



Overview of Data & Analytics at the Department of Veterans Affairs

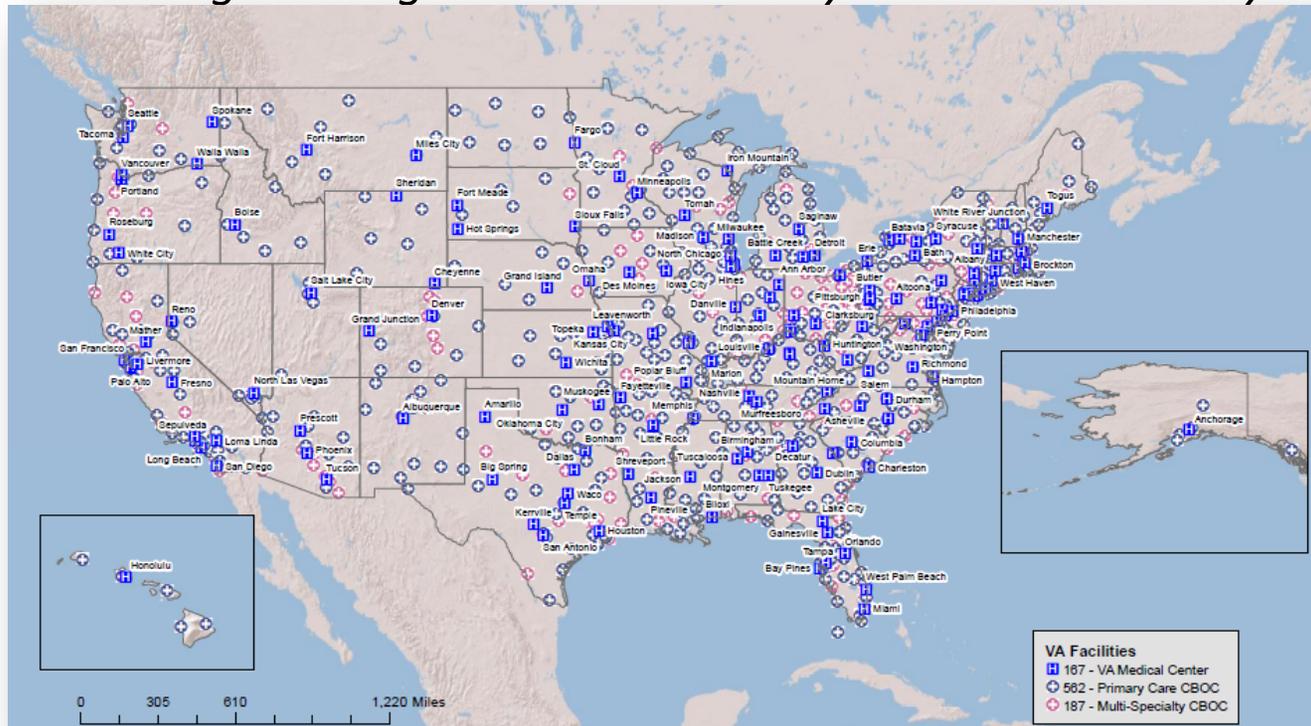
Jack Bates

VA Office of Information & Technology

August, 2016

Veterans Health Administration (VHA)

The Largest Integrated Healthcare System in the Country



VHA Points of Care (1,748)

- Integrated Healthcare Networks: 21
- Major Medical Centers: 152
- Outpatient Clinics: 990
- Vet Centers: 370
- Domicillaries: 102
- Community Living Centers: 134

Patient Population

- Enrollees: 8.8M
- Active Patients: 6M
- All Time Patients: 22M
- FY15 Outpatient Visits: 84M
- FY15 Inpatient Admissions: 703K

VA Electronic Health Record (EHR)

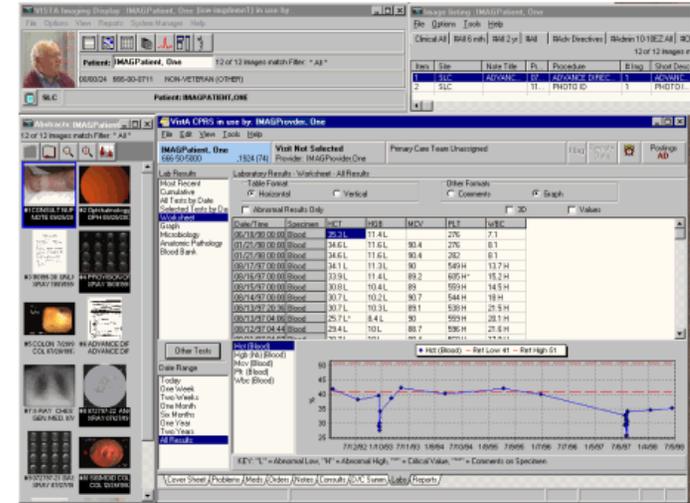
EHR – Provider Generated Data

- **Vista (Veterans Health Information Systems and Technology Architecture)**

- An integrated Electronic Health Record (EHR) data system with application modules that share a common data store and common internal services
- There are **130 hospital based instances of Vista**
- There are approximately 200 modules comprising Vista
 - Clinical Modules (e.g. Lab, Pharmacy, Radiology, Vital Signs)
 - Administrative Modules (e.g. Enrollment, Scheduling, Admission-Discharge-Transfer, Outpatient Encounters)
 - Financial Modules (e.g. Payroll, Beneficiary Travel, Fee Basis)

Horizon:

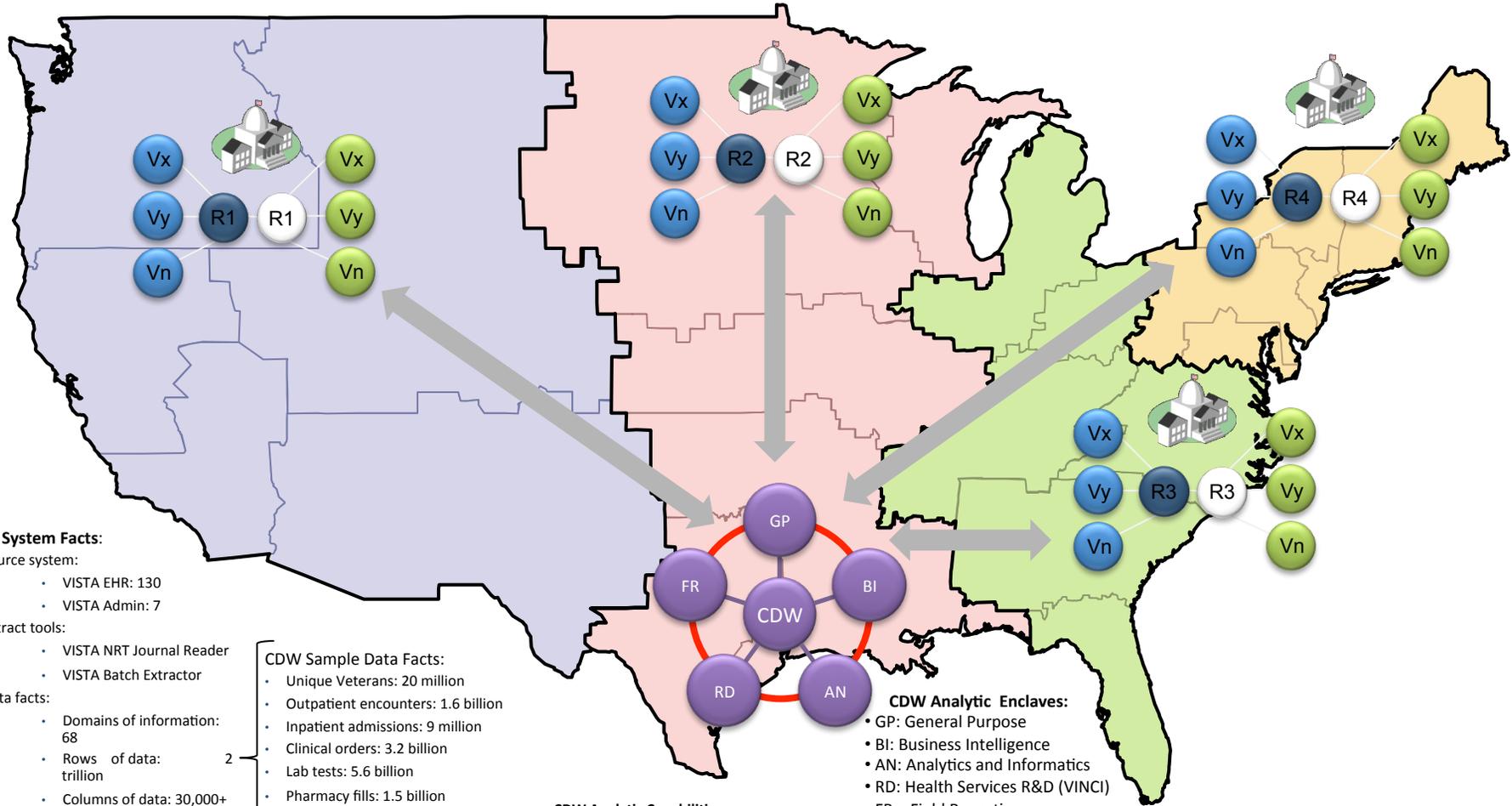
- PGD – Patient Generated Data (Mobile, Web, Kiosk)
- MGD – Machine Generated Data (Connected Devices, IoT)





VA Analytic Ecosystem

Common Data ♦ Common Infrastructure ♦ Common Tools ♦ Common Security



CDW System Facts:

- Source system:
 - VISTA EHR: 130
 - VISTA Admin: 7
- Extract tools:
 - VISTA NRT Journal Reader
 - VISTA Batch Extractor
- Data facts:
 - Domains of information: 68
 - Rows of data: trillion
 - Columns of data: 30,000+
 - Tables of data: 900+
- Data quality program
- Active Users: 80,000/Month
- Vibrant user community

CDW Sample Data Facts:

- Unique Veterans: 20 million
- Outpatient encounters: 1.6 billion
- Inpatient admissions: 9 million
- Clinical orders: 3.2 billion
- Lab tests: 5.6 billion
- Pharmacy fills: 1.5 billion
- Radiology procedures: 162 million
- Vital signs: 2.3 billion
- Text notes: 2.0 billion

CDW Analytic Capabilities:

- Primary/Secondary/Data Mart Structures
- Data Standardization
- Metadata Services
- Business Intelligence Reporting & Dashboards Tools
- Geospatial Mapping Tools and Images
- SAS/Grid High Performance Compute Grid
- Natural Language Processing Engines
- Hadoop Cluster

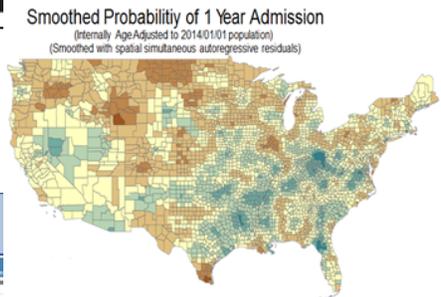
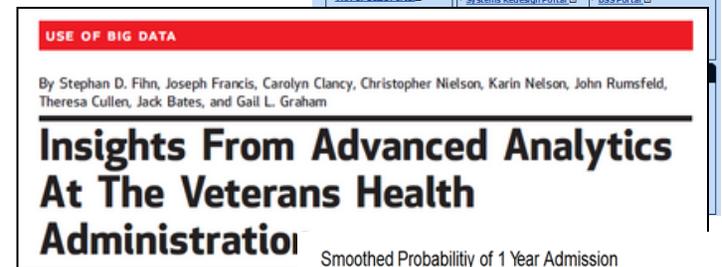
CDW Analytic Enclaves:

- GP: General Purpose
- BI: Business Intelligence
- AN: Analytics and Informatics
- RD: Health Services R&D (VINCI)
- FR – Field Reporting

Corporate Data Warehouse (CDW)

Facts & Figures

- (>16 years) Longitudinal data on >22.3M veterans
 - > 200 million rows of data/day → 2.4 trillion - (80% of all VA data, doubling annually)
 - Business : Accounting (AP/AR/GL/MCA), Acquisition/ Logistics, Assets, HR, Emergency Management,
 - Clinical : VistA (100%), EDIS, MVP Genomic
 - Veteran: Patient generated, Enrollment, Demographics, Telephony, CRM, Survey
- **80,000 users throughout VA**
 - >650 user-driven analytic solutions
 - 125,000 reports and 2,500 dashboards
 - 125 training events in FY15
 - e.g. BI, Analytics, Data Mgmt, Best Practices
 - Community Collaboration/Crowdsourcing
 - Subject Matter Expertise
 - Sharing of lessons learned & best practices
 - Data sharing (e.g. Data Object Exchange)



V20 Opioid Risk - Patient Detail Report

Data as of: 5/28/2014 6:09:00 AM

Client	PCP	Resident Status	VETIC	Preferred Name	Local	Age	Sex
PUGET SOUND	NOI						

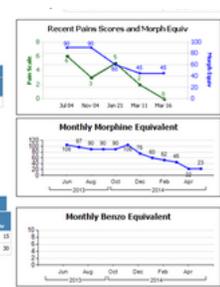
Month/Year	Prescription Fills	Weggs	FY13	ADULT C	12 Flag Active	12 High Risk ID	Drug Symbols	Adviser	CD Status
7/08	176								

First Date	Second Date	Third Date	Fourth Date	Fifth Date	Sixth Date	Seventh Date	Eighth Date	Ninth Date	Tenth Date
5/28/14	5/28/14	5/28/14	5/28/14	5/28/14	5/28/14	5/28/14	5/28/14	5/28/14	5/28/14

Prescription	Last Order	Last Date	Last Qty	Last PC	Last MOP	Last SATP
ORIPROXIFEN	5/28/14	5/28/14	5/28/14	5/28/14	5/28/14	5/28/14

Current Active Opioid & Benzodiazepine (If applicable) Prescriptions	Quantity	Days Supply	Equivalency
Local Disp Name: WEGG DISP	412/14 EXPIRED	28	28
WEGG DISP: WEGG DISP	5/28/14 DISCONTINUED	28	28

Total Disp	Equivalency	Days Supply	Prescription
412/14 EXPIRED	28	28	28
5/28/14 DISCONTINUED	28	28	28



CDW EHR Based Data Examples

Patients: 22 M/16 years		
Lab Results 7.7B	Clinical Orders 4.5B	Immunizations 71 M
Pharmacy Fills 2.2B	Clinical Notes 3.2B	Health Factors 2.2B
Radiology Proc 202 M	Vital Signs 3.3B	Consults 315 M
Appointments 1.4B	Encounters 2.4 B	Admissions 17 M

Production

- Allergy
- Appointment
- Consult
- CPRS Orders
- Data Profiling
- Dental
- Health Factors
- Immunization
- Inpatient
- Lab Microbiology
- LabChem
- Mental Health Assessment
- Non-VA Meds
- Outpatient
- Patient
- Patient Associated
- Patient Enrollment
- PCMM (Primary Care Management Module)
- Pharmacy BCMA (Bar Code Medication Administration)
- Pharmacy Outpatient
- Pharmacy Patient
- Purchased Care (formerly fee)
- Pyramid GIS
- Recall Reminders
- SPatient
- Staff
- SStaff
- Surgery PRE Table (Not the Surgery RAW domain)
- VistA Compensation & Pension
- VistA Waitlist
- Vital Signs

RAW*

- Bill Claims (CBO)
- CAPRI Audit Trail table
- CliniComp
- Echocardiogram
- Emergency Dept. Int. Software (EDIS)
- Equipment Inventory
- FBCS (Fee Basis Claim System)
- IFCAP (Integrated Funds Control, Accounting, and Procurement)
- Intravenous meds (IV)
- Oncology
- PAID (Personnel and Accounting Integrated Data System)
- Prosthetics
- Pulmonary Function Test (PFT)
- Radiology
- RxUD (Unit Dose)
- Surgery
- Travel
- VACAA (Veterans Choice Program Eligibility)

CDW Use Case Examples

- Operational Business Intelligence

- Access Management
- Resource Management
- Electronic Clinical Quality Measures

- Operational Analytics

- Hospital Readmission Predictive Models
- Patient Fall Predictive Models

- Point of Care Decision Support

- Panel Management
- Disease or Condition Management

- Population Health & Surveillance

- Hospital Acquired Infections
- Biosurveillance

- Health Services Research

- Research on patient care, care delivery, health outcomes, cost, and quality

V20 Opioid Risk - Patient Detail Report

Data as of: 5/28/2014 6:06:00 AM

USE OF BIG DATA

By Stephen D. Fihn, Joseph Francis, Carolyn Clancy, Christopher Nelson, Karin Nelson, John Rumsfeld, Theresa Cullen, Jack Bates, and Gail L. Grabau

Insights From Advanced Analytics At The Veterans Health Administration

DOI: 10.1371/journal.pone.0120404
HEALTH AFF AFF 33, NO. 5 (2014), 1253-1261
©2014 Fihn et al.
This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Feature Article

Infusion Pump Reliability and Usability: Veterans Health Administration Examines Differences Between Manufacturers

Reference Question: R05 (Biomedical Engineering)
From the John Cochran VA Medical Center, St. Louis, Missouri.

Through recent efforts aimed toward consolidation of information for medical devices, the Veterans Health Administration (VHA) is able to track the reliability and usability of medical devices through a single point of measurement. This type of data is critical to the number of device-related injuries, such as those from infusion pumps. More importantly, cost and time associated with these devices are also being tracked. This information is being used to help inform decision-making with regard to the purchase of new devices and to help inform the number of devices that are purchased.

None, alcohol, sedation, and benzodiazepine are the most common drugs used in infusion pumps. The VHA is currently in the process of reviewing the reliability and usability of these devices. The VHA is currently in the process of reviewing the reliability and usability of these devices. The VHA is currently in the process of reviewing the reliability and usability of these devices.

Projected Hurricane Impact Application

3,348,155

1,158,370

50,573

Number of Divisions within 1 day error case

Number of Divisions within 1 day error case

Number of Divisions within 1 day error case

Encounters and 3rd Party Bills

Encounters by Vehicle Encounters on 3rd Party Bills

Encounter Insurance Type

Encounter Insurance Type

Encounter Insurance Type

Station	Fiscal Year	Fiscal Quarter	Fiscal Month	Axis Code
V01	Fiscal Calendar 2013	Fiscal Quarter 1, 2013	Fiscal October, 2013	A
V02	Fiscal Calendar 2014	Fiscal Quarter 2, 2015	Fiscal November, 2015	B
V03	Fiscal Calendar 2015	Fiscal Quarter 3, 2015	Fiscal December, 2015	C
V04	Fiscal Calendar 2016	Fiscal Quarter 4, 2015	Fiscal January, 2015	D

Station	Count	Unique AME Events	Unique Patients	AME Equipment Type	Cost
V01	54,468,475.46	255	236	Accelerator/Brake	\$60,000,000.00
V02	\$1,484,111.97	118	92	Air Conditioning	\$10,000,000.00
V03	\$65,127.65	33	31	Automatic Transmission	\$10,000,000.00
V04	\$1,406,279.22	98	84	Battery System/Back up	\$10,000,000.00
V05	\$6,884,478.77	46	44	Brakes-Other	\$10,000,000.00
V06	\$6,262,202.46	425	344	Cell Phone	\$10,000,000.00
V07	\$6,262,202.46	425	344	Control System-Outside/Inside	\$10,000,000.00
V08	\$6,262,202.46	425	344	Conversion Van/Truck	\$10,000,000.00
V09	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V10	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V11	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V12	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V13	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V14	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V15	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V16	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V17	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V18	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V19	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V20	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V21	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V22	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V23	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V24	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V25	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V26	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V27	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V28	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V29	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V30	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V31	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V32	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V33	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V34	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V35	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V36	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V37	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V38	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V39	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V40	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V41	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V42	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V43	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V44	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V45	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V46	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V47	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V48	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V49	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V50	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V51	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V52	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V53	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V54	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V55	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V56	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V57	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V58	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V59	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V60	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V61	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V62	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V63	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V64	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V65	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V66	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V67	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V68	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V69	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V70	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V71	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V72	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V73	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V74	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V75	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V76	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V77	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V78	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V79	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V80	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V81	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V82	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V83	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V84	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V85	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V86	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V87	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V88	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V89	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V90	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V91	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V92	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V93	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V94	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V95	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V96	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V97	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V98	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V99	\$6,262,202.46	425	344	Crash	\$10,000,000.00
V100	\$6,262,202.46	425	344	Crash	\$10,000,000.00

Data Quality Highlights

Null LOINC Codes in Vista

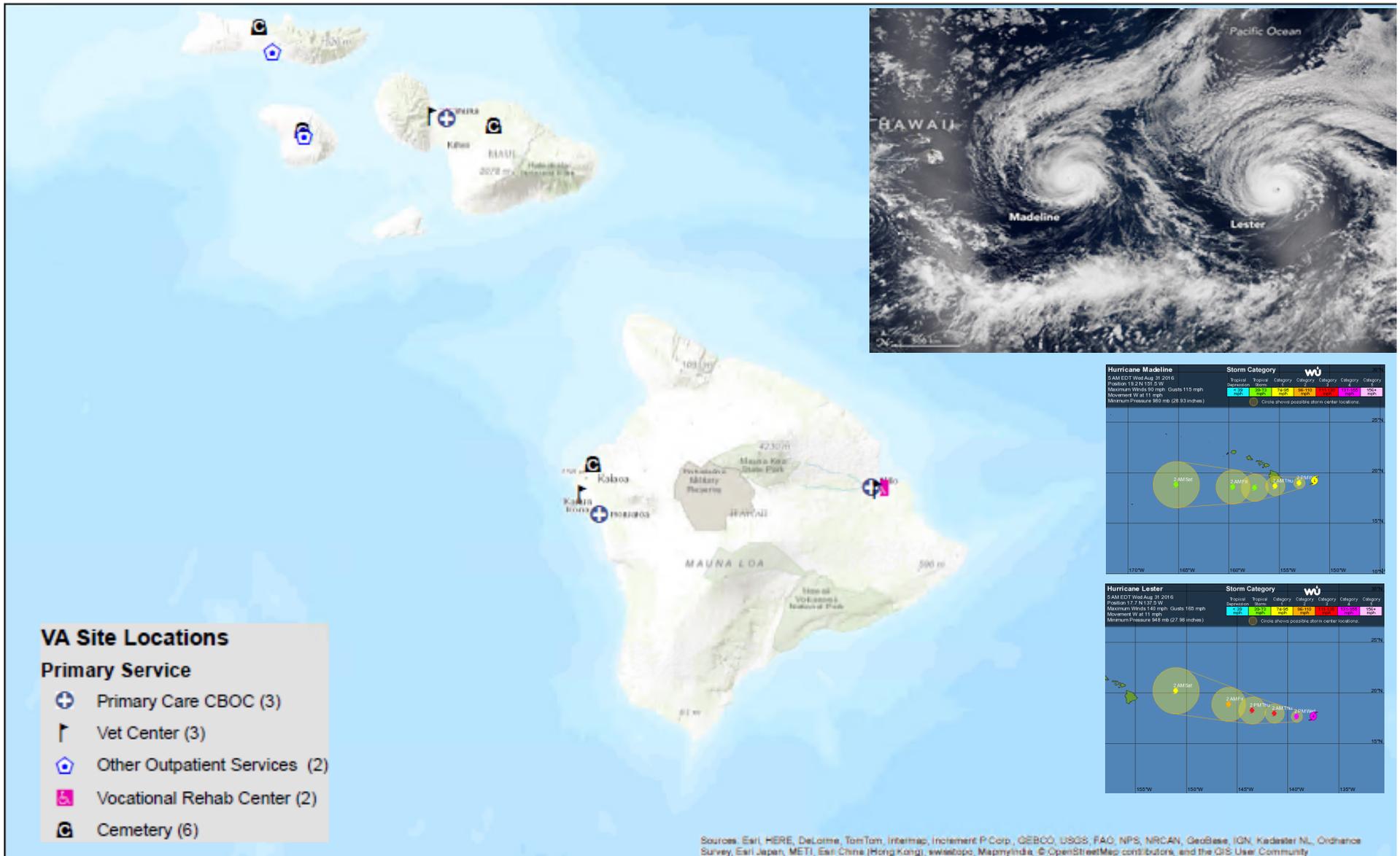
Overview: Null LOINC Codes in Vista

Actual Same Day (Preferred Date)

App	Same Day AM/AM Clinic Group	Count	Percentage
V01	Same Day AM/AM Clinic Group	1,475,000	38.87%
V02	Same Day AM/AM Clinic Group	1,000,000	25.62%
V03	Same Day AM/AM Clinic Group	500,000	12.81%
V04	Same Day AM/AM Clinic Group	250,000	6.41%
V05	Same Day AM/AM Clinic Group	125,000	3.20%
V06	Same Day AM/AM Clinic Group	62,500	1.60%
V07	Same Day AM/AM Clinic Group	31,250	0.80%
V08	Same Day AM/AM Clinic Group	15,625	0.40%
V09	Same Day AM/AM Clinic Group	7,812	0.20%
V10	Same Day AM/AM Clinic Group	3,906	0.10%
V11	Same Day AM/AM Clinic Group	1,953	0.05%
V12	Same Day AM/AM Clinic Group	976	0.02%
V13	Same Day AM/AM Clinic Group	488	0.01%
V14	Same Day AM/AM Clinic Group	244	0.00%
V15	Same Day AM/AM Clinic Group	122	0.00%
V16	Same Day AM/AM Clinic Group	61	0.00%
V17	Same Day AM/AM Clinic Group	31	0.00%
V18	Same Day AM/AM Clinic Group	16	0.00%
V19	Same Day AM/AM Clinic Group	8	0.00%
V20	Same Day AM/AM Clinic Group	4	0.00%
V21	Same Day AM/AM Clinic Group	2	0.00%
V22	Same Day AM/AM Clinic Group	1	0.00%
V23	Same Day AM/AM Clinic Group	1	0.00%
V24	Same Day AM/AM Clinic Group	1	0.00%
V25	Same Day AM/AM Clinic Group	1	0.00%
V26	Same Day AM/AM Clinic Group	1	0.00%
V27	Same Day AM/AM Clinic Group	1	0.00%
V28	Same Day AM/AM Clinic Group	1	0.00%
V29	Same Day AM/AM Clinic Group	1	0.00%
V30	Same Day AM/AM Clinic Group	1	0.00%
V31	Same Day AM/AM Clinic Group	1	0.00%
V32	Same Day AM/AM Clinic Group	1	0.00%
V33	Same Day AM/AM Clinic Group	1	0.00%
V34	Same Day AM/AM Clinic Group	1	0.00%
V35	Same Day AM/AM Clinic Group	1	0.00%
V36	Same Day AM/AM Clinic Group	1	0.00%
V37	Same Day AM/AM Clinic Group	1	0.00%
V38	Same Day AM/AM Clinic Group	1	0.00%
V39	Same Day AM/AM Clinic Group	1	0.00%
V40	Same Day AM/AM Clinic Group	1	0.00%
V41	Same Day AM/AM Clinic Group	1	0.00%
V42	Same Day AM/AM Clinic Group	1	0.00%
V43	Same Day AM/AM Clinic Group	1	0.00%
V44	Same Day AM/AM Clinic Group	1	0.00%
V45	Same Day AM/AM Clinic Group	1	0.00%
V46	Same Day AM/AM Clinic Group	1	0.00%
V47	Same Day AM/AM Clinic Group	1	0.00%
V48	Same Day AM/AM Clinic Group	1	0.00%
V49	Same Day AM/AM Clinic Group	1	0.00%
V50	Same Day AM/AM Clinic Group	1	0.00%
V51	Same Day AM/AM Clinic Group	1	0.00%
V52	Same Day AM/AM Clinic Group	1	0.00%
V53	Same Day AM/AM Clinic Group	1	

Use Case: Situational Awareness/Disaster Response

Twin Hurricanes Lester (C4) and Madeline (C1)



Operational Business Intelligence Use Case: Veteran Choice Act Support

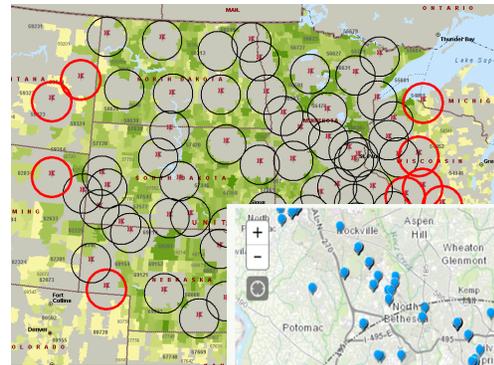
- Data/Infrastructure**

- VACAA Patient Geocoding and Drive Times
- VACAA Choice Patient Wait Times
- VACAA Patient Eligibility
 - Registry of VACAA Eligible Patients and related demographics per CBO
- Fee Basis Claims (FBCS) for Community Care

- Reports/Tools**

- Drive Time Analysis
- Community Provider Lookup
- VISN/Facility Patient Eligibility
- Patient Wait Times

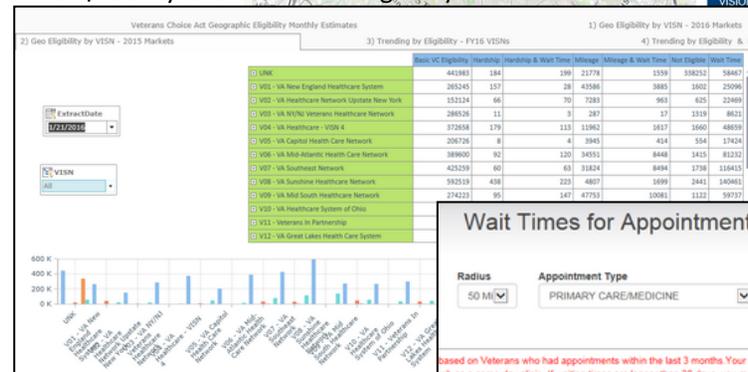
Veteran Drive Time Analysis



Veteran Community Provider Lookup



VISN/Facility Choice Patient Eligibility and Status

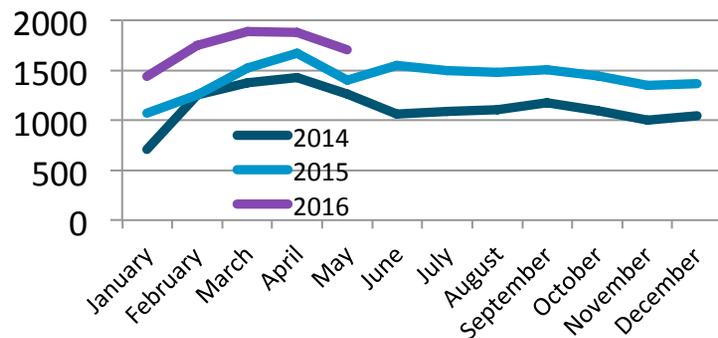


Wait Times for Appointments at VA Facilities

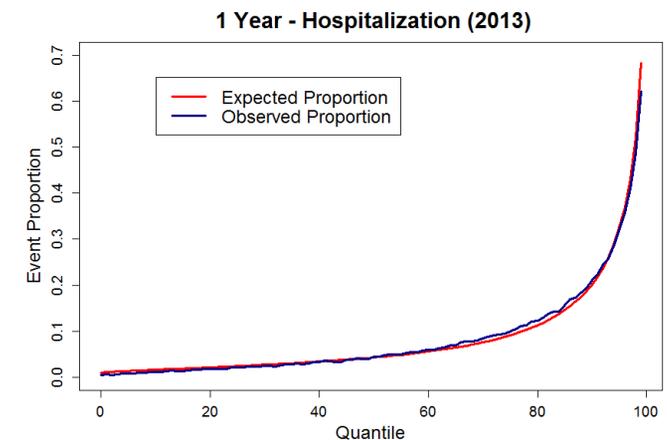
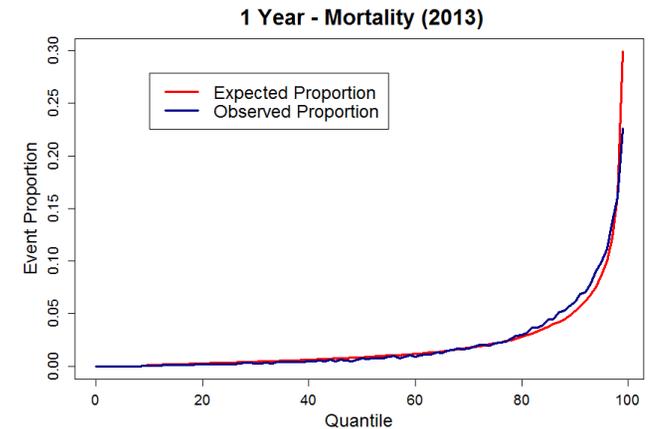


Predictive Analytics Use Case: Care Assessment Need Score

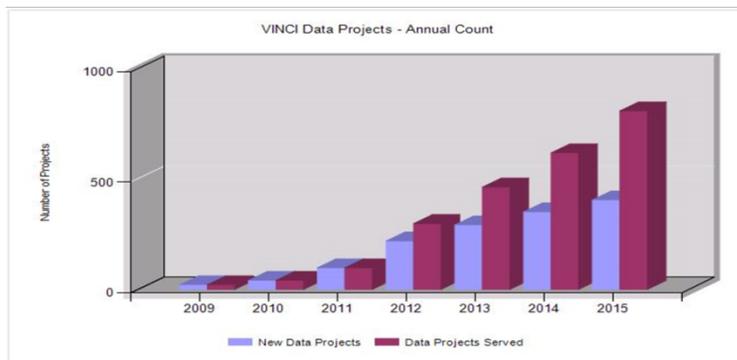
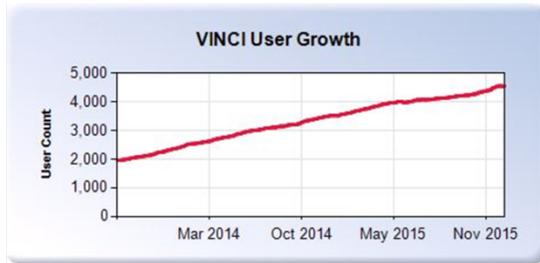
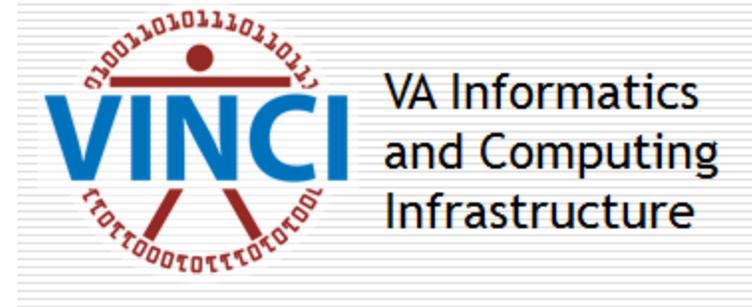
- Complex predictive model - 40 variables incl. demographics, diagnoses, lab, vitals, meds, health care utilization, DoD data, census data
- Highly accurate predictions of death and hospitalization for all VA primary care pts (~6M)
- Updated weekly and available to all primary care providers via VistA
- Tightly linked with web-based care coordination software – Patient Care Assessment System



Veterans in highest %ile of risk have 58% probability of admission, 23% probability of death, and 64% probability of either event.



Use Case: Health Services Research



FY15 Statistics:

- 5,547 Users (50% increase)
- 871 Projects (30% increase)

Capabilities:

- National Scale Data
- Prep to Research Tools
- Data Support
- SAS/Grid for Statistical Processing

Project Examples:

- *Collaborative for Improving Hypertension Management*
- *Primary Care Quality and Service Customization of the Homeless*
- *Using Knowledge Discovery Strategies to Identify Fall-Related Injuries*
- *Post-Deployment Sentinel Event Surveillance*
- *Monitoring and Management of Metabolic Side Effects of Antipsychotics*
- *In-Theater Medical Treatment and Long-Term Health Outcomes of Recent Military Amputees*
- *Outcomes, Costs and Trends in Dialysis Timing in VA*
- *PTSD, Depression and Substance Use Disorders among U.S. Veterans Returning from Iraq and Afghanistan*
- *Traumatic Brain Injury among Homeless Veterans.*



Precision Medicine/Genomics Use Case: MVP-CHAMPION



MVP-CHAMPION

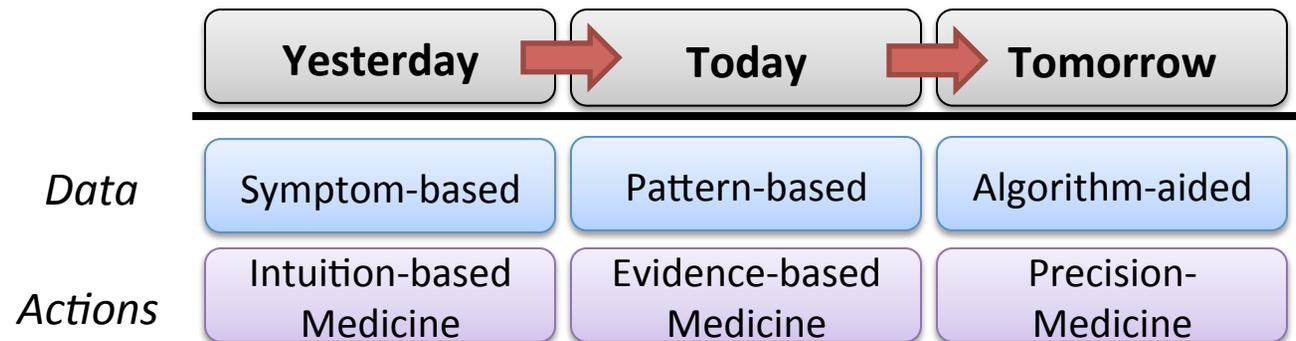
VA Million Veteran Program (MVP)

Computational Health AnalYTics for Medical
Precision to Improve Outcomes Now



MVP-CHAMPION: The Vision

1. To improve the lives and well-being of our Nations veterans by discovering new or improved treatment regimens for diseases and conditions that are prevalent in the Veteran community.
2. To usher in a new era of personalized and precision medicine that will change the science and practice of medicine.
3. To make quantum leaps in the knowledge and treatment of our most prevalent diseases and conditions.
4. To advance the state of the art in data science, informatics, and analytics in medicine.
5. To build a sustainable data science workforce that have the skill to tackle the most complex problems in medicine.



MVP-CHAMPION *FastTrack* Goal:

Make a discovery that advances the state of the art in prostate cancer treatment within 18-months

VA Cases of Common Cancers FY2005-FY16

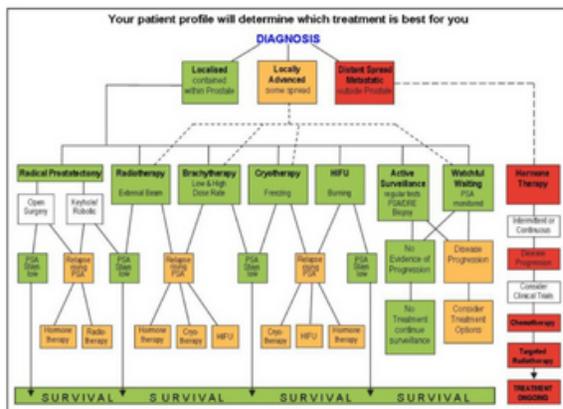
As of 3/16/2016

Rank	ICD-9	Common Cancers	VA Cases (Patients)
1.	185	Malignant neoplasm of prostate	362,256
2.	162	Malignant neoplasm of trachea, bronchus, and lung	105,731
3.	153	Malignant neoplasm of colon	61,797
4.	188	Malignant neoplasm of bladder	61,262
5.	200-202	Lymphoma	40,165
6.	198	Secondary malignant neoplasm of other specified sites	35,141
7.	189	Malignant neoplasm of kidney and other and unspecified urinary organs	32,694
8.	204-208	Leukemia	31,354
9.	172	Malignant melanoma of skin	27,184
10.	197	Secondary malignant neoplasm of respiratory and digestive systems	20,058

VA Prevalence of Prostate Cancer

Prostate Cancer:
 >100 new cases/day;
 > 60 deaths/day
 (2016: New cases: 9,241; Deaths: 5,234)

Prostate Cancer: The Paradox of Choice



- Prostate Cancer is a difficult and complex disease, and the choices facing a man who is diagnosed with it — what treatment to have, indeed whether to be treated at all — are complex, perhaps more complex than in any other major cancer.
- The goal of this project is within 18-months to make significant progress in being able to categorize prostate cancer patients based upon genetic factors to be able to be more precise (fewer, more precise) options in their treatment and treatment regimens.

Watchful Waiting • Active Surveillance • Radical Prostatectomy • Radiation Therapy • Brachytherapy • Proton Therapy

Methodological Challenges

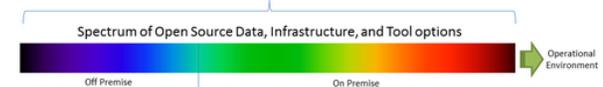
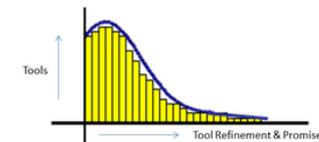
- How to scale “deep” phenotyping to large data sets (how-to use data to characterize):
 - Stage
 - Severity
 - Social context
 - Disease course
 - Treatment response
 - Differential diagnosis
- Need for NLP plus structured data!
- Need to integrate phenotype and genotype data
 - Scale of genotype data

Big Data Challenges

- **Things we want to do but haven't**
 - Analytic work flow “factory” (democratize analytics)
 - Real-Time Point of Care Decision Support (“Amazon” like patient care “suggestions”)
 - Mainstream genomic data into analytic pipelines
- **Things we want to do but can't**
 - Open source community collaboration with VA data
 - Population scale Health Services Research
 - Rapid analytic model prototyping
- **Things we are doing and need to do better**
 - **360° View of the Veteran (all interactions)**
 - Advanced analytics education and training of workforce (Data, Technique, Tools)
 - New generation computing platforms (processor, memory, storage, networking)
 - Information protection (Rogue States and individuals, Ransomware)



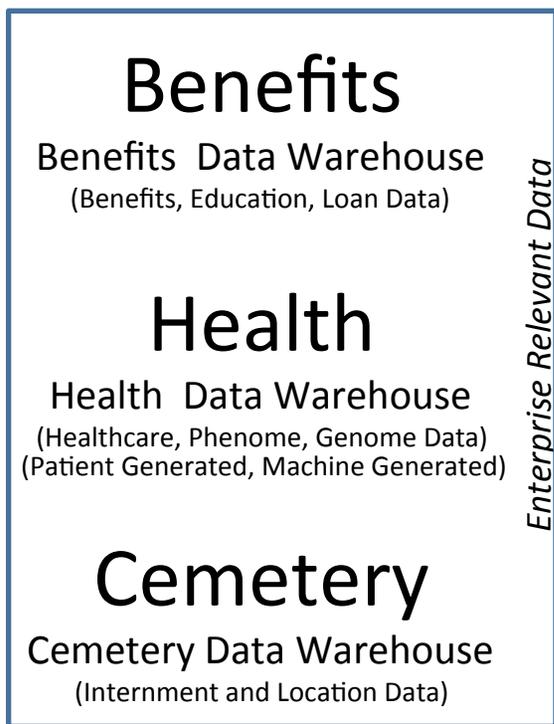
Open Source Collaborative
Data Lab Concept





Toward a Veteran Centric “Big Data” Capability

VA Line of Business Data



Enterprise Relevant Data
Common Veteran Identifier

360° Analytic View of the Veteran (and the VA)



Customer Master Data

- Master Veteran Demographics

Customer Experience Data

- Customer Sentiment
- Customer Relationship Management
- Telephony
- Vets.gov

VA Business Data

- Accounting (AP/AR/GL/MCA)
- Acquisition
- Assets (Facilities/Equipment)
- Human Resources/Payroll
- Emergency Management
- Information Technology

Non VA Data

- Department of Defense
- Department of Homeland Security
- Department of Energy
- Department of Commerce
- Health & Human Services
- Community Care
- Lenders, Schools
- Private Sector

Analytic Capabilities:

- Descriptive
- Diagnostic
- Predictive
- Geospatial
- Natural Language Processing



High Performance Computing
(DoE Oak Ridge National Lab)
(Infrastructure, Knowledge, Skill, AoA)



CDW



Infrastructure



Cloud Computing
(IaaS, PaaS, SaaS)
(Infrastructure)

Conclusion

VA's goal is to leverage our vast array of clinical and genomic data to:

1. Improve the lives and well-being of our Nations veterans by discovering new or improved treatment regimens for diseases and conditions that are prevalent in the Veteran community.
2. Usher in a new era of personalized and precision medicine that will change the science and practice of medicine.
3. Make quantum leaps in the knowledge and treatment of our most prevalent diseases and conditions.
4. Advance the state of the art in data science, informatics, and analytics in medicine.
5. Build a sustainable data science workforce that have the skill to tackle the most complex problems in medicine.
6. All of the above will NOT be possible without a strong partnership with the Department of Energy and their domain expertise in high performance computing and big data analytics.

