Measuring Performance in Physician Organizations in California P4P 2005: Distinctions without Differences?

Neil Maizlish, PhD, MPH, Director Quality Improvement
Stanley Padilla, MD, Chief Medical Officer, VP Med. Services
Fiona Wilson, MD, VP for Quality Initiatives

Brown & Toland Medical Group

Presentation at Second National Pay For Performance Summit
Beverly Hills, CA - February 16, 2007

Definitions of Quality in Health Care

- The current “paradigm” defines quality as:

  The degree to which services directed at individuals and populations increase the probability of achieving desired health outcomes and are consistent with current professional knowledge (Institute of Medicine, 1990)

- Evidence-based medicine and retrospective application of clinical guidelines imply an absolute benchmark (100%), qualified by explicit inclusion and exclusion criteria based on patient demographics, contraindications, medical conditions

Dimensions of Quality Measurement

- Measurement System
  - ✓ Validity
    - Conceptual foundation: Are we truly measuring what we purport to measure? (Evidence-based guidelines supported by randomized controlled trials)
    - Execution: Is the measurement free of systematic bias?
  - ✓ Precision
    - Random sampling error (margin of error) is small <u>relative to a difference worth measuring</u>
      - ✓Clinical measures (year to year variation)
      - ✓PAS (statistical sample within year)

Dimensions of Quality Measurement

- Benchmarking
  - ✓ Absolute
    - 90%-95% may be practical maximum in administrative data
      - ✓ Denominator: difficult to eliminate all medical exclusions, refusals, etc.
      - ✓ Numerator: difficult to get historical outcomes when look-back period exceeds membership in PO or health plan (e.g. Pap, 3 years, colon cancer, 10 years)
  - ✓ Relative
    - Self-comparison
      - ✓ Change in baseline between measurement years

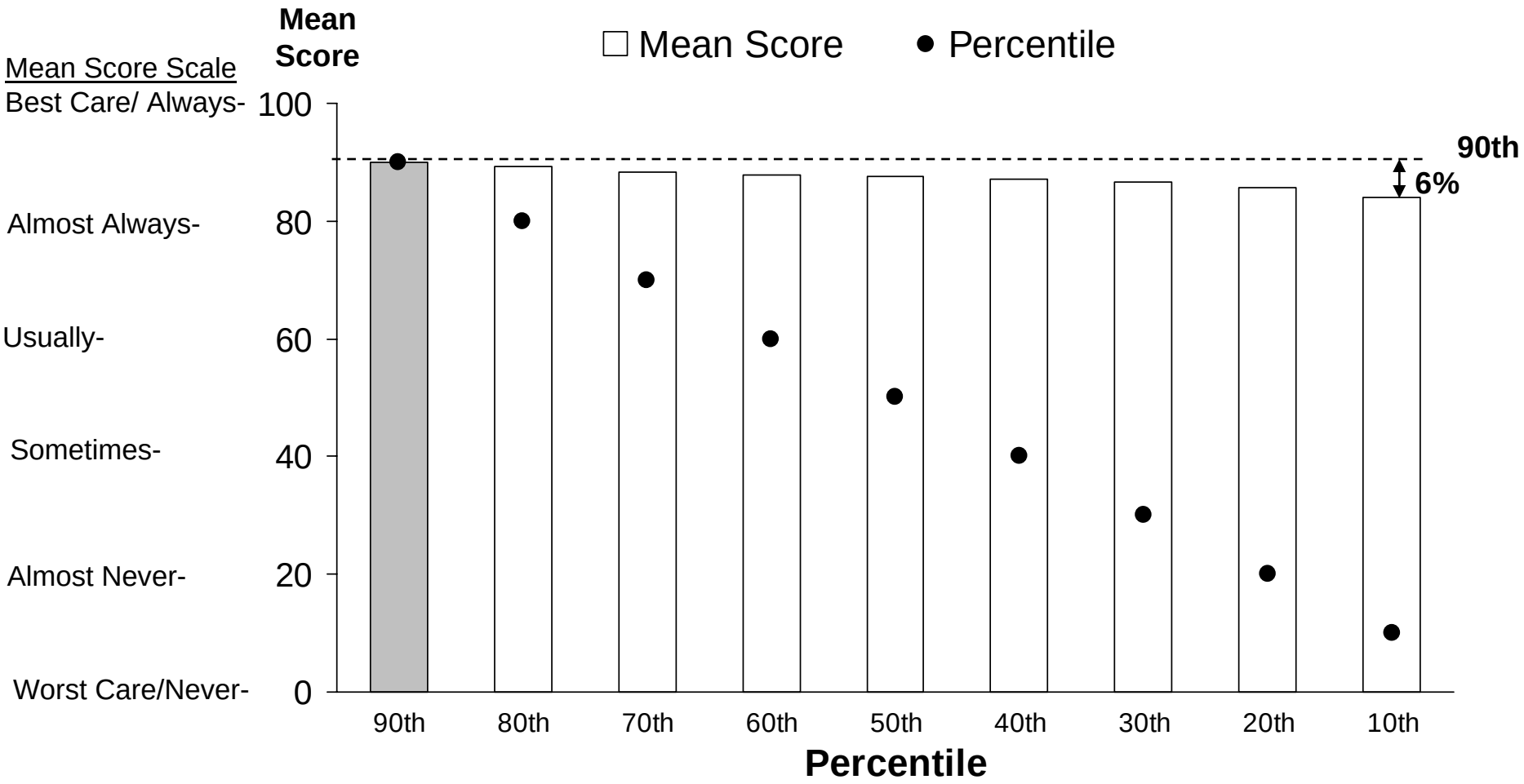
$$\frac{\text{Outcome}_{2006} - \text{Outcome}_{2005}}{\text{Outcome}_{2005}}$$

      - ✓ Change in difference to goal (Jencks et al., 2003)

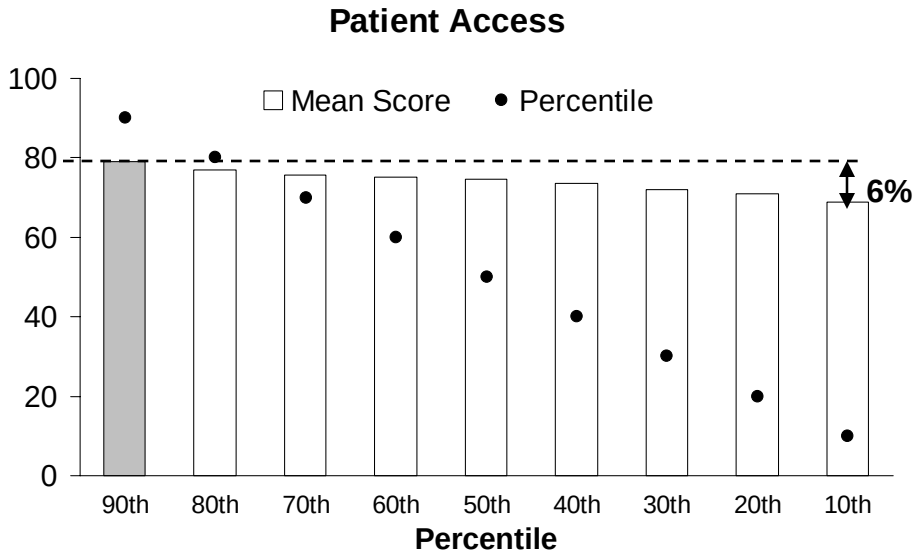
$$\frac{\text{Outcome}(\%)_{2006} - \text{Outcome}(\%)_{2005}}{100 - \text{Outcome}(\%)_{2005}}$$

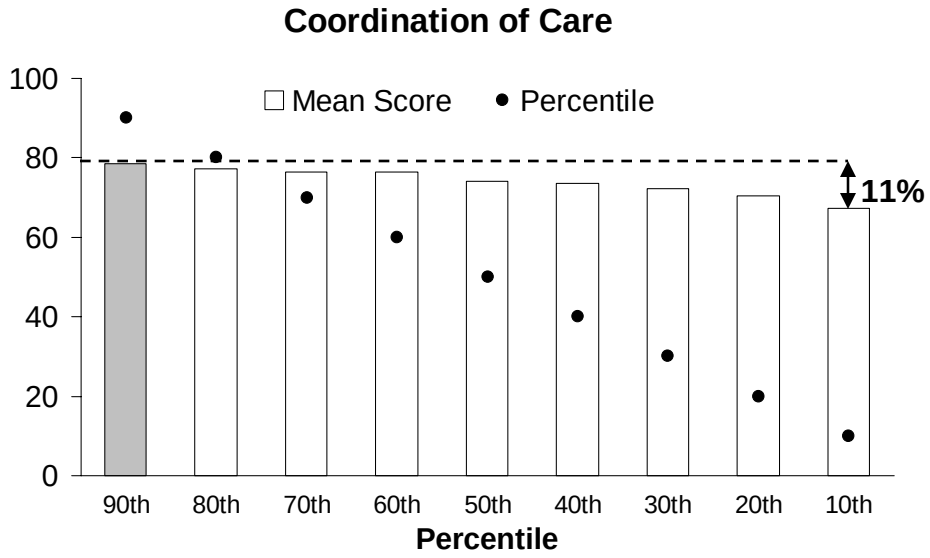
    - External comparison (percentile ranking)

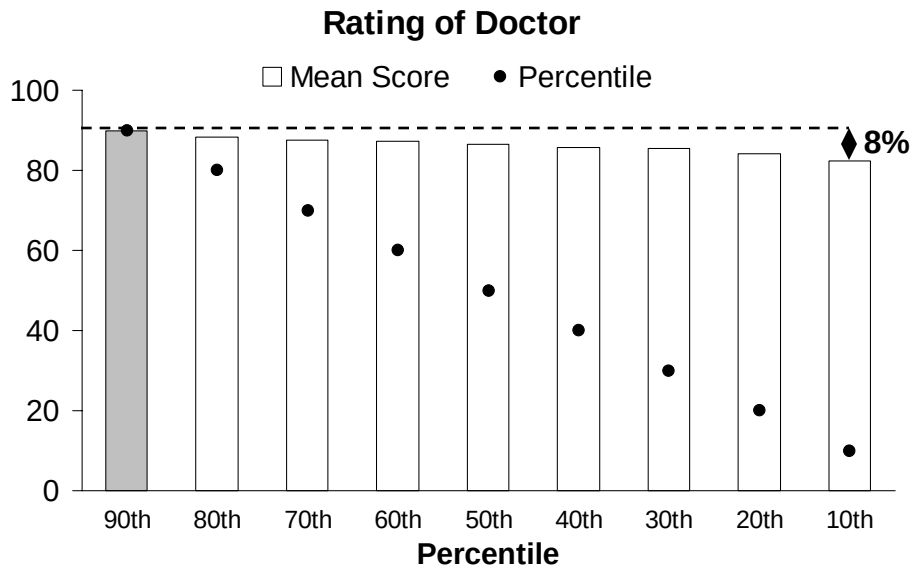
Mean Score and Percentile Distribution of Doctor-Patient Interaction Composite, Group-Level PAS, 137 Physician Organizations, 2005

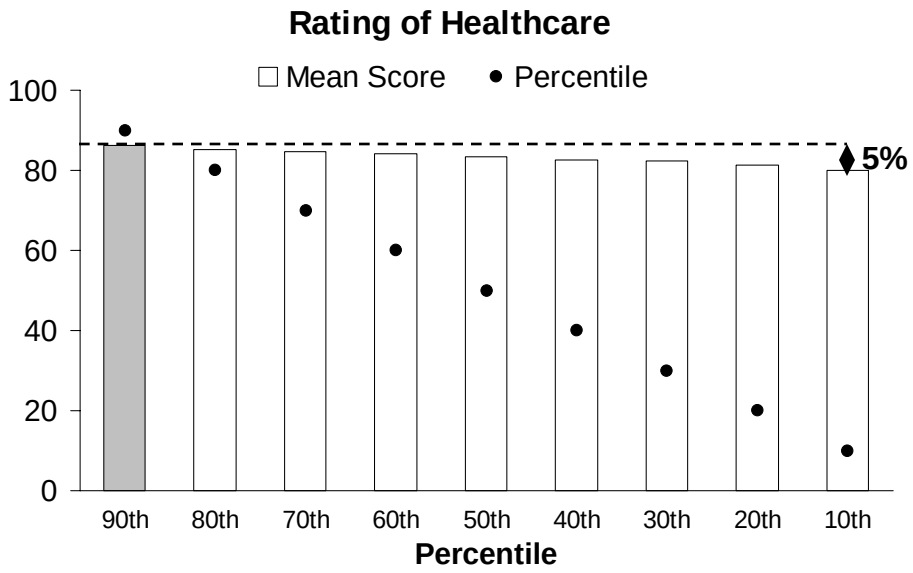


Mean Score and Percentile Distribution of Patient Experience Composites, Group-Level PAS, 137 Physician Organizations, 2005 Measurement Year

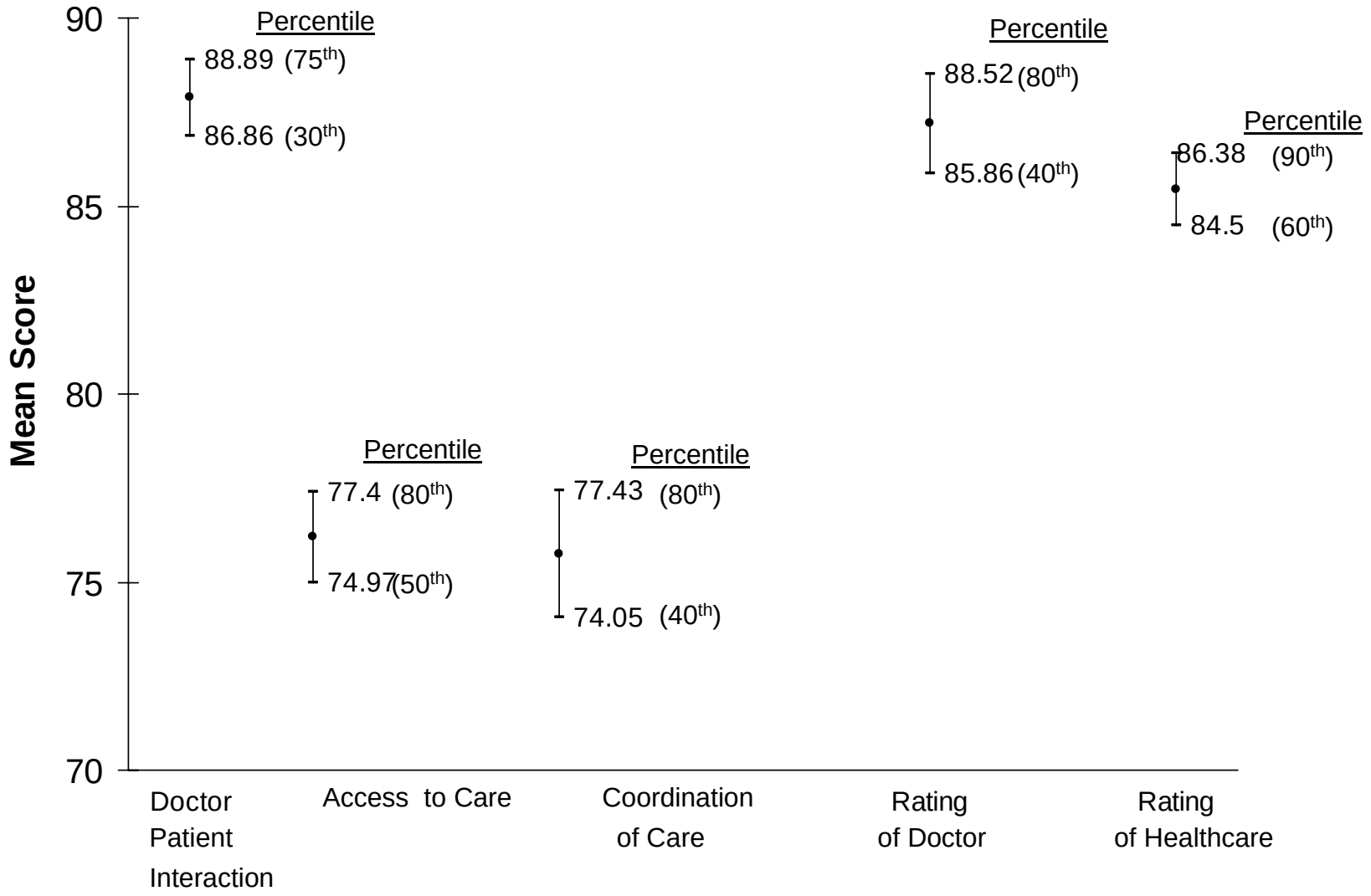




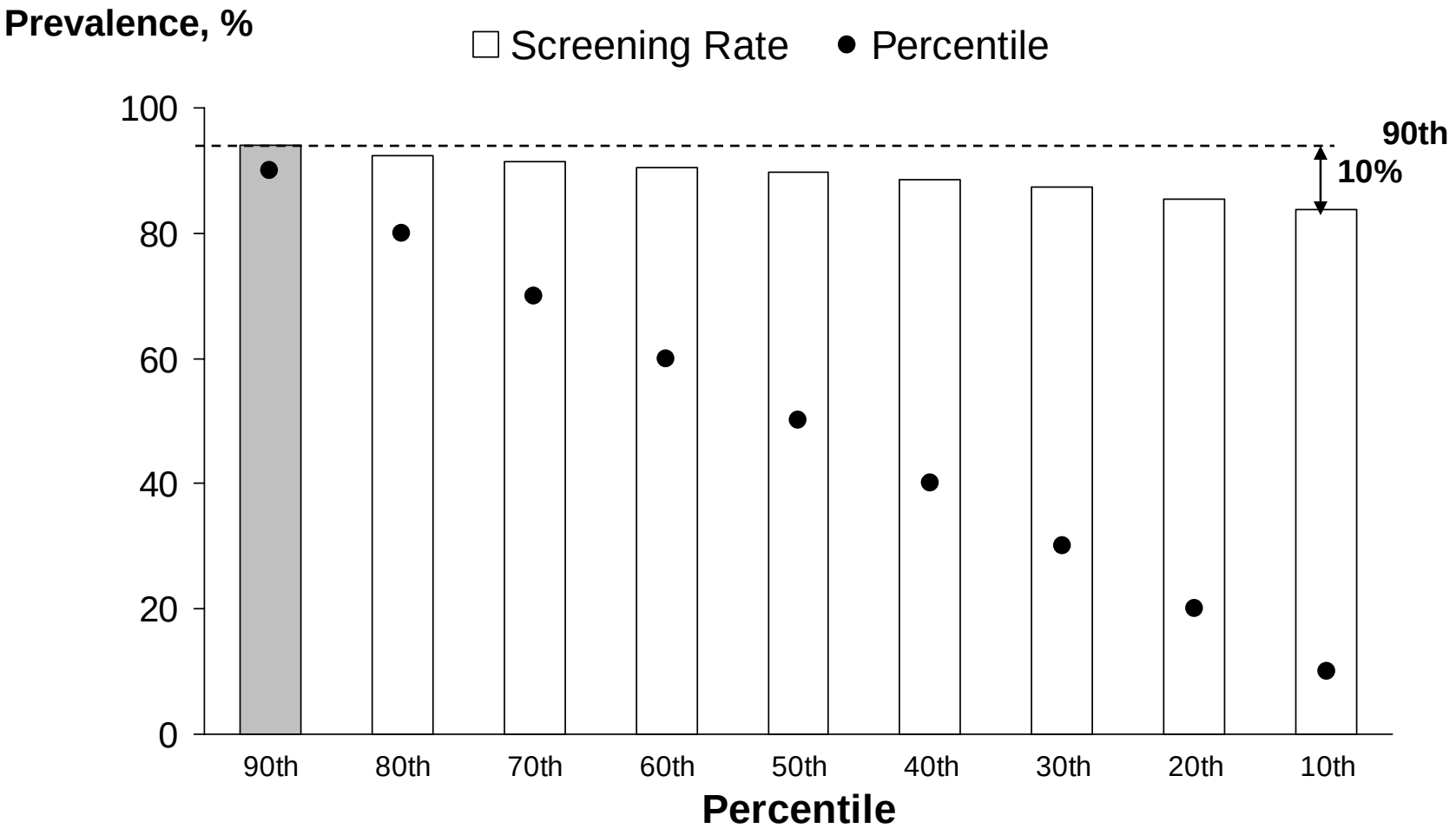




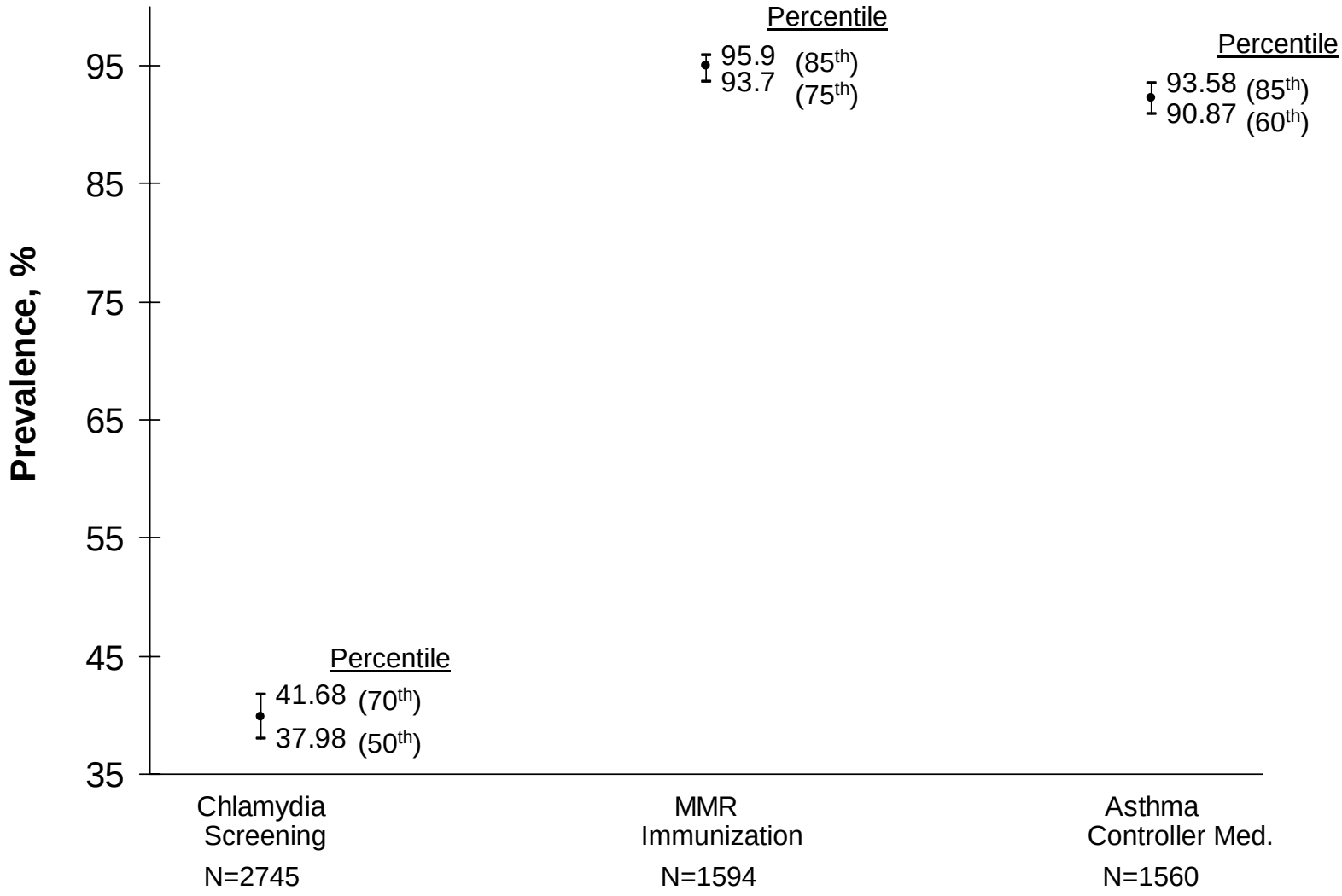
95% Confidence Interval of Mean Scores (Percentile Ranges)
PAS Composites, Group-Level PAS, Brown & Toland, 2005



Prevalence of Controller Medication Use in Patients with Persistent Asthma and Percentile Distribution, 131 Physician Organizations, P4P, 2005



95% Confidence Intervals of Selected Clinical Performance Measures and their Percentile Ranges, Brown & Toland, 2005



Summary

- Small absolute differences separate “high” (90th percentile) and “low” (10th percentile) performance
  - ✓ All group-level PAS composites
  - ✓ Some P4P clinical measures

- Because of clustering, percentile rankings PAS/Clinical measures are sensitive to sample variation even at moderately large sample sizes (N >1000).

- Many POs have high absolute achievement in controller meds (asthma), some immunizations, LDL screening

- Irrespective of percentile rank, most organizations need to improve access and coordination of care, and glycemic and lipid control in patients with diabetes

Summary

- Differences between Provider Organizations are not so great that alternative explanations might play a role in PAS group-level results:
  - ✓Selection bias (e.g. different sampling rules)
  - ✓Non-response bias

- What may account for such homogeneity in PAS results?
  - ✓ Statistical over-adjustment
  - ✓ Overlapping of some provider networks
  - ✓ Homogeneity in processes (e.g. low level of innovation such as adoption of open access)

- What are the implications of such homogeneity?
  - ✓ Variability within physician organizations may be more important to consumers than variability between physician organizations

Aligning Incentives and Quality

- When clustering of POs is pronounced and performance is generally high:
  - ✓ What might be alternatives to percentile distribution to align incentives with quality improvement? Do the Jencks et al (2003) and other approaches magnify the importance of small absolute differences?
  - ✓ Should measures be retired from incentives, but not from reporting?

- When clustering of POs is pronounced and performance is low:
  - ✓ What might be alternatives to percentile distribution to align incentives with quality improvement? Will high performers still say “why should any kind of poor performance be rewarded no matter the relative improvement?”

- Lake Woe-be-gone Effect: As performance improves on some measures and nearly all POs achieve a high level of performance, should incentives emphasize the areas of lowest performance?

References

1. Jencks SF, Huff ED, Cuerdon T. Change in the quality of care delivered to Medicare beneficiaries, 1998-1999 to 2000-2001. JAMA 2003; 289:305-312.
2. Elliott MN, Swartz R, Adams J, Spritzer KL, Hays RD. Case-mix adjustment of the national CAHPS benchmarking data 1.0: Violation of model assumptions. Health Serv Res 2001;36:555-573.