

The Wisdom and Follies of Numbers in Healthcare

**Complex Adaptive Systems Conference
November 15, 2013**

Joseph Francis, MD, MPH
Director, Clinical Analytics and Reporting
Office of Analytics and Business Intelligence
Veterans Health Administration

Goals of this Presentation

- **Identify the power and value of measurement in driving health system improvement**
- **Identify known pitfalls of measurement based on**
 - **Limits of Humans**
 - **Limits of Numbers**
- **Give examples of how Health IT might make a difference**
- **Suggest a “balanced approach” for leaders of healthcare systems**

A Few Disclaimers...

- **I like measurement**
- **I won't be addressing the scientific controversies underlying how quality is defined**
- **I make no moral judgment about the unintended behaviors resulting from measurement**
- **The examples I share are all in available literature**
- **The opinions in this talk are my own**
- **I am not a complexity scientist**

Why Measure in Healthcare?

- **More reliable than peer review**
- **Provides “objective” basis for accountability**
- **Can change provider behavior**
 - **Accelerate uptake of evidence-based practices**
 - **Reduce practice variation**
 - **Reduce low-value care**
- ***May* drive patients towards higher value health care**

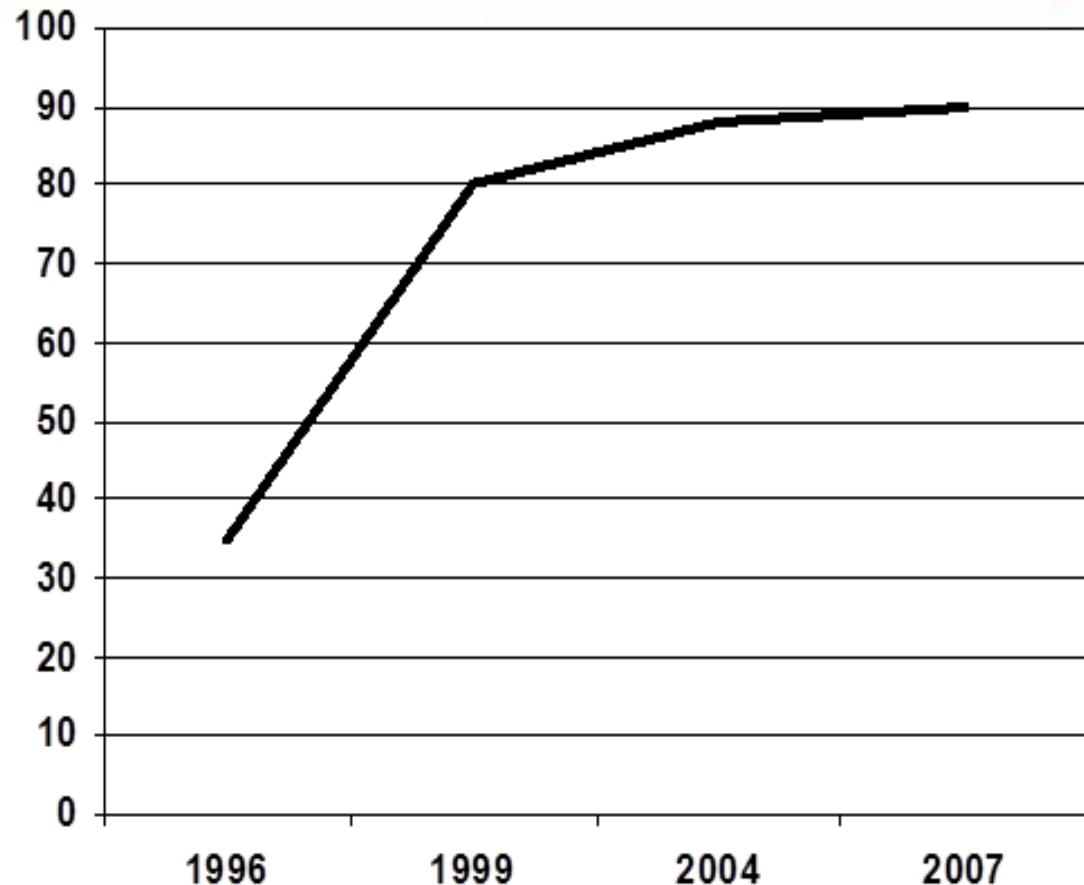
Measurement, Analysis, & Reporting in VA

- **Began in 1996 with 10 manual measures**
 - **Linked to Senior Executive ratings**
 - **Focus of local clinical QI**
- **A key driver of VA's quality transformation**
- **VA now tracks over 500 performance measures from multiple sources including our electronic health record system, VistA**
- **HITECH and ACA discussions are catalyzing shift to electronic measures of quality in VA and across the nation.**

VA Performance Transformation

Prevention Index

- Flu & pneumo vaccine
- Breast, cervical, colorectal screening
- Tobacco & alcohol screening/intervention
- Lipid screening
- Prostate counseling



IMPROVING PATIENT CARE

Improving Patient Care is a special section within *Annals* supported in part by the U.S. Department of Health and Human Services (HHS) Agency for Healthcare Research and Quality (AHRQ). The opinions expressed in this article are those of the authors and do not represent the position or endorsement of AHRQ or HHS.

Comparison of Quality of Care for Patients in the Veterans Health Administration and Patients in a National Sample

► Steven M. Asch, MD, MPH; Elizabeth A. McGlynn, PhD; Mary M. Hogan, PhD; Rodney A. Hayward, MD; Paul Shekelle, MD, MPH; Lisa Rubenstein, MD; Joan Keesey, BA; John Adams, PhD; and Eve A. Kerr, MD, MPH

21 December 2004 | Volume 141 Issue 12 | Pages 938-945

Background: The Veterans Health Administration (VHA) has introduced measurement, and other system changes directed at improving care. Results are based on a small set of indicators.

Objective: To compare the quality of VHA care with that of care in a national sample.

Design: Cross-sectional comparison.

Setting: 12 VHA health care systems and 12 communities.

Patients: 596 VHA patients and 992 patients identified through random-digit dialing.

Measurements: Between 1997 and 2000, quality was measured by using a validated instrument. Results were adjusted for clustering, age, number of visits, and medical conditions.

Results: Patients from the VHA scored significantly higher for adjusted chronic disease care (72% vs. 59%; difference, 13 percentage points [CI, 12 to 28 percentage points]), but not for acute care. The VHA advantage was consistent across all 12 systems and communities. The VHA advantage was also consistent across all 12 communities. The VHA advantage was also consistent across all 12 systems and communities.

- **Article**
 - ▶ Table of Contents
 - ▶ Abstract of this article
 - ▶ PDF of this article (PDFs free after 6 months)
 - ▶ Summary for Patients
 - ▶ Figures/Tables List
 - ▶ Appendix Table
 - ▶ Related articles in Annals
- **Services**
 - ▶ Send comment/rapid response letter
 - ▶ Notify a friend about this article
 - ▶ Alert me when this article is cited

EDITORIAL

Creating a Culture of Quality: The Remarkable Transformation of the Department of Veterans Affairs Health Care System

For decades, fairly or unfairly, the Department of Veterans Affairs (VA) health care system had a suboptimal image in the quality of care it provided and in the evaluation of that care. About 10 years ago, the VA leadership

came, diabetes severity, and other comorbid conditions) uniformly across systems and used these measures to adjust for differences other than sex between the VA and commercial managed care.

IMPROVING PATIENT CARE

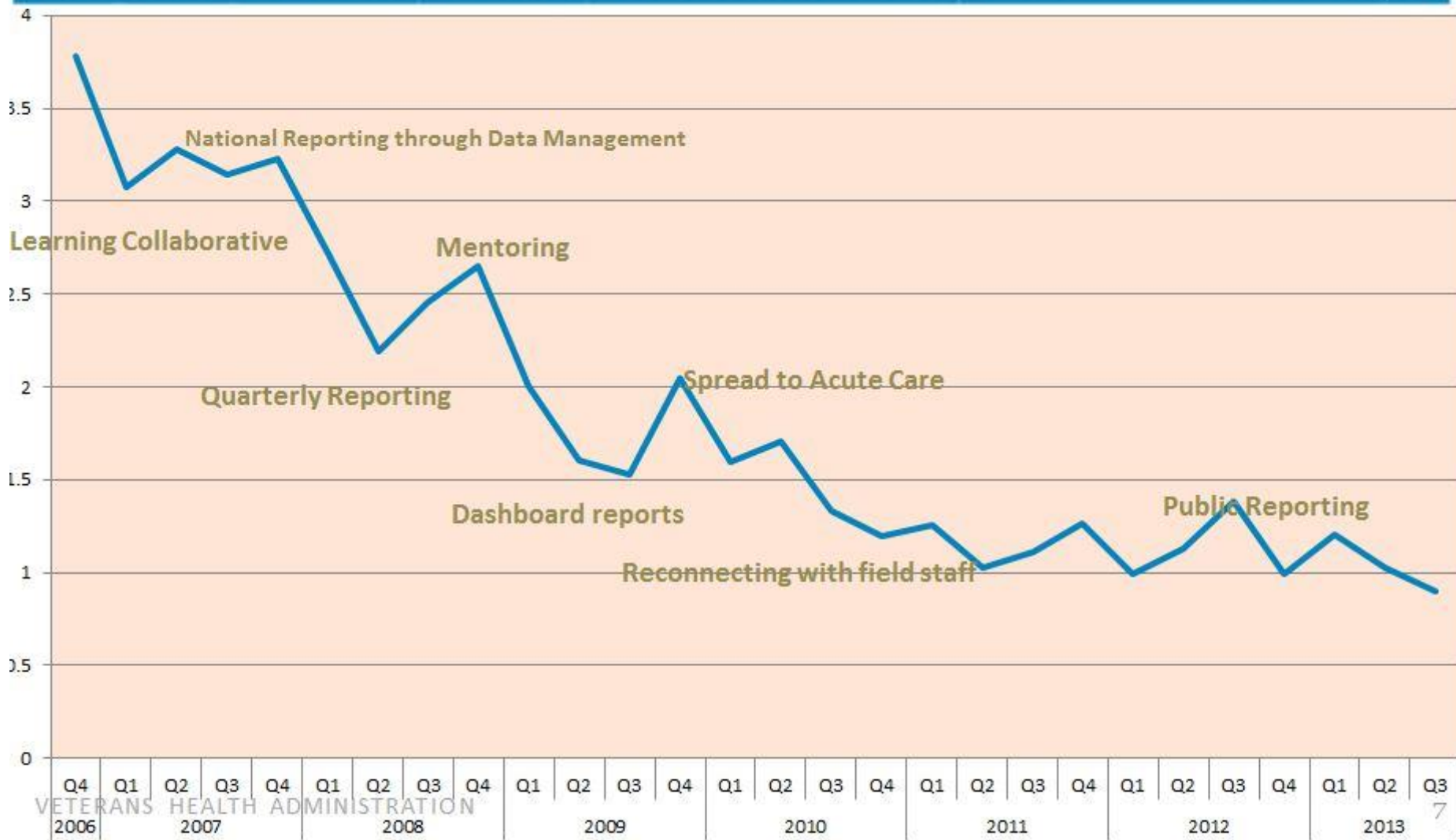
Diabetes Care Quality in the Veterans Affairs Health Care System and Commercial Managed Care: The TRIAD Study

Eve A. Kerr, MD, MPH; Robert B. Gerzoff, MS; Sarah L. Krein, PhD, RN; Joseph V. Selby, MD, MPH; John D. Piette, PhD; J. David Curb, MD, MPH; William H. Herman, MD, MPH; David G. Marrero, PhD; K.M. Venkat Narayan, MD, MSc, MBA; Monika M. Safford, MD; Theodore Thompson, MS; and Carol M. Mangione, MD, MSPH

Background: No studies have compared care in the Department of Veterans Affairs (VA) with that delivered in commercial managed care organizations, nor have studies focused in depth on care comparisons for chronic, outpatient conditions.

Results: Patients in the VA system had better scores than patients in commercial managed care on all process measures (for example, 93% vs. 83% for annual hemoglobin A_{1c}; $P = 0.006$; 91% vs. 75% for annual eye examination; $P < 0.001$). Blood

Power of Measurement, Feedback, and QI: Central Line Associated Bloodstream Infections



Universal Measurement Formula

SQR of Sample
Size

Information Richness

$$\textit{Confidence} = (\textit{Signal} \div \textit{Noise}) \times \sqrt{N}$$

How good is my decision?

“Messiness of Life”

Measurement Properties

- Measures only approximate reality
 - Signals are imperfect
 - Noise is always present
- Measures don't need to be perfect, just good enough to make our decisions better
- “Bigger data” is not necessarily better

The measure should be the starting point for deeper conversation



Human Limits and their consequences

Cognitive Biases when using Numbers

- Underestimating the likelihood of randomness
- Seeing patterns in randomness (clustering illusion)
- Seeing what we want to see (confirmation bias)
- Seeing what we're used to seeing (availability bias)
- Extrapolating beyond what we should

Gilovich, *How We Know What Isn't So*

Behavioral Biases with Numbers

- Measure is confused with reality
- We ignore the realities that are harder to quantify
- Feeling *losses* more sharply than gains
- ### + \$\$\$ →
 - Less willing to be creative or innovative
 - Doing what everyone else is doing
 - Manipulating measures (often unconsciously)

Ariely, *Predictably Irrational; The Upside of Irrationality*
Likierman, *The 5 traps of performance measurement*

Unintended Consequences of Well-Intended Measures (1)

Unintended Consequences of Implementing a National Performance Measurement System into Local Practice

Adam A. Powell, PhD^{1,2}, Katie M. White, EdD³, Melissa R. Partin, PhD^{1,2}, Krysten Halek, MA¹, Jon B. Christianson, PhD³, Brian Neil, MD⁴, Sylvia J. Hysong, PhD^{5,6}, Edwin J. Zarling, MD⁷, and Hanna E. Bloomfield, MD^{1,2}

“Facility-level strategies undertaken to implement national PM systems may result in inappropriate clinical care, can distract providers from patient concerns, and may have a negative effect on patient education and autonomy”

Unintended Consequences of Well-Intended Measures (2)

Journal of Health Economics 32 (2013) 341–352



Contents lists available at SciVerse ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase



Shipping out instead of shaping up: Rehospitalization from nursing homes as an unintended effect of public reporting

R. Tamara Konetzka^{a,*}, Daniel Polsky^b, Rachel M. Werner^{c,1}

Unintended Consequences of Well-Intended Measures (3)

ORIGINAL INVESTIGATION

ONLINE FIRST

The Cost of Satisfaction

A National Study of Patient Satisfaction, Health Care Utilization, Expenditures, and Mortality

Joshua J. Fenton, MD, MPH; Anthony F. Jerant, MD; Klea D. Bertakis, MD, MPH; Peter Franks, MD

Conclusion: In a nationally representative sample, higher patient satisfaction was associated with less emergency department use but with greater inpatient use, higher overall health care and prescription drug expenditures, and increased mortality.

Why Rating Your Doctor Is Bad For Your Health

SUFFERING FROM A TOOTHACHE, a [South Carolina](#) woman headed to her local emergency room a few months ago. The doctor there responded by administering Dilaudid, a powerful intramuscular narcotic typically reserved for cancer-related pain. Why, his nurse queried, was he killing a flea with a sledgehammer? Afraid of malpractice? No, the doc replied, Press Ganey. “My scores last month were low.”

**“Afraid of malpractice?” “No,”
the doc replied, “Press Ganey.
My scores last month were low”**

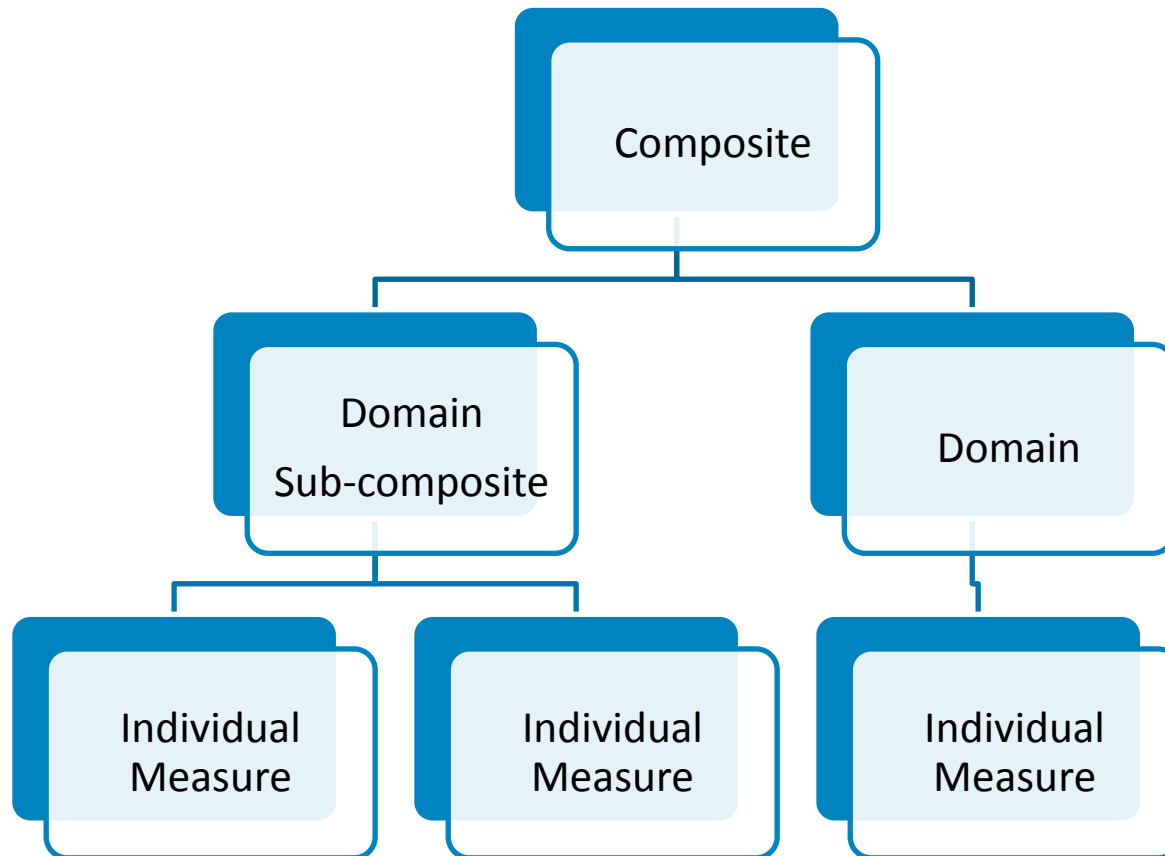
Measurement Overload



VA Measure Count

- **586 National Performance Indicators:**
 - 285 for Accountability
 - 271 for Quality Improvement
 - 30 pilot indicators
- **Measures by Domain:**
 - Clinical care: 360
 - Access/Timeliness: 117
 - Functional Status/Healthy Communities: 10
 - Operations: 85
 - Veteran Experience: 14

Composite Measures



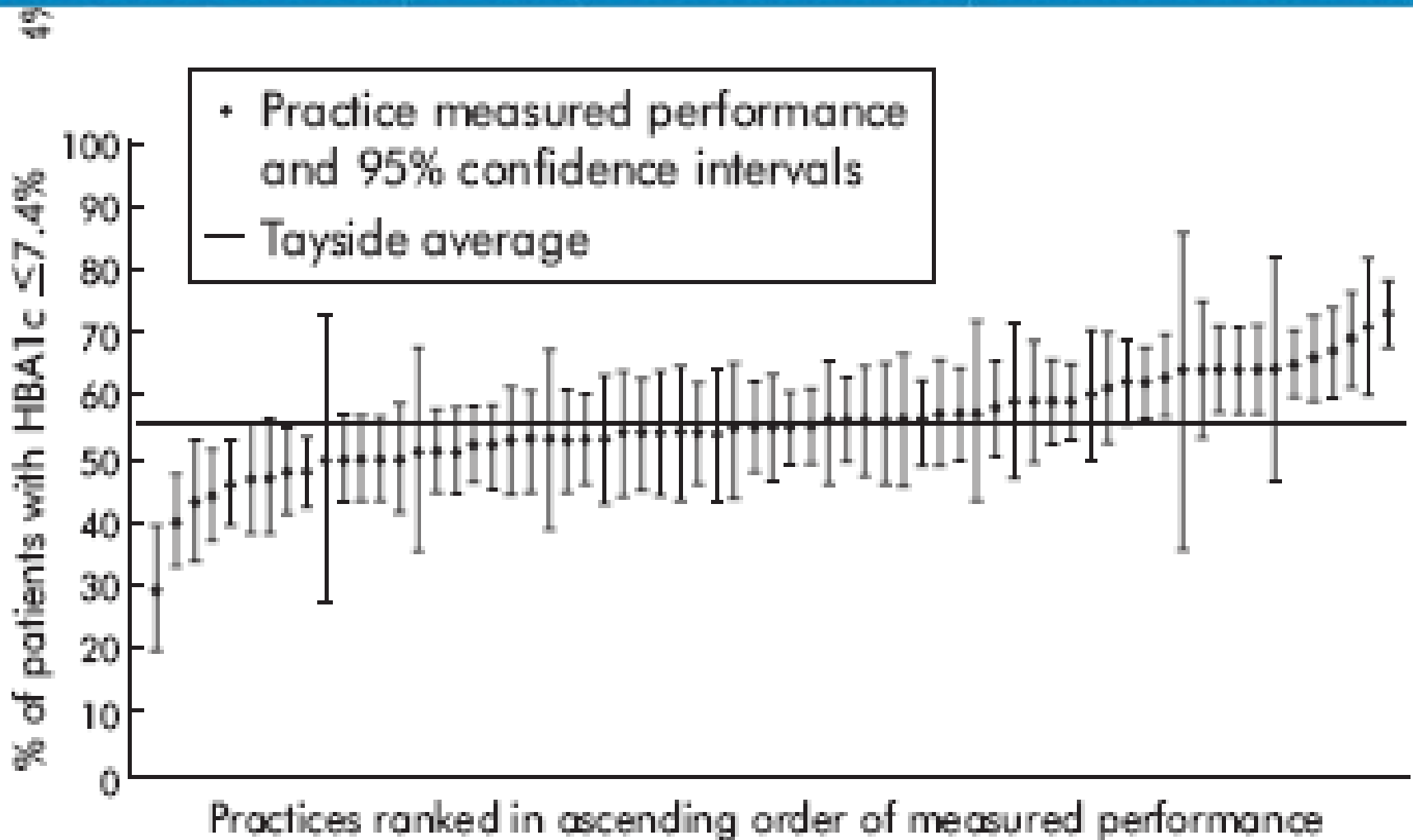
Why Composites?

- Used in other areas for years
 - Business, economics, education, “most liveable cities”
 - Often converted to Ranks or “Star Scores”
- Benefits:
 - A way to reduce complexity
 - May stimulate improvement or highlight weakness areas (especially if easy to “decompose”)

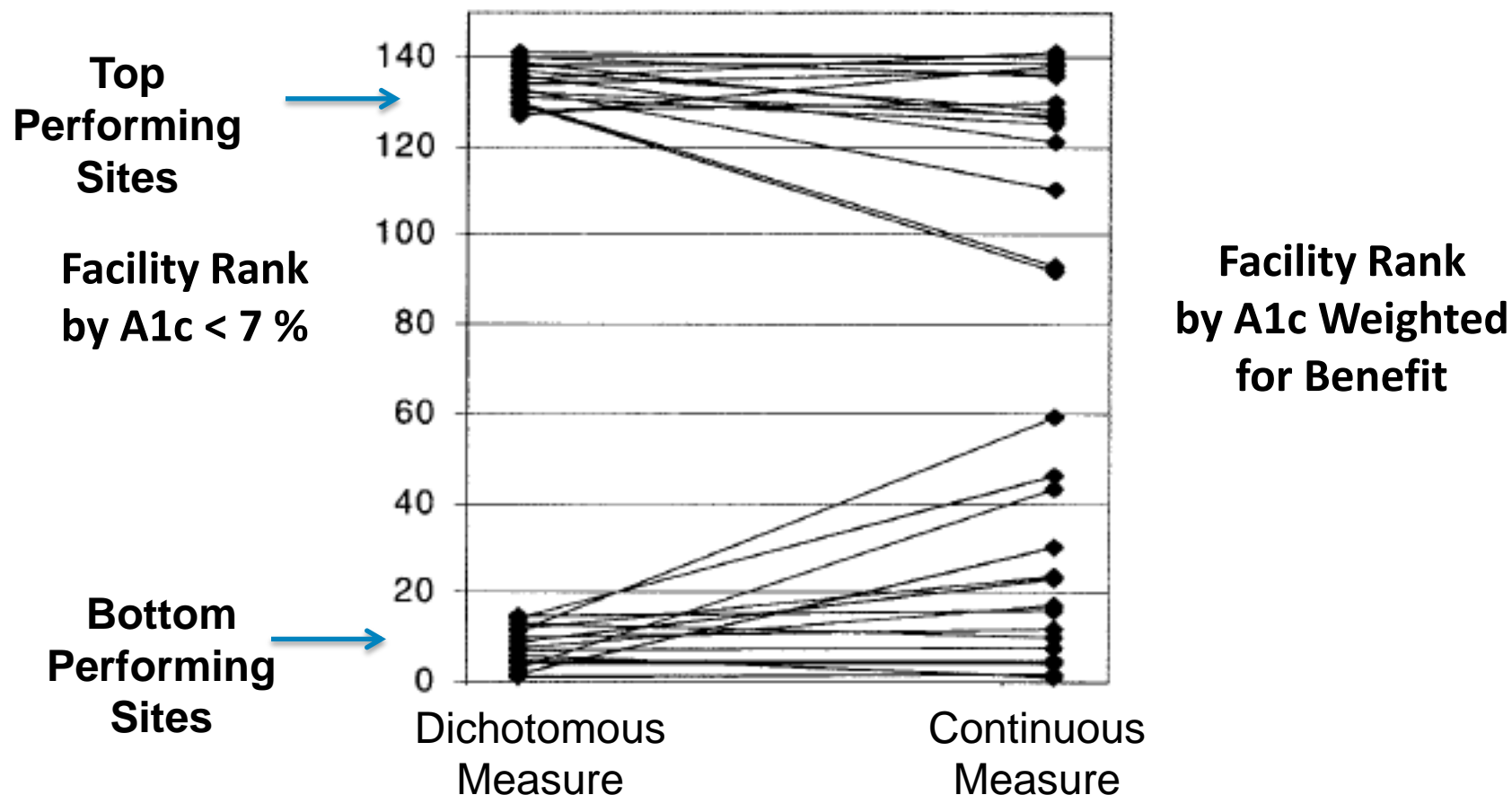
Concerns with Composites and Ranks

- **Too little variation:** high and low performers do not differ by a “clinically meaningful” amount
- **Too much variation:** “Overdispersion” can create goose chases
- **Funny things that numbers do** when aggregated (Simpson’s paradox, etc)
- **Sensitivity** to rules of construction

Ranking: For most indicators, actual differences in performance (signal) are small, and error (noise) is high



Ranking: Depends on Measure Construction, which can be arbitrary



Top Ranked Hospitals in 2010

U.S. News Top 5

- Hopkins
- Mayo Clinic
- UCLA
- Cleveland Clinic
- Mass General

“Why Not the Best” Top 5

- Hackettstown NJ
- NYC Community
- West Anaheim
- Flowers
- La Palma

“Top 100 Hospitals” – None of the Above!!

Ranking can provoke wasteful actions

The New York Times

Education

Baylor Rewards Freshmen Who Retake SAT

By SARA RIMER

Published: October 14, 2008

Baylor University in Waco, Tex., which has a goal of rising to the first tier of national college rankings, last June offered its admitted freshmen a \$300 campus bookstore credit to retake the SAT, and \$1,000 a year in merit scholarship aid for those who raised their scores by at least 50 points.



Health IT

Measurement and Health IT – potential benefits

- “Large N” samples & subsamples
- Potentially stronger Signal:
 - Hone in on defined populations (registries)
 - Capture continuous values and clinically meaningful actions
 - Track change at the level of individual patient (repeated measures)
- May reduce noise by capturing important sources of variation (e.g. co-morbidity)

Measurement and Health IT – potential pitfalls

- Key data are often missing
- Key data are often in unstructured text
- Structured data are often not coded consistently
- Capturing data in structured format is time consuming
- Structured data are often not interoperable between IT systems
- Data may not be consistent over time



Case Study 1: Clinical Action Measures

Traditional Performance Measurement Has Driven “Treat to Target”

- VHA has attained high levels of blood pressure (BP) control among patients with diabetes
 - BP < 140/90 approaching 80%
 - Less than 10 years ago it was 50%
- VHA has attained high levels of LDL lipid control
 - National private sector rates of LDL are <100mg/dl 46-56%
 - VHA attained 68% in first year of metric (2008) and 70% in 2010

What Dichotomous Measures Miss

- Dichotomous measures have “low signal”
- Dichotomous target measures fail to give credit for:
 - Intensification of medications
 - On “maximal” or reasonable medications
 - Contraindications to intensifying treatment
 - Actual lowering of risk

Alternate approach: Tightly linked Clinical Action Measure

- Care is deemed acceptable if:
 - The target is reached OR
 - Appropriate treatment is in place OR
 - Appropriate treatment changes have been made OR
 - A follow-up in 90 days shows improvement
- Such an approach provides a longitudinal view of a patient's care
- Note: even better to calculate risk reduction in risk!

Example: BP Control in Diabetic Patients

<i>Status</i>	Total	
	<i>N</i>	<i>%</i>
Index BP < 140/90	568,857	81.6%
Index SBP < 150 and DBP < 65	149,928	21.5%
Index SBP < 150 and ≥ 3 mod BP meds	106,612	15.3%
Increase dose existing med	47,474	6.8%
Start new med class or switch class	183,736	26.3%
Repeat BP < 140/90	149,168	21.4%
<i>Does not pass any of the above</i>	44,147	6.3%

Measuring Potential Overtreatment:

Potential Overtreatment for Hypertension

Index BP <140/65 mm Hg; *AND*

- Increase in BP medication dose within 90 days; OR
- Addition of BP medication class within 90 days and no medication drop within 120 days; OR
- On ≥ 4 BP medications with no medication drop within 120 days

Could we be overtreating BP in diabetics?

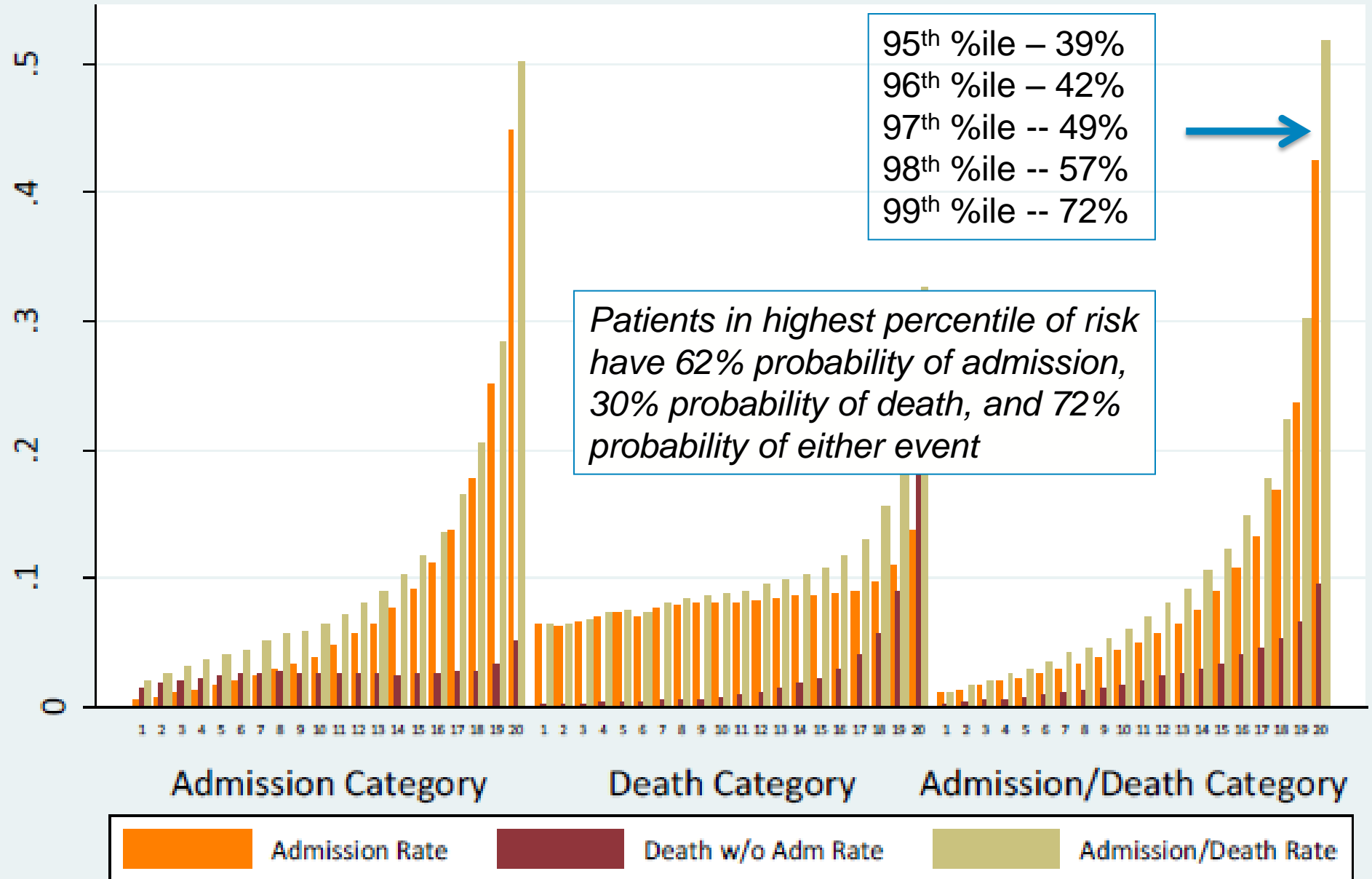
Reason	Hierarchical		Total	
	N	%	N	%
Increase BP med dose	5,999	0.9%	5,999	0.9%
Start new med class (no drop)	26,594	3.8%	27,775	4.0%
On \geq 4 BP meds (no drop)	12,270	1.8%	15,166	2.2%
Total overtreatment	44,863	6.4%		

Rate of over-treatment equals that of under-treatment!



Case Study 2: Assessing Patient Care Needs

1-Year Adverse Event Rates by Risk Categories

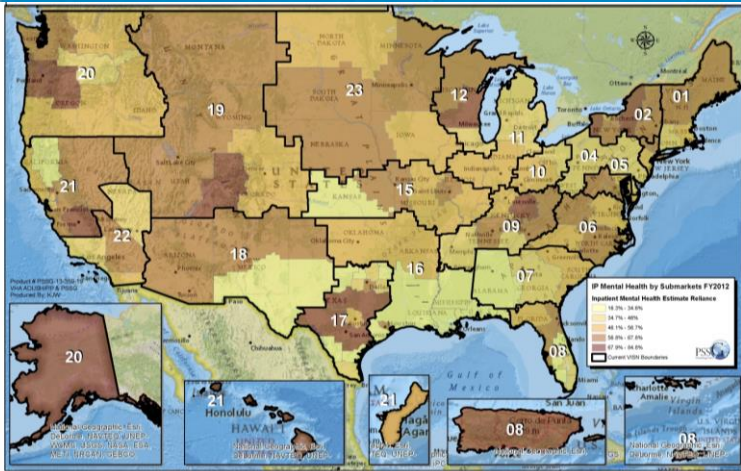


Patient Care Assessment System (PCAS)

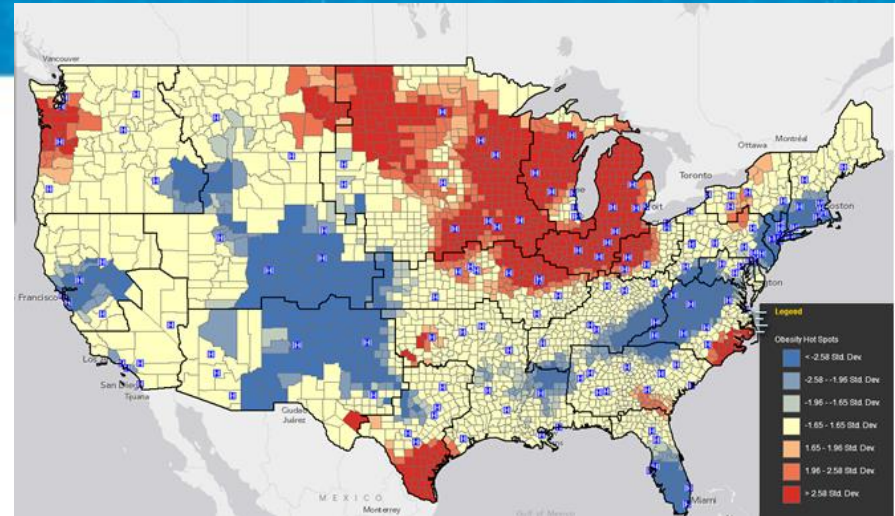
- Integration of key data from multiple sources
- Summary of patient risk factors
- Task Lists and notifications
- Multiple VAMCs & Community info
- Ability to create a care plan and write it back to CPRS as a standardized note



Geospatial Analysis

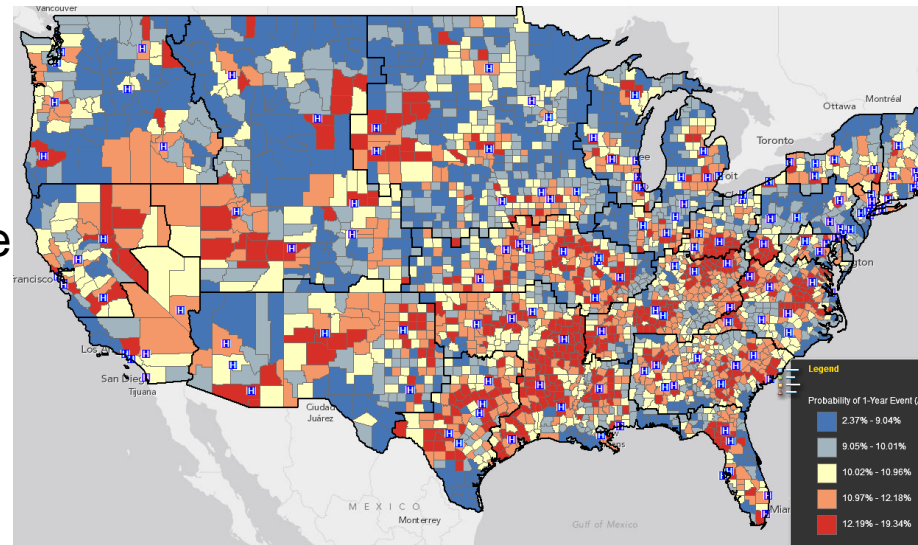


Inpatient Mental Health Utilization



Obesity "Hot Spots"

Care Assessment
Need (CAN) Score



<http://vhacinweb3i/flexviewer/>

VETERANS HEALTH ADMINISTRATION



Coda:

Moving from the “iPatient” to Context Sensitive Quality

Verghese, NEJM 2008;359:2749



Ceci n'est pas une pipe.

Magnit

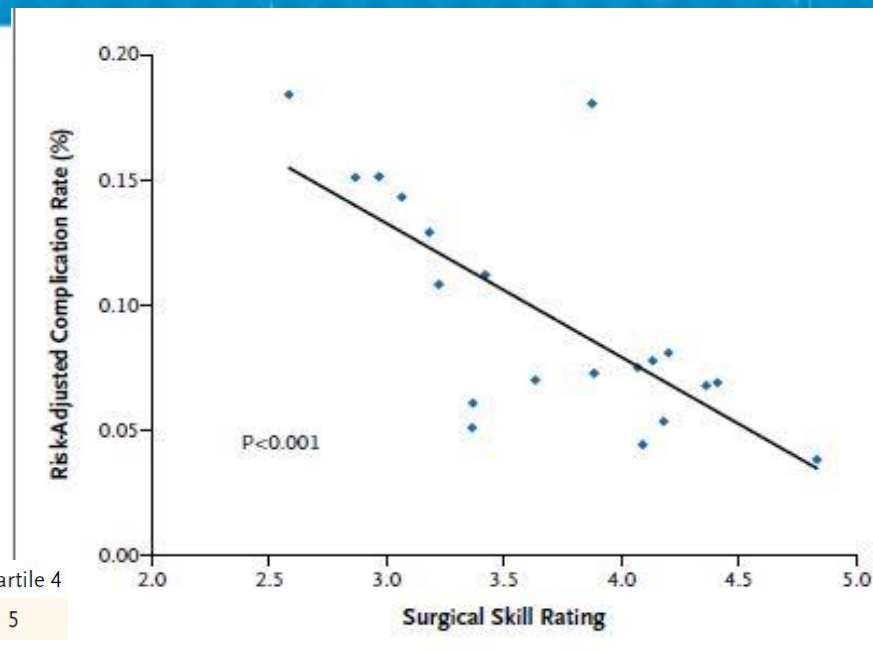
Direct observation of quality (1)



Clinical Scenario	Overall care appropriateness*
Straightforward Problem	73%
Co-morbid Biomedical Problems	38%
Psychosocial Issues	22%
Combined	9%

* using standard patients and clinical scenarios of asthma, diabetes, hip replacement, weight loss

Direct observation of quality (2)



	Quartile 1	Quartile 2 or 3	Quartile 4
Surgeons (no.)	5	10	5
Mean peer rating of technical skill*			
Gentleness	3.3	3.9	4.4
Time and motion	2.6	3.4	4.3
Instrument handling	2.9	3.7	4.4
Flow of operation	3.1	3.8	4.5
Tissue exposure	3.0	3.9	4.4
Overall technical skill	2.7	3.6	4.4
Summary rating	2.9	3.7	4.4

A Balanced Perspective

- Performance Measures are imperfect but powerful tools for shaping behavior.
- Professionalism is needed to mitigate the well-recognized downsides of imperfect measures
- Because measures only approximate reality, other sources of “strong signal” are needed.
- “Signals” can come from many sources
 - Direct observation should not be discounted
- A team willing to be self-critical can use measures as a springboard for the discussions that promote deeper understanding

Some Wisdom about Numbers

- “Maia” (Hindi): “measure” or “illusion”
- George Box (b. 1919): “All models are wrong, but some are useful”
- Korzybski (1879-1950): “The map is not the territory”
- Deming (1900-1993): *“Management by numerical goal is an attempt to manage without knowledge of what to do, and in fact is usually management by fear.”*