The Use of Large Data Sets to Respond to Policy Relevant Questions in Health Services Research

Lewis E. Kazis, Sc.D.
Professor and Director
Center for the Assessment of Pharmaceutical Practices (CAPP)
Department of Health Law Policy and Management
Boston University School of Public Health

May 9, 2018
CTSI/CRRO Clinical Research Seminar
Boston University

THE 5 MOST SPECTACULAR PHOTOS IN THE WORLD

5th

4th

3rd

2nd

1st


© Kurt Jones 2003

Objectives of Presentation

- Use of Big Data
- Context within the US Health Care system
- Applications (VA Example)
- Vision for the future

Institute of Medicine 2012 Report: *Best Care at Lower Cost*

- Context
  - Unmanageable complexity of care
  - Escalating costs
  - High volume of under-used data
  - Increasingly affordable computing power
  - Connectivity and unprecedented diffusion of information
- Vision for a Continuously Learning Healthcare System
  - *3 imperatives:*
    - *(1) managing rapidly increasing complexity;*
    - *(2) achieving greater value in health care; and*
    - *(3) capturing opportunities from technology, industry, and policy*
- IOM's 2012 Report outlines how to harness new technologies, innovations, and approaches to overcome challenges of the US healthcare system

Prospect Of *Big Data*

- Predicting Personalized Health and Health Risk
- 'Real Time' Management of Complexity of Health
- Personalizing Prevention Strategies
- Innovative research based on stitching disparate and complex data sources

"Tonight, I'm launching a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes and to give all of us access to the personalized information we need to keep ourselves and our families healthier."

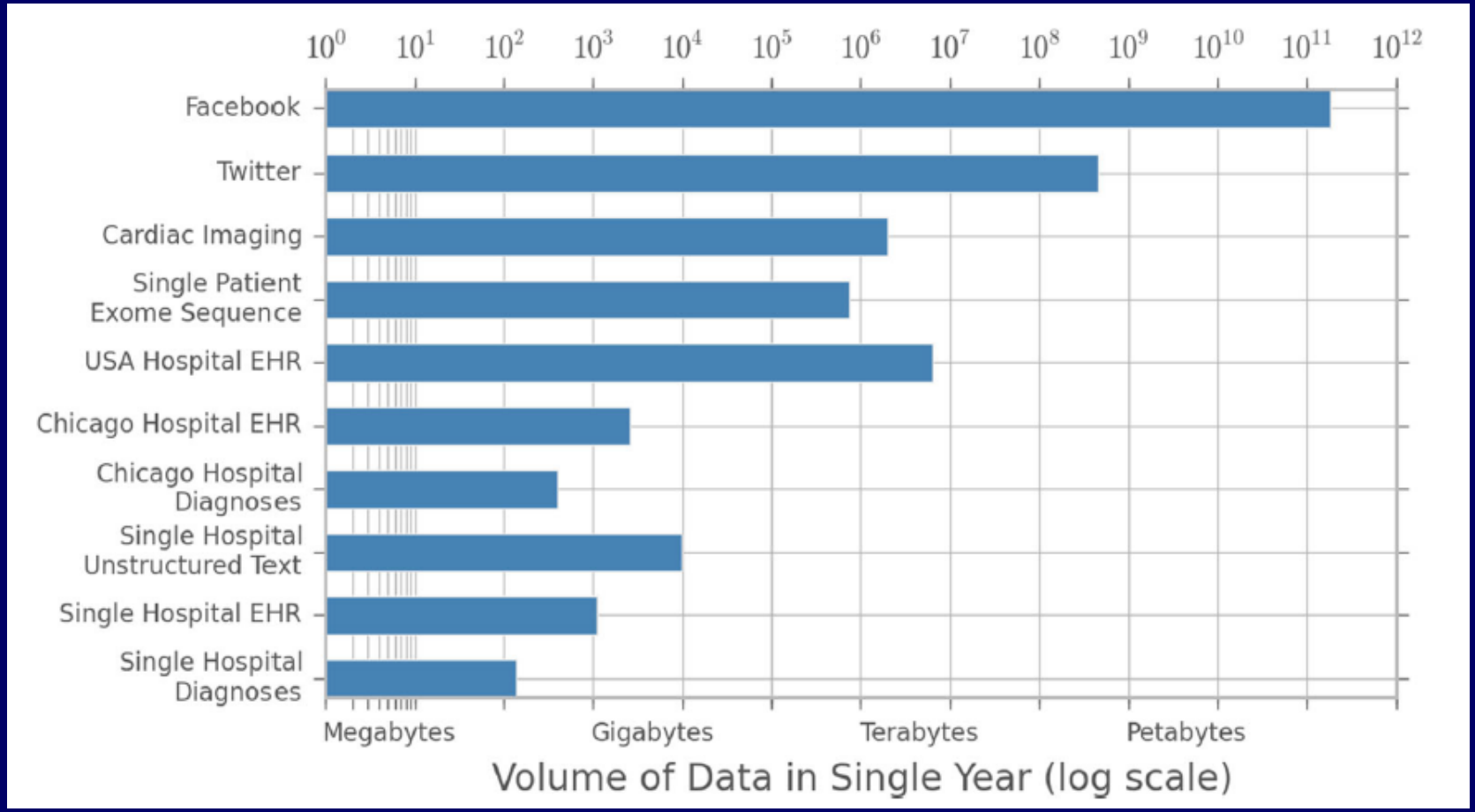
— President Barack Obama, State of the Union Address, January 20, 2015

Institute of Medicine (U.S.). 2012 Report *Best Care at Lower Cost: Role of Information Technology*

- Institute of Medicine 2012. ***Best Care at Lower Cost: The Path to Continuously Learning Health Care in America***. Washington, D.C: National Academies Press.
- download.nap.edu/cart/download.cgi?&record_id=13444

Institute of Medicine 2012. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America

Volume: Data Generated In A Single Year For Various Sources



Source: Pah AR, Rasmussen-Torvik LJ, Goel S, Greenland P, Kho AN. Big Data: What Is It and What Does It Mean for Cardiovascular Research and Prevention Policy. Current Cardiovascular Risk Reports]. 2015 Jan

Velocity, Complexity and Variety

Velocity

- Era of streaming data
- Timeliness of collection, analysis and translation
- Data overload

Complexity/veracity

- Data pollution
- Eliminating noise; detecting faint signals
    - Data quality and integrity
    - Missingness
    - Statistical significance: P hacking

Variety: Sources Of *Health Big Data*

- Electronic health records (EHRs)
    - Medical imaging
    - Clinical and anatomic pathology
- Administrative records
- Patient genomics
- Device, log, and sensor data
- Crowd-sourced environmental and behavioral data
- Other unstructured text data:
    - Provider notes
    - Biomedical literature

Some of the leading 'Big Data' Environments used for Health Services Research in the US

Big Data Environment	Established	Founded by	Setting	Population base	High Level Definition
VHA's Corporate Data Warehouse (CDW)	2006 (data available since 1991) **	Department of Veterans Affairs	Federal organization	Administrative data for VHA-provided health care utilized by veterans and some nonveterans (e.g., VHA employees and research participants)	• Live administrative records, and linkable EMR (some use of NLP) • Linkage with CMS and death data • Well-developed meta data • Intended primarily for intramural research
UnitedHealth (OptumLabs Data Warehouse)	Jan 2013 (data available since 1994)	UnitedHealth Optum, Mayo Clinic, AARP (closely related to UHC Ingenix and Lewin Group)	Private multi stakeholder research collaboration	Privately insured and Medicare Advantage patients and their families	• Administrative claims, linkable EMR (extensive use of NLP#), include patient surveys • Powerful analytic platform • Collaborative environment • Both intramural and extra mural research
IBM Watson Health (Truven Health Marketscan)	1988* (data available since 1995)	Thomson Reuters MedStat	Private data vendor	Privately insured patients and their families	• Administrative and EMR • Well-developed meta-data • Well tested data quality and integrity

#- NLP: Natural Language Processing, *- B.R.I.D.G.E. TO DATA® 2015;
**- https://catalog.data.gov/dataset/national-patient-care-database-npcd#access-use, 2015;
***- Mitchell et al. 1994

Historical Perspective

- Over the past 25 years new approaches for health outcomes have been developed from the patient perspective

- The Shattuck Lecture by Paul Ellwood

- Mortality, morbidity and standard clinical outcome assessments are not sufficient for measurement of health outcomes

Perspectives using PROs

- **Applications of PROs on Several Levels:**

- **Population Based Studies:** for purposes of gauging the effectiveness of patient reported outcomes on large populations.

- **Clinical studies:** using quasi and randomized controlled designs for purposes of demonstrating efficacy/effectiveness of interventions using PROs as endpoints. Conducted in more homogeneous populations.

- **PROs at the individual subject level:** "N of 1" studies as a clinical assessment.

Population Based Studies

Population Based Studies

- How systems of care perform among different hospitals throughout a network of care from the consumers perspective
- Comparison of new systems and existing systems of care
- Use general or generic PRO measures
- Example: CMS Medicare Advantage System with the Veterans Administration

VR-12

- Veterans RAND 12 Item Health Survey (VR-12)
- Developed from the MOS SF-36 (v.1.0) and the Veterans RAND 36 Item Health Survey (VR-36).
- More than 15 years of use with over 7 million administrations internationally.
- Applied to Quality Improvement and management programs in the VA and CMS Medicare Advantage Program.

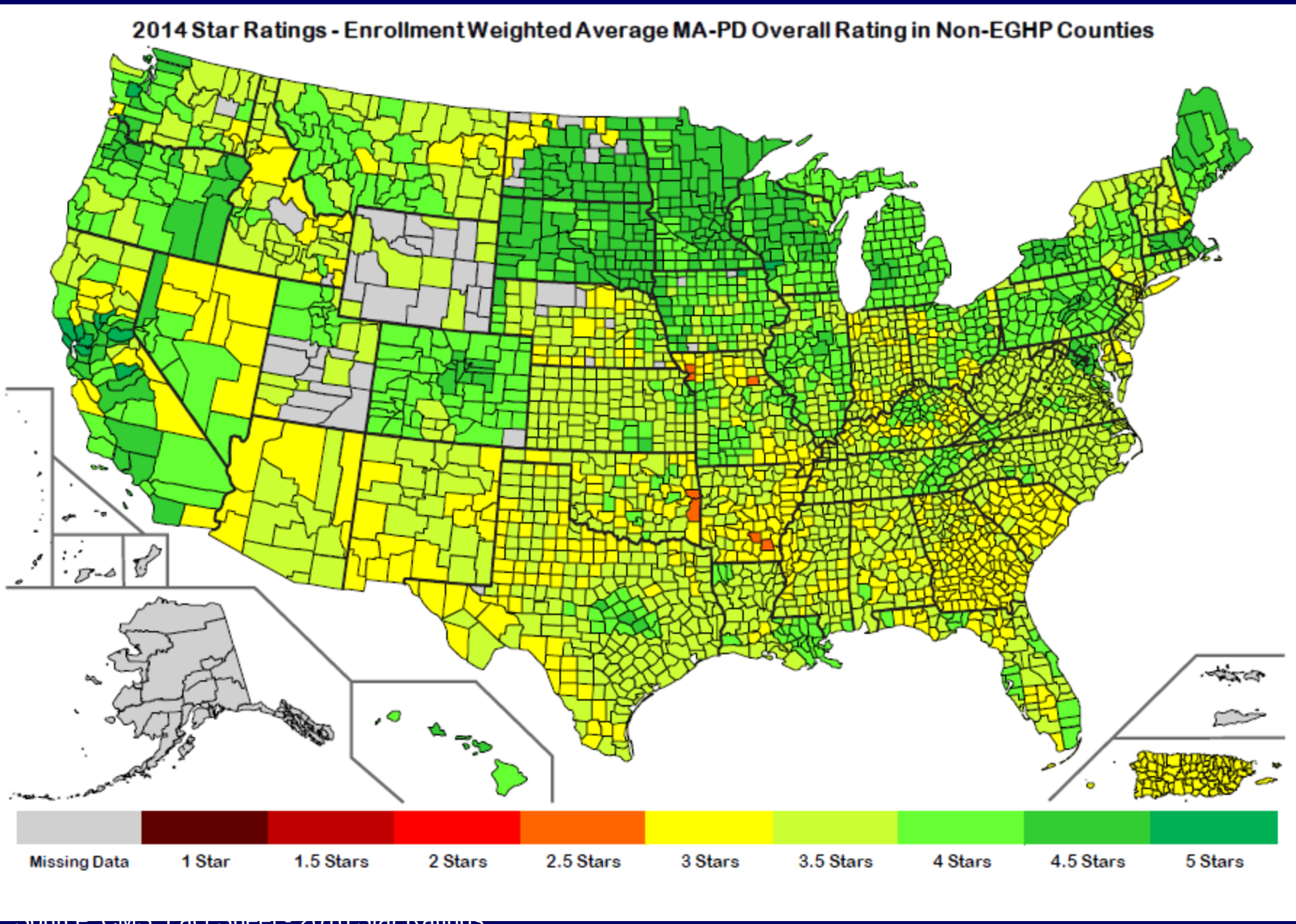
Veterans VR-12

- VR-12 adopted as a HEDIS Measure (2004)

- VR-12 has been adopted as one of the major endpoints for the Center for Medicare and Medicaid Services (CMS) Health Outcomes Survey (HOS) using the physical and mental summaries.

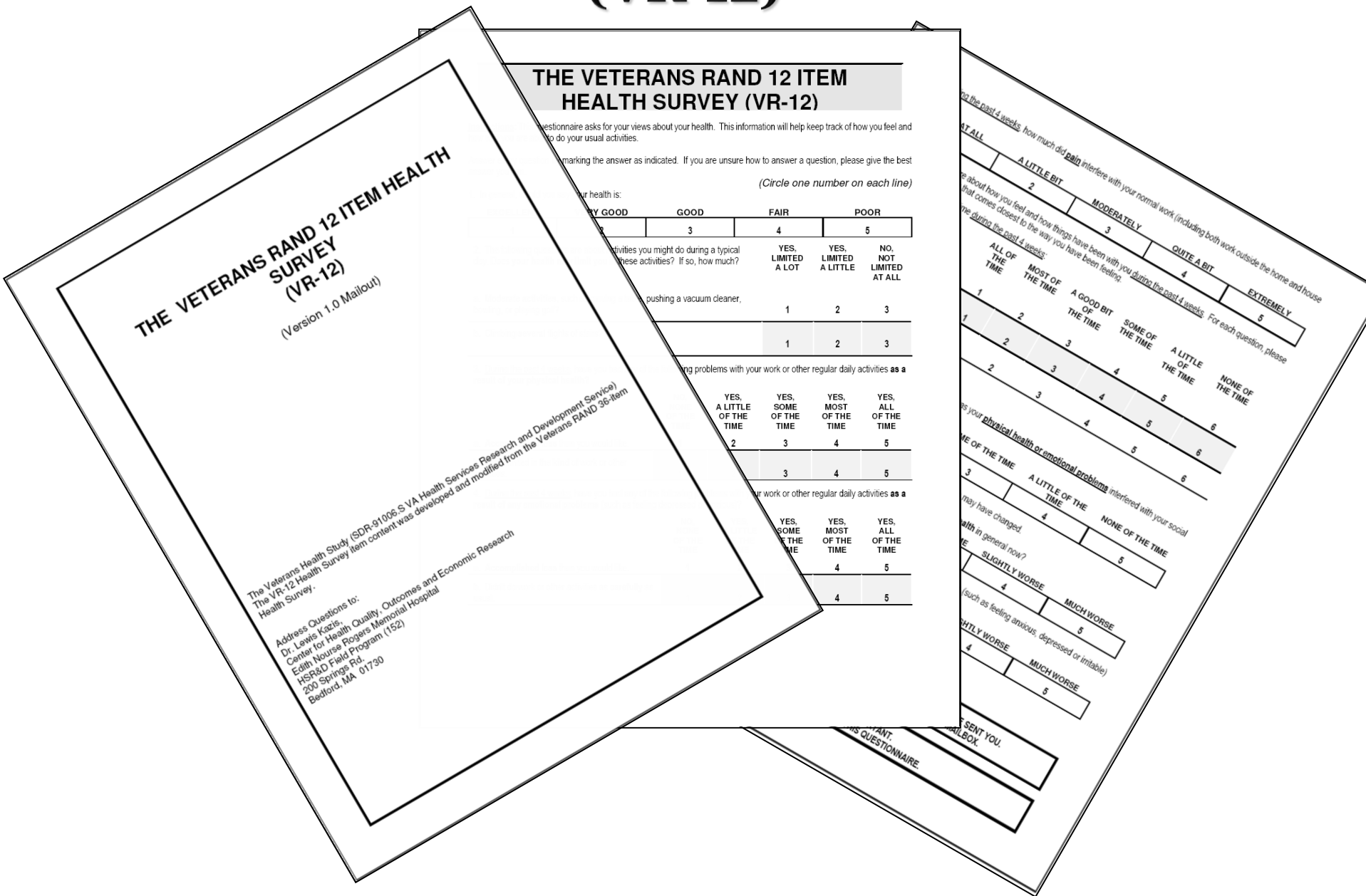
Veterans VR-12

- CMS introduced the STAR System within past few years for purposes of consumer evaluation of the Medicare Advantage Plans (each plan is rated from 1 to 5 stars).
- VR-12 from the HOS is one of the principal components of the STAR system for this evaluation. The other components include HEDIS measures, CAHPS and satisfaction measures.



Source: CMS, Fact Sheet - 2016 Star Ratings

VETERANS RAND 12 ITEM HEALTH SURVEY (VR-12)



General Health Perception

Physical Function Inventory

Role Limitations Due to Physical Problems

Role Limitations Due to Emotional Problems

THE VETERANS RAND 12 ITEM HEALTH SURVEY (VR-12)

Instructions: This questionnaire asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities.

Answer every question by marking the answer as indicated. If you are unsure how to answer a question, please give the best answer you can.

(Circle one number on each line)

1. In general, would you say your health is:

EXCELLENT	VERY GOOD	GOOD	FAIR	POOR
1	2	3	4	5

2. The following questions are about activities you might do during a typical day. Does **your health now limit you** in these activities? If so, how much?	YES, LIMITED A LOT	YES, LIMITED A LITTLE	NO, NOT LIMITED AT ALL
a. **Moderate activities,** such as moving a table, pushing a vacuum cleaner, bowling, or playing golf?	1	2	3
b. Climbing **several** flights of stairs?	1	2	3

3. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities **as a result of your physical health?**

	NO, NONE OF THE TIME	YES, A LITTLE OF THE TIME	YES, SOME OF THE TIME	YES, MOST OF THE TIME	YES, ALL OF THE TIME
a. **Accomplished less** than you would like.	1	2	3	4	5
b. Were limited in the **kind** of work or other activities.	1	2	3	4	5

4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities **as a result of any emotional problems** (such as feeling depressed or anxious)?

	NO, NONE OF THE TIME	YES, A LITTLE OF THE TIME	YES, SOME OF THE TIME	YES, MOST OF THE TIME	YES, ALL OF THE TIME
a. **Accomplished less** than you would like.	1	2	3	4	5
b. Didn't do work or other activities as **carefully** as usual.	1	2	3	4	5

Bodily Pain

5. During the past 4 weeks, how much did **pain** interfere with your normal work (including both work outside the home and house work)?

NOT AT ALL	A LITTLE BIT	MODERATELY	QUITE A BIT	EXTREMELY
1	2	3	4	5

These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling.

Mental Health/ Energy-Fatigue

6. How much of the time during the past 4 weeks:

	ALL OF THE TIME	MOST OF THE TIME	A GOOD BIT OF THE TIME	SOME OF THE TIME	A LITTLE OF THE TIME	NONE OF THE TIME
a. Have you felt **calm and peaceful**?	1	2	3	4	5	6
b. Did you have **a lot of energy**?	1	2	3	4	5	6
c. Have you felt **downhearted and blue**?	1	2	3	4	5	6

Social Functioning

7. During the past 4 weeks, how much of the time has your **physical health or emotional problems** interfered with your social activities (like visiting with friends, relatives, etc.)?

ALL OF THE TIME	MOST OF THE TIME	SOME OF THE TIME	A LITTLE OF THE TIME	NONE OF THE TIME
1	2	3	4	5

Now, we'd like to ask you some questions about how your health may have changed.

Physical/ Mental Change

8. Compared to one year ago, how would you rate your **physical health** in general now?

MUCH BETTER	SLIGHTLY BETTER	ABOUT THE SAME	SLIGHTLY WORSE	MUCH WORSE
1	2	3	4	5

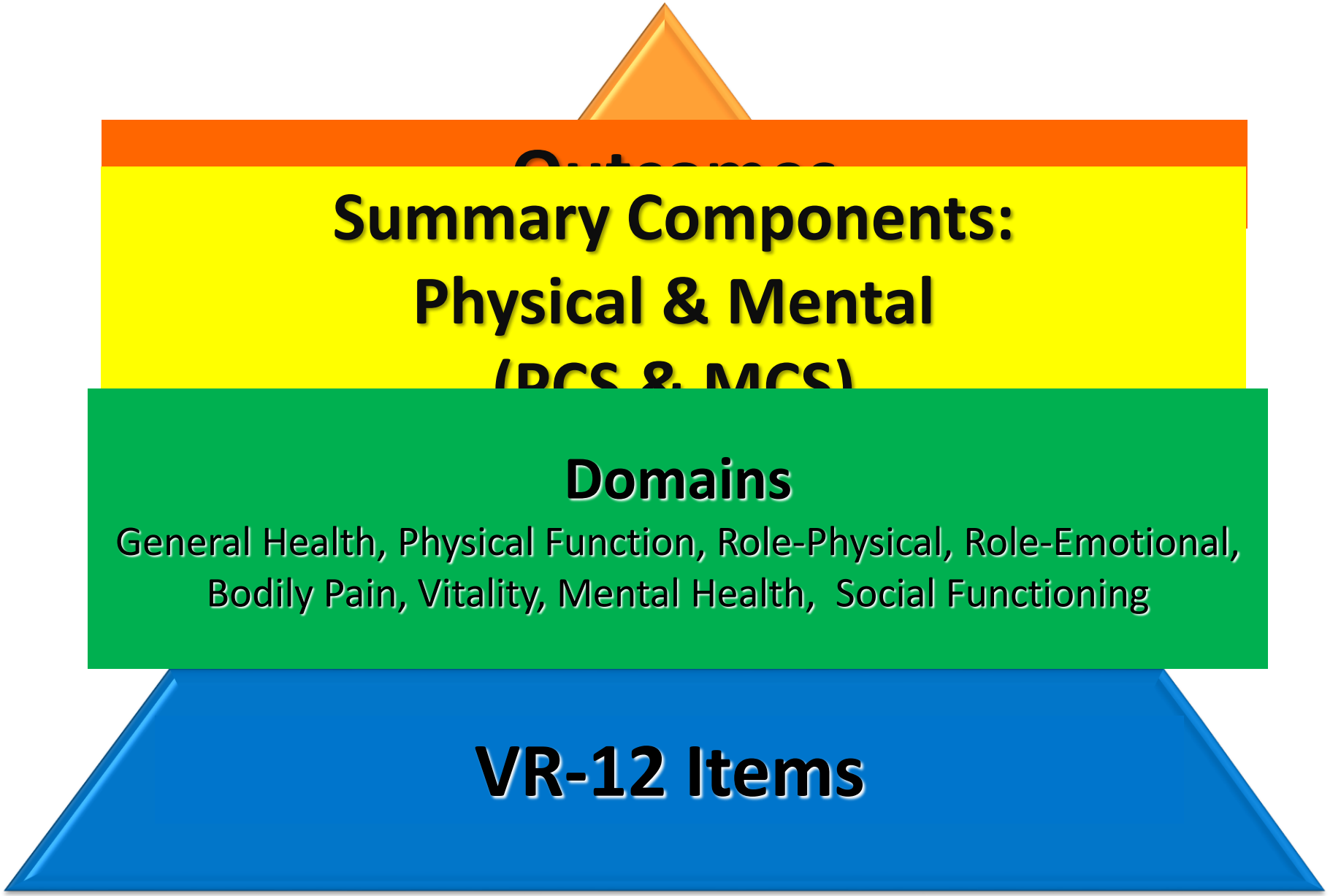
9. Compared to one year ago, how would you rate your **emotional problems** (such as feeling anxious, depressed or irritable) **now**?

MUCH BETTER	SLIGHTLY BETTER	ABOUT THE SAME	SLIGHTLY WORSE	MUCH WORSE
1	2	3	4	5

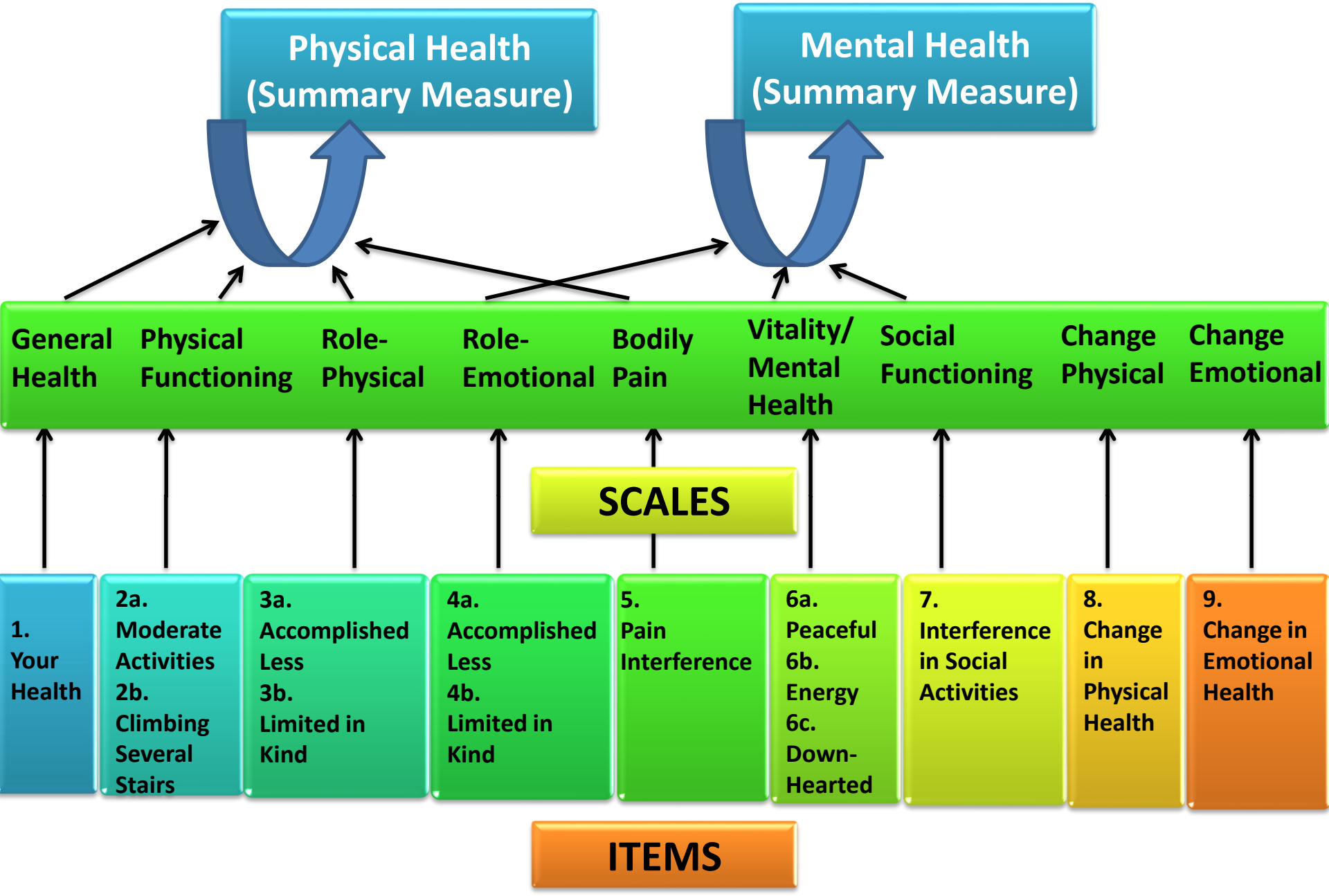
PLEASE PLACE THE COMPLETED QUESTIONNAIRE IN THE ENVELOPE WE SENT YOU.
NO STAMP IS REQUIRED: SIMPLY PLACE THE ENVELOPE IN ANY MAILBOX.

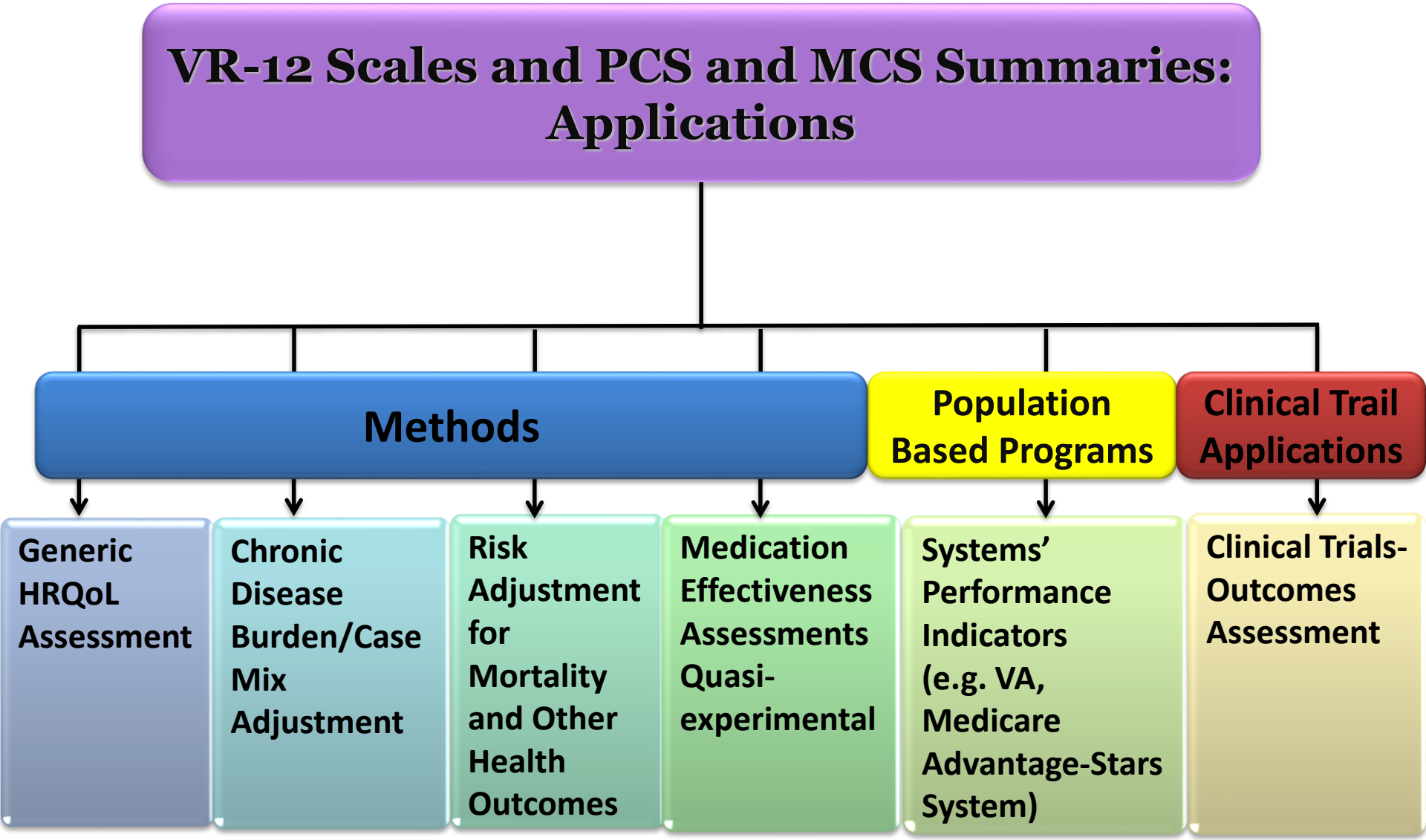
YOUR ANSWERS ARE IMPORTANT.
THANK YOU FOR COMPLETING THIS QUESTIONNAIRE.

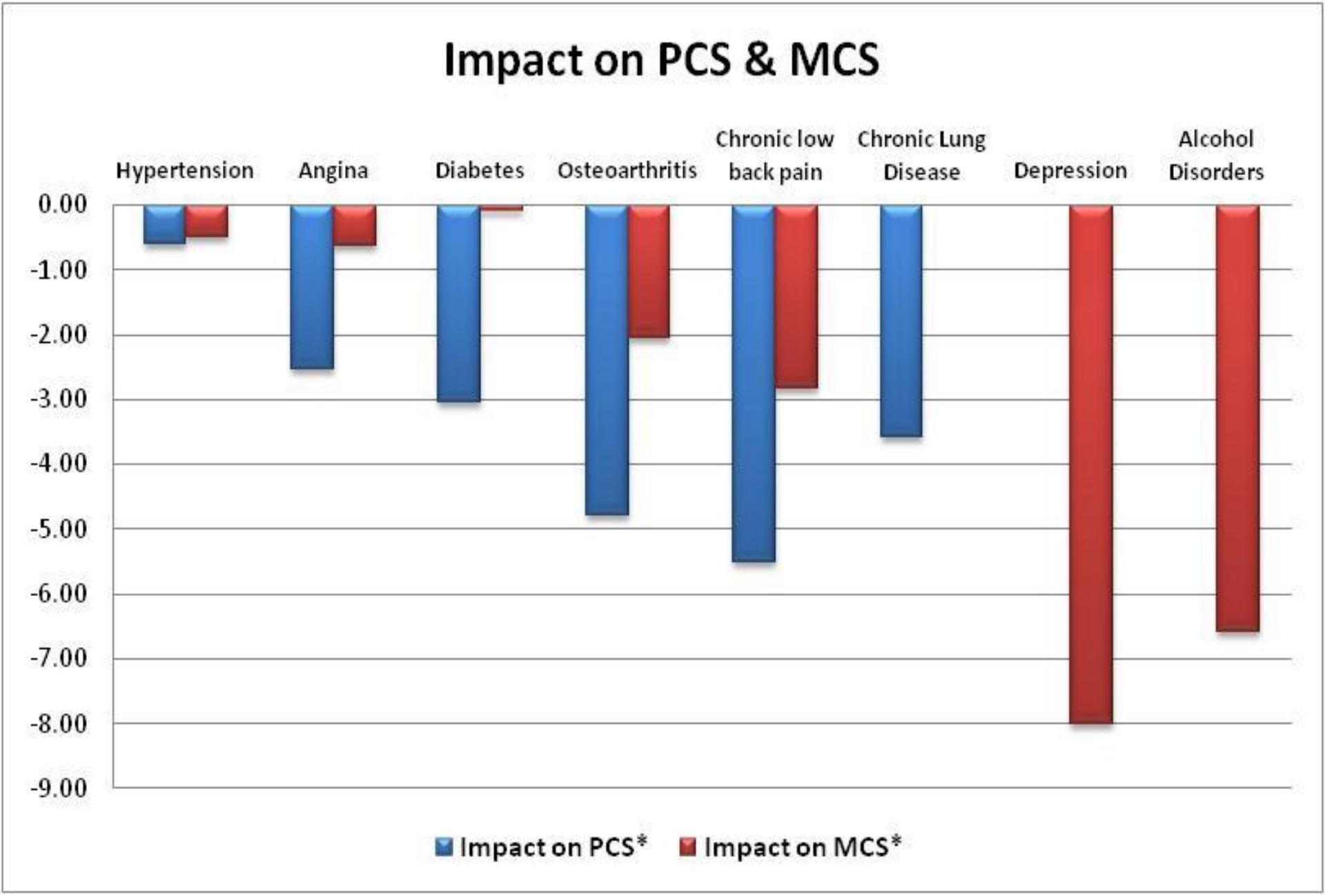
VETERANS RAND 12 ITEM HEALTH SURVEY



VR-12 Schematic Model




VR-12 Scales and PCS and MCS Summaries: Applications
Methods
Population Based Programs
Clinical Trail Applications
Generic HRQoL Assessment
Chronic Disease Burden/Case Mix Adjustment
Risk Adjustment for Mortality and Other Health Outcomes
Medication Effectiveness Assessments Quasi-experimental
Systems' Performance Indicators (e.g. VA, Medicare Advantage-Stars System)
Clinical Trials-Outcomes Assessment


Impact on PCS & MCS
Hypertension
Angina
Diabetes
Osteoarthritis
Chronic low back pain
Chronic Lung Disease
Depression
Alcohol Disorders
0.00
-1.00
-2.00
-3.00
-4.00
-5.00
-6.00
-7.00
-8.00
-9.00
Impact on PCS*
Impact on MCS*

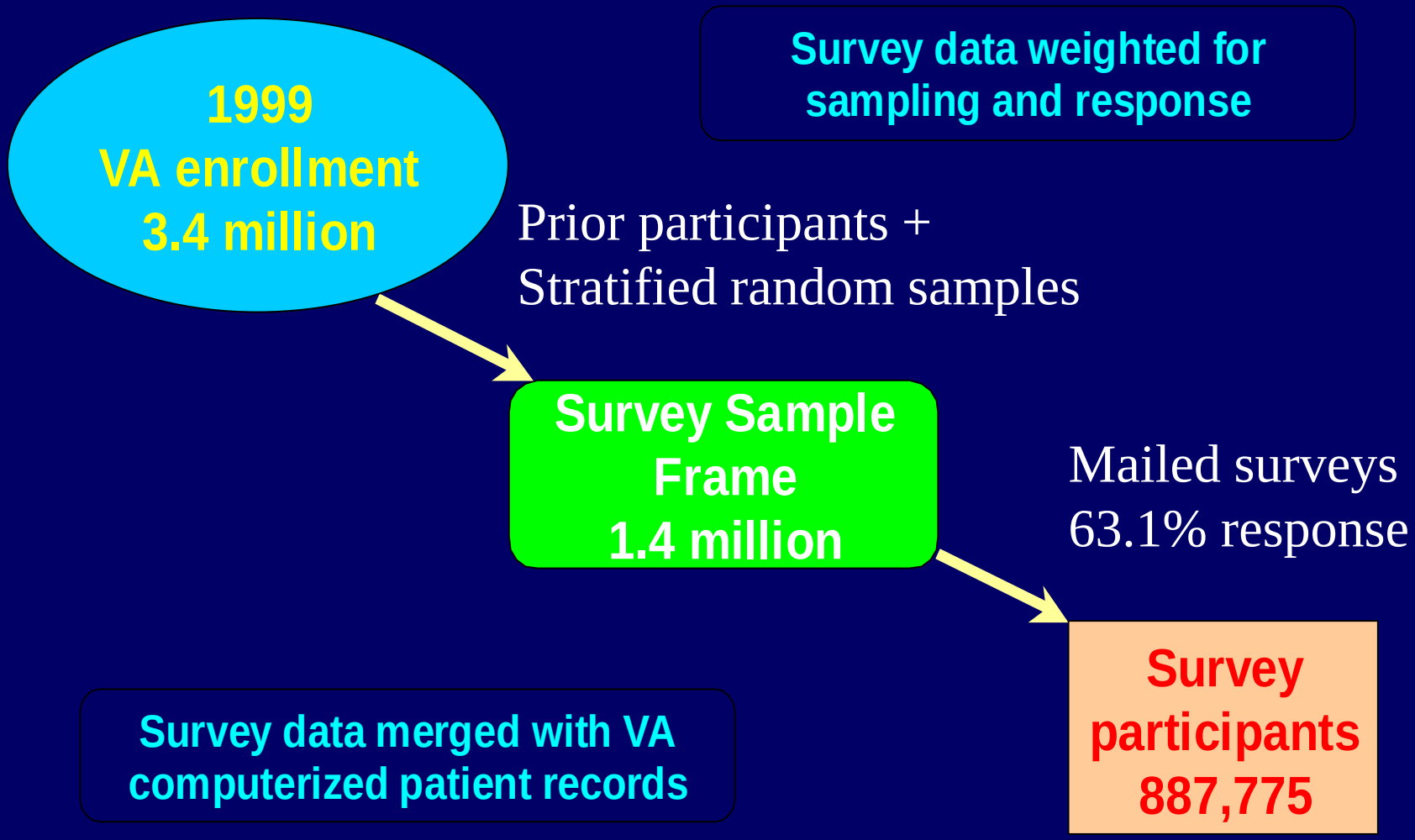
2000 VA National Health Survey of Veteran Enrollees Background

- The 2000 VA national health survey of veteran enrollees
  - The largest and most ambitious survey ever conducted by the VHA
  - To provide assessments of veteran's health status and health behaviors for gauging the veteran's health care needs and outcomes of care.

2000 VA National Health Survey of Veteran Enrollees Objectives

- To provide baseline norms of health using the VR 36 based upon enrollees in the VHA.
- To monitor the health of veterans at the Network level (VISN) and facility levels.
- To provide system comparisons (VA and non-VA).
- To develop the basis for implementing a performance measurement system for evaluating changes in health outcomes in the VHA.

1999 VA National Health Survey



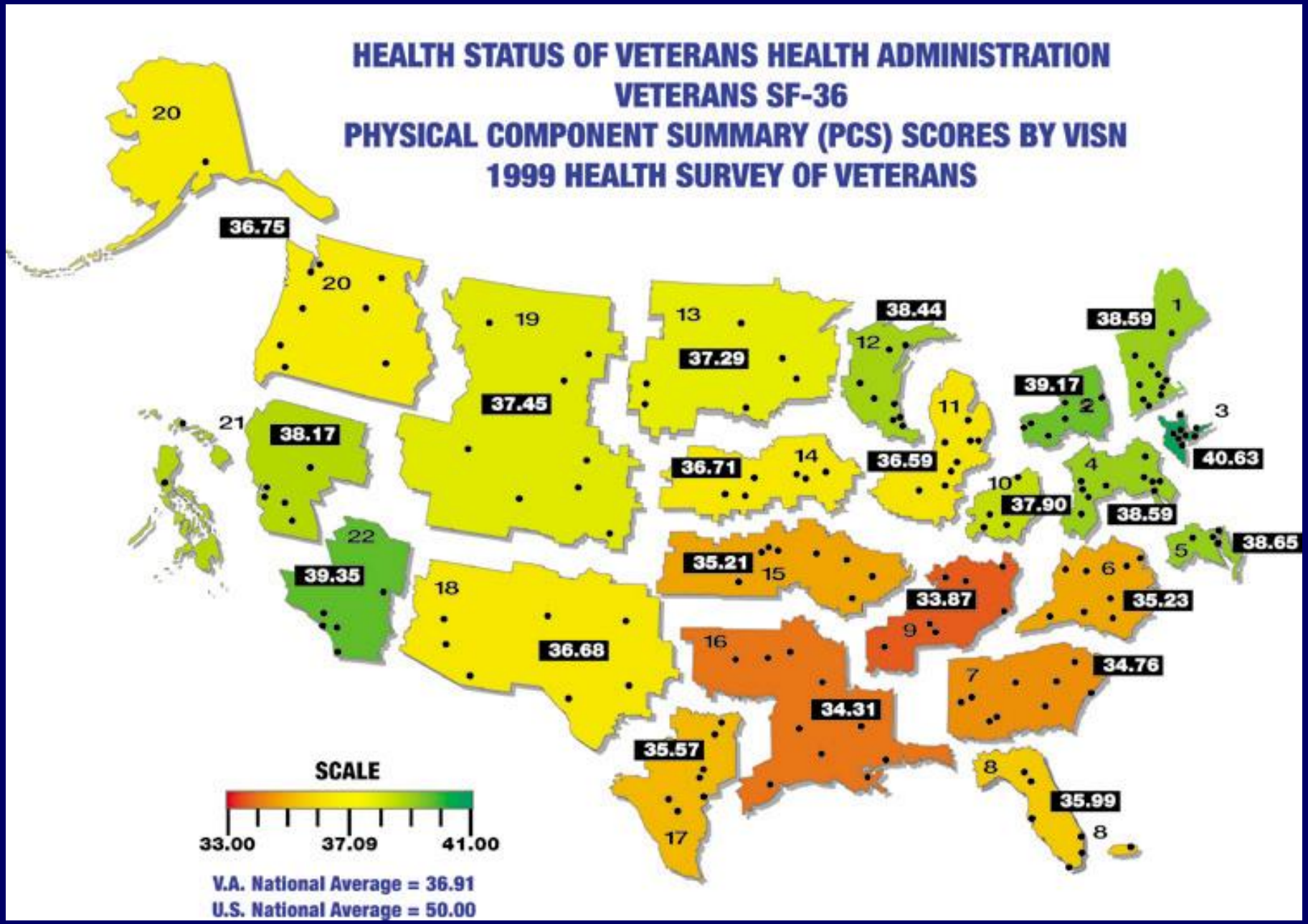
1999 VA National Health Survey of Veteran Enrollees Methods

- Survey responses merged with computerized medical record files (VISTA files) to obtain socio-demographics and diagnoses

1999 VA National Health Survey of Veterans

Demographics

Variable	Category	Percent
Age (mean = 59.9; S.D. = 15.1)	18-49 50-64 65-98	25.8 29.7 44.5
Race	White Black Hispanic Other	72.9 15.4 5.8 5.9
Gender	Male	95.2
Marital Status	Married	58.3
Education	≤ 12 years	53.8


HEALTH STATUS OF VETERANS HEALTH ADMINISTRATION
VETERANS SF-36
PHYSICAL COMPONENT SUMMARY (PCS) SCORES BY VISN
1999 HEALTH SURVEY OF VETERANS
20
36.75
20
19
37.45
13
37.29
38.44
12
38.59
1
39.17
2
3
40.63
11
36.59
14
36.71
21
38.17
10
37.90
4
38.59
5
38.65
22
39.35
18
36.68
35.21
15
6
35.23
33.87
9
16
34.31
7
34.76
35.57
17
8
35.99
8
SCALE
33.00
37.09
41.00
V.A. National Average = 36.91
U.S. National Average = 50.00

Comparison of the Medicare Advantage Program with the Veterans Health Administration

Background

- **U.S. Veterans Administration Health Care System.**
  - The largest integrated health care system in the United States with about 7 million veteran enrollees
- **The Centers for Medicare and Medicaid Services** offers the Medicare Advantage Program which is modeled after a capitated managed care system.
  - About 28% of all subscribers in Medicare are enrolled.

Comparison of the Medicare Advantage Program with the Veterans Health Administration

Overview

- In order to assess quality of care provided by Medicare Advantage (MCO plans), CMS has conducted the Medicare Health Outcomes Survey (HOS) a longitudinal initiative of health outcomes using Patient Reported Outcome Measures nationwide.

Comparison of the Medicare Advantage Program with the Veterans Health Administration

Overview

- Since 1998 and each subsequent year 2 year cohorts have been created on the basis of all plans (managed care organizations) nationally as a longitudinal evaluation of the health outcomes of Medicare Advantage managed care patients.
- The VA has conducted national surveys using PRO Measures for more than a decade for their QI program.

Comparison of the Medicare Advantage Program with the Veterans Health Administration

Overview

- The VR-12 is used in both systems of care
- These nationwide initiatives among the two systems of care offer the unique opportunity to compare the VA with CMS.
- Using appropriate risk adjustments of PRO outcomes between these two systems of care

Comparison of the Medicare Advantage Program with the Veterans Health Administration

Objectives

- We sought to examine the Patient Reported Outcomes and mortality of patients in the Medicare Advantage program compared to the VHA using nationally representative data.

Comparison of the Medicare Advantage Program with the Veterans Health Administration

Objectives

- The objectives were:

1. To examine differences in patient outcomes with the VR-36/12 at baseline and follow-up between the Medicare Advantage Program and the VHA.
2. To determine if there are differences between the Medicare Advantage Program and the VHA after controlling for differences in case-mix.

Comparison of the Medicare Advantage Program with the Veterans Health Administration

Methods

- Samples included
  - National VA data from 1999 to 2003 that identified patient cohorts from the Large Health Survey of Veteran Enrollees and the SHEP (Survey of Health Care Experiences) with an N= 35,876 of veterans 65 to 99 years of age with calculable PCS and MCS scores at baseline and follow-up.
  - For CMS Medicare advantage three HOS cohorts (1999 – 2003) also with calculable PCS and MCS scores giving an N= 71,424.

Comparison of the Medicare Advantage Program with the Veterans Health Administration

Methods

- Dependent variables included:
- 1. the probability of being alive with the same or better PCS at 2 years (+/-5.66) (more than would be expected by chance)

  (1- prob death)*(prob PCS the same or better)

- 2. the probability of being alive with the same or better MCS at 2 years (+/- 6.72) (more than would be expected by chance)

  (1- prob death)*(prob MCS the same or better)

- 3. Mortality uses a two year window and is based upon the Death Master File from the Social Security Administration and the VA death files (BIRLS files) .

Comparison of the Medicare Advantage Program with the Veterans Health Administration

Methods

- Independent Variables:
  - Demographics: age, race/ethnicity, marital status, education and income.
  - Medical conditions: CAD/MI, Angina, CHF, Stroke, Hypertension, Diabetes, COPD, asthma and cancer (self-report).

Comparison of the Medicare Advantage Program with the Veterans Health Administration

Statistical Methods (continued)

- Adjusted probability of being alive (Cox Models) were combined with the adjusted probability of same/better using the physical summary score (PCS) and separately mental summary score (MCS).

- Propensity matching was also used for comparing like patients across the systems of care.

Comparison of the Medicare Advantage Program with the Veterans Health Administration

Statistical Methods (continued)

- We also examined special populations separately (older subjects 75 and older, whites and minorities (hispanics and blacks) and selected chronic conditions
- To give special consideration to population groups with a higher risk of death and likelihood of poorer health outcomes in order to examine the degree to which the systems differ for the most vulnerable patients.

Sociodemographic Characteristics of Medicare Advantage and Veterans Health Administration Patients

Variable	Medicare Advantage (N=198,421)	Veterans Health Administration (N=360,316)
Age, years (SD)	74.0 (±6)	73.7 (±5)
Race/Ethnicity		
Whites	89.0%	82.2%
African Americans	6.4%	8.7%
Hispanics	1.8%	4.8%
Other	2.8%	4.3%
Marital Status (Married)	77.4%	68.9%
Education (<12 years)	30.7%	38.9%
Income (<$20,000)	41.9%	65.6%

- The social disadvantage index includes minority, unmarried, less than 12 years of education and income less than $20,000. A higher score indicates greater disadvantage.

All comparisons between MA and VHA were significant at <0.0001

Clinical Features of Medicare Advantage and Veterans Health Administration Patients

Variable	Medicare Advantage (N=198,421)	Veterans Health Administration (N=360,316)
Diabetes	19.80%	28.20%
Hypertension	52.20%	65.70%
Angina	20.70%	34.50%
Coronary Artery Disease/Myocardial Infarction	15.70%	28.30%

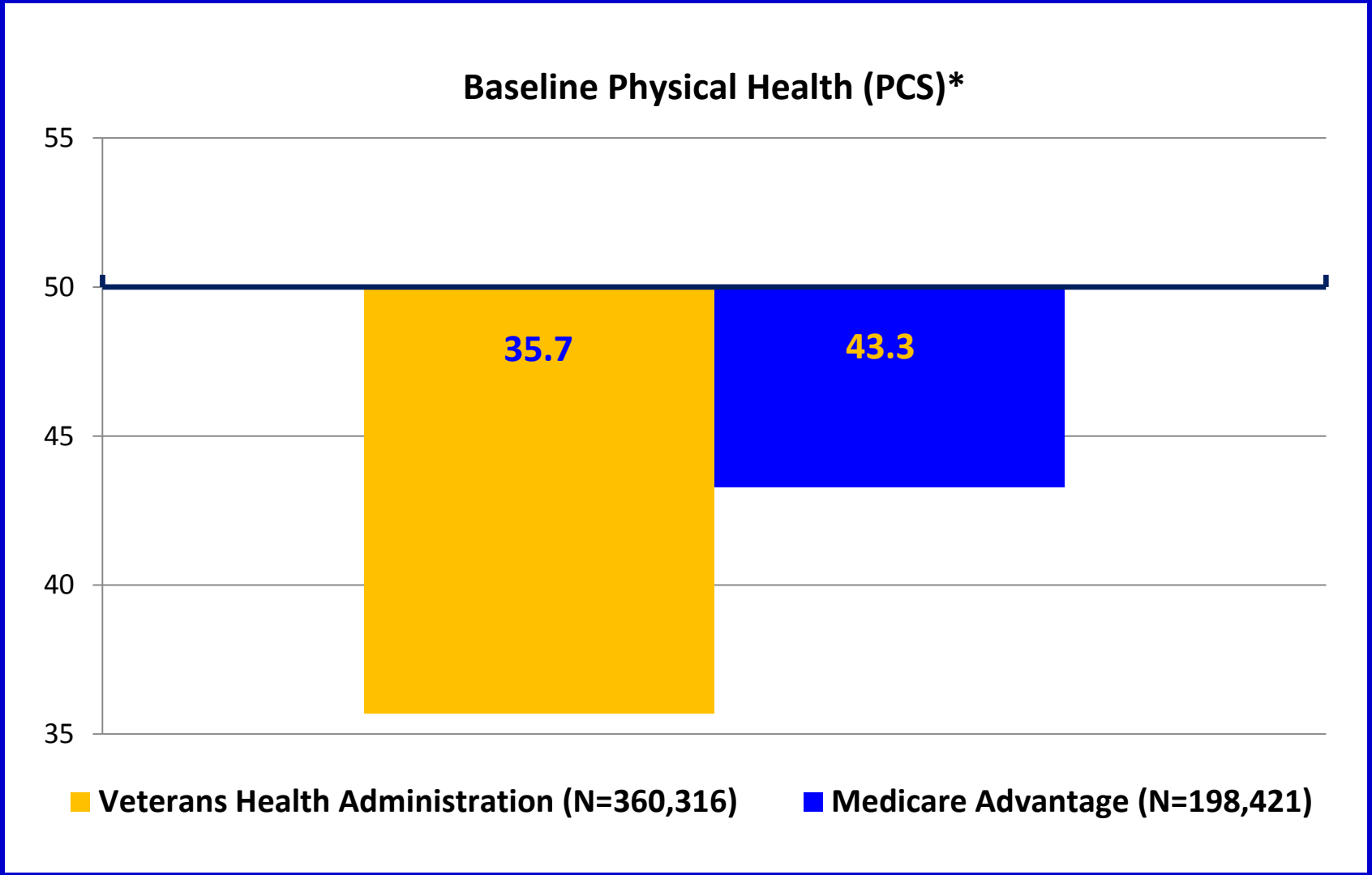
*A lower number is indicative of poor health status for MCS and PCS
All comparisons between MA and VHA were significant at <0.0001

Clinical Features of Medicare Advantage and Veterans Health Administration Patients

Variable	Medicare Advantage (N=198,421)	Veterans Health Administration (N=360,316)
Congestive Heart Failure	8.70%	24.90%
Stroke	9.30%	15.30%
Chronic Obstructive Pulmonary Disease/Asthma	13.50%	25.80%
Cancer	15.10%	19.70%

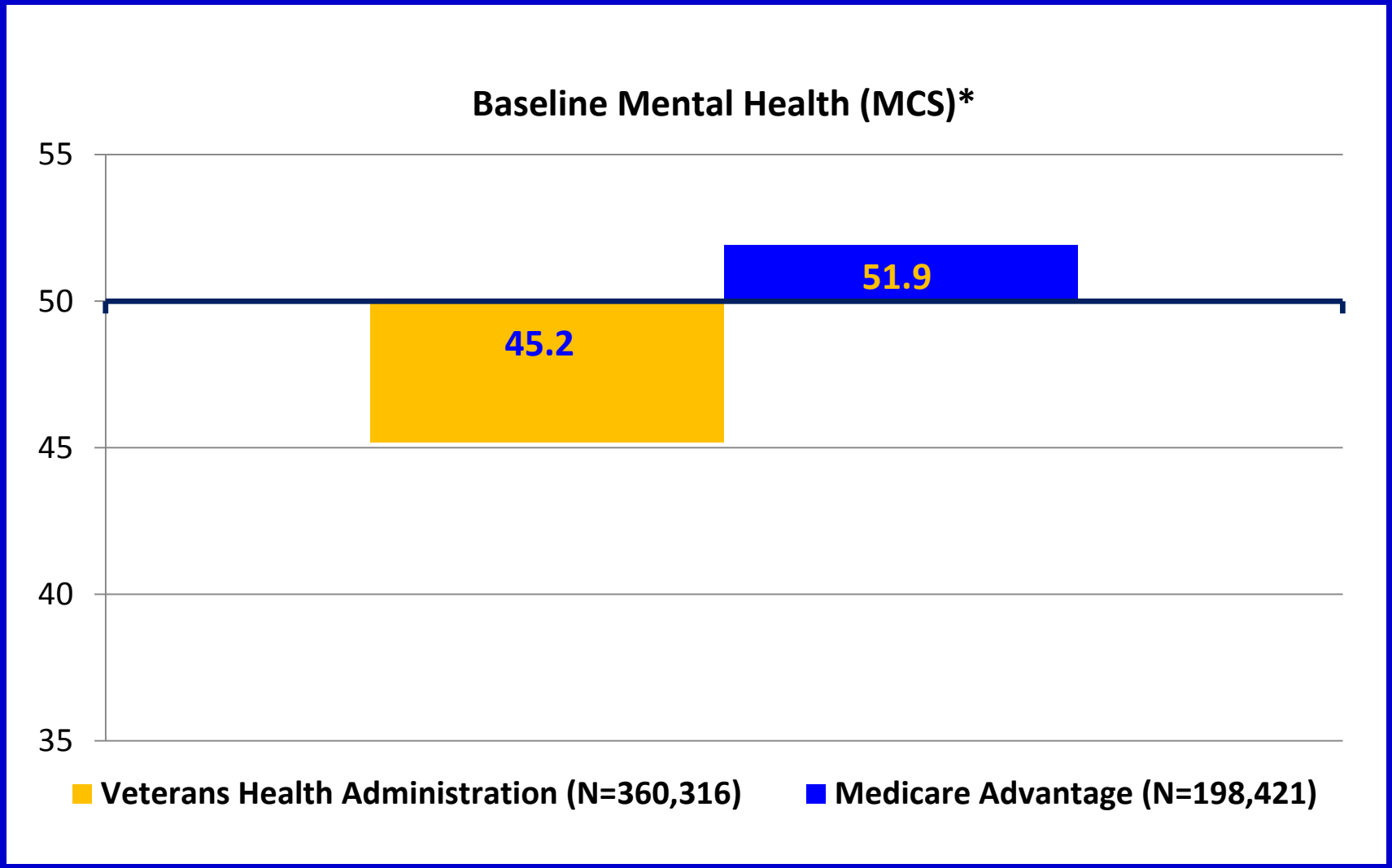
*A lower number is indicative of poor health status for MCS and PCS
All comparisons between MA and VHA were significant at <0.0001

Clinical Features of Medicare Advantage and Veterans Health Administration Patients



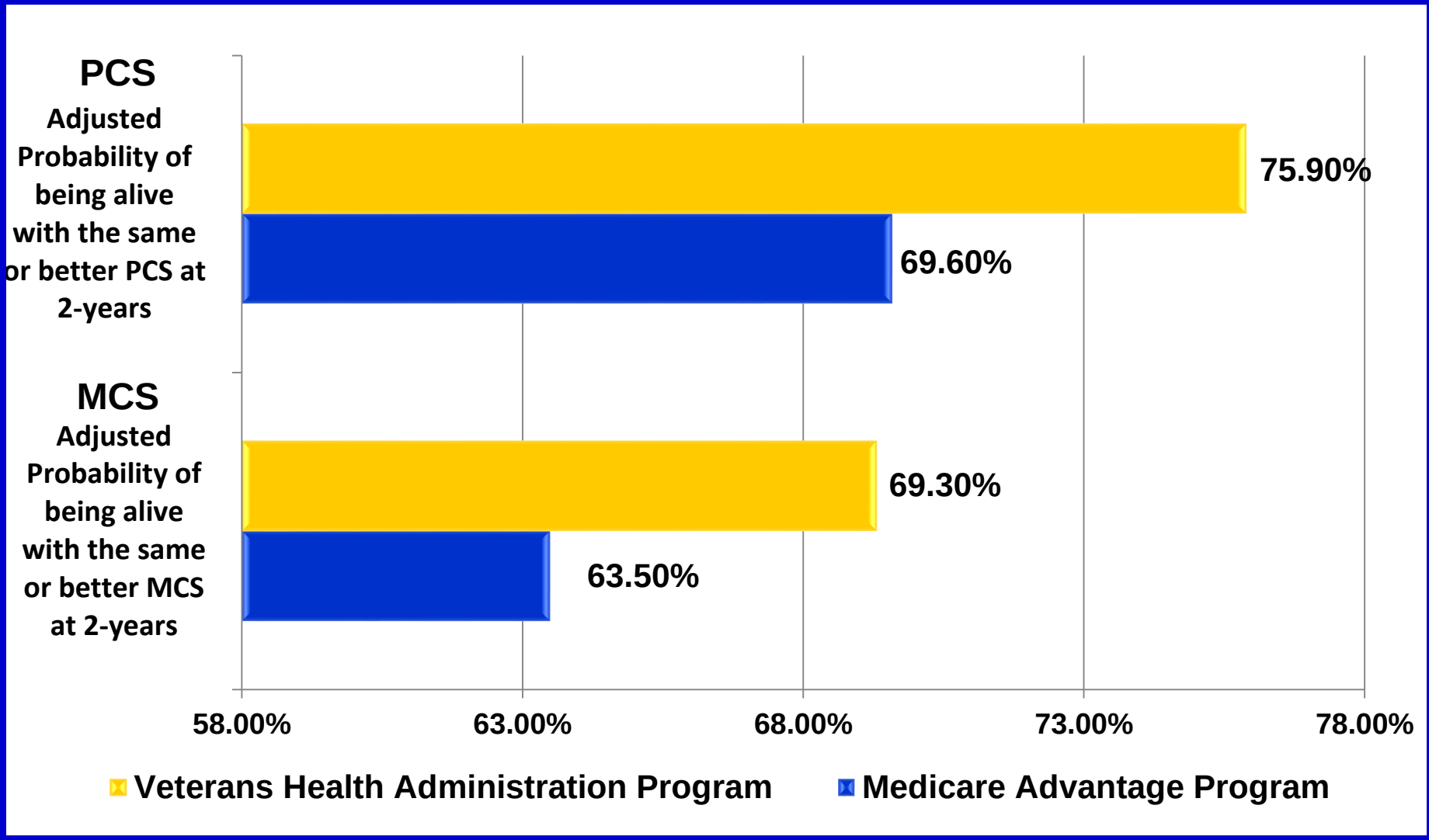
*A lower number is indicative of poor health status for MCS and PCS
All comparisons between MA and VHA were significant at <0.0001
*One standard deviation is equal to 10 points.

Clinical Features of Medicare Advantage and Veterans Health Administration Patients

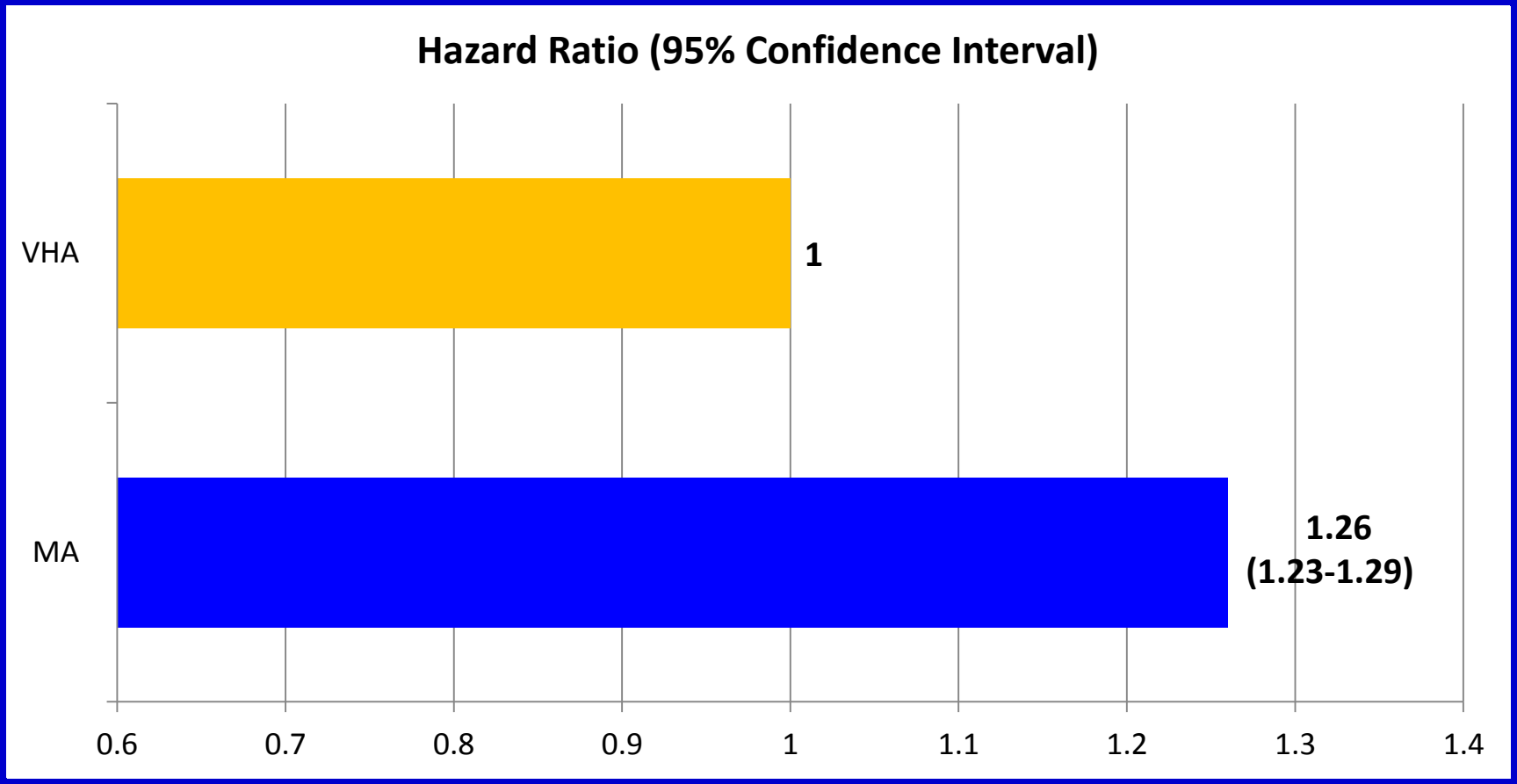


*A lower number is indicative of poor health status for MCS and PCS
All comparisons between MA and VHA were significant at <0.0001
*One standard deviation is equal to 10 points.

Change in Health Status in the Medicare Advantage and the Veterans Health Administration over 2 Years



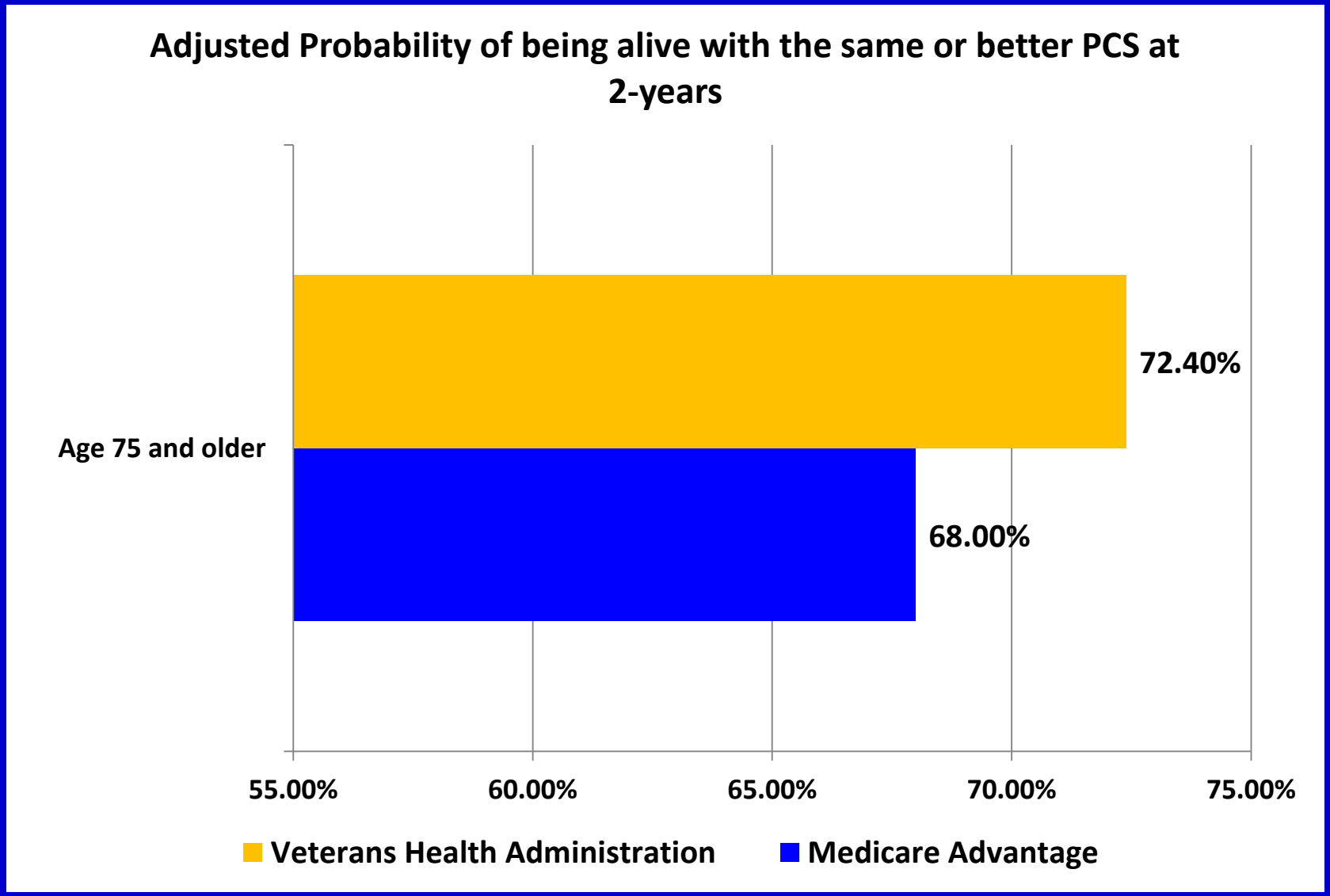
Risk of Death in 2 years Measured by Hazard Rate Comparing the Medicare Advantage (MA) and the Veterans Health Administration (VHA)



¶ All comparisons between MA and VHA were significant at p-value <0.001.

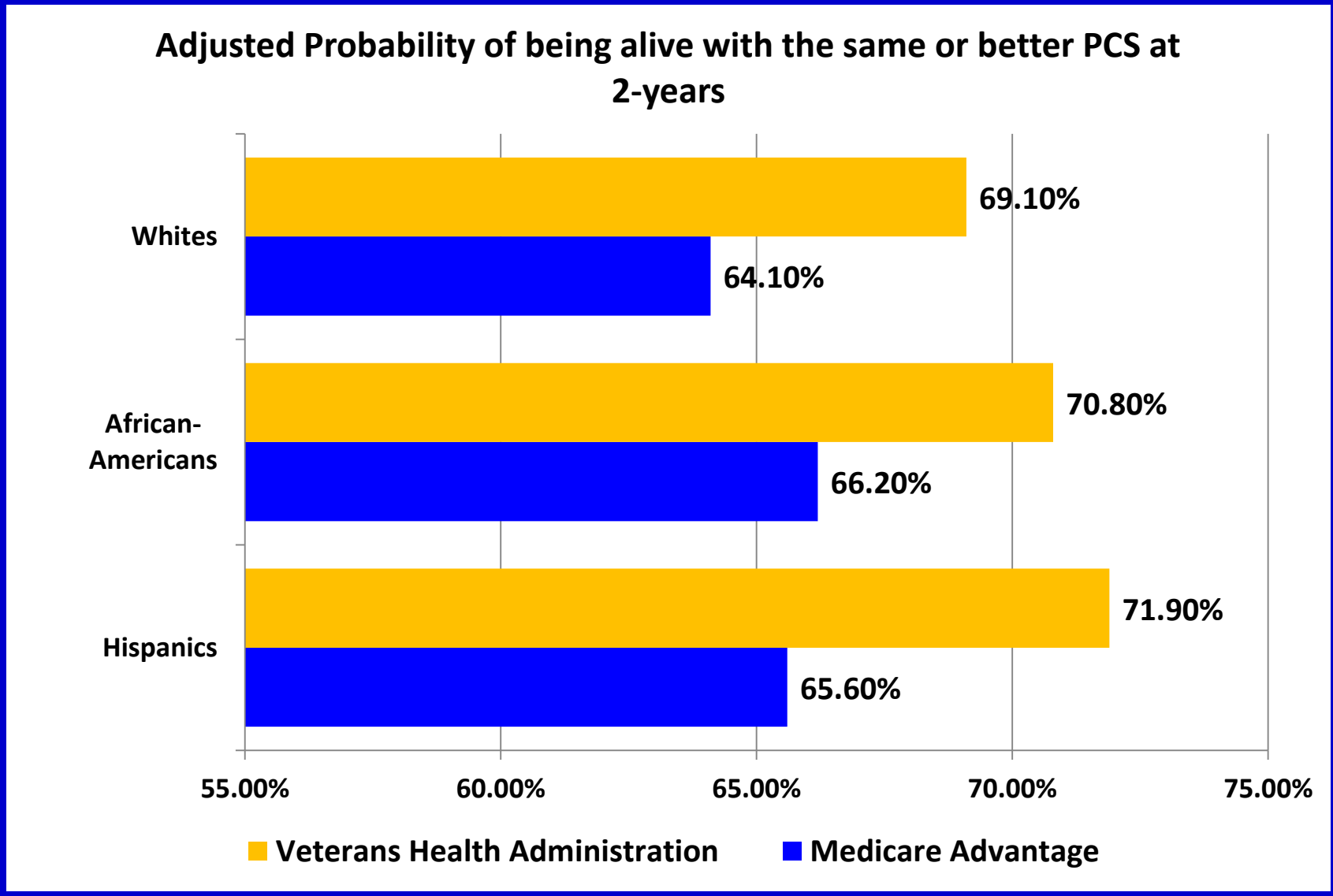
Change in Health Status in the MA and the VHA by Vulnerable Populations

Physical Health (PCS)



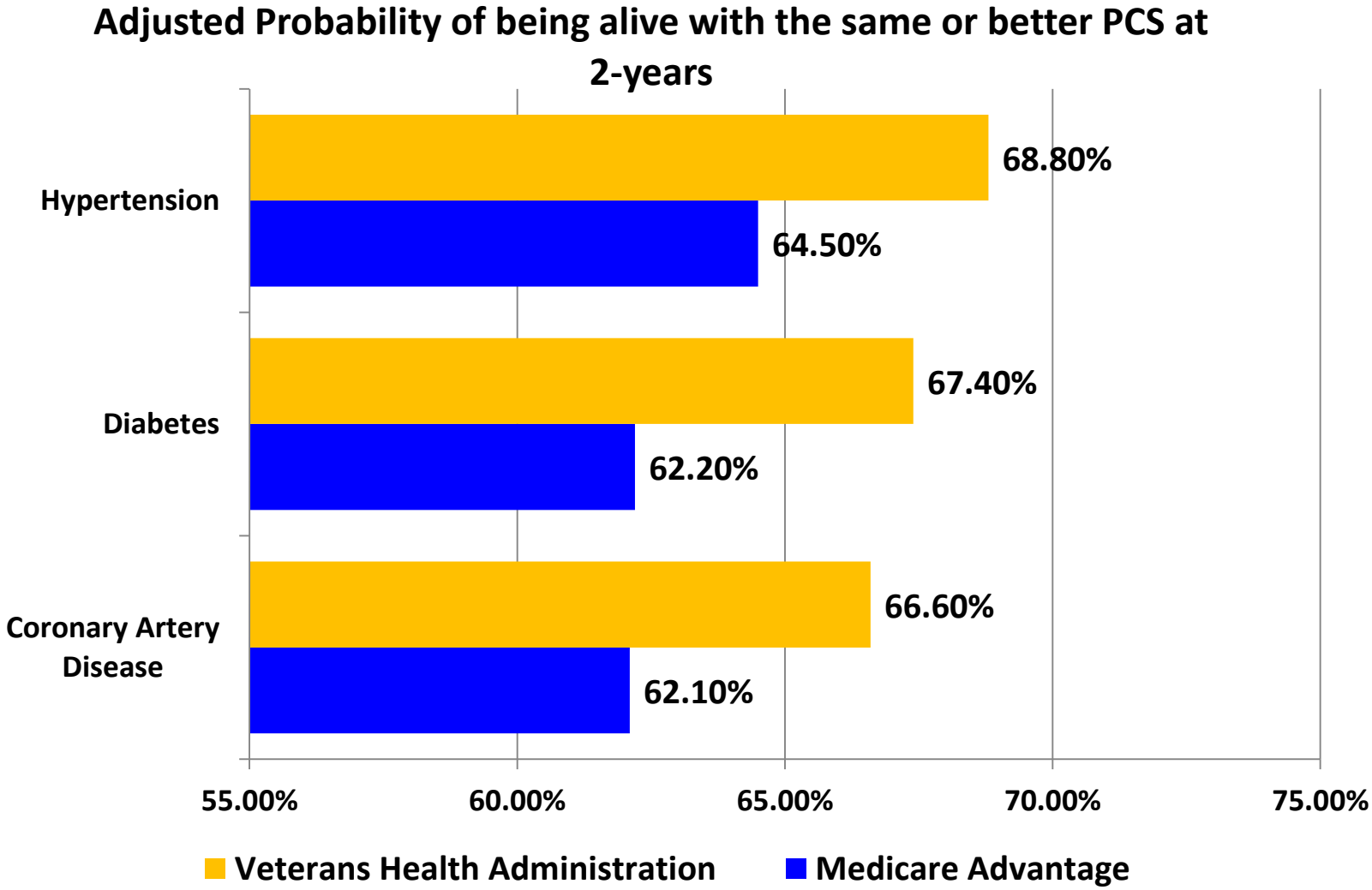
Change in Health Status in the MA and the VHA by Vulnerable Populations

Physical Health (PCS)



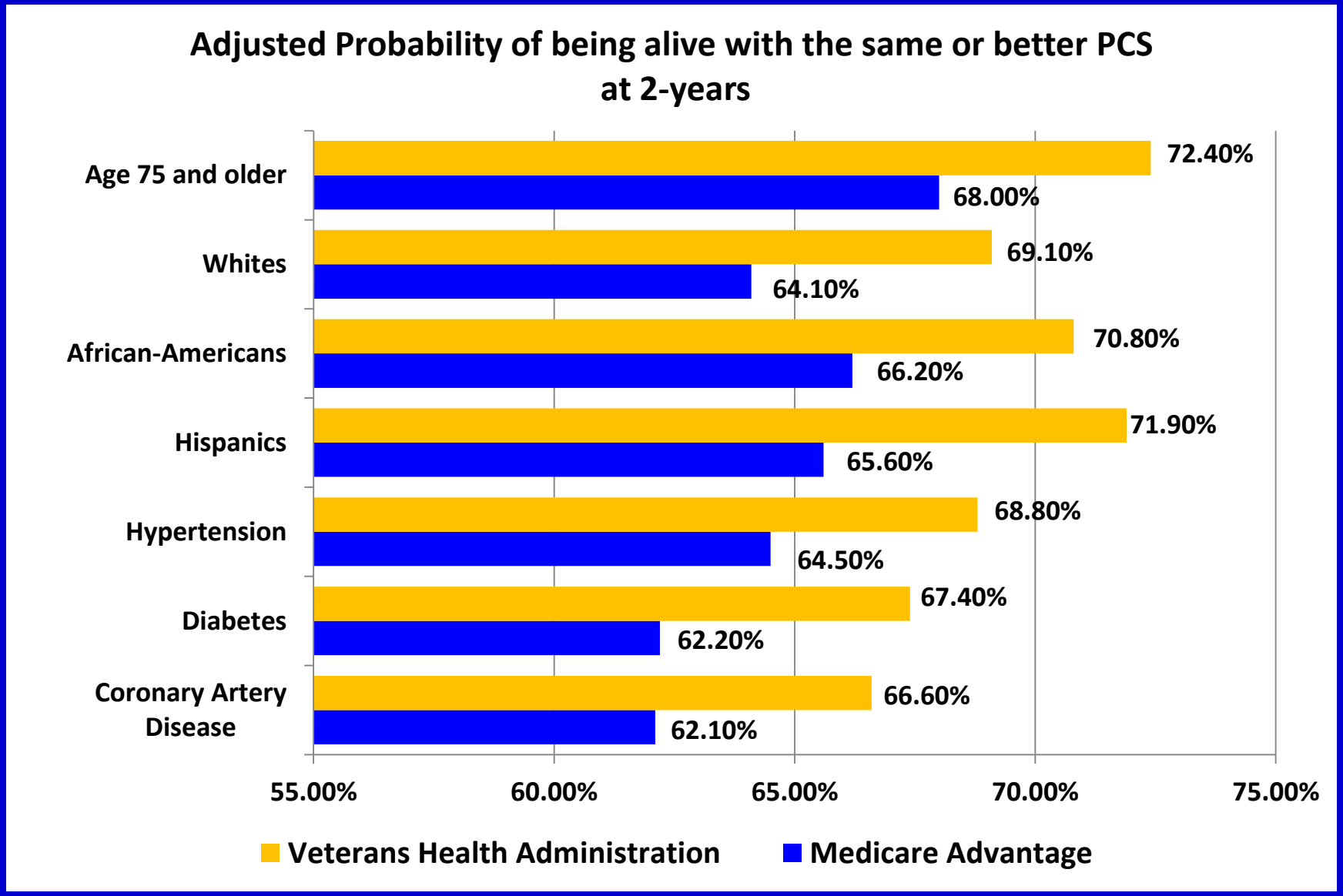
Change in Health Status in the MA and the VHA by Vulnerable Populations

Physical Health (PCS)



Change in Health Status in the MA and the VHA by Vulnerable Populations

Physical Health (PCS)



Population Based Studies

- Comparisons between the VHA and the Medicare Advantage Program using VR-12 outcomes and mortality rates favor the VHA overall.

- These findings are fairly robust as they are consistent across a number of vulnerable demographic and chronic disease subgroups.

- This suggests that future work can consider factors other than case-mix in explaining variations in outcomes, including access to health care and specific processes of care.

Discussion

- Initiation of the VHA System Transformation in the 1990's and early 2000 may have contributed to the better health outcomes found in this study.

Discussion

- This VA transformation included:
- A new organizational structure (VISN)
- State of the art Electronic Medical Record System
- Active Monitoring System to achieve system wide coordination
- Accountability (one of the first systems to develop a QI program with metrics that were linked to resource allocation at the VISN levels with bonuses (not unlike a pay for performance system).
- Emphasis on access to care and an integrated patient referral system.
- Use of big data to monitor health care processes and assist in the development of a system transformation

Enriching data via mining electronic health records
Natural Language Processing (NLP)

- Sources:
  - Unstructured clinical text, such as progress notes, radiology reports, and pathology reports
- Applications:
  - Automated data extraction (eg ejection fraction, clinical assessments such as Framingham Criteria, PHQ-2, & PHQ-9)
  - Disease phenotyping
    - Assess accuracy of diagnoses/coding
    - Evaluate severity
  - Cohort selection
  - Risk prediction (eg predictors of suicide among veterans)
  - Track adverse events (eg pharmacovigilance, complications of surgery)

(1) Future and Vision

- Big Data has important applications such as monitoring the processes of health care systems and can be applied at a number of different levels starting with the population and then targeting clinical outcomes and individual patients.

(2) Future and Vision

- The use of computer technologies has many hazards associated with its use.
  - Results can only be as good as the data used to address a problem. There are many problems with the use of claims data, clinical information and other personal data.
  - Study design is also an issue. Data is often cross-sectional and can be prospective. Naturalistic studies (Quasi-experimental designs) might lead to erroneous conclusion.
  - Confidentiality of data is also important challenge.
- As we move forward the use of these measures is bound to permeate health care at all levels.

(3) Future and Vision

- Big data provides enormous opportunities for reshaping and providing monitors for health care quality and better understanding processes of care to improve the health care system.

- As we move forward the use of big data is bound to permeate health care at all levels.

"Before we get started, who's your health-care provider?"