

# TAVR Performance Requirements: How many, how well, or both?

David M. Shahian, MD

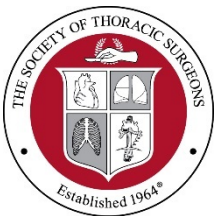
Professor of Surgery, Harvard Medical School

Chair, STS Council on Quality, Research, and Patient Safety

July 25, 2018

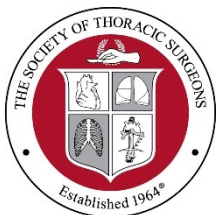
# Disclosures

- No financial disclosures or COI
- STS: No disclosures



# Performance measurement

- STS has been a leader in advocating direct measures of quality and public reporting
- STS has 35 NQF-endorsed quality metrics, most of which are risk-adjusted outcomes
- We prefer outcomes over structure or process measures, although we have used all three



## Search AVR Data by Hospital

Hospital

Filter by name

State

- Any -

Apply

Name

Overall  
Composite  
Score\*

Absence of  
Operative Mortality

Absence of Major  
Morbidity

[St. Joseph Medical Center](#)

Tacoma, Washington

[El Camino Hospital](#)

Mountain View, California

[Kaiser Sunnyside Medical Center](#)

Clackamas, Oregon

[MedStar Heart Institute MedStar Washington Hospital Center](#)

Washington, District of Columbia

### CABG Results

Year	Overall Composite Score*	Absence of Operative Mortality	Absence of Major Morbidity	Use of Internal Mammary Artery	Receipt of Required Perioperative Medications
July 2015 - June 2016	★★★★ 97.7	★★★★ 97.7	★★★★ 91.1	★★★★ 99.5	★★★★★ 99.5
July 2016 - June 2017	★★★★ 97.7	★★★★ 97.6	★★★★ 92.2	★★★★ 99.5	★★★★★ 99.2

### AVR Results

Year	Overall Composite Score**	Absence of Operative Mortality	Absence of Major Morbidity
July 2014 - June 2017	★★★★ 97.0	★★★★ 98.1	★★★★ 92.1

### AVR + CABG Results

Year	Overall Composite Score***	Absence of Operative Mortality	Absence of Major Morbidity
July 2014 - June 2017	★★★★ 93.2	★★★★ 95.7	★★★★ 85.0

**Table 3a : TAVR Program Performance Requirements:  
Minimum quality benchmarks for TAVR sites**

**2018 Criteria**

<b>Primary Outcome Metrics</b>	<b>Performance Measure</b>
<b>In-hospital risk-adjusted all-cause mortality</b>	<b>Based on 95% CI and national benchmark data, program's performance "as expected" or "better than expected"</b>
<b>30-day risk-adjusted all-cause mortality</b>	<b>Based on 95% CI and national benchmark data, program's performance "as expected" or "better than expected"</b>
<b>30-day all-cause neurologic events including TIAs</b>	<b>Funnel plots: performance within 95% upper control limits (outlier); programs exceeding 90% upper control limits (warning) merit further internal study *</b>
<b>30-day major vascular complication</b>	<b>Funnel plots: performance within 95% upper control limits (outlier); programs exceeding 90% upper control limits (warning) merit further internal study *</b>
<b>30-day major bleeding</b>	<b>Funnel plots: performance within 95% upper control limits (outlier); programs exceeding 90% upper control limits (warning) merit further internal study *</b>
<b>30-day moderate or severe AR</b>	<b>Funnel plots: performance within 95% upper control limits (outlier); programs exceeding 90% upper control limits (warning) merit further internal study *</b>

**Primary Outcome Metrics In Development**

**1-year risk-adjusted all-cause mortality**

**\*Risk-adjusted measures for all major complications (transition to statistical hypothesis testing)**

**Patient reported health status (KCCQ) at 30 days and 1 year versus baseline**

**30-day and 1-year risk-adjusted mortality and morbidity composite measure**

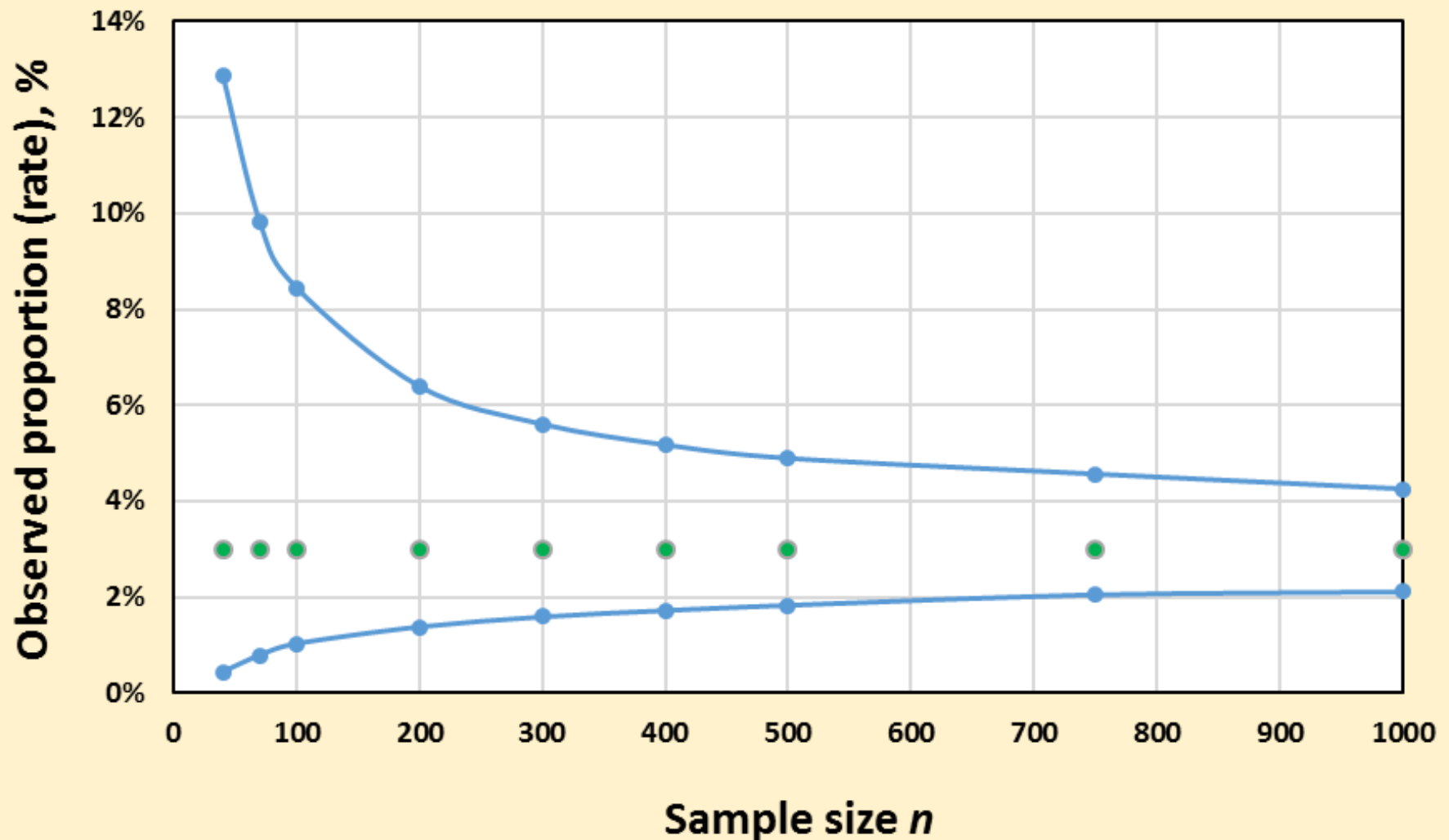
# Why volume thresholds for TAVR?

- Expertise: the volume-outcome association
  - Volume during procedural adoption--learning curve
  - Volume of established procedures--generally correlates with outcomes for complex procedures
- Measurement challenges with low volumes
  - Randomness—inherent uncertainty of small sample estimates
  - Measure reliability
  - Statistical power to detect outliers



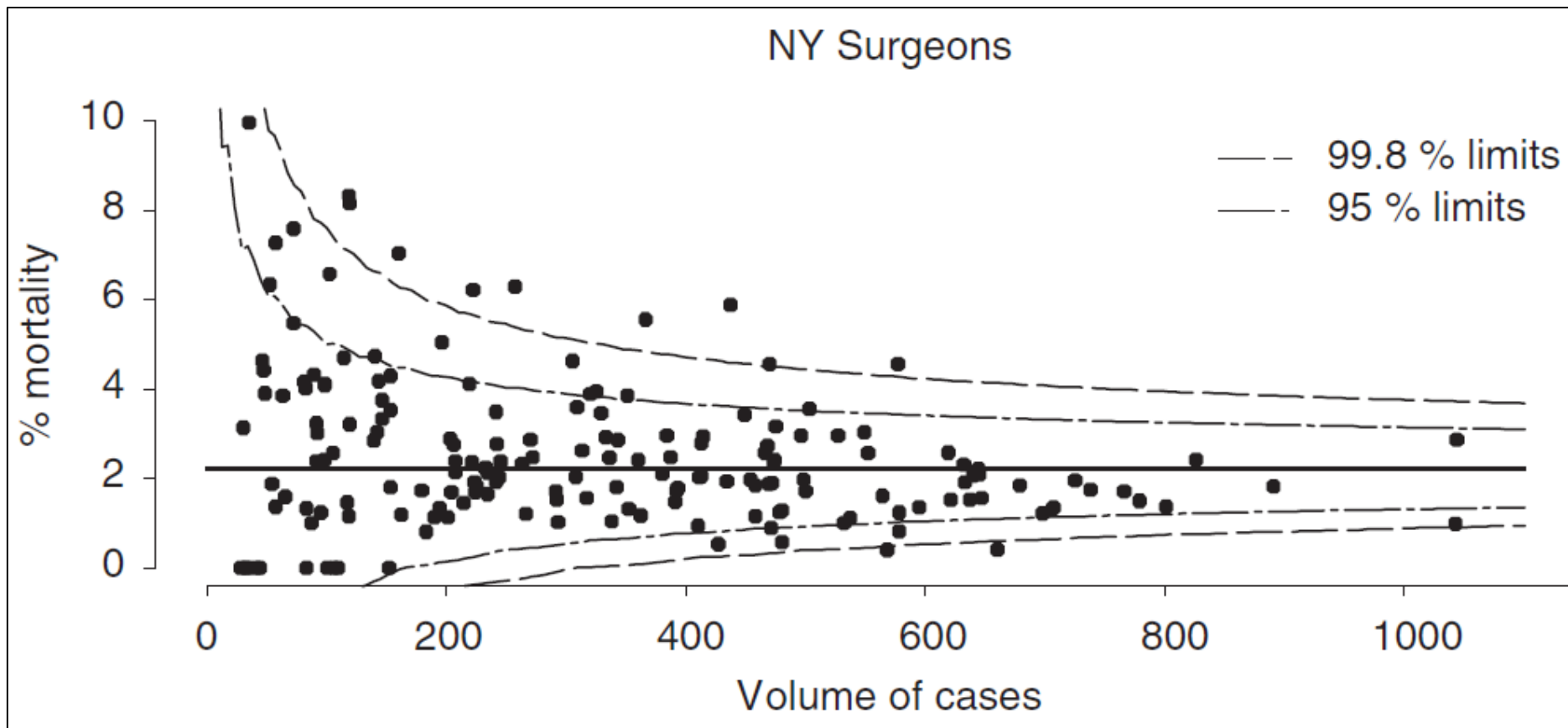
## 95% Confidence Intervals

Given a sample of size  $n$  and its estimate (e.g., 3%), how certain can we be about the true population proportion?



# Prediction Intervals (used for funnel plots)

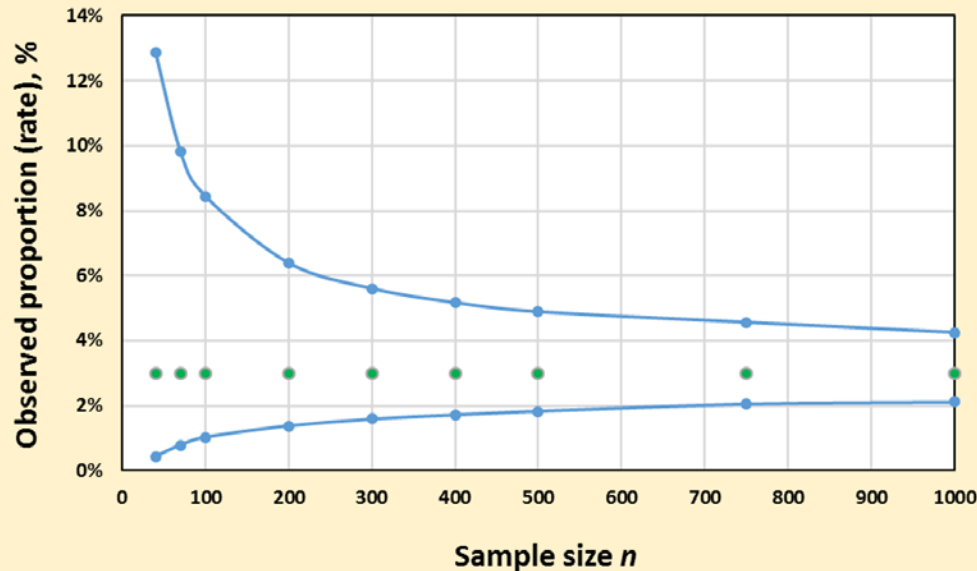
Given known population parameters (mean, distribution), what can we say about estimates from future samples of size  $n$ ?





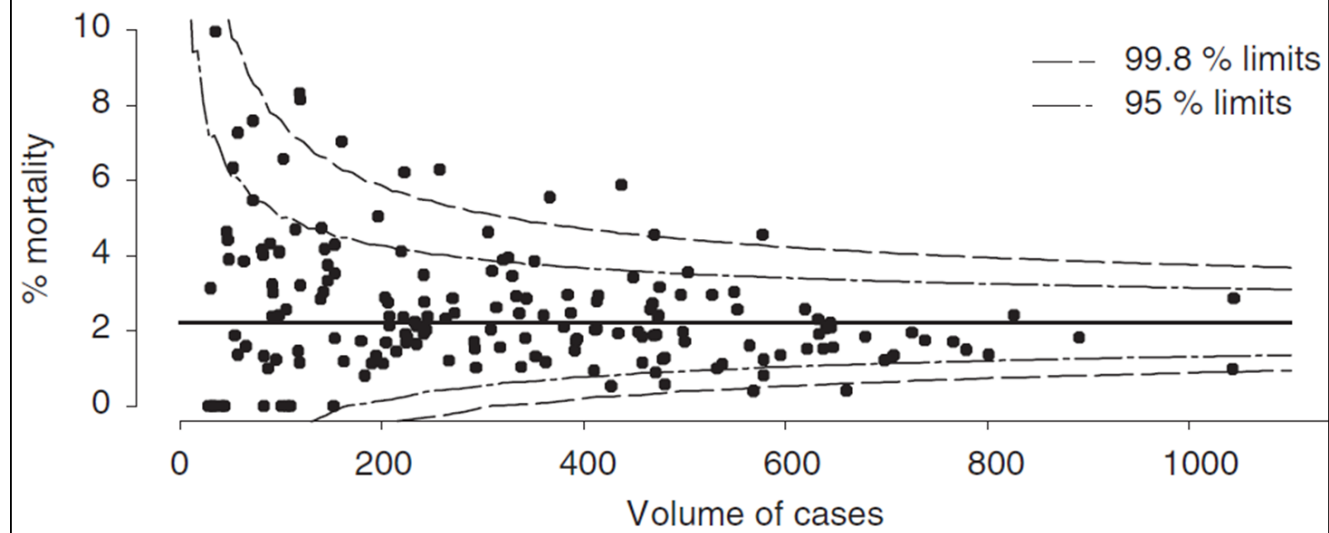
### 95% Confidence Intervals

Given a sample of size  $n$  and its estimate (e.g., 3%), how certain can we be about the true population proportion?

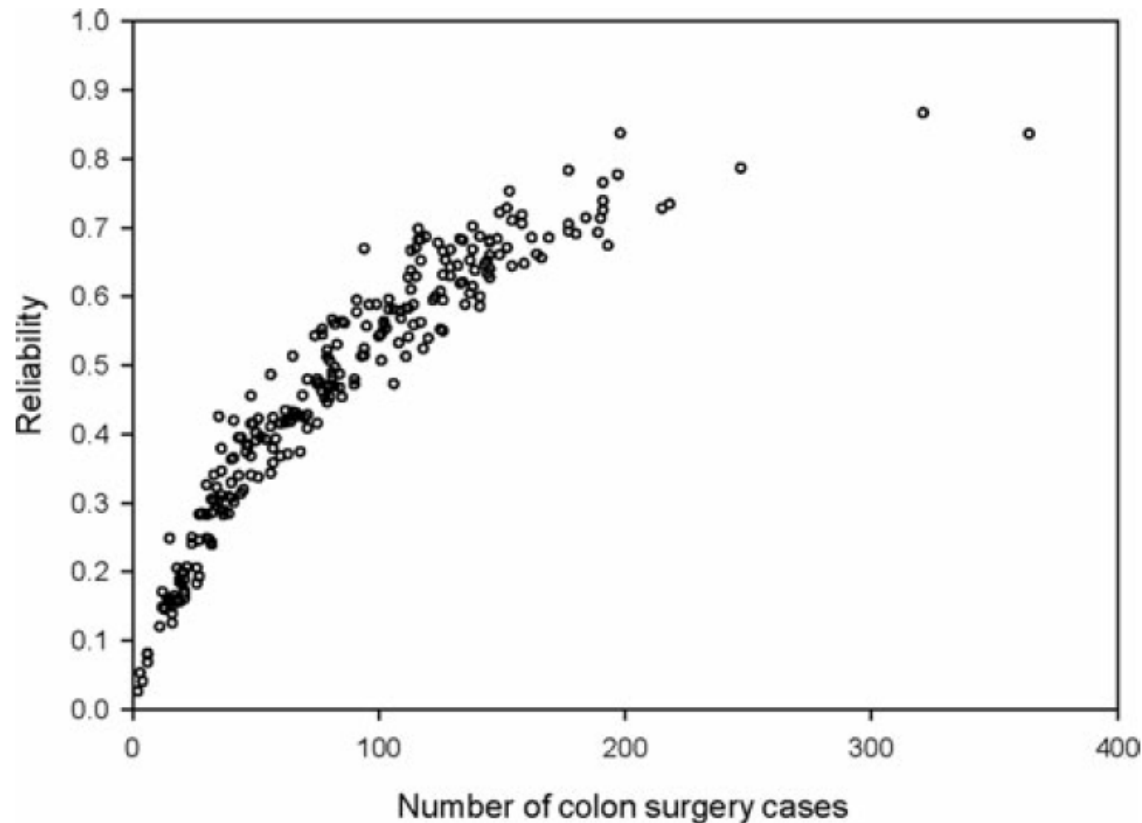


Two statistical tools, same message  
Estimates from small samples have substantial random variation

### Funnel Plot NY Surgeons (Prediction intervals)



# Measure reliability (signal to noise ratio) highly dependent on sample size



**FIGURE 1.** Hospital-level reliability estimates by colectomy annual caseloads based on American College of Surgeons National Surgical Quality Improvement Program-sampled cases in 2010 for the mortality or any serious morbidity composite outcome measure. (Event rate 20%)

# Statistical Power Decreases with Smaller Sample Size

## Type II errors more likely

	National postoperative mortality (%)	Median annual number*	Number of procedures necessary to detect poor performance		
			60% power	70% power	80% power
Hip fracture surgery	8.4%†	31	56	75	102
Oesophagectomy or gastrectomy	6.1%‡	11	79	109	148
Bowel cancer resection	5.1%§	9	95	132	179
Cardiac surgery	2.7%¶	128	192	256	352

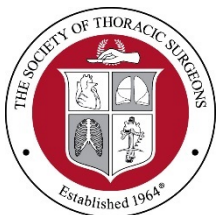
5% significance level. Poor performance defined as double the national overall mortality rate. \*On the basis of hospital episode statistics<sup>5</sup> for the 3-year period from April, 2009, to March, 2012 (except for cardiac surgery, for which reported numbers<sup>2</sup> are used). †30-day mortality (March 1, 2010–Feb 28, 2011).<sup>6</sup> ‡90-day mortality (Oct 1, 2007–June 30, 2009).<sup>7</sup> §90-day mortality (Aug 1, 2010–July 31, 2011).<sup>8</sup> ¶In-hospital mortality (April 1, 2008–March 31, 2011).<sup>9</sup>

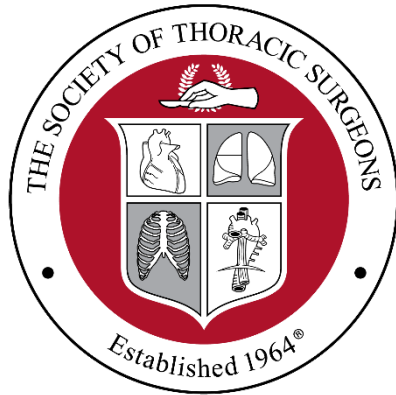
**Table 1: Mortality after four surgical procedures, the number of procedures that occur annually, and how many would be necessary to detect poor performance with different statistical powers**

# TAVR Quality:

## How many, how well, or both?

- Outcome measure statistics are problematic with small sample sizes (low volume)
- Adequate volume is associated with expertise and facilitates accurate, reliable outcome measurement





AMERICAN  
COLLEGE *of*  
CARDIOLOGY

