

2009-2010 VIREC Database and Methods Cyber Seminar Series



Using VA Corporate Data Warehouse to Assess Vitals

Session 11
March 1, 2010

Presented by:
Polly Hitchcock Noël, PhD
Laurel Copeland, PhD
VERDICT REAP & UTHSCSA
San Antonio, Texas



Session Objectives

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

- **Identify current approaches to assessing vitals using the Corporate Data Warehouse**
- **Describe limitations of using the Corporate Data Warehouse to assess vitals**
- **Describe recent research and methods used to assess vitals that have used the Corporate Data Warehouse**



Session Outline

- **Overview of CDW**
- **Possible sources of error in VA anthropometric data**
- **Evaluation of quality of CDW anthropometric data**
- **Case study of research using CDW anthropometric 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)



White Board

- **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 VA anthropometric data**
- **Evaluation of completeness & accuracy of CDW anthropometric data**
- **Case study of research using CDW anthropometric data**
- **Recommendations**
- **Where to Go for More Help**



Examples of research topics requiring information on vital signs & anthropometric data

■ Quality of Care

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

■ Healthcare disparities

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

■ Implementation research

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

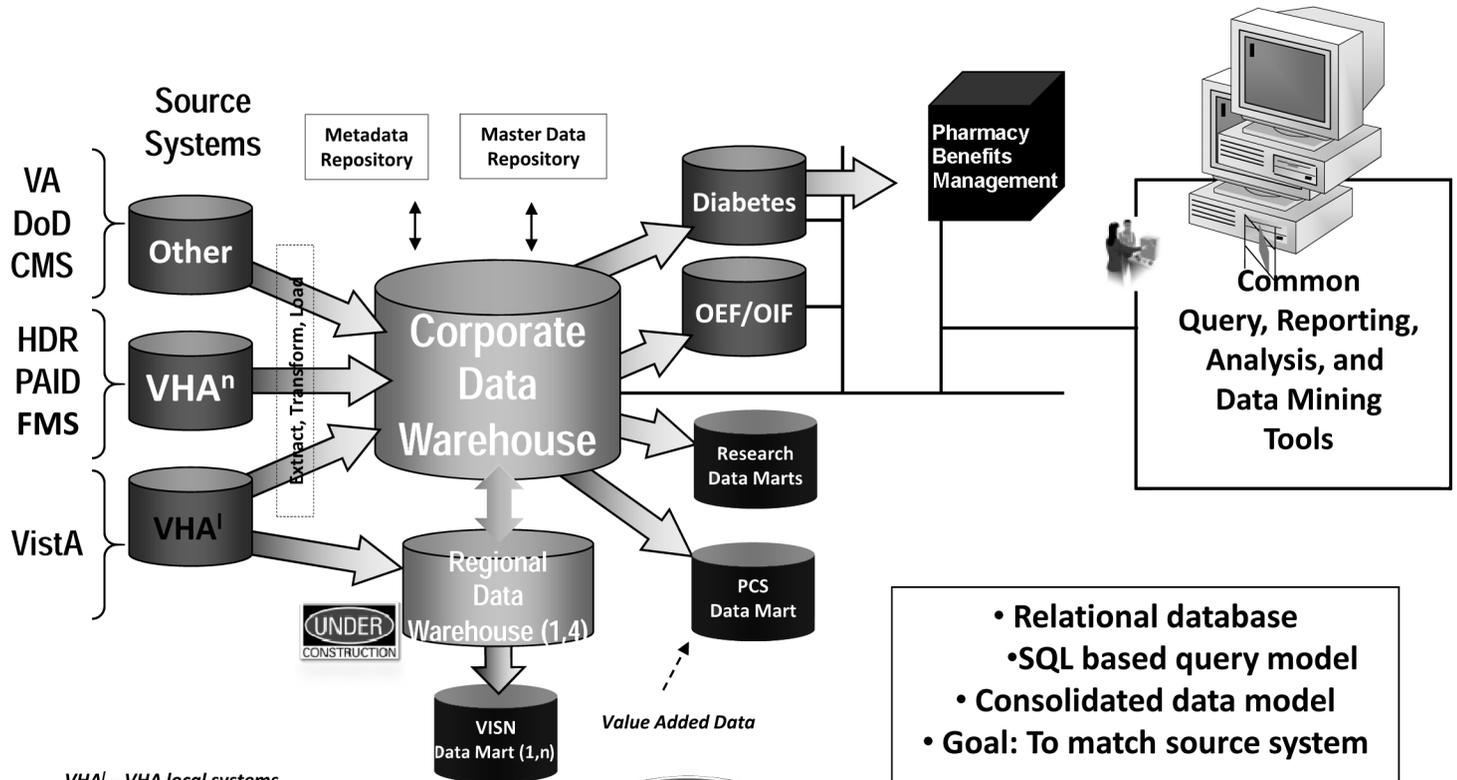


Corporate Data Warehouse

- **draws from several VHA clinical & administrative systems**
- **created to support administrative & research objectives**
- **historical data from FY 1999, current data added nightly**
- **additional domains will be added in future**
- **relational database, unlike some other national repositories in VHA**



VHA Data Warehousing Architecture



VHA^l – VHA local systems
VHAⁿ – VHA national systems



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.

Data found in the CDW

- Patient demographics (VistA/All time)
- Staff demographics (VistA/All time)
- Outpatient encounters (NPCD/10/1/1999)
 - All ICD9 Codes
 - All CPT Codes
- Vital Signs (HDR/From 10/1/1999)
- Allergies (HDR/From 10/1/1999)
- Outpatient Pharmacy (HDR/From 10/1/1999)
 - Loading
- Surgical Procedures (VistA/All time)
- Non VA Meds (VistA/All time)
- Prosthetics (VistA/All time)

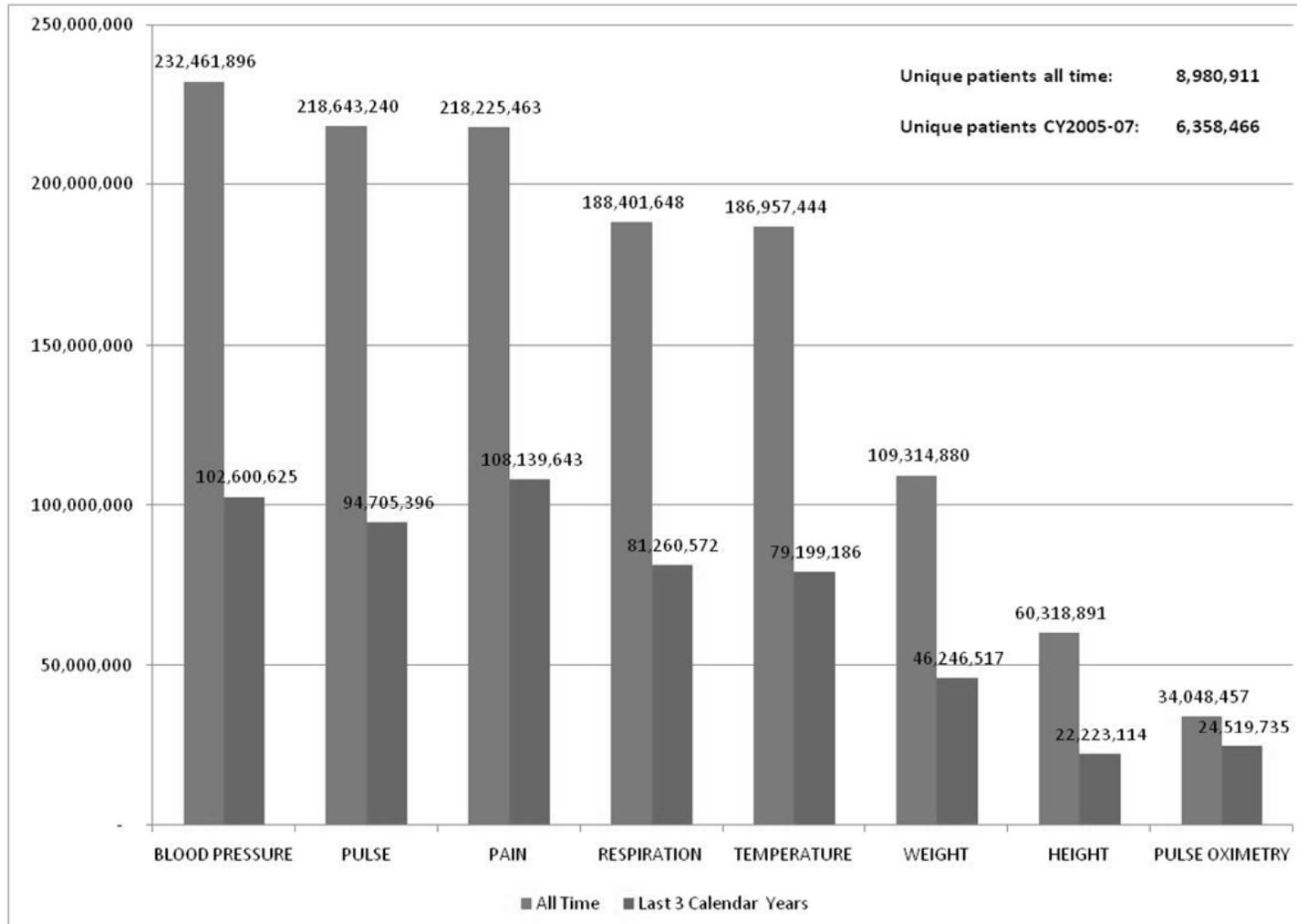


Vital Signs Available from CDW

(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/HDR)



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 VA anthropometric data**
- Evaluation of completeness & accuracy of CDW anthropometric data
- Case study of research using CDW anthropometric data
- Where to Go for More Help



Tracing Path of CDW Data

- **Vital sign & anthropometric data assessed & entered into CPRS**
- **Data stored in local Vista systems**
- **Transmitted by HL 7 messages to Health Data Repository (HDR)**
- **CDW extracts, transforms, and loads data from HDR 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 HDR)



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 VA anthropometric data
- **Evaluation of completeness & accuracy of CDW anthropometric data**
- Case study of research using CDW anthropometric data
- Recommendations
- Where to Go for More Help



Evaluation of CDW Data Quality: 1st Example

■ 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 Quality: are CDW data fields populated as expected over time?

Comparison of anthropometric vs. other 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 Quality: 2nd Example

- **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 implausible 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 Quality: 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 Quality: 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)
$0 < \text{difference} \leq 1$	1,771 (5.30)	42,186 (7.82)
$1 < \text{difference} \leq 2$	764 (2.28)	14,302 (2.65)
$2 < \text{difference} \leq$ 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**



Session Outline

- **Overview of CDW**
- **Possible sources of error in VA anthropometric data**
- **Evaluation of completeness & accuracy of CDW anthropometric data**
- **Case study of research using CDW anthropometric data**
- **Recommendations**
- **Where to Go for More Help**



Study Objectives

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

- 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

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:

- **id 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 FYs into quarters
- calculated median weight for each quarter

■ **Heights**

- used modal ht
- 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

Modified plan:

- **BMI s across 5 year study period**
 - median (instead of average) 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 VA anthropometric data**
- **Evaluation of completeness & accuracy of CDW anthropometric data**
- **Case study of research using CDW anthropometric 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?**



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?

References for Further Info

- Noël PH, Copeland LA, Pugh MJ, Kahwati L, Tsevat J, Nelson K, Wang CP, Bollinger M, Hazuda H. Obesity Diagnosis and Care Practices in the Veterans Health Administration. JGIM (in press).
- 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. (under review)



Session Outline

- **Overview of CDW**
- **Possible sources of error in VA anthropometric data**
- **Evaluation of completeness & accuracy of CDW anthropometric data**
- **Case study of research using CDW anthropometric data**
- **Recommendations**
- **Where to Go for More Help**



Current Research Data Request Process to Access/Obtain Data from CDW

- **Contact CDW to confirm data availability**
- **Request access through VHA National Data Systems**
 - Contact Irma McCaffrey (Irma.McCaffrey@va.gov)
 - Process involves a review by the VHA Privacy Office



Current Research Data Request Process

- **Once approved by VHA**
 - Arrange data set creation with CDW team
 - Establish Data Transfer Agreement (DTA) with CDW team
 - File format (e.g. SQL, .txt)
- **Secure transfer of data**



VIReC Help

- **VIReC Webpage**

<http://www.virec.research.va.go>

- Information on VA data sources and how to access data
- Documentation on some VA datasets

- **CDW Meta-data Webpage**

<http://vhacdwa09.vha.med.va.gov:8080/>



VIReC Help (cont'd)

■ HSRData Listserv

- Join at the VIReC Web site
- Discussion among >500 data stewards, managers, and users
- Past messages in archive (on intranet)

■ VIReC Help Desk

- VIReC staff will answer your question and/or direct you to available resources on topics
- VIReC@va.gov
- (708) 202-2413



Questions?

