



## Database & Methods Cyberseminar Series

---

Studying Utilization with MedSAS & CDW Data: How Do  
Changing Data Structures Affect Health Services  
Research Processes?

**December 5, 2016**

**Peter A. Richardson, PhD**

**Center for Innovations in Quality, Effectiveness and  
Safety (IQeSt)**

**Houston, Texas**



# Database & Methods Cyberseminar Series

*Informational seminars to help VA researchers access and use VA databases.*

## Topics

- VA data sources & data access systems
- Application of VA data to research and quality improvement questions
- Limitations of secondary data use
- Resources to support VA data use



## FY '17 Database & Methods Schedule

First Monday of the month\* | 1:00pm-2:00pm ET

Date	Topic
10/3/16	Overview of VA Data & Research Uses
11/7/2016	Requesting Access to VA Data
<b>12/5/2016</b>	<b>Utilization with MedSAS &amp; CDW</b>
1/9/2017*	VA Medicare Data (VA/CMS)
2/6/2017	Measuring & Assessing Utilization
3/6/2017	Mortality Ascertainment & Cause of Death
4/3/2017	Assessing Race & Ethnicity
6/5/2017	Pharmacy Data
7/10/2017*	CAPRI/VistAWeb for EHR Access
8/7/2017	Comorbidity Measures Using VA and CMS Data
8/21/2017	Advanced Topics in Comorbidity Measures
9/11/2017*	CDW microbiology, lab, & pharmacy domains

Visit our Education page for more information & registration links.

[www.virec.research.va.gov](http://www.virec.research.va.gov)



## Database & Methods Cyberseminar Series

---

Studying Utilization with MedSAS and CDW Data: How Do Changing Data Structures Affect Health Services Research Processes?

## 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:*

- Increase awareness of the opportunities for using MedSAS and CDW in utilization research.
- Discuss the differences between the paradigms represented by these data sources.

# Session roadmap

- Introduction
- Where Does VA Data Come From
- MedSAS: The Traditional Paradigm
- CDW: the New Paradigm
- MedSAS vs CDW Use-Case Examples



# Session roadmap

- Introduction
- Where Does VA Data Come From
- MedSAS: The Traditional Paradigm
- CDW: the New Paradigm
- MedSAS vs CDW Use-Case Examples

# Session roadmap

- Introduction
- Where Does VA Data Come From
- MedSAS: The Traditional Paradigm
- CDW: the New Paradigm
- MedSAS vs CDW Use-Case Examples

## VA Data: What is it for?

- Main purpose is healthcare operations

This includes

- billing
  - clinical care
  - decision support.
- 
- Health Services Research (HSR) “borrows” this data.

## VA Data: Where does it come from?

- Info from patient records → database by way of MUMPS VISTA programming at local VA sites.
- Data from local VISTAs is aggregated into national VA database- National Patient Care Database (NPCD).
- Data source used to be paper-based and is now electronic (Computerized Patient Record System - CPRS)

# Session roadmap

- Introduction
- Where Does VA Data Come From
- MedSAS: The Traditional Paradigm
- CDW: the New Paradigm
- MedSAS vs CDW Use-Case Examples

## Poll #3: Your experience with MedSAS Datasets

***How would you rate your overall knowledge of the VA Medical SAS datasets?***

1 Never Used MedSAS Datasets

2

3

4

5 Frequently used MedSAS



# Overview: Traditional MedSAS Data

- Originally, these were available to researchers only as SAS data sets on the Austin Automation Center mainframe system.
- Now, they are available in the VINCI Microsoft SQL Server, just as the CDW structures are, and despite the name, are manipulated with SQL instead of SAS.
- This makes the MedSAS data available to other analytic platforms on VINCI such as Stata, R and SPSS (as well as SAS).

# Overview: Traditional MedSAS Data

- Consists of Inpatient and Outpatient datasets:
- Inpatient (acute care):
  - Main: one record per stay (demographics, up to 26 diagnoses, discharge destination etc.)
  - Bedsection: a segment of a stay
  - PS: surgical procedures (in the OR, not available for observation care)
  - PP: other procedures (not in the OR)
- Similar structures for inpatient extended care, observation care and non-VA care.



# Traditional MedSAS Data: Inpatient Encounter

## Inpatient Encounter

- Inpatient professional services provided to patients during an inpatient stay (clinic stop codes, up to 10 diagnoses, 20 procedures etc.)

# Traditional MedSAS Data: Outpatient Encounter

## Outpatient

- SF: one record per visit day (demographics, list of clinic stops visited)
- SE: one record per clinic stop (up to 20 procedures and 10 diagnoses)

# Traditional MedSAS Data

HSR questions could be addressed by investigating:

- Patient demographics
- Geographic information
- Diagnosis histories
- Procedure histories
- Clinic stop histories (the ER officially counts as an outpatient clinic stop)

# Mortality: An Important Outcome

- Within hospital death was recorded in the inpatient encounter file (PM)
- Other deaths were recorded in the separate VHA Vital Status file
- For more details see the cyberseminar: [\*Ascertaining Veterans' Vital Status: VA Data Sources for Mortality Ascertainment and Cause of Death\*](#)

# Data Querying Tools

- Local (VISTA) data was created using MUMPS database language
- This was aggregated nationally (NPCD) in Oracle databases (using SQL as query language)
- Finally, selected fields of the Oracle data were translated into SAS versions of the NPCD.
- In general, MUMPS and (pre-CDW) SQL-based data querying were not available for research.

## Other Data Sources

- Pharmacy data:
  - Available on request for pre-defined patient cohorts from PBM (Pharmacy Benefits Management)
- Lab data:
  - Limited set of lab tests in DSS (now MCA)
  - Fuller lab test coverage in some disease-specific clinical case registries.

## Other Data Sources, continued

- So, pharmacy utilization in general or drug-specific were not available for general incorporation in HSR questions
- Lab testing (apart from actual lab results), as another utilization measure, was also not generally available for HSR.

# Session roadmap

- Introduction
- Where Does VA Data Come From
- MedSAS: The Traditional Paradigm
- CDW: the New Paradigm
- MedSAS vs CDW Use-Case Examples



## Poll #4: Your experience with CDW

***How would you rate your experience using CDW?***

1 Never Used CDW

2

3

4

5 Frequently used CDW



# A New Paradigm: CDW

- CDW is the VHA Corporate Data Warehouse
- Resides on Business Intelligence Service Line (BISL) servers
- Can be accessed by researchers on the VINCI system

# A New Paradigm: CDW

## Data include:

- Inpatient and outpatient encounters, diagnoses and procedures
- Lab histories
- Pharmacy histories
- Vital signs (height, weight, BP etc.)
- Many more

# A New Paradigm: CDW

- Data is housed in Microsoft SQL Server
- It is queried with SQL
- Involves extensive use of lookup tables (“Dim tables”)
- Constructed data can be exported to
  - SAS
  - R
  - Stata
  - And more ...

# CDW Data Layout Examples: Inpatient Data

- In inpatient MedSAS, all diagnosis data for a hospital stay is all in one observation (10 diagnosis code fields).
- In inpatient CDW, each diagnosis gets its own observation.
- This makes searching for diagnoses in the data easier.

# CDW Data Layout Examples: Outpatient Data

- In outpatient MedSAS, all diagnosis data for a clinic stop is all in one observation (10 diagnosis code fields).
- In outpatient CDW, each diagnosis gets its own observation.
- Again, this makes searching for diagnoses in the data easier.

# CDW and Lookup Tables: Diagnoses

- In MedSAS, diagnoses are given explicitly in ICD-9 form.
- In CDW, diagnoses are given by a CDW-specific code (an SID), which is decoded into ICD9 (or ICD10) by linking to a lookup (Dim) table.
- This looks cumbersome but ...

# CDW and Lookup Tables: Dim Tables

- Use of the information in the Dim table allows cross-reference to other attributes of the coded diagnosis than just the translation into ICD (9 or 10).
- Dim tables can be updated without disturbing the structure of the actual diagnosis data.



## More on CDW and Lookup Tables

- Similar use of Dim tables applies to procedures (with translation of procedure SIDs into ICD or CPT).
- Lab tests can be identified by names, LOINCs or other identification systems.
- Drug dispensings can be identified by names, NDCs or other identifiers.

# Using CDW and Lookup Tables

- Lookup Tables make for versatility.

## MedSAS vs CDW Use-case Example (1): HCV Diagnosed Subpopulation

- In this example, we define a cohort of patients diagnosed with hepatitis C (HCV)
- We then examine utilization in the 365 days prior to first HCV diagnosis.
  - All outpatient utilization
  - Visits to GI or hepatology clinics

# MedSAS vs CDW Use-case Example (1): HCV Diagnosed Subpopulation

In MedSAS...

- One searches over the **PM** inpatient and **SE** outpatient tables.
- Within an encounter, search across the 10 dx code fields to find occurrences of the HCV codes.
- To make this easier, one can spill all the dx codes into a vertical list (one dx code per data row) before searching for HCV codes

# MedSAS vs CDW Use-case Example (1): HCV Diagnosed Subpopulation (continued)

In MedSAS...

- To use in defining utilizations, save 3 fields for inpatient and for outpatient: patient identifier, first HCV code date and code count

# MedSAS vs CDW Use-case Example (1): HCV Diagnosed Subpopulation

In CDW...

- Create your own HCV code lookup table from the ICD9 or ICD10 Dim tables.
  - This will decode the CDW-specific dx code identifiers into ICD9 and ICD10 coding systems.
- It will only include HCV-relevant codes.

# MedSAS vs CDW Use-case Example (1): HCV Diagnosed Subpopulation

In CDW...

- Find the dx codes (already verticalized), one per row in **InpatientDiagnosis** and **VDiagnosis** tables.
- Link these records to your HCV code lookup table (so you can retain only the HCV code occurrences).
- No searching through the codes; it's done automatically!

# MedSAS vs CDW Use-case Example (1): HCV Diagnosed Subpopulation

In CDW...

- As before, to use in defining utilizations, save 3 fields for inpatient and for outpatient: patient identifier, first HCV code date and code count



# MedSAS vs CDW Use-case example (1): outpatient visits in the year before HCV diagnosis

In MedSAS...

- Link the HCV patient cohort with the SF table
- Search through the linked observations, making a count of visit days during the 365 period preceding the HCV “index” diagnosis
- Assign a count of 0 if no matching visit days are found

# MedSAS vs CDW use case example (1): outpatient visits in the year before HCV diagnosis

In CDW...

- Link the HCV patient cohort with the Visit table, restricting to those records during the 365 period preceding the HCV “index” diagnosis
- On this table, do a count query by patient (this is easy!)
- Assign a count of 0 if patient has no records in that table

## Compare: MedSAS vs CDW Use-case Example (1): Outpatient Visits in the Year before HCV Diagnosis

- Which one is easier?
- Which one takes fewer steps?

## MedSAS vs CDW Use-case Example (1): Outpatient Visits with GI or Hepatology Clinic Stops

- In MedSAS
  - Link the HCV patients with the SF table
  - Search through the clinic stop code fields looking for GI or hepatology clinic stops. Keep each one in a new table.
  - Go through the GI/hepatology clinic stop table, making a count per patient of the visit days during the 365 period preceding the HCV “index” diagnosis
  - Assign a count of 0 if no matching visit days are found)

## MedSAS vs CDW Use-case Example (1): Outpatient Visits with GI or Hepatology clinic (CDW)

In CDW...

- Three-way link between
  - The HCV patient cohort
  - The **Visit** table (that links patient and visit date)
  - The **Workload** table (that links visit date and clinic stop)
- Restricting to visits in the 365 days prior to HCV “index”
- Do a per-patient count query of visit dates.
- Assign a count of 0 if no matching visit days are found)

## MedSAS vs CDW Use-case Example (1): Outpatient Visits with GI or Hepatology Clinic Stops in the Year Before HCV diagnosis

- Which one is easier?
- Which one takes fewer steps?

# CDW: Lab and Pharmacy Utilization

- Using with SQL proceeds in an analogous way to what we've already seen
- As a partial alternative to CDW lab and pharmacy tables, the MCA tables retain the structure of the pre-CDW DSS datasets.
- Consult other cyberseminars in this series for details about CDW and MCA lab and pharmacy sources.

# In Conclusion

- MedSAS data is an earlier data paradigm that lives on within the VINCI SQL server.
  - Lots of legacy SAS programming is available for data construction in utilization research.
  - Many analysts find it easier to “think” in MedSAS terms.
- CDW data is a newer paradigm
  - It has wider data coverage
  - It takes advantage of powerful SQL querying methods.
  - However, some tasks that are easy in SAS can be more challenging in SQL.



# Additional Resources

---

The screenshot shows the VIREC INTRANET website. The header features the VIREC logo and the word "INTRANET" in a stylized font. A search bar is located in the top right corner. Below the header, the main content area is titled "VA INFORMATION RESOURCE CENTER (VIREC)". On the left side, there is a navigation menu with links: "VIREC Home", "VA/CMS Home", "About Us", "New Users of VA Data", "FAQs", "Acronyms", and "HelpDesk". The main content area is titled "Medical SAS Datasets Documentation" and includes an "Overview" section with a paragraph explaining the documentation's purpose. Below this is a "Data Documentation" section with a list of links: "Research User Guides", "Data Contents", "Frequencies", "Historical Variable Attributes", and "Historical Stop Codes". On the right side, there is a sidebar with a "Medical SAS Datasets" section containing links to "Overview", "Inpatient", "Inpatient Encounters", "Outpatient", and "Documentation". Below this is a "General Resources" section with links to "Data Access", "Data Sources", "Data Tools", "Data Topics", "Products & Services", and "Special Projects". A red-bordered box highlights a section titled "ICD-10 Transition" with the text: "Visit the 'Medical SAS Datasets ICD-10 Transition' page to learn more." At the bottom of the page, there is a footer with a list of links: "VIREC Home", "VA/CMS Home", "VHA Data Portal", "VA R&D Home", "VA HSR&D Service", "Viewers", "Sitemap", and "Contact Us".

**VIREC** *INTRANET*

Search All VA Web Pages   [Open Advanced Search](#)

**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](#)

VIREC Home | VA/CMS Home | VHA Data Portal | VA R&D Home | VA HSR&D Service | Viewers | Sitemap | Contact Us

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



Search All VA Web Pages  
   
[» Open Advanced Search](#)

**VA INFORMATION RESOURCE CENTER (VIREC)**

VIREC Home  
 VA/CMS Home  
 About Us  
 New Users of VA Data  
 FAQs  
 Acronyms  
 HelpDesk

## CDW Documentation

### Overview

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.

### Sign-up for Product News & Updates

E-mail the VIREC [HelpDesk](#) to receive notification of our new CDW products and product updates.

---

### Data Documentation

Expand each type of documentation below to view these resources.

- + **Getting Started with Using CDW**
- + **NEW! Factbooks**
- + **NEW! Domain Layout & Descriptions**
- + **Data Contents**
- + **Discrete Frequencies**
- + **Record & Null Counts**

#### CDW

- Overview
- Data Transition to CDW
- Documentation

---

#### ICD-10 Transition

Visit the "CDW ICD-10 Transition" page to learn more.

---

#### General Resources

- + Data Access
- + Data Sources
- + Data Tools
- + Data Topics
- + Products & Services
- + 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.vhadataportal.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.hsrld.research.va.gov/for\\_researchers/cyber\\_seminars/archives/video\\_archive.cfm?SessionID=101](http://www.hsrld.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=101)

# VIReC options for specific questions

## HSRData Listserv

- Community knowledge sharing
- ~1,200 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**

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

Houston, Texas

Peter.Richardson2@va.gov

**Next session:  
January 9, 2017  
1:00 PM Eastern**



## Database & Methods Cyberseminar Series

---

### VA Medicare Data (VA/CMS)

**Kristin de Groot, MPH  
VA Information Resource Center**

**[Register for session](#)**



# Appendix: Inpatient Use-Case Example

---

## MedSAS vs CDW Use-case Example (2): Subpopulation with Colorectal Resection

- In this example, we defined a subpopulation of patients receiving colorectal resection surgery.
- We examine 90-day readmissions after surgery.

## MedSAS vs CDW Use-case Example (2): Subpopulation with Colorectal Resection

In MedSAS...

- One searches over all encounters in the PS table
- Within an encounter, search across 5 surgery code fields to find occurrences of the colorectal resection codes
- To make this easier, one can vertically list all the surgery codes before searching for colorectal resection.
- To use in defining utilizations, save 2 fields: patient identifier, colorectal resection and discharge dates

## MedSAS vs CDW Use-case Example (2): Subpopulation with Colorectal Resection

In MedSAS...

- To use in looking for post-surgery utilizations, save 3 fields: patient identifier, colorectal resection surgery and discharge dates

## MedSAS vs CDW Use-case Example (2): Subpopulation with Colorectal Resection

In CDW...

- Create your own lookup table of colorectal resection codes from the ICD9 or ICD10 procedure Dim tables.

# MedSAS vs CDW Use-case Example (2): Subpopulation with Colorectal Resection

In CDW...

- Find the surgery codes are already verticalized, one per row in the [InpatientICDProcedure](#) table
- Link this table to your colorectal resection lookup table
  - This keeps only the colorectal resections
  - It decodes the CDW-specific surgery identifiers into ICD9 or ICD10 codes

## MedSAS vs CDW Use-case Example (2): Subpopulation with Colorectal Resection

In CDW...

- To use in defining post-surgery utilizations, save 3 fields: patient identifier, colorectal resection surgery dates and hospital stay discharge dates

# MedSAS vs CDW Example: 90-day Readmission for Patients with Colorectal Resection

- Which one is easier?
- Which one takes fewer steps?



# MedSAS vs CDW Example: 90-day Readmission for Patients with Colorectal Resection

In MedSAS...

- Link your colorectal resection patient cohort to the **PM** table
- Keep those with admit date within 90 days of the colorectal resection discharge.
- In case of multiples, keep the earliest readmit.

# MedSAS vs CDW Example: 90-day Readmission for Patients with Colorectal Resection

In CDW...

- Link your colorectal resection patient cohort to the **Inpatient** table (with admit and discharge information but not diagnoses and procedures), restricted to those with admit date within 90 days of the colorectal resection discharge.
- In case of multiples, keep the earliest readmit.

# MedSAS vs CDW Example: 90-day Readmission for Patients with Colorectal Resection

- Which one is easier?
- Which one takes fewer steps?