

Ensuring Low Risk of Misclassification in Reporting Physician and Group-Level Results

Dana Gelb Safran, ScD
The Health Institute
Institute for Clinical Research and Health Policy Studies
Tufts-New England Medical Center

Presented at:
CAHPS® User Group Meeting

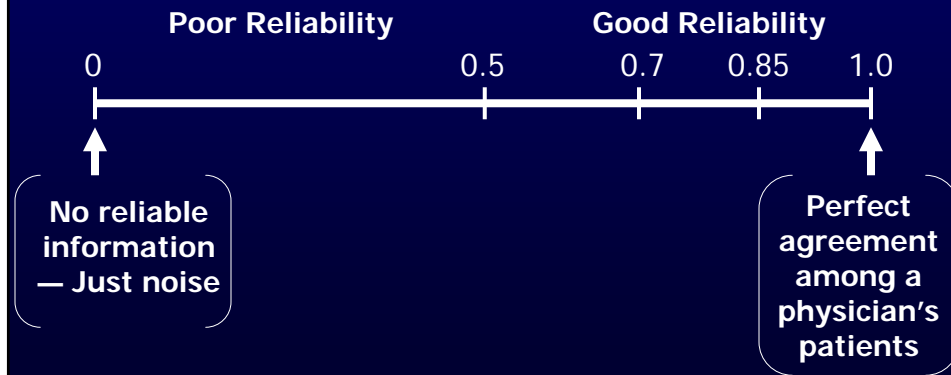
31 March 2006



Measuring Patients Experiences with Individual Physicians and their Practices

- ◆ Careful attention is given to determining the sample sizes required to ensure high measurement reliability at the physician-level
- ◆ A widely accepted standard for reliability has been ≥ 0.70
- ◆ What does this level of reliability imply about the “risk of misclassifying” an individual physician (or practice)?

Physician-Level Reliability: A Measure of Concordance Among Patients



3

Sample Size Requirements for Varying Physician-Level Reliability Thresholds

	Number of Responses per Physician Needed to Achieve Desired MD-Level Measurement Reliability		
	Reliability: 0.7	Reliability: 0.8	Reliability: 0.95
ORGANIZATIONAL ACCESS	16	26	123
Schedule urgent care	29	49	231
Schedule routine care	26	44	209
Call back – regular hrs.	23	39	181
Call back –after hrs.	22	37	175
< 15 minutes wait	11	18	82
COMMUNICATION	32	55	258
Explains clearly	41	70	330
Listens carefully	36	62	291
Clear instructions	42	72	342
Shows respect	37	63	300
Enough time	33	56	263
Knows medical hx.	28	48	224

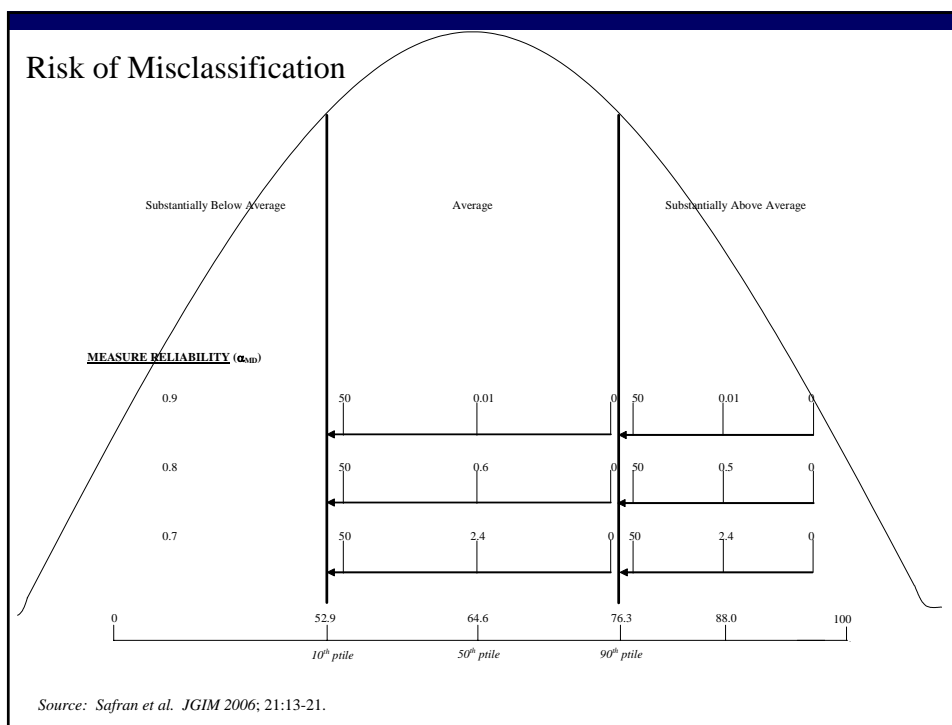
4

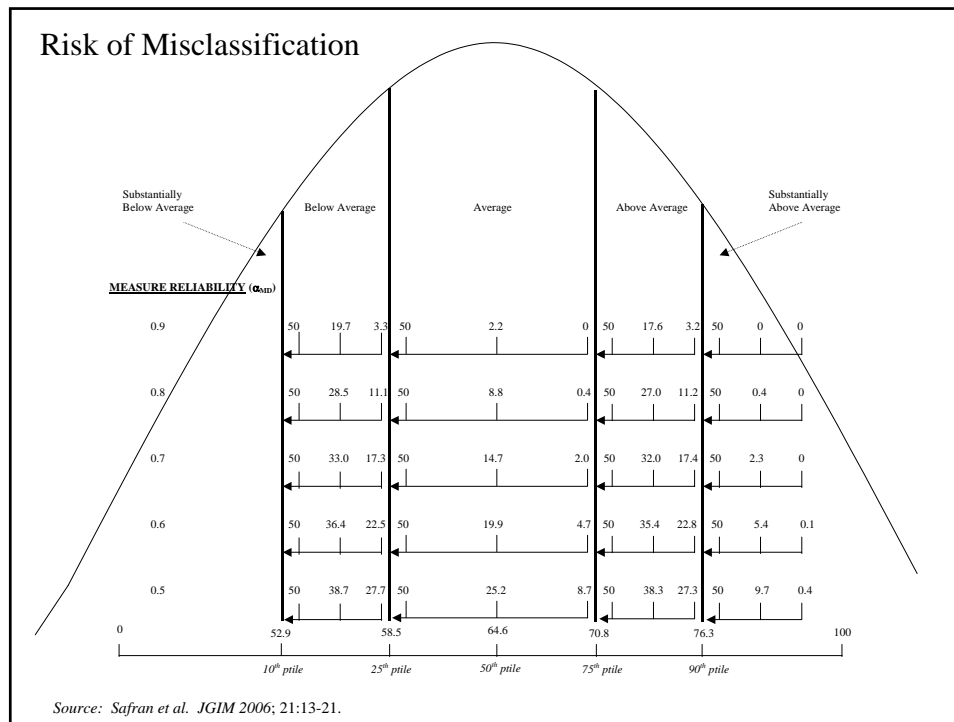
What is the Risk of Misclassification?

- ◆ Not simply $1 - \alpha_{MD}$
- ◆ Depends on:
 - ◆ Measurement reliability (α_{MD})
 - ◆ Proximity of score to the cutpoint
 - ◆ Number of cutpoints in the reporting framework

Source: Safran et al. JGIM 2006; 21:13-21.

5





MHQP 2005 Statewide Survey

- ◆ Physician-level survey format
- ◆ Site-level sampling to support site-level reporting
- ◆ Estimated samples required to achieve ≥ 0.70 site-level reliability

Number of MDs per site	Target number of completes per site	Starting sample
3	90	257
4-9	100	286
10-13	125	357
14-19	150	429
20-28	175	500
29+	200	571

Site-Level Reliability by Practice Size

◆ Sample required for site-level reliability ≥ 0.70 depended on practice size and varied by measure

◆ Example: Communication

Practice Size (Number of Doctors)	Site-Level Reliability			
	A	B	C	D
	≥ 0.7	0.5 - 0.69	0.34 - 0.49	< 0.34
3	53	22-52	11-21	10
4	67	28-66	14-27	13
5	79	34-78	17-33	16
6	91	39-90	20-38	19
7	101	43-100	22-42	21
8	110	47-109	24-46	23

9

Summary Chart: Internal Reports

Summary Measures	Score	Mean Score	Reliability
Quality of Doctor-Patient Interaction:	0 10 20 30 40 50 60 70 80 90 100		
Communication		95.5	A
Integration of Care		77.0	B
Knowledge of Patient		88.1	A
Health Promotion		83.0	C
Organizational/Structural Features of Care:	0 10 20 30 40 50 60 70 80 90 100		
Organizational Access		87.8	A
Visit-Based Continuity		92.7	A
Clinical Team		86.8	C
Office Staff		92.1	A
Global Rating:	0 10 20 30 40 50 60 70 80 90 100		
Willingness to		92.2	B

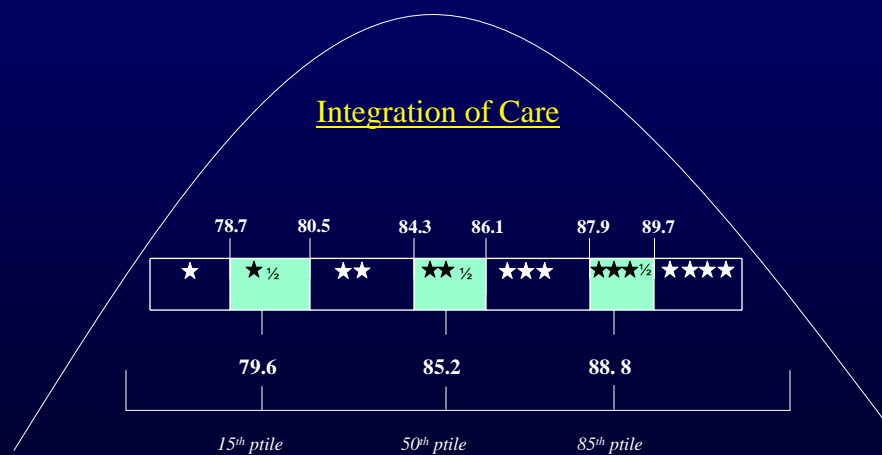
10

Percent of Sites with A-Level Reliability by Measure and Survey-Type

	Adult PCP %	Pediatric %
MD – Patient Interactions		
Communication	98	97
Knowledge of patient	91	86
Health Promotion	46	97
Integration of care	79	61
Organizational/Structural Features of Care		
Access	99	100
Visit-based continuity	100	100
Office Staff	95	99
Clinical Team	37	86
Willingness To Recommend	62	59

11

Framework for Public Reporting



12

Summary Performance Chart: Internal Reports

Summary Measures	Summary Performance
Quality of Doctor-Patient Interaction:	
Communication	★★★★
Integration of Care	★
Knowledge of Patient	★★½
Health Promotion	★★½
Organizational/Structural Features of Care:	
Organizational Access	★★½
Visit-Based Continuity	★★★★
Clinical Team	★★½
Office Staff	★★★★
Global Rating:	
Willingness to Recommend	★★★★

13

Summary and Implications

- ◆ With sufficient sample sizes, data obtained using CAHPS Clinician & Group approach yields data with MD- and site-level reliability ≥ 0.70
- ◆ For site-level reliability, number of MDs per site influences requires sample sizes
- ◆ Risk of misclassification can be held to <5% with by
 - ❖ Limiting number of performance categories
 - ❖ Creating buffer (“zone of uncertainty”) around performance cutpoints
- ◆ Trade-offs are likely around data quality standards (e.g., acceptable “risk”) vs. data completeness

14

For More Information:

Dana Gelb Safran, ScD
The Health Institute
Tufts-New England Medical Center

Department of Medicine
Tufts University School of Medicine

dsafran@tufts-nemc.org

15