CHALLENGES & OPPORTUNITIES IN CURATING LARGE ADMINISTRATIVE & CLINICAL DATASETS FOR RESEARCH USE

VIReC’S EFFORTS TO ADVANCE VA INFORMATICS AND RESEARCH PROJECTS

Presented by:
Maria Souden, MSI, PhD
Chandrea Culbreath, MPH, BSN, CHTS-TR
AGENDA

1. INTRODUCTION TO THE VA
2. RESEARCH IN THE VA
3. VA DATA ENVIRONMENT/EHR MODERNIZATION
4. OPPORTUNITIES AND CHALLENGES
5. VIReC’S ROLE IN THE DATA LANDSCAPE
6. PARTNERING WITH VIReC & THE VA
Veterans Benefit Administration

The Veterans Benefits Administration (VBA) provides a variety of benefits and services to Servicemembers, Veterans, and their families.

National Cemetery Administration

The National Cemetery Administration (NCA) provides burial services in a VA national cemetery to all members of the armed forces and Veterans who have met minimum active-duty service requirements.

Veterans Health Administration

The Veterans Health Administration is America’s largest integrated health care system, providing care at 1,298 health care facilities, including 171 medical centers and 1,113 outpatient sites of care of varying complexity (VHA outpatient clinics).

U.S. Department of Veterans Affairs
Comprised of 3 organizations to address the unique needs of our nations Veterans
Veterans Health Administration

Over the years....

- Established as a national program in 1930
- Early pioneer of the electronic health record
- Major reforms starting in 1990’s including
  - investments in electronic health record
  - shift from emphasis on inpatient and specialty care to emphasis on primary care
- Contains 9M Veterans enrolled in the VA health care system.

Evolved Healthcare Delivery Model

- Personalized Care
- Proactive and Patient-Driven
- Team-based Primary Care
- Evidence-Based
- Prevention / Population Health
- Data-Driven

Learning Healthcare System
Learning Healthcare System

“science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the delivery process and new knowledge captured as an integral byproduct of the delivery experience.”
VA RESEARCH

IMPROVING THE LIVES OF VETERANS AND ALL AMERICANS THROUGH HEALTH CARE DISCOVERY AND INNOVATION
Program supporting pre-clinical research to understand life processes from the molecular, genomic, and physiological levels.

Intramural program for improving the quality of life of impaired and disabled Veterans.

Program advancing ideas along the translational pathway from scientific discovery to clinical application in order to advance the healthcare of our Veterans.

Program supporting the identification, evaluation, and implementation of evidence-based strategies that improve the quality and safety of care delivered to Veterans.
<table>
<thead>
<tr>
<th>Health Services Research</th>
<th>Evaluates the <strong>effectiveness of clinical procedures or practices and processes of care</strong> in the &quot;real world.&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Compares <strong>VA outcomes, cost, and quality of care</strong> among VA sites or between VA and non-VA providers</td>
</tr>
<tr>
<td></td>
<td>Evaluates <strong>utilization patterns and costs</strong> associated with practice patterns, interventions, policy changes, etc.</td>
</tr>
<tr>
<td></td>
<td>Develops and evaluates <strong>new measures and methods</strong> for use in health services research.</td>
</tr>
<tr>
<td></td>
<td>Focuses on <strong>implementing research findings into practice.</strong></td>
</tr>
<tr>
<td></td>
<td>Looks at <strong>variations in care</strong> (e.g., ethnic, cultural, provider, geographic based) and their determinants</td>
</tr>
<tr>
<td></td>
<td>Examines <strong>organization, management, and leadership</strong> in health care</td>
</tr>
</tbody>
</table>

[www.hsrdr.research.va.gov](http://www.hsrdr.research.va.gov)
VA DATA ENVIRONMENT

BIG DATA ENVIRONMENTS PULL DATA FROM A VARIETY OF SOURCES
VA data is continuously collected in the context of healthcare operations and delivery.
CPRS
Computerized Patient Record System

- Front end for most data collected in clinical care and care delivery
- Provides an interface for data entry and viewing of EHR

VistA
Veterans Health Information Systems & Technology Architecture

- Backend storage of data; integrated system of software applications
- VistA “system” is actually 130 separate instances

CPRS/VistA system has been used in the VA since the 1980s and has evolved to suit Veteran-specific care with heavy local customization.
VA data is entered in CPRS and other systems and curated into various data sources. These sources are commonly used for QI, research, or informatics projects.

**Corporate Data Warehouse (CDW)**
National repository of relational data from the EHR and several other VHA clinical and administrative systems.

**VA Millennium (VA Mill)**
VA's new EHR (a commercial Cerner system) that includes a user interface and backend database for data storage analogous to VistA.

**Pharmacy Benefit Management (PBM)**
National database containing details and extensive information about all prescriptions dispensed within VHA systems.

**VA Medicare & Medicaid (VA/CMS)**
Data acquired from the Centers for Medicare and Medicaid Services (CMS) and United States Renal Data System (USRDS) specific to Veterans.

**Managerial Cost Accounting (MCA)**
Data source providing activity-based cost information and clinical information.
VA’s New EHR System

- VA has begun a decade-long transition to the commercially developed Cerner EHR
- First instance of VA Millennium went live at the Mann-Grandstaff VA Medical Center (Spokane, WA) in October 2020
- Rollout is expected to continue in 2023 completing in ~2030
- New EHR = new data
OPPORTUNITIES
AND CHALLENGES

BALANCING DEMAND
Strengths of VA Data

- Size/Volume - largest integrated healthcare system in US, with 20 years of data of longitudinal data on 24 million patients
- Breadth/Scope of data – includes clinical data, assessment tools, administrative etc.
- Timely – some data available in almost real time
- Can be linked to external data (i.e., CMS, NDI, DoD)
- Data created for VHA operations can also be used for research
Challenges & Limitations

• Data landscape can be complicated – multiple sources for similar data elements; likely to get more complicated with EHRM
• Analyses may not be generalizable to general population
  • Mostly male, sicker than average, Veteran-prevalent conditions (PTSD, mental health, traumatic brain injury, limb loss)
• Not all Veterans enrolled in VHA
• Many Veterans also receive care outside VHA
• Some data elements are less reliably available, e.g., income, education
Striking a Balance

Opportunities

- Years of data available
- Rich computational power
- Many sources all in one place
- Lots of data!

Challenges

- New skills and processes
- Less documentation
- More complexity
- Lots of data!
This is where VIReC comes in...

Building a community of efficient data users
VIREC’S ROLE IN THE DATA LANDSCAPE

RESEARCHER’S GUIDE TO VA DATA
www.virec.research.va.gov
VIReC is...

- one of four Health Services Research and Development Service (HSR&D) Resource Centers established in 1998 to service the VA research community.

- a key resource providing information, training and other services to support the use of VA data. We accomplish this by:
  - Developing Education & Data Documentation Resources
  - Disseminating Data Knowledge
  - Strengthening VA Data Capacity

About Us

The VA Information Resource Center (VIReC) supports the effective use of data by VA’s analytic and informatic communities to transform Veteran health and health care as part of a learning health system.
VIReC Functions and Expertise

Continuum of initiatives supporting research data needs

- Data Availability & Access Policy
- Data Knowledge & Documentation
- Education & Dissemination
- HelpDesk & Knowledge-Sharing
- Liaisons within Data Ecosystem
- Provisioning & Oversight of CMS Data
- VA REDCap Data Collection & Management
How does VIReC Support Research, Informatics or other Data Needs?

VIReC assists a variety of users at different points of their project lifecycle:

• Navigate the VA Data Landscape
• Prepare to Access the Data
• Understand Data Contents and Nuances with the Data
• Stay Abreast of New Data or Updates in a Changing Data Environment
Navigate the VA Data Landscape

- Locate Resources About VA Data
- Browse Commonly Used VA Data Sources
- Explore Specific Topic Areas
- View Presentations About VA Data and Its Use
VIReC Supported Websites

www.virec.research.va.gov
- Public-facing
- VIReC overview
- Publications, Cyberseminars & events

vaww.virec.research.va.gov
- VA network only
- Resources for working with data
- News & new resources – home

vawww.vhadataportal.med.va.gov
- VA network only
- Collaborative with NDS, VINCI, & HERC
- Data source & access information

Locate resources about VA data
View Presentations about VA Data and its use in Research

- Database & Methods
- Corporate Data Warehouse
- Using Data & Information Systems in Partnered Research
- Research & EHR Synergy
- Good Data Practices Miniseries
- VIReC CMS Data Miniseries

https://www.virec.research.va.gov/Resources/Cyberseminars.asp#Upcoming
Browse Commonly Used VA Data Sources

http://vaww.vhadataportal.med.va.gov/DataSources/DataSourcesOverview.aspx
(VA Intranet)
Explore Specific Topic Areas

Data Topics
- Accuracy of Death Dates
- Cause of Death
- Comorbidities
- Cost of Care
- Data Quality
- Emergency Department & Urgent Care Encounters
- Electronic Health Record Modernization (EHRM) & Implications for Data
- Geocoded Data
- VA Health and Retirement Study (HRS) Data
- Home Telehealth
- ICD-10 Implementation
- Laboratory
- Long Term Care
- Mortality Data
- Nursing
- Patient Satisfaction
- Pharmacy
- Polytrauma
- Preparatory to Research
- Priority Groups
- Race & Ethnicity
- VA Care in the Community
- Veteran Homelessness
- Veteran Income Data
- Veteran Statistics
- VistA
- Vital Status Ascertainment
- Women Veterans’ Health Data

https://vaww.virec.research.va.gov/Intro/Working-with-VA-Data.htm
(VA Intranet)
Prepare to Access the Data

- Apply for Access to VA Data
- Request Medicare & Medicaid Data for Veterans
- Learn How to Work with the Corporate Data Warehouse (CDW)
- VA REDCAP for Primary Data Collection & Management
Apply for Access to VA Data

(VA Intranet)
Request Medicare & Medicaid Data for Veterans

https://vaww.virec.research.va.gov/VACMS/Requests/Overview.htm
Learn How to Work with the Corporate Data Warehouse

New to using CDW?

https://vaww.virec.research.va.gov/CDW/Documentation.htm (VA Intranet)

Email virec@va.gov to be notified of new CDW products and seminars.
Primary Data Collection & Management

Learn More:
http://vaww.virec.research.va.gov/REDCap/Overview.htm
(VA Intranet)
Understand Data Contents and Nuances with the Data

UNDERSTAND DATA ELEMENTS AND FORMAT

METHODS AND TECHNICAL DOCUMENTATION

INDIVIDUALIZED SUPPORT (HELPDESK)
Understand Data Contents and Structure

https://vaww.virec.research.va.gov/CDW/Documentation.htm
(VA Intranet)
CDW Factbooks provide detailed descriptions of tables, columns, and values in select CDW domains.

VIReC Product Spotlight

**VIReC Factbook**

*Corporate Data Warehouse (CDW)*

*VA Millennium Immunization 2.1 Domain*

April 2021

**Part 2. Introduction to Immunization 2.1 Domain**

*VistA-based Data*

CDW is a relational database organized into a collection of data domains. Domains represent logically or conceptually related sets of data tables. Domain themes generally indicate the application in the VistA.

**Part 3. Clinical and Technical Context**

*VistA Data Management/Security*

According to the VistA Metadata Repository [3], a VistA Directive mandates the lock down of the Immunization Dimension File using Data Standardization (DS). Additions, edits or deletions will only be completed by

**Part 4. Table and Column Descriptions**

Part 4 provides descriptions of tables and selected columns in the Immunization Domain. The name of each table is presented in large font with a description of its general content immediately following. After each table description, selected columns within that table are described as indicated below.

**Content of Column Descriptions**

**Part 5. Primary and Foreign Key Connections Query**

Once you have been granted access to CDW data, this query can be used to generate an up-to-date list of primary and foreign keys associated with this domain.

```sql
SELECT * FROM [tables] WHERE [foreign_key] = [value]
```

**Part 6. Example SQL Code for Data Exploration**

This section contains a simple question that may be answered by applying basic SQL queries to the data in this domain. It does not involve other domains.

1. How many immunizations were given at Mann-Grandstaff VA Medical Center for Hepatitis B in the past year?

2. What are the most commonly reported reactions to a COVID 19 vaccine in the VA?
Methods & Technical Documentation

The Researcher’s Notebook: Practical information on using data from researchers in the field.

https://vaww.virec.research.va.gov/Notebook/Overview.htm
(VA Intranet)
Exploring Relationships between Procedure and Procedure Diagnosis Data in the CDW Outpatient Domain

Introduction

This notebook is the first of a series that explores relationships between procedure and Procedure Diagnosis data in the CDW Outpatient Domain. In Part I of this series, we aim to broadly characterize the types of procedures and diagnoses associated with these data. This includes examining the occurrence of these procedures and diagnoses and the relationships between them. The focus is on identifying patterns in the data that may be relevant to clinical practice and research. The data is taken from the CDW Outpatient Visit and Procedure Diagnosis tables and the associated procedures and diagnoses found in the CDW Outpatient Visit and Procedure Diagnosis tables (see Table 1). Using a sample of encounters from Outpatient Visit and Procedure Diagnosis tables, we can determine the number of diagnosis, procedure, and procedure diagnosis records that link to each other. This will help us identify which data patterns exist in our sample. In Table 1, a “0” indicates that no record exists, while a “1” indicates that at least one record was populated. For example, encounters that link to at least one procedure record, and zero procedure diagnosis records have a data pattern of “1 0 0 Procedure Diagnoses”.

Table 1. Potential Patterns of Procedure and Procedure Diagnosis Data for Encounter Records in the Outpatient Domain

<table>
<thead>
<tr>
<th>Pattern</th>
<th># Diagnoses</th>
<th># Procedures</th>
<th># Procedure Diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: The data in this table is based on a sample of encounters and may not be representative of the entire dataset. Further analysis is needed to validate these findings.

Step 1 | Identify Tables and Columns of Interest

Table 3 contains a list of tables and columns used to capture the topic of interest. The primary keys from Outpatient Visit and Procedure Diagnosis tables will be used to determine the number of procedures, procedure and procedure diagnosis records that are populated for encounter records pulled from Outpatient Visit. The need to link diagnosis, procedure, and procedure diagnosis records to encounter records requires the inclusion of the VisitDateTime column from each of these tables. The VisitDateTime column from these tables will be used to filter the data to a selected timeframe. Additionally, ServiceCategory and EncounterType from Outpatient Visit will be used to further delineate and clarify the data patterns identified in this notebook (e.g., ICD/CPIC). The current notebook has   | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 |
| 1 | 1 | 1 | 1 |
| 1 | 2 | 1 | 1 |
| 1 | 3 | 1 | 1 |
| 1 | 4 | 1 | 1 |

Note: The data in this table is based on a sample of encounters and may not be representative of the entire dataset. Further analysis is needed to validate these findings.

Step 2 | SQL Code

Note: The SQL code for this notebook including text and results can be found in Appendix A.

Step 2.1 | Reducing Large Fact Tables and Creating a Cohort of Encounters

This step has two main purposes:
1. Reduce large fact tables containing patient encounter records (Outpatient Visit, diagnosis records)

   The patient table is joined to Outpatient Visit (specified in the FROM clause) by linking the VisitDateTime column from the Outpatient Visit table to the Diagnosis table. The WHERE clause filters records from Outpatient Visit to exclude historical visits (ServiceCategory = 'H') and to identify records that link to test patients (as indicated by the PatientVisit table's possibleTestPatientFlag column). The WHERE clause also filters records from Outpatient Visit to only include those within a specified encounter visitDateTime range.

   ```sql
   SELECT * FROM PatientVisit JOIN Encounter ON PatientVisit.visit_id = Encounter.visit_id
   WHERE EncounterServiceCategory = 'H' AND PatientVisit.possibleTestPatientFlag = 1
   ```

Step 2.2 | Identifying Patterns of Procedure and Diagnosis Data

Using the tallied data from Step 2.2 above, this step determines the number of encounters [if any] from the #Encounters temp table that display each of the 8 procedure/diagnosis data patterns shown in Table 1 above.

```sql
SELECT * FROM #Encounters
```
The Researcher’s Notebook: Millennium Data Discovery series provides guides to data collected in the Cerner EHR.

Identifying Immunizations Administration

Cerner EHR Transition

Introduction

This notebook provides two potential methods for identifying immunization data in the Cerner EHR. The methods are based on the Cerner EHR at the VA.

The first method, pre-transition immunization records, is pulled from Cerner Millennium data. Cerner Millennium data contains data collected in the Millennium EHR. The data is stored in a table called ImmunizationsVista.

In the second method, both pre- and post-transition immunization records are pulled. CDW Work3 consists of all CDW Vista data, stored alongside Cerner Work2 Millennium data.

Additional information about immunization data in the CDW Vista is available in the VIReC’s VA Millennium Immuno Factbook.

CDW Work and CDW Work2 Method

Step 1 | Tables and Columns of Interest

Table 1 contains a list of tables and columns used in the demonstration that follows. Main columns include those that appear in the query SELECT statement. Linking keys include primary and foreign key columns that are used for joining tables together, and filtering columns are those that appear in query filtering logic.

Step 2 | SQL Code

The query below pulls immunizations from CDW Vista that were administered at station 608 during the VisitDate range 9/1/19 to 3/1/21. The output from the query is stored in a temporary table called ImmunizationsVista.

CDW Work3 Method

Step 1 | Identify Tables and Columns of Interest

Table 1 contains a list of tables and columns used in the demonstration that follows. Main columns include those that appear in the query SELECT statement. Linking keys include primary and foreign key columns that are used for joining tables together, and filtering columns are those that appear in query filtering logic.

Step 2 | SQL Code

The query below pulls immunizations from CDW Vista that were administered at station 608 during the VisitDate range 9/1/19 to 3/1/21. The output from the query is stored in a temporary table called ImmunizationsVista.

Conclusion

In this notebook, we provide two potential methods for identifying immunizations administered before and after the Cerner EHR transition, using 1) CDW Vista and CDW Millennium data, and 2) converted Vista/MDX. Immunizations administered after the Cerner EHR transition are reported in the CDW data models.

Appendix A: SQL Code

Use of the methods in this appendix require access privileges for CDW Work2 and/or CDW Work3.

```sql
/* Pulling immunizations from CDW Vista data */

-- DROP TABLE if exists ImmunizationsVista;

FROM CDWVista.Visit a, CDWVista.Diagnosis b 
WHERE a.VisitID = b.VisitID 
AND a.Diagnosis = b.Diagnosis 
AND a.DiagnosisType = 'IMMUNIZATION' 
AND a.Vistime >= '2020-01-21' 
AND a.Vistime <= '2021-03-31';
```
VIReC Data Reviews examine VA-relevant data to support their use in VA research, evaluation, and quality improvement analyses.

Table 1. Laboratory Events Most Commonly Recorded in Millennium for Station 668 June 29, 2021

<table>
<thead>
<tr>
<th>Parent Event Description</th>
<th>Event Description</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab View</td>
<td>Chemistry</td>
<td>Creatinine Level</td>
</tr>
<tr>
<td>Lab View</td>
<td>Chemistry</td>
<td>eGFR Non-AA</td>
</tr>
</tbody>
</table>

HDL and LDL Tests

The "Lab View" event hierarchy included 95 laboratory tests containing the character strings 'HDL', 'LDL' or ' lipoprotein'. Only 10% of these tests had values recorded in Millennium on June 29, 2021. Table 2 contains those tests with the Logical Observation Identifiers Names and Codes (LOINC) code, normal low, normal high, result units, topography, and number of the tests. Code for the analyses of the HDL and LDL tests is in Appendix B.

LOINC is a universal standard for identifying medical laboratory and clinical test results. The LOINC codes in Table 2 were validated at [LOINC.org](http://LOINC.org).

Table 3 contains the minimum, maximum, and average result values for HbA1c tests. The American Diabetes Association provides the following criteria in interpreting the results of an HbA1c test:

- below 5.7% is normal

Appendix A: SQL Code to List Lab View Event Hierarchy and Identify Commonly Recorded Events

```sql
/*----------------------------*/
select EventSet
from CDWork2.HDIHMTestEventSet
where ParentEventSet = 'Lab View'
order by 1;
/*----------------------------*/

/*----------------------------*/
select distinct
  ev.CodeValueDescription as [Parent Event Description]
/*----------------------------*/
```
VIReC HelpDesk for Individualized Support

Assists researchers with understanding VA data and information systems within the context of their specific projects.

- VIReC subject matter experts respond directly to HD questions
- Use organizational knowledge and liaison relationships to connect requestors with the expertise they need

virec@va.gov
Stay Abreast of New Data or Updates

NEWS FOR USERS OF VA DATA AND INFORMATION SYSTEMS

PEER LEARNING & COMMUNITY KNOWLEDGE-SHARING

TRANSITION TO NEW ELECTRONIC HEALTH RECORD
News for Users of VA Data and Information Systems

- Quarterly updates on clinical and administrative data resources, VA databases, clinical informatics, electronic health record modernization
- Timely topics, spotlight tips, and training opportunities

Current and archived issues:
https://vaww.virec.research.va.gov/DIB/Overview.htm (VA Intranet)
E-mail virec@va.gov to subscribe
HSRData Listserv for Community Knowledge-Sharing

- Members ask and answer specific questions about using VA data
- 1,600+ participants
- Multiple perspectives from across the VA data ecosystem: PIs, fellows, analysts, data managers, statisticians, operations and program staff, data stewards, resource center experts, data architects
- Shared experience and tacit knowledge about historical and nuanced aspects of VA data and their application to research questions

More information about HSRData
http://vaww.virec.research.va.gov/Support/HSRData-L.htm (VA Intranet)
PARTNERING WITH THE VA

OPPORTUNITIES FOR USING AND WORKING WITH VA DATA
VA Partnerships, Collaborations, Affiliations

All VA research is intramural, meaning only VA employees can conduct research under VA’s sponsorship. Non-VA researchers can access VA data through collaborative research studies or by becoming a VA employee.

Get Involved in VA Research

• Collaborate with a VA researcher
  • Search the VA Health Services Research & Development (HSR&D) Directory to find VA researchers who share your research interests. [www.hsrd.research.va.gov](http://www.hsrd.research.va.gov)
  • Reach out to the Associate Chief of Staff (ACOS-R) for Research at your local VA Medical Center (VAMC) to ask about collaboration opportunities. [www.research.va.gov/about/national_directory.cfm](http://www.research.va.gov/about/national_directory.cfm)
  • Unpaid, without compensation (WOC) research positions will allow you to work with a VA investigator.

Become a VA employee

• Visit the VA Office of Research and Development website for a list of open Research positions and fellowship opportunities. [www.research.va.gov](http://www.research.va.gov)
• Visit USAJobs.gov

VIReC has positions available!
Visit our website to learn more about open positions!
THANK YOU!

Visit our website at:
www.virec.research.va.gov

Contact our HelpDesk for questions on VA data at:
virec@va.gov

Additional questions may be sent to:
Maria Souden, MSI, PhD
National Chair, EHR Modernization Research Sub-Council
PI, ORD Strategic Initiative for Research & EHR Synergy (OSIRES)
Director, VA Information Resource Center
maria.souden@va.gov

Chandrea Culbreath, MPH, BSN, CHTS-TR
Customer Engagement, Assessment and Portfolio Management
Lead, VA Information Resource Center
chandrea.culbreath@va.gov