

2012 VIReC Database and Methods Cyber Seminar Series



2012 VIREC Database and Methods Cyber Seminar Series

Using VA Corporate Data Warehouse for Health Services Research

June 4, 2012

Polly Hitchcock Noël, PhD

VERDICT & UT Health Science Center at San Antonio

San Antonio, Texas

Laurel Copeland, PhD

Center for Applied Health Care – CTVHCS & Scott & White

Temple, Texas



Session Objectives

At the end of this session, the participant will be able to:

- **Identify data from the Corporate Data Warehouse that can be used for Health Services Research**
- **Describe limitations of using the Corporate Data Warehouse data**
- **Describe recent research that has used Corporate Data Warehouse data**



Session Outline

- **Overview of CDW**
- **Possible sources of error in CDW data**
- **Evaluations of CDW data**
- **Case study: CDW anthropometric data**
- **Case study: CDW blood pressure data**
- **Recommendations**
- **Where to Go for More Help**



Audience Poll

- **How would you rate your overall knowledge of the Corporate Data Warehouse (CDW)?**
 - 1 (No knowledge)
 - 2
 - 3
 - 4
 - 5 (Expert-level knowledge)



Audience Poll

- **Are you currently using and/or have you ever used data from the CDW for your research?**
 - **1** I have never used CDW data, but hope to use it for my research in the future
 - **2** I have designed a study that uses CDW data, but am awaiting funding decisions or IRB/data security approvals
 - **3** I am currently using data from the CDW for my research
 - **4** I have a paper that uses CDW data submitted for review.
 - **5** I have used CDW data in 1 or more articles that are published or in press.



White Board

- If you have never used the CDW, what do you want to learn about the CDW or what data do you want to get from the CDW?

Session Outline

- **Overview of CDW**
- **Possible sources of error in CDW vitals data**
- **Evaluations of CDW data**
- **Case study: CDW anthropometric data**
- **Case study: CDW blood pressure data**
- **Recommendations**
- **Where to Go for More Help**



Examples of research topics requiring information on “vitals” data

■ Quality of Care

- *Does patient BMI class predict receipt of recommended preventive screenings?*

■ Healthcare disparities

- *Is hypertension an independent predictor of survival disparity among African American and white veterans with schizophrenia?*

■ Implementation research

- *Can quality of chronic pain management be improved by implementation of the PCMH?*



Corporate Data Warehouse

- draws from several VHA clinical & administrative systems
- like a nationwide view of all CPRS / VISTA
- created to support administrative & research objectives
- historical data from FY 1999, current data added nightly
 - Warning: this means there is no static file as with Medical SAS datasets; studies may not be precisely replicable if data must be re-extracted
- additional domains will be added in future
- relational database, unlike some other national repositories in VHA



Data found in the CDW

See CDW metadata on the VA Intranet:

<http://vaww.cdw.med.va.gov/Pages/welcome.aspx>

As of May 2012, these data domains are available:

- Consults (10/1/1999 and forward)
- Health Factors (FY00+)
- Immunizations (FY00+)
- Mental Health [Mental health scores list: HSRData]
- PCMM (Primary Care Management Module)
- Vital Signs (FY00+)



Data found in the CDW

As of May 2012, these data domains are available:

- DSS National Data Extracts [no longer at AITC]
- Inpatient (PTF, Movement)
- Lab, Chemistry
- Outpatient Encounters (FY00+) [All ICD9 & CPT Codes]
- Pharmacy, Outpatient (FY00+)
- Patient demographics (All time)
- Staff demographics (All time)

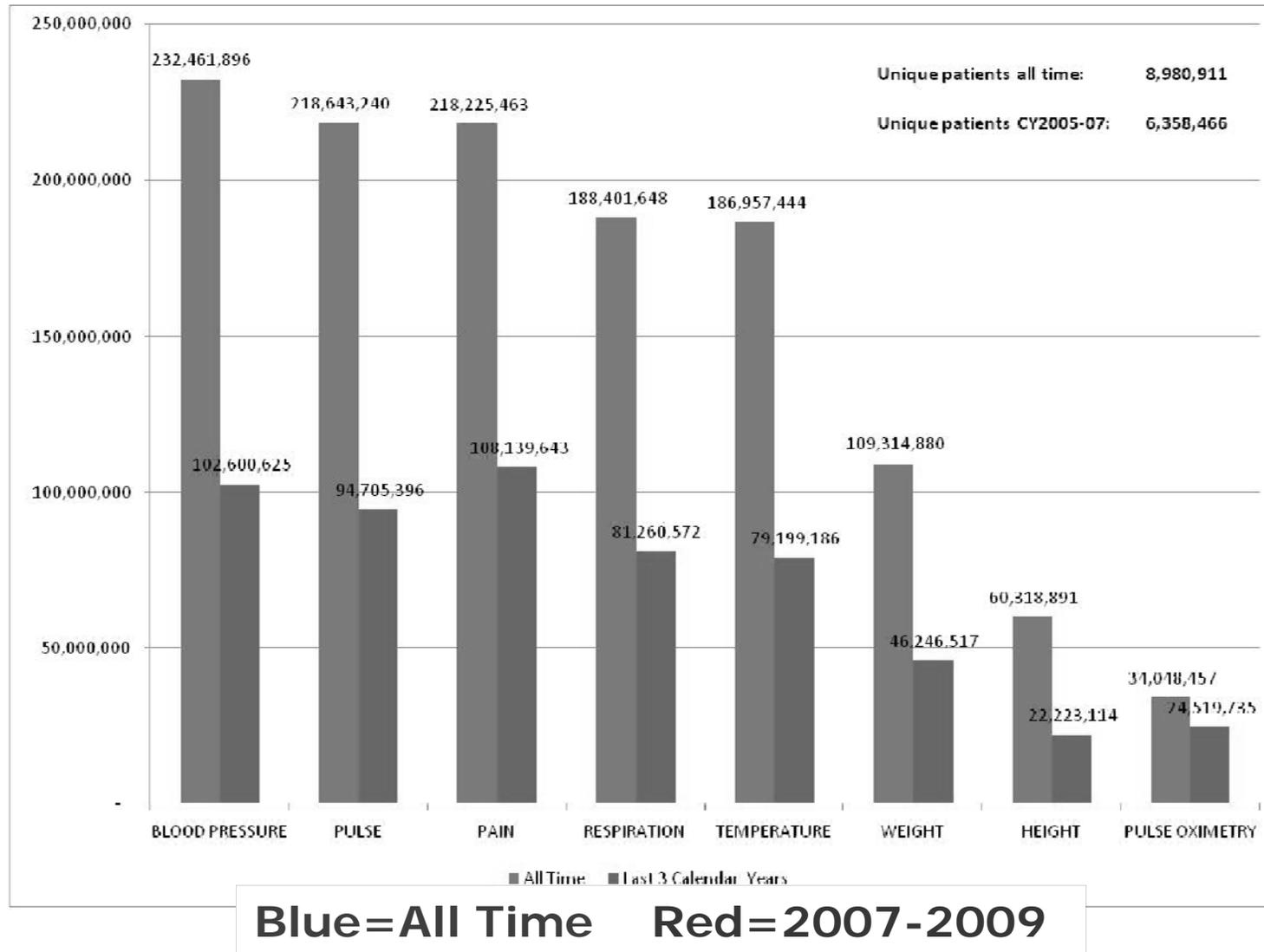


Vital Signs Available from CDW

(with abbreviations used in the CDW database)

- Audiometry (AU)
- Blood pressure (BP)
- Circumference/girth (CG)
- Central venous pressure (CVP)
- Fundal height (FH)
- Fetal heart tones (FHT)
- Hearing (HE)
- Height (HT)
- Pain (PA)
- Pulse oximetry (PO)
- Respiration (RE)
- Temperature (TE)
- Tonometry (TO)
- Vision corrected (VC)
- Ventilator minute volume (VMV)
- Ventilator tidal volume (VTV)
- Vision uncorrected (VU)
- Weight (WT)

Vital Sign History (VistA/RDW)



Perrin R, Bates J, Noel PH, Copeland LA, Lancaster B. National Clinical Data for VA Research: the VA Corporate Data Warehouse. Workshop. February, 2008 National HSR&D Meeting, Baltimore, MD.



Session Outline

- Overview of CDW
- Possible sources of error in CDW data
- Evaluations of CDW data
- Case study: CDW anthropometric data
- Case study: CDW blood pressure data
- Recommendations
- Where to Go for More Help



Tracing Path of CDW Data

- **Vitals data assessed & entered into CPRS**
- **Data stored in local VistA systems**
- **Transmitted by HL 7 messages to 1 of 4 Regional Data Repositories (RDW's)**
- **CDW extracts, transforms, and loads data from RDW's into its SQL data fields**



Tracing Path of CDW Data

- **CDW is regularly updated (refreshed)**
- **Does not hold stable reference files**
- **Errors & out of range values are not cleaned**

Possible Sources of Error

- **Measurement/Reporting**
- **Data Entry**
- **Data Transfer or Extraction**

Possible Sources of Error: Measurement/Reporting

- Equipment incorrectly calibrated
- Reliance on patient self-report vs. clinician measurement
- Values rounded “up” or “down”
- Patients measured inconsistently with or without shoes, clothing, etc.
- Last entered height carried forward
- Weights and/or heights less likely to be measured for specific patient populations

Possible Sources of Error: Data Entry

- **Transposition of numbers**
- **Keying of number adjacent to intended target**
- **Addition/deletion of digits**
- **Erroneous transformation of values**

Possible Sources of Error: Data Transfer or Extraction

- **Specific to the VA, the “same” anthropometric data entered into CPRS may exist at several levels**
 - local VistA systems
 - VISN data warehouses (directly from VistA or from a collector/feeder database on a SQL server connected to VistA)
 - CDW (from VistA via RDW)



Possible Sources of Error: Data Transfer or Extraction

- **Variations between sources can arise**
 - Data can be lost in transmission
 - Different filters can result in inclusion of slightly different subsets of data
 - Refresh differences - VistA/CDW constantly changing; updates not done simultaneously by all warehouses

Possible Sources of Error: Data Transfer or Extraction

- **Numeric data can be redefined as character data or “rounded” if stored with smaller number of decimal places**
 - Out-of-range & outlier values not cleaned in CDW, data appear in both text & numeric form
 - Text field “as is” from VistA extraction
 - Numerical field based on a very conservative transformational algorithm

Possible Sources of Error: Data Transfer or Extraction

■ Programmer error

- Programmer may misinterpret variables or cases needed or investigator may mis-specify needs
- Research team may fail to request variable to differentiate inpatient from outpatient measurements

Session Outline

- **Overview of CDW**
- **Possible sources of error in CDW data**
- **Evaluations of CDW data**
- **Case study: CDW anthropometric data**
- **Case study: CDW blood pressure data**
- **Recommendations**
- **Where to Go for More Help**



Evaluation of CDW Data: Administrative Project

- **VHA Support Service Center (VSSC)**
 - 1) Compared CDW data fields to one another and to overall patient utilization over time to see if populated as expected
 - 2) Examined data for biologically implausible values
 - 3) Compared data from 10 facilities (CDW vs. VISN warehouse)

Evaluation of CDW Data: are CDW data fields populated as expected over time?

Comparison of VHA utilization and CDW data over 4 years				
	2004 N	2005 N	2006 N	2007 N
Unique Patients	4,976,773	5,094,494	5,188,836	5,230,452
Blood Pressure	31,598,526	33,029,724	34,293,140	35,276,799
Circumference/Girth	11,004	16,933	82,482	125,577
Central Venous Pressure	17,344	16,018	64,988	79,737
Height	8,534,729	8,521,504	7,435,312	6,266,231
Pain	32,130,438	34,926,122	36,150,234	37,062,161
Pulse Oximetry	3,233,096	3,869,444	7,746,200	12,903,893
Pulse	29,468,149	30,613,107	31,590,314	32,501,113
Respiration	25,680,395	26,623,272	27,117,212	27,519,501
Temperature	25,358,389	26,100,279	26,382,625	26,715,839
Weight	14,764,754	15,258,657	15,490,210	15,497,385

Adapted from Perrin R, Bates J, Noel PH, Copeland LA, Lancaster B. National Clinical Data for VA Research: the VA Corporate Data Warehouse. Workshop. February, 2008 National HSR&D Meeting, Baltimore, MD.



Evaluation of CDW Data: Research Project

- **HSR&D-funded IIR (*Obesity Care Practices in Veterans Healthcare Administration*)**
 - 1) Compared number & values of CDW vs. VISN warehouse height & weight data
 - 2) Examined im plausible variation in repeated measurements recorded in the CDW on the same day, and on the same year, for the same individuals



Evaluation of CDW Data: CDW vs. VISN data warehouse

Percent concordance in # of wt records for N* individuals appearing in both CDW & VISN data warehouses for 3 Fiscal Years (FY)

	VISN A	VISN C	VISN E	VISN F
	N (%)	N (%)	N (%)	N (%)
FY2002	97,375 (98.6)	81,125 (62.3)	106,010 (99.3)	158,088 (99.5)
FY2004	34,815 (98.3)	31,988 (98.5)	47,572 (99.2)	77,233 (99.4)
FY2006	30,812 (97.5)	28,289 (97.5)	41,944 (98.4)	68,592 (98.2)

*Note: FY2002 data based on patients with > 1 pc visits in FY2002, FY2004 & FY2006 data based on pc patients identified as obese in FY2002



Evaluation of CDW Data: Implausible variation in repeated measurements

Frequency distribution of differences in minimum & maximum CDW hts recorded same day & same year (FY2006)		
Difference in inches	occurrences of individuals with ≥ 2 hts recorded same day N=33,424	individuals with ≥ 2 hts recorded same year N=539,489
	N (%)	N (%)
difference = 0	29,319 (87.72)	469,226 (86.98)
0 < difference \leq 1	1,771 (5.30)	42,186 (7.82)
1 < difference \leq 2	764 (2.28)	14,302 (2.65)
2 < difference \leq 10	1,207 (3.61)	11,063 (2.05)
Difference > 10	363 (1.09)	2,712 (0.50)

Overall findings from administrative & research project evaluations

- **Weights $>$ Heights $>$ WC**
- **Some anomalies identified & corrected as result of administrative project**
- **Concordance between the number & values of hts & wts stored in CDW & VISN warehouses generally 97.5%-99.9%**
- **Biologically implausible values noted, as well as biologically improbable variation**



Summary

- In spite of occasional anomalies, CDW appears to reliably reflect hts & wts recorded in VistA/VISN-level sources
- Probable data errors appear to be present in both data sources
- CDW valuable & useful source of nationwide vital sign & anthropometric data

Noel et al., JRRD 2010;47(8):739-250



Session Outline

- Overview of CDW
- Possible sources of error in CDW data
- Evaluations of CDW data
- Case study: CDW anthropometric data
- Case study: CDW blood pressure data
- Recommendations
- Where to Go for More Help

Case Study: Anthropometric Data

Obesity Care Practices in the Veterans Health Administration (HSR&D IIR 05-121)

Study Objectives

- describe variations in VHA obesity care practices
- identify factors that predict these variations
- examine the impact of care practices on BMI trends & other clinical outcomes

Noel et al., JGIM 2010; 25(6):510-516.

Challenges

- **Height (& weight) data inherently variable over time**
- **To detect changes in height & weight that are “real” vs. those due to error**
- **When errors detected or suspected**
 - find valid/consistent way to eliminate or neutralize OR to understand how it may affect findings/interpretation

Methods

Original plan:

- **Identify obese cohort BMI ≥ 30 in FY2002 (and follow through FY2006):**
 - last recorded weight in FY02
 - last recorded height in FY02 (or 1st available in FY03-FY06)
- **Included use of filtering scheme to remove “biologically implausible” weights (Das et al., 2005)**
 - deleted all weights ≥ 700 lbs & ≤ 75 lbs
 - deleted all heights ≥ 84 in & ≤ 48 ins

Limitations of Data: Impact on Cohort Identification

Patients with:	N (%)
≥ 1 primary care visits in FY2002 in 6 VISNs	1,053,228
≥ 1 wts & ≥ 1 hts recorded in FY2002	844,066 (80.1%)
≥ 1 wts in FY2002 & ≥ 1 hts in FY2003-FY2006	89,018 (8.5%)
sufficient ht & wt data to calculate BMI	933,084 (88.6%)

Limitations of Data: Impact on Cohort Identification

Due to extreme BMI values, modified plan to id obese cohort:

- **Weights**

- divided baseline FY into quarters
- calculated median weight for each quarter

- **Heights**

- used modal ht FY2002-FY2006
- for patients with ≥ 2 or more modal hts, averaged if difference btw modes was ≤ 3 ins

- **Baseline BMI**

- Used median of median wts for FY2002 & modal ht FY2002-2006

Limitations of Data: Impact on Outcome Assessment

Original plan:

- **BMI s across 5-year study period**
 - BMI based upon average quarterly wt for each of 20 quarters

Modified plan:

- **BMI s across 5-year study period**
 - BMI based upon median quarterly wt for each of 20 quarters
 - same modal ht (from FY2002-FY2006) applied across 5 years

Session Outline

- Overview of CDW
- Possible sources of error in CDW data
- Evaluations of CDW data
- Case study: CDW anthropometric data
- Case study: CDW blood pressure data
- Recommendations
- Where to Go for More Help



Blood Pressure Data

- **Publications using CDW data becoming more numerous; no specific example for BP**
- **Kerr et al J Qual Improv 2002; 28:555-565 used BP values from VA Ann Arbor from VistA**
- **Valid values for QUERI-DM:**
 - 75 <= systolic <=250
 - 25 <= diastolic <=180
 - Usually use last value in period



Late-life Patients with Schizophrenia

Of **2.5M records** for BP from FY02-FY05, representing 36,876 unique patients [IIR-05-326]:

- **14,000 w/o valid numeric data for both diastolic and systolic pressures**
- **0.7% had extreme non-missing values**
 - systolic 0-39 (n=9,081), systolic 0-75 (n=12,956)
 - diastolic 0-10 (n=13,396), diastolic 0-25 (n=14,070)
- **Very low values were primarily 0**
 - Systolic : n=8,927 of 12,956; diastolic: n=13183 of 14070
- **Few high values** (n=119 sys 250-299, n=13 dias 180-299)

Late-life Patients with Schizophrenia

the cohort originally matched to CDW : N=39,226
the cohort with valid BP & location in CDW: N=33,271

LOCATION:

- 70% of readings from inpatient encounters
- 30% of readings from outpatient encounters
 - (6% were missing location)
- 11,606 inpatients with valid location + valid BP
- 32,069 outpatients with valid location + valid BP



Case Study: Blood Pressure Data

Is hypertension an independent predictor of survival among black and white patients with schizophrenia?

Depends on what you look at.

Average measures generally protective (inverse association) or n.s.
 Maxima were positively associated with relative odds of death
 Minima were inversely associated with relative odds of death

Effect	OR	
Black	0.73	0.68 - 0.78
Female	0.59	0.49 - 0.70
Age10	2.09	2.03 - 2.15
Max OP Sys x10	1.03	1.02 - 1.04
<i>... in another model --</i>		
Min IP Dias x10	0.79	0.76 - 0.82
<i>... in another model --</i>		
Avg OP Dias x10	0.86	0.90

NOTE: PROC LOGISTIC is modeling the probability that died=1 (n=34094)



0.82 -

Session Outline

- **Overview of CDW**
- **Possible sources of error in CDW data**
- **Evaluations of CDW data**
- **Case study: CDW anthropometric data**
- **Case study: CDW blood pressure data**
- **Recommendations**
- **Where to Go for More Help**



Recommendations

- Consider pros & cons of using repository of data derived from routine clinical encounters
- Strategies may be helpful for controlling or minimizing impact of error - impossible to eliminate all
- “Missing” data may limit usefulness of CDW for certain types of questions or patient populations
- Incumbent upon researchers to be aware of limitations of data & potential impact on findings



Recommendations

Carefully examine each data file supplied by any repository

- **Do number of records and values make clinical sense?**
- **Are there unusual variations over time or by facility?**
- **Many measures to choose from – consider clinical implications of each**



Recommendations

Approaches for dealing with error or “missing” data may vary depending upon clinical population or research question

- **Select range filters that make sense for particular patient population**
- **Delete same day measurements or inpatient measurements?**
- **Is measure needed for specific time frame and/or are repeated measurements needed over time?**



Suggestions When Requesting CDW Data

■ Programming Issues

- SQL view is simplest option for transferring (if SQL programmer on study team)
- Be specific in request & re-iterate specifications if time passes or CDW programmer changes
- Research programmer may require guidance on which cases to throw out or how to summarize multiple values in meaningful way

■ Other Issues

- Because of the size of these files, assess server space needs in advance of receipt of data; how many copies in what formats will you need to create?

How to Acquire CDW Extracts

- **Start with DART**
- **The NDS website on DART [Data Access Request Tracker] -- (<http://vaww4.va.gov/NDS/DataAccess/Researchers/RESCDW.asp>) describes these 2 methods for accessing CDW data:**
- **Note: Data access forms** include Form 9957, Privacy and DSS Non-Disclosure if required, CDW Rules of Behavior
- **Note: Other documents required** may include IRB Approval Memo, R&D Approval Memo, Research Request Memo, IC/HIPAA Waiver Approval(s)
- **Data used in VINCI** – Data Access Agreement, Rules of Behavior
- **Data delivered as an extract for use on local secure VA research servers** – Data Access forms, CDW Rules of Behavior
 - NDS expects the PI to control the data locally and track who has access to the data. If research staff will assist in obtaining the extract by downloading or file transfer from the CDW platform, these individuals will need to complete the access forms and sign CDW Rules of Behavior.



References for Further Info

- Noël PH, Copeland LA, Perrin RA, Lancaster AE, Pugh MJ, Wang CP, Bollinger MJ, Hazuda HP. VHA Corporate data warehouse height and weight data: Opportunities and challenges for health services research. JRRD. 2010;47(8):739-750.
- Noël PH, Copeland LA, Pugh MJ, Kahwati L, Tsevat J, Nelson K, Wang CP, Bollinger M, Hazuda HP. Obesity Diagnosis and Care Practices in the Veterans Health Administration. JGIM 2010;25(6):510-516.
- Noël PH, Wang CP, Bollinger MJ, Pugh MJ, Copeland LA, Tsevat J, Nelson KM, Dundon MM, Hazuda H. Intensity and duration of obesity-related counseling: Association with 5-year BMI trends among obese primary care patients. Obesity 2012;20(4):773-782.
- Copeland LA, Pugh MJ, Hicks PB, Noël PH. Use of obesity-related care by psychiatric patients. Psychiatric Services 2012;63(3):230-236.

Session Outline

- **Overview of CDW**
- **Possible sources of error in CDW data**
- **Evaluations of CDW data**
- **Case study: CDW anthropometric data**
- **Case study: CDW blood pressure data**
- **Recommendations**
- **Where to Go for More Help**



Research Data Request Process to Obtain Data from CDW

- Visit CDW site to confirm data availability
- Request access through VHA National Data Systems / DART
 - <http://vaww4.va.gov/NDS/DataAccess/Researchers/RESCDW.asp>
 - Contact the help personnel at that URL or within DART
 - Process involves a review by the VHA Privacy Office
 - Process requires patience



Current Research Data Request Process

- **Once approved by VHA**
 - Arrange data set creation with CDW team
 - Establish File format (e.g., SQL, .txt)
 - Honor your agreements regarding data safekeeping
 - Do not delete research data without first obtaining your IRB, PO, ISO, and Research Office approval (data destruction embargo)
- **Is direct access to CDW possible? No.**
- **Secure transfer of data**



VIREC Help

- **VIREC Webpage**

- <http://vaww.virec.research.va.gov/>

- <http://www.virec.research.va.gov>

- Information on VA data sources and how to access data, including CDW
- Documentation on some VA datasets
- HSRData Listserv sign-up_{info}

- **CDW Information Webpages**

- <http://vaww.cdw.med.va.gov/Pages/welcome.aspx>

- <http://vaww.virec.research.va.gov/DataSourcesName/CDW/CDWdocumentation.htm>

- <http://vaww.dwh.cdw.portal.va.gov/Pages/welcome.aspx>



VIReC Help (cont'd)

■ HSRData Listserv (join this!)

- Join at the VIReC Web site
- http://vaww.virec.research.va.gov/Support/HSRData-L_Listserv.htm
- Discussion among >500 data stewards, managers, and users including Sophie Lo, Elliott Lowy, Yiwen Yao
- Past messages in Archive (on intranet)

■ VIReC Help Desk

- VIReC staff will answer your question and/or direct you to available resources on topics
- (708) 202-2413 [8:00am to 4:30pm Central]
- VIReC@va.gov



Questions?

Polly.Noel@va.gov

Laurel.Copeland@va.gov



Session Outline

- **Overview of CDW**
- **Possible sources of error in CDW data**
- **Evaluations of CDW data**
- **Case study: CDW anthropometric data**
- **Case study: CDW blood pressure data**
- **Recommendations**
- **Where to Go for More Help**

