> R Desiderio,<sup>1</sup> D Sanchez<sup>1</sup> and EY Chiao<sup>1,2,3</sup>

<sup>1</sup>Center for Innovations in Quality, Effectiveness and Safety (IQeES), Michael E. DeBakey VA Medical Center, Houston, TX, USA, <sup>2</sup>Department of Medicine, Baylor College of Medicine, Houston, TX, USA, <sup>3</sup>Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX, USA and <sup>4</sup>Center for Translational Research in Inflammatory Diseases (CTRID), Michael E. DeBakey VA Medical Center, Houston, TX, USA

### Objectives

The US Department of Veterans Affairs (VA) is the largest integrated health care provider for HIV-infected patients in the USA. VA data for HIV-specific clinical and quality improvement research are an important resource. We sought to determine the accuracy of using the VA Corporate Data Warehouse (CDW), a fully automated medical records database for all VA users nationally, to identify HIV-infected patients compared with a gold-standard VA HIV Clinical Case Registry (CCR).

### Methods

We assessed the test performance characteristics of each of our CDW criteria-based algorithms (presence of one, two or all of the following: diagnostic codes for HIV, positive HIV laboratory tests, and prescription for HIV medication) by calculating their sensitivity (proportion of HIV-positive patients in the CCR accurately detected as HIV-positive by the CDW algorithm) and positive predictive value (PPV; the proportion of patients identified by the CDW algorithm who were classified as HIV-positive from the CCR).

### Results

We found that using a CDW algorithm requiring two of three HIV diagnostic criteria yielded the highest sensitivity (95.2%) with very little trade-off in PPV (93.5%).

### Conclusions

A two diagnostic criteria-based algorithm can be utilized to accurately identify HIV-infected cohorts seen in the nationwide VA health care system.

**Keywords:** cohort, Department of Veterans Affairs, epidemiology, HIV, validation

Accepted 18 April 2019

### Introduction

The US Department of Veterans Affairs (VA) was at the forefront of implementing electronic medical records, and its large and automated clinical and administrative databases have become useful tools for quality assessment as well as research [1]. In addition, the VA is the largest single integrated health care provider for HIV-infected patients in the USA [2]. Thus, VA data that can be utilized for HIV-specific clinical research remain an

important resource. As HIV-infected patients are living longer as a result of improvements in antiretroviral therapy [3], chronic diseases such as heart disease, diabetes and cancer are becoming an increasing burden for HIV-infected patients [4]. Large sample sizes and detailed clinical data make the VA an ideal setting in which to study the incidence and outcomes of HIV-associated comorbidities. In order to address these research needs, the VA developed an adjudicated VA HIV Clinical Case Registry (CCR) [5] that was utilized to conduct large-scale HIV-related epidemiological and clinical research. In recent years, the VA has released extensive individual-level electronic medical record (EMR) data via the Corporate Data Warehouse (CDW) for use in research; therefore,

Correspondence: Dr Jennifer R. Kramer, Center for Innovations in Quality, Effectiveness and Safety (IQeES), Michael E. DeBakey VA Medical Center, 2002 Holcombe Blvd (150), Houston, TX 77030, USA. Tel: 713 440 4423; fax: 713 748 7359; e-mail: jkramer@bcm.edu

# Session roadmap

- The mission
- Before CDW
- Finding first diagnoses
- Finding first medication exposure
- Finding first positive lab result
- Putting these together for further use
- **Would this go better with OMOP?**

# Poll #6: Your experience with OMOP

*How would you rate your experience using OMOP?*

1 Never Used

2

3

4

5 Frequently Used



# Would This Go Better with OMOP?

## Dim Tables

- In OMOP these are called CONCEPT tables
- There are very few of them
- No site specific SIDs!

# Would This Go Better with OMOP?

## Fact Tables

- For diagnoses, there is only one fact table:  
CONDITION\_OCCURRENCE
- For drug records, there is only one fact table:  
DRUG\_EXPOSURE
- For labs, there is only one fact table: MEASUREMENT

# Would This Go Better with OMOP?

## Fact Tables

- Having one table for each data source means
  - one-stop shopping
  - enormous tables to run queries on

# In Conclusion

- Using diagnoses, medication records and lab results to ascertain first dates of a condition is a lot of work.
- But, it allows use of more evidence than by diagnoses alone.
- Exactly how the 3 sources are integrated should be based on validation work that would be study-specific.

# Additional Resources

---



## VA INFORMATION RESOURCE CENTER (VIREC)

- [VIREC Home](#)
- [VA/CMS Home](#)
- [About Us](#)
- [New Users of VA Data](#)
- [FAQs](#)
- [Acronyms](#)
- [HelpDesk](#)

### Medical SAS Datasets Documentation

#### Overview

VIREC's documentation on the Medical SAS Inpatient, Inpatient Encounters, and Outpatient Datasets is intended to help new and seasoned data users with understanding the content and variables in these datasets.

#### Data Documentation

- [+ Research User Guides](#)
- [+ Data Contents](#)
- [+ Frequencies](#)
- [+ Historical Variable Attributes](#)
- [+ Historical Stop Codes](#)

#### Medical SAS Datasets

- [■ Overview](#)
- [■ Inpatient](#)
- [■ Inpatient Encounters](#)
- [■ Outpatient](#)
- [■ Documentation](#)

#### ICD-10 Transition

Visit the "Medical SAS Datasets ICD-10 Transition" page to learn more.

#### General Resources

- [+ Data Access](#)
- [+ Data Sources](#)
- [+ Data Tools](#)
- [+ Data Topics](#)
- [+ Products & Services](#)
- [+ Special Projects](#)

The screenshot shows the VIREC INTRANET website. At the top left is the VIREC logo, and at the top right is the word 'INTRANET' in a stylized font. A search bar is located in the top right corner with the text 'Search All VA Web Pages' and a 'Search' button. Below the header is a navigation menu with the following items: VIREC Home, VA/CMS Home, About Us, New Users of VA Data, FAQs, Acronyms, and HelpDesk. The main content area is titled 'VA INFORMATION RESOURCE CENTER (VIREC)' and features a section for 'CDW Documentation' with an 'Overview' link. The overview text states: 'VIREC's CDW documentation is designed to help new and seasoned CDW users with understanding the structure and contents of the CDW. This summary information is available by domain.' Below this is a 'Sign-up for Product News & Updates' section, which asks users to email the VIREC HelpDesk for notifications. A 'Data Documentation' section follows, with the instruction to 'Expand each type of documentation below to view these resources.' This section contains a list of expandable links: 'Getting Started with Using CDW', 'NEW! Factbooks', 'NEW! Domain Layout & Descriptions', 'Data Contents', 'Discrete Frequencies', and 'Record & Null Counts'. On the right side of the page, there is a sidebar with a 'CDW' section containing links for 'Overview', 'Data Transition to CDW', and 'Documentation'. Below this is an 'ICD-10 Transition' section with a link to a page about the transition. At the bottom of the sidebar is a 'General Resources' section with links for 'Data Access', 'Data Sources', 'Data Tools', 'Data Topics', 'Products & Services', and 'Special Projects'.

<http://vaww.virec.research.va.gov/CDW/Documentation.htm>  
(VA Intranet)

## Archived VIREC cyberseminars

### **CDW Fundamentals**

[CDW: A Conceptual Overview](#)

[CDW: Locating Its Documentation](#)

[Building Your Dataset in CDW: Joining Tables within a Domain](#)

[Getting the Information You Need From CDW: SQL Starter Language](#)

[Getting CDW Back Together: Joining CDW Tables \(Continued\)](#)

[Data Management in SQL: Selected Intermediate SQL Skills](#)

# Quick links for VA data resources

## **Quick Guide: Resources for Using VA Data**

<http://vaww.virec.research.va.gov/Toolkit/QG-Resources-for-Using-VA-Data.pdf> (VA Intranet)

**VIReC:** <http://vaww.virec.research.va.gov/Index.htm> (VA Intranet)

**VIReC Cyberseminars:** <http://www.virec.research.va.gov/Resources/Cyberseminars.asp>

**VHA Data Portal:** <http://vaww.vhadatportal.med.va.gov/Home.aspx> (VA Intranet)

**VINCI:** <http://vaww.vinci.med.va.gov/vincicentral/> (VA Intranet)

**Health Economics Resource Center (HERC):** <http://vaww.herc.research.va.gov> (VA Intranet)

**CDW:** <https://vaww.cdw.va.gov/Pages/CDWHome.aspx> (VA Intranet)

**Archived cyberseminar: What can the HSR&D Resource Centers do for you?**

[http://www.hsrd.research.va.gov/for\\_researchers/cyber\\_seminars/archives/video\\_archive.cfm?SessionID=101](http://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=101)

# VIReC options for specific questions

## HSRData Listserv

- Community knowledge sharing
- ~1,400 VA data users
- Researchers, operations, data stewards, managers
- Subscribe by visiting <http://vaww.virec.research.va.gov/Support/HSRData-L.htm> (VA Intranet)



## HelpDesk

- Individualized support



[virec@va.gov](mailto:virec@va.gov)

(708) 202-2413

# Contact information

**Peter Richardson, PhD**

**Kathryn Royse, PhD**

Center for Innovations in Quality,  
Effectiveness and Safety (IQuEST)

Houston, Texas

[Peter.Richardson2@va.gov](mailto:Peter.Richardson2@va.gov)

[Kathryn.Royse@va.gov](mailto:Kathryn.Royse@va.gov)



**Next session:  
July 1<sup>st</sup>, 2019  
1:00 PM Eastern**



## Database & Methods Cyberseminar Series

---

# Advanced Uses for Joint Legacy Viewer

Reese K. Omizo, MD

VA Pacific Islands Health Care System