



American Board
of Internal Medicine

What do Professional, Accreditation Organizations, and Regulators Need to Assess Clinical Performance Across the Continuum?

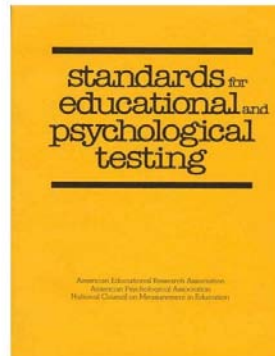
Rebecca S. Lipner, PhD
**NCVHS SUBCOMMITTEE ON QUALITY
HEARING**
October 18-19

Clinical Performance Assessment is Complex

- Physician clinical performance is a function of multiple competencies
 - Diagnostic reasoning, clinical care, communication with patients and peers, ability to work within a system, professionalism
- Need different types of data/measures to assess these
 - And at multiple levels: Patient, physician, system
 - Patients nested within physician
 - Physicians nested within system
- Need to make most accurate decisions about physician's clinical performance
- Need to evaluate performance improvement over time

High Quality Data and Information

- Data
 - Accurate, complete, comparable, timely
- Measures
 - Reliable, valid, feasible
- Classifications/decisions and consequences
 - High classification accuracy - reproducible, valid, meaningful and fair
 - Based on appropriate and planned sample design



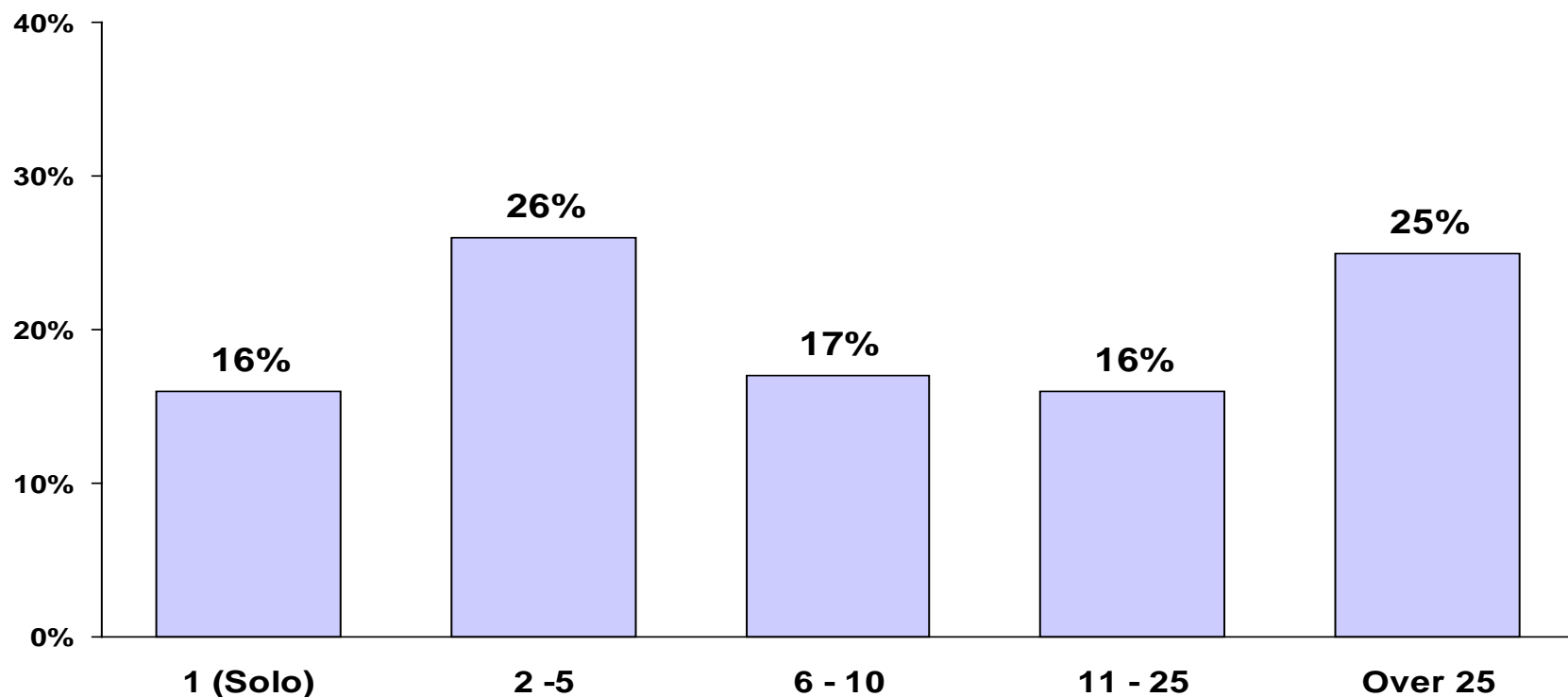
Specific Data Elements (Examples)

- Clinical data
 - Processes and outcomes of care
 - Medications, problem lists, laboratory findings
 - Care transitions
- Patient data
 - Basic demographics (e.g., age)
 - Patient risk adjustors (e.g., insurance coverage, compliance)
 - Patient self-care (experience of care)
- Physician data
 - Basic demographics (e.g., gender, specialization)
 - Diagnostic reasoning (e.g., errors)
 - Unrestricted medical license (e.g., disciplinary actions)
- System data
 - Type of data collection (e.g., EHR)
 - Type of setting (e.g., ambulatory – small practice)

ABIM's Defines the Field of Internal Medicine

- Advanced Heart Failure
- Cardiovascular Disease
- Endocrinology, Diabetes and Metabolism
- Gastroenterology
- Hematology
- Infectious Disease
- Medical Oncology
- Nephrology
- Pulmonary Disease
- Rheumatology
- Adolescent Medicine
- Clinical Cardiac Electrophysiology
- Critical Care Medicine
- Geriatric Medicine
- Hospice & Palliative Care
- Interventional Cardiology
- Sleep Medicine
- Sports Medicine
- Transplant Hepatology

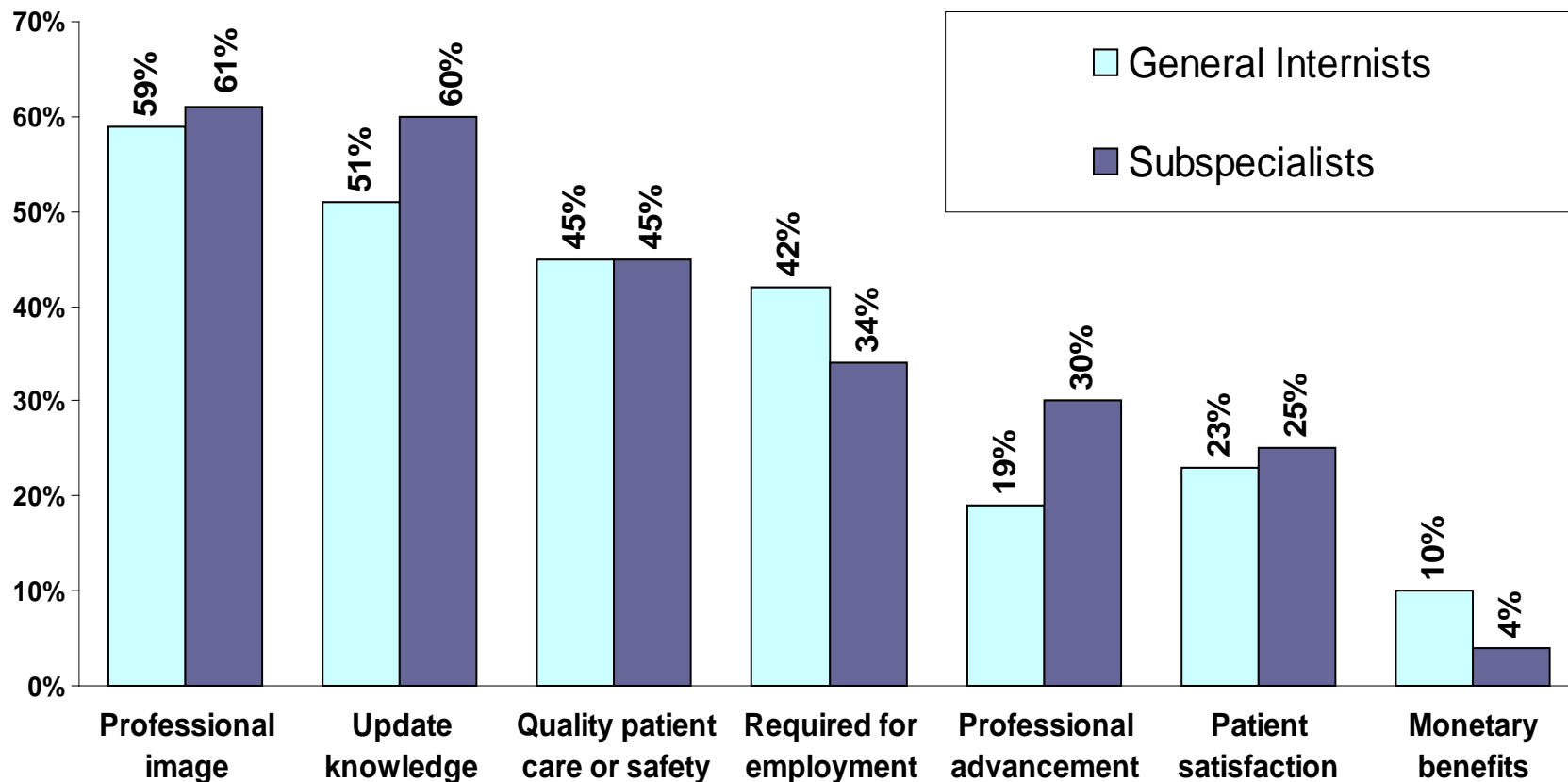
How many physicians work in your practice?



Number of physicians in practice

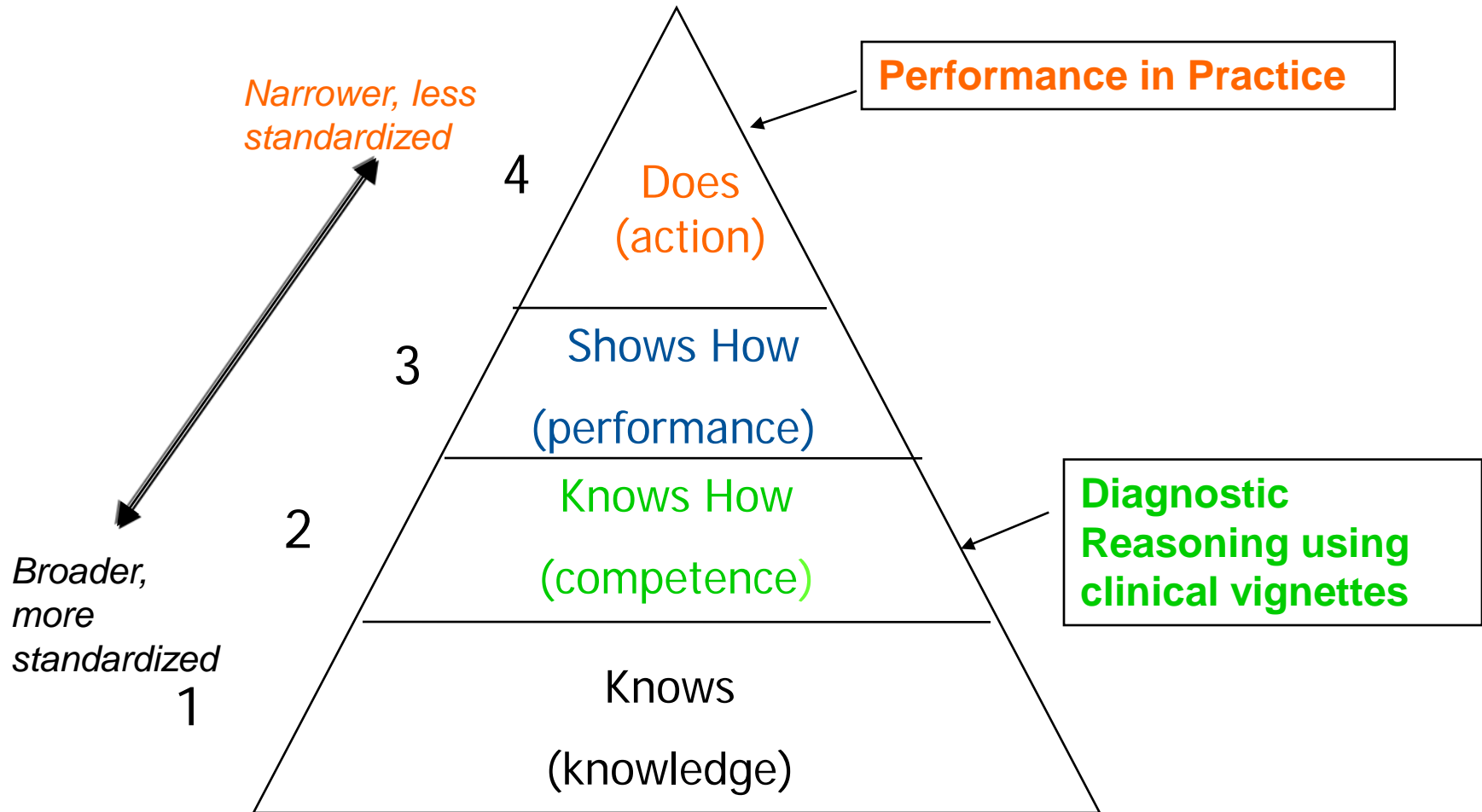
Results through 2009

What Motivates Physicians to Participate in a Voluntary Regulatory Program?

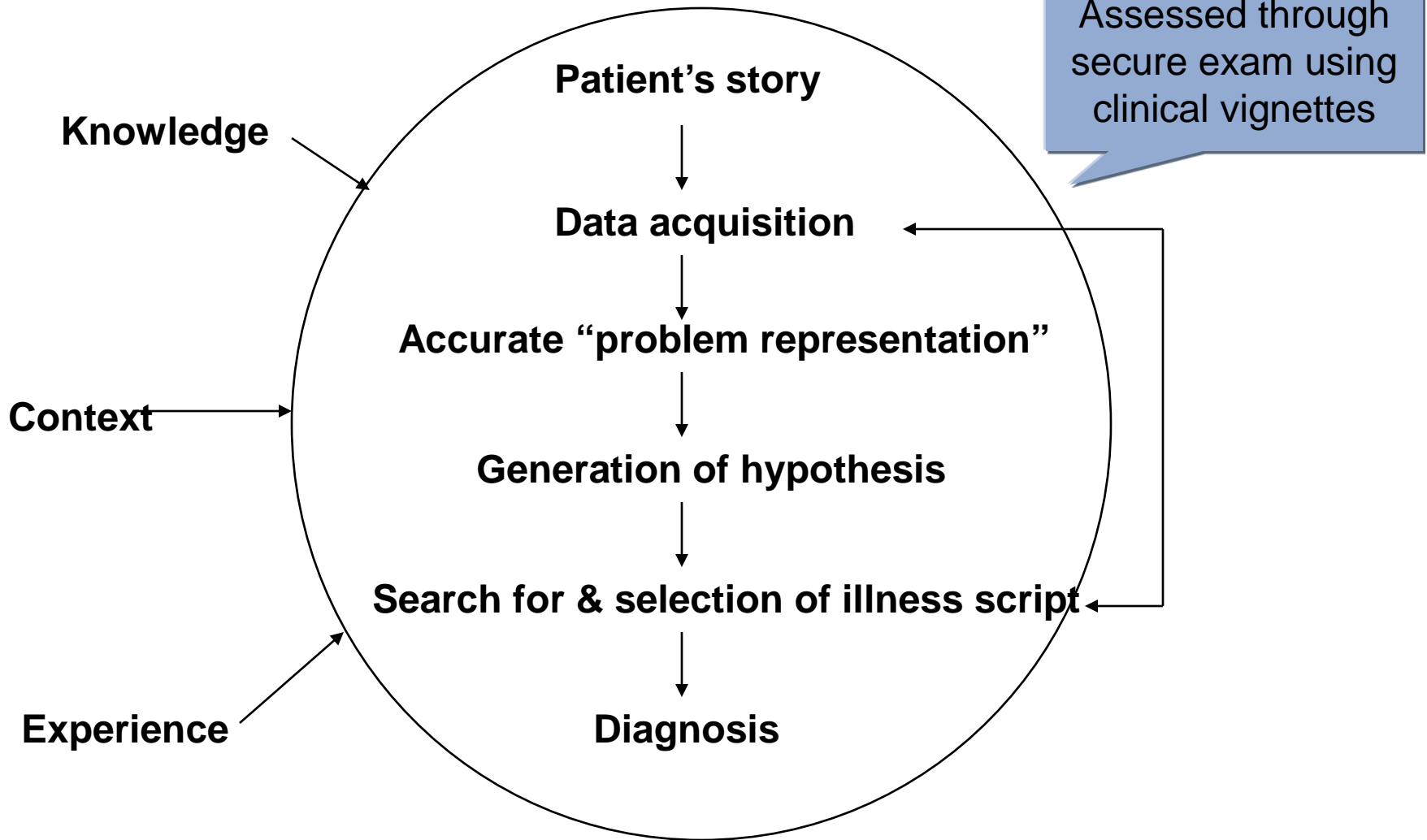


Lipner, Ann Int Med, 2006

Miller's Framework for Clinical Assessment (1990)

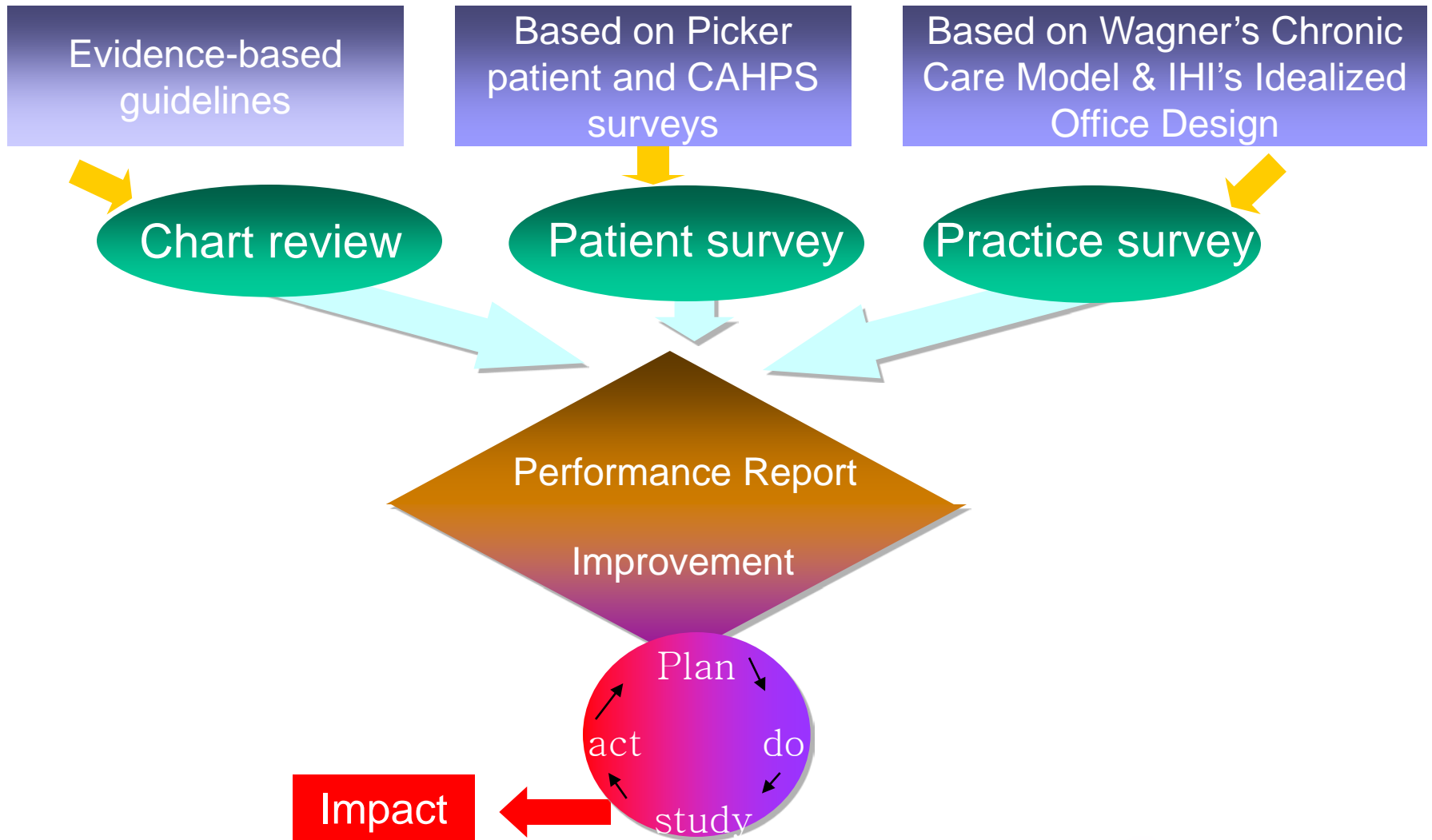


Clinical Diagnostic Reasoning Process*



* Bowen JL, NEJM 2006;355:2217-25.

Performance in Practice (PIMSM)



Diabetes Composite Score – at Physician Level

- Composite measure is more reliable than individual measures
- Classification/decision based on a composite measure is more reproducible than a decision based on an individual measure (fewer false positive and negatives)
- Composites allow for a more comprehensive assessment but performance feedback on individual measures is also important
- Classification/decision that is based on a scientific approach to standard setting is credible and defensible

Process for Developing a Composite Score

- Started with the raw data from a sample of 957 physicians completing *Diabetes PIM*
 - 81% general internists, 13% endocrinologists
 - 20,131 patient charts (21.0 patients per physician)
 - 18,974 patient surveys (19.8 patients per physician)
- Review actual performance on individual measures
- Review reliability of individual measures
- Select clinical and patient experience measures
- Apply modified Angoff standard setting method
 - Convene an expert panel
 - Define a “Borderline Candidate”
 - Develop performance thresholds for individual measures
 - Weight importance of individual measures (Dunn-Rankin method)
- Review reliability of composite & classification accuracy
- Review actual performance on composite

Computation of Pass/Fail Standard for Competent Diabetes Care

Measure	Threshold		Importance Weights		Points
Intermediate Outcome Measures					
HgBA1c ~poor control (≤ 9.0)	72.5%	X	10	=	7.25
Blood pressure ~poor control (<140/90)	53.7%	X	10	=	5.37
LDL ~poor control (<130mg/dl)	58.7%	X	10	=	5.87
HgBA1c at goal (<8.0 or <7.0)	36.0%	X	7	=	2.52
Blood pressure superior control (<130/80)	16.9%	X	9	=	1.52
LDL superior control (<100 mg/dl)	23.8%	X	8	=	1.90
Clinical Process Measures*					
Eye exam	28.8%	X	9	=	2.59
Test for urine protein	73.1%	X	10	=	7.31
Foot exam	35.6%	X	4	=	1.42
Smoking status & cessation advice	67.5%	X	7	=	4.73
Patient Experience Measures					
Patient satisfaction with diabetes care	46.3%	X	7	=	3.24
Patient self-care support	53.1%	X	9	=	4.78
Standard (passing score)			SUM	=	48.51

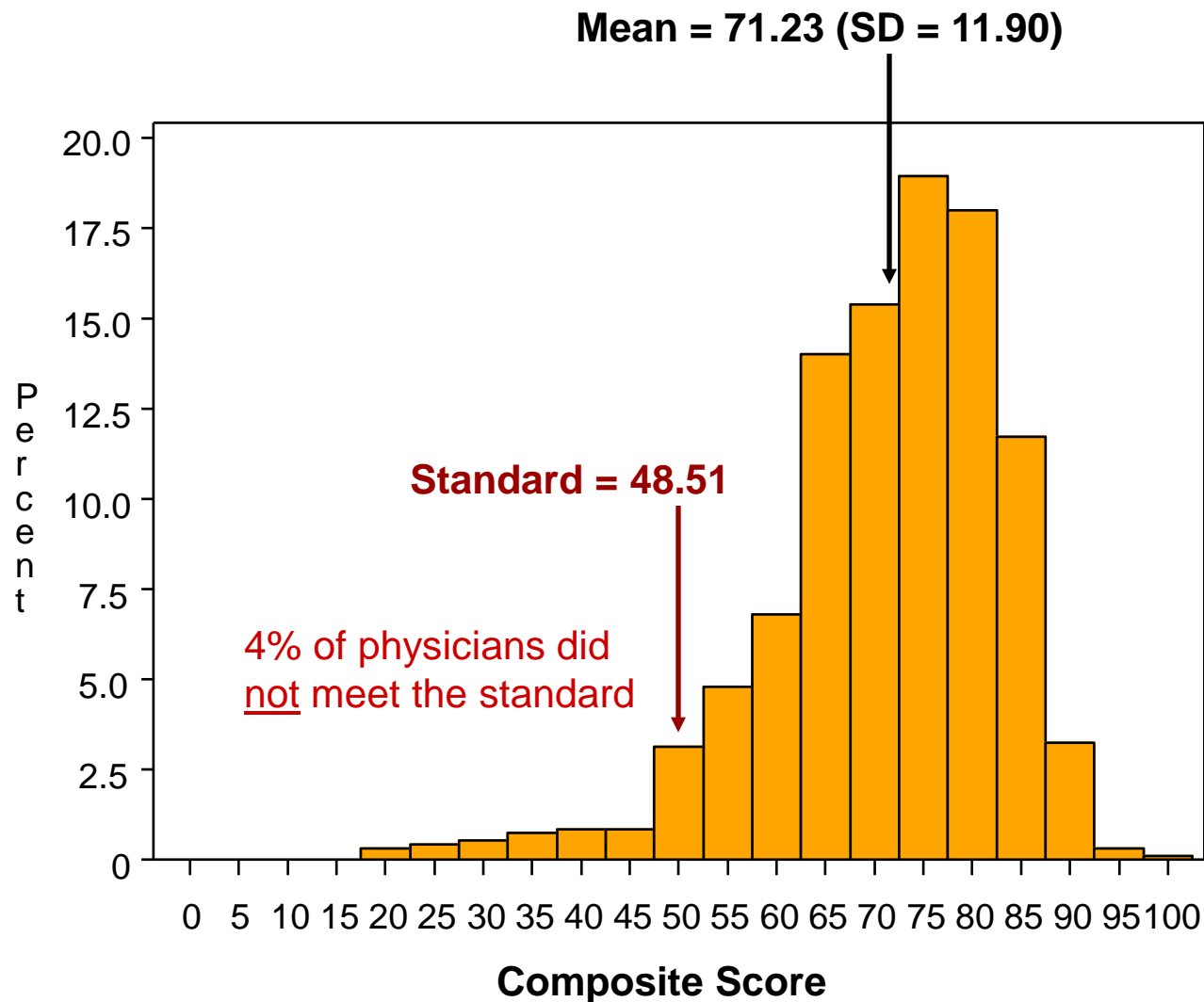
* If physician is below threshold, no points are awarded

Feedback: Dr. Smith's Performance Score

Measure	Performance Rate		Importance Weights		Points
Intermediate Outcome Measures					
HgBA1c ~poor control (<= 9.0)	82.6%	X	10	=	8.26
Blood pressure ~poor control (<140/90)	69.6%	X	10	=	6.96
LDL ~poor control (<130mg/dl)	87.0%	X	10	=	8.70
HgBA1c at goal (<8.0 or <7.0)	65.2%	X	7	=	4.56
Blood pressure superior control (<130/80)	34.8%	X	9	=	3.13
LDL superior control (<100 mg/dl)	73.9%	X	8	=	5.91
Clinical Process Measures					
Eye exam*	26.1%	X	9	=	0.00
Test for urine protein	100.0%	X	10	=	10.00
Foot exam	60.9%	X	4	=	2.44
Smoking status & cessation advice	78.3%	X	7	=	5.48
Patient Experience Measures					
Patient satisfaction with diabetes care	60.0%	X	7	=	4.20
Patient self-care support	80.0%	x	9	=	7.20
Dr. Smith's Composite Score			SUM	=	66.84

*Below the threshold (28.8%) so no points (0.000) are awarded.

Feedback: Distribution of Diabetes Composite



Reliability of Composite Score = 0.91*

Classification Accuracy = .98*

*Bootstrap method (Weng et al.)

N=957; updated for 2,823 physicians

Feedback: “Your Performance Quartiles”

	Low		High	
	Bottom (bottom 25th percentile)	Bottom Middle (26th- 49th percentile)	Top Middle (50th-74th percentile)	Top (top 25th percentile)
A1C poor control				X
Blood pressure poor control				X
LDL poor control			X	
A1C at Goal				X
Blood Pressure Superior Control				X
LDL Superior Control			X	
Eye exam			X	
Test for urine protein				X
Foot exam				X
Smoking Status & Cessation Advice				X
Overall Diabetes Care Satisfaction			X	
Patient Self-care Support				X
Total composite score				X

Composite and Standard Setting

- ✓ Reliability and classification accuracy – Good!
- ✓ Standard setting approach – Credible!
- ✓ Competence standard – Reasonable! but high pass rate
- ✓ Composite score interpretation – Valid (meaningful)!
 - Endocrinologists performed better than internists
 - Those classified as “incompetent”
 - Scored lower on diagnostic reasoning exam
 - Had lower overall ratings in residency
 - Were more likely to be in solo practice

Comprehensive Care – Meaningful Use of Health IT

- 7 Chronic conditions:
 - Coronary artery disease
 - Acute myocardial infarction
 - Congestive heart failure
 - Atrial Fibrillation
 - Diabetes
 - Hypertension
 - Osteoarthritis (knee and/or hip)
- 4 Acute care conditions:
 - Acute depression
 - Low back pain
 - Upper respiratory infection
 - Urinary tract infection
- 6 Preventive care measures:
 - Influenza and pneumococcal vaccinations
 - Mammography and colorectal cancer screening
 - Osteoporosis screening
 - Smoking cessation counseling

Comprehensive Care Study Summary

- Complexity increases across multiple conditions and for a particular time frame
- Measurement of chronic disease care & preventive services using composites is feasible and reliable
- Acute care conditions were not well documented and were not measured well
- Performance in practice was correlated with diagnostic reasoning skills

High Quality Clinical Assessments Should Include...

- High quality data and access to raw data!!
- Evidence-based measures
- Multiple sources of data
- Connectivity to electronic databases
- Data safeguards to ensure privacy of patients
- Enhancements through research
 - Relationships among these data should be examined through qualitative and quantitative research
- Feedback to encourage quality improvement – move the curve to the right-> better patient care

Related References

Diagnostic Reasoning

Lipner RS, Lucey CR. Putting the Secure Exam to the Test. *JAMA*, 2010; 304(12):1379-1380.

Composite Scoring and Standard Setting

- Weng W, Hess BJ, Lynn LA, Holmboe ES, Lipner RS. Measuring Physicians' Performance in Clinical Practice: Reliability, Classification Accuracy, and Validity. *Evaluation in the Health Professions*. 2010; 33(3): 302-20.
- Hess BJ, Weng W, Lynn LA, Holmboe ES, Lipner RS. Setting a Fair Performance Standard for Physicians' Quality of Patient Care. *Journal of General Internal Medicine*. In press, 2010.

Comprehensive Care and Systems

- Holmboe ES, Weng W, Arnold GK, Kaplan SH, Normand SLK, Greenfield S, Lipner RS. Measuring Physician Performance Comprehensively in Ambulatory Practice. *Health Services Research*. In press; 2010.
- Holmboe ES, Arnold GK, Weng W, Lipner RS. Current Yardsticks May be Inadequate for Measuring Quality Improvements From the Medical Home. *Health Affairs*. 2010; 29 (5): 859-66.