



Physician Clinical Performance Measurement

AEI Conference on
Delivering Better Value in Health Care

November 5, 2007

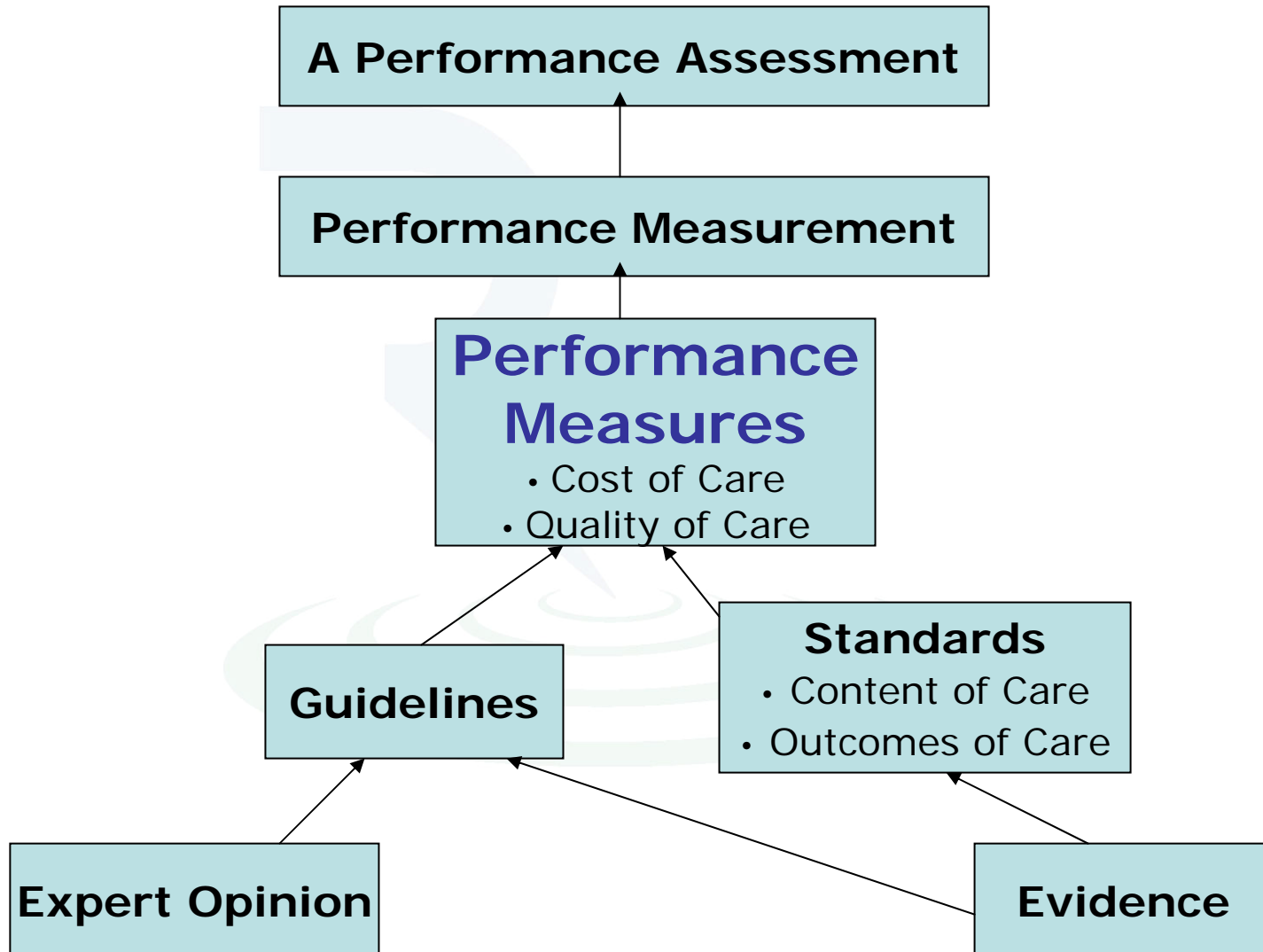
Earl Steinberg, MD, MPP
President and CEO, Resolution Health Inc.
Adjunct Professor of Medicine and of
Health Policy & Management,
Johns Hopkins University

- 1. Types of available data**
- 2. Quality of available data**
- 3. Measures**
- 4. Physician identifiers**
- 5. Physician attribution**
- 6. Measurement**
- 7. Statistical Issues/Scoring**
- 8. Reporting**
- 9. How results are used**
- 10. Privacy**

Focus Today

- What do we mean by clinical performance measurement (“physician quality profiling”)?
- Data
- Measures
- Statistical Issues/Scoring

The “Geology” of Clinical Performance Measurement



Potential Data Sources

- Claims
- Hospital Billing Data
- Lab results (in electronic form)
- Independent Chart Review
- Electronic Medical Record
- Self-reported by MD
- Patient Surveys
- Combinations of above

Claims Data vs Chart Abstraction

Claims Data	Chart Abstraction
Missing claims (e.g., OON utilization)	Info. re care provided by other MDs often missing
Incomplete coding	Incomplete documentation
Errors in coding (both inadvertent and deliberate)	NA
Variations in coding	Variation in terminology, abbreviations, symbols and criteria
Limited breadth and depth of codes	NA
NA	Problems with legibility
Low cost	High cost

RHI Measure Set

- 102 measures in the current set were culled from approx. 500 candidate measures; together, they “address” 19 specialties
- A candidate measure was excluded if:
 - a) it could not be applied to claims data with reasonable accuracy; or
 - b) it required line item detail from hospital claims
- Of the candidate measures, the number that can be used depends on:
 - tolerance for false positives
 - whether restrict to “evidence-based” measures
- Probably only have sufficient measures to adequately profile care provided by primary care physicians and several types of specialists

Measure Sources

- AQA (15 of 26 approved ambulatory care measures are included in our measure set)
- NCQA
- AHRQ
- Evidence-based clinical practice guidelines from major organizations (e.g. AHA/ACC, ADA)
- RAND quality indicators
- Drug safety literature (Zahn's Criteria and Updated Beer's Criteria)
- Medical literature

Practice Guidelines Don't = Quality Measures

- Relationship between a practice guideline and a valid algorithmic representation of the guideline (i.e. a valid measure based on the guideline) is “1 to many”, not “1 to 1”
- Potential measures vary in terms of false negative vs false positive rate
- We've deliberately constructed our measures to try to minimize false positives
- The devil's in the details!

Examples of Ways to Deal with Shortcomings of Claims Data

Shortcoming	How to Address the Shortcoming
Certain tests and services are (may) not be reflected reliably in claims data	Do not use measures that focus on use of those tests and services (e.g. spirometry; influenza vaccine; smoking cessation programs)
Use of OTC drugs is not captured in claims data	Do not use measures that focus on use of meds that could be purchased OTC (e.g. Use of aspirin post-MI)
Inaccuracies in diagnoses that are listed on claims	Require more than one indication that a member has a disease/condition; don't use diagnosis codes submitted by labs

Scoring/Statistical Issues

- Types of scores
 - Single (aggregated) score per MD (or MD group)
 - A score for certain types of care (e.g. management of diabetes; cardiac care)
 - Measure-specific scores
- Weighting
 - unweighted sum of numerators and denominators across all pt-measures applicable to a given MD (implies all measures are equally important)
 - weighted sum (implies some measures are more important than others)
- Statistical issues
 - Minimum number of observations per MD for a given measure for that measure to be considered?
 - Minimum number of observations per MD for that MD to receive a score?
 - Confidence intervals around point estimates
 - When is a score for a group practice a reasonable surrogate for an individual MD score?
- Distinguishing between MDs
 - Numerical score vs %-ile vs quartile (or other group)
 - How big a difference in score is clinically meaningful?