



How Close to Perfect is Good Enough When MD Performance Measurements are Used by Health Care Customers?

Arnold Milstein MD, MPH
November 5, 2007
AEI

How Big are the Problems We Must Solve?



**Unaffordability due to
40% muda and <2.5
annual efficiency gain**



The Dorsett Family

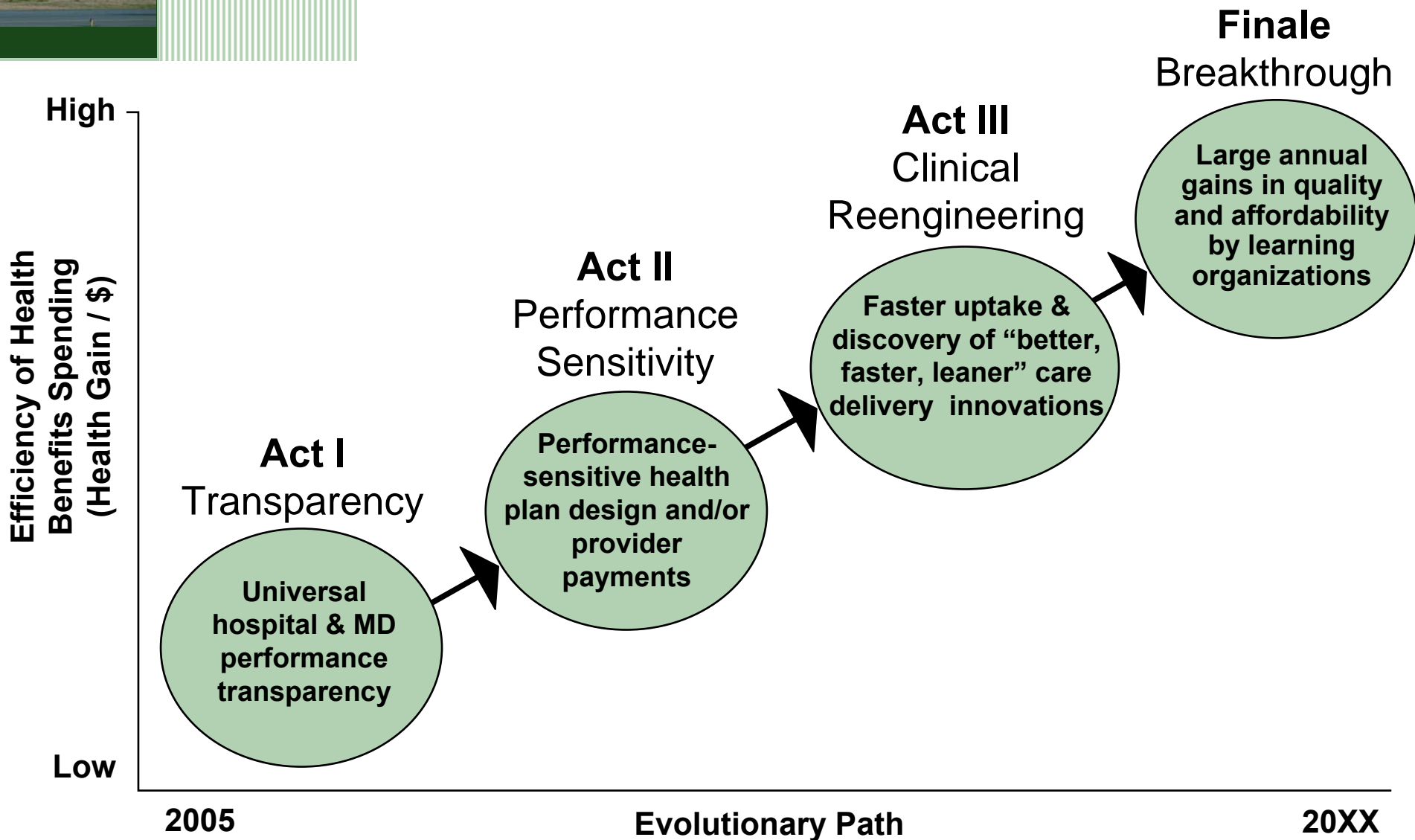
**1.2 Σ defect rate due to
work-around culture and
paper records**



Marie Dotseth



Clarifying the Path From Transparency to Moore's Law in Health Care





How Close to Perfect Measurement is Good Enough?

- **A values question, not a technical question**
- **Most stakeholders' perspectives are significantly rooted in the public interest**
- **Disagreement has emerged, especially for measurement of individual physician efficiency and effectiveness via claims data**



Advancing More Quickly Will Require Tolerance for Use of Claims Data and Other Sources of Imperfection

...there are serious limitations and opportunities for inaccuracies when using administrative data collected primarily for billing purposes to determine the quality of care physicians provide.

–State Medical Society President

...there are as many measurement challenges associated with medical record review as with claims data. Careful record review would cost > \$10 billion for a small quality starter set.

- Health Services Researcher



Since Excellence Requires Multi-MD Systems & Friction Centers on Individual MD Measurement and Tiering, Can't We Limit Transparency to MDs in Groups?

Variability in the effectiveness of care is driven by medical group *and* individual physician performance.

–Jeffery Levin-Scherz MD and Thomas H. Lee MD

Improving the quality of care ultimately requires changes in the behavior of individual physicians, even if systems to improve the quality of care play an important role.

–Arnold Epstein MD, Thomas H. Lee MD & Mary Beth Hamel MD



Social Science 101: Groups Outperform Individuals Only When Individual Identifiability is Preserved

Number of Positive Uses of a Brick

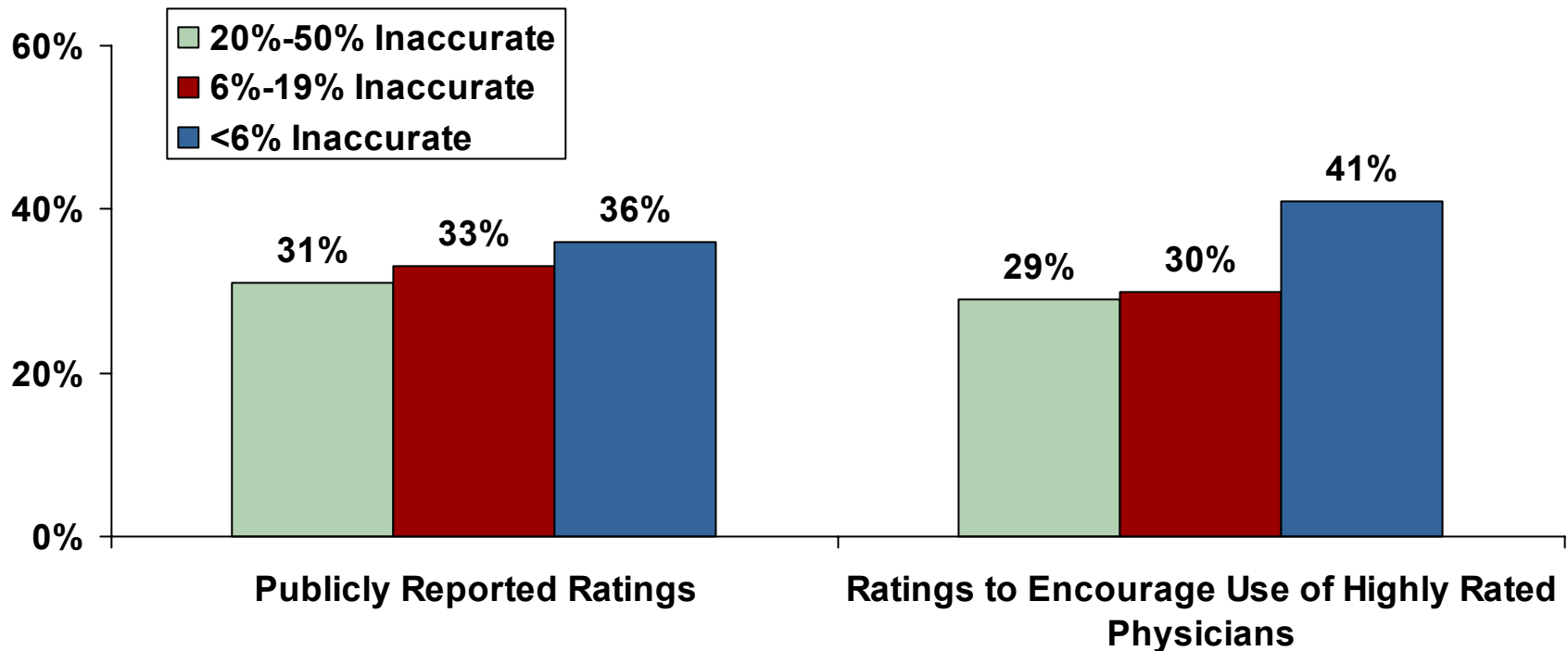
<u>Individual identifiability</u>	<u>Group</u>	<u>No group</u>
Identifiable	21.1	15.1
Anonymous	14.8	15.8

Condensed from: Guerin, B. (1999). Social behaviors as determined by different social consequences: social loafing, social facilitation, deindividuation, and a modified social loafing. *The Psychological Record*, 49, 565-578.



Consumers and Purchasers Prefer Imprecise MD Measurements to Flying Blind (So Do Statisticians When it's Personal)

**Consumer Tolerance for Inaccuracy of Physician Performance Ratings,
by Rating Application: "One Size Fits None"**

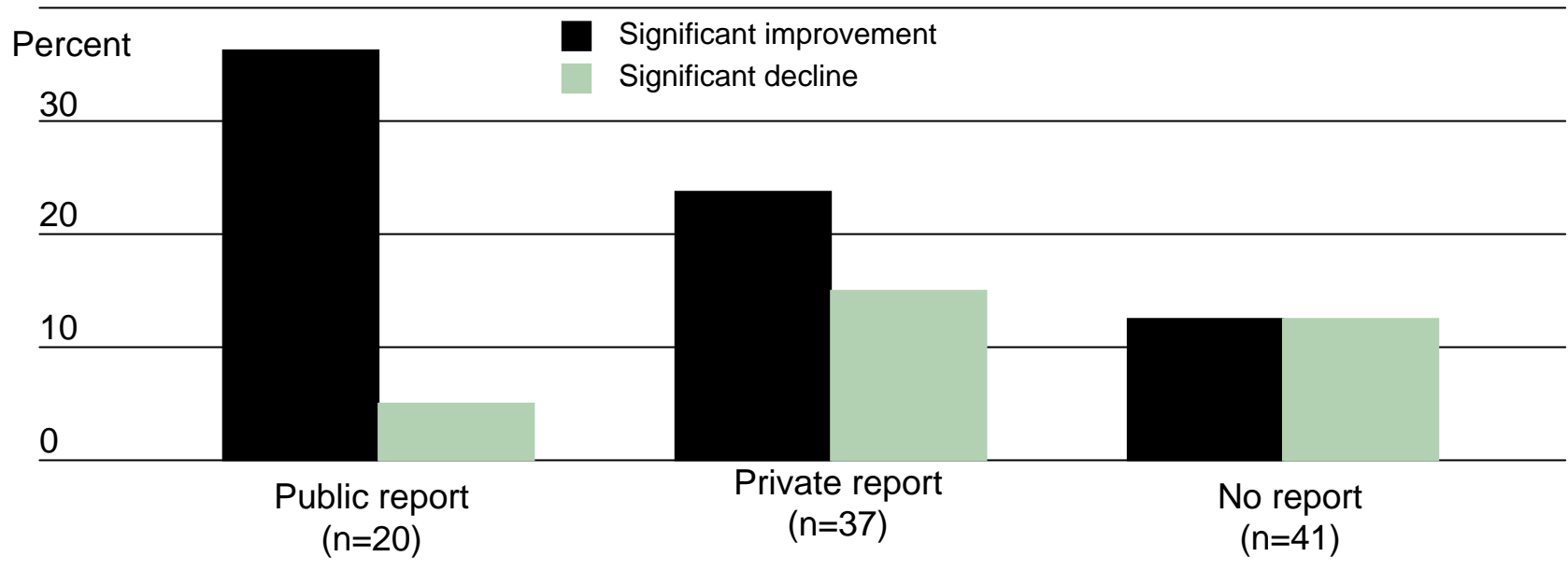


Source: Davis, Hibbard & Milstein, "CONSUMER TOLERANCE FOR INACCURACY IN PHYSICIAN PERFORMANCE RATINGS: ONE SIZE FITS NONE," Center for Studying Health System Change IB 110, March 27, 2007.



Imperfect Quality Measurements Stimulate Major Quality Gains (~ LA Restaurant Hygiene)

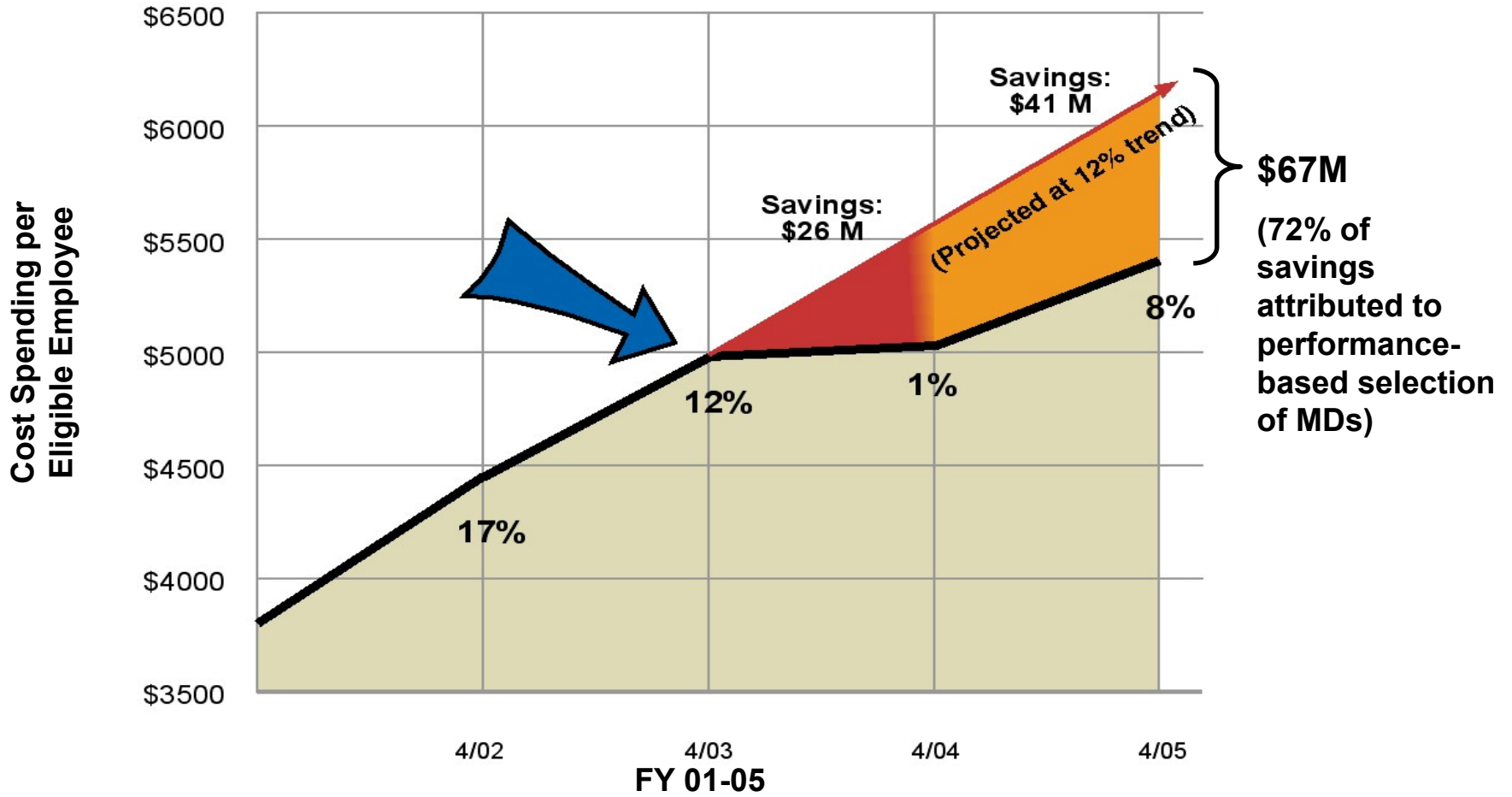
Percentage Of Hospitals With Statistically Significant Improvements Or Declines In Obstetrics Performance In The Post-Report Period (2001-2003)



Source: J. Hibbard, et al., Health Affairs, July 2005; and Wisconsin Bureau of Health Information, risk-adjusted by Medstat.



Imperfect Efficiency Measurements Stimulate Major Efficiency Gains (and Wage Increases for Low Wage Workers)





What Measurement Dragons Do Americans (Especially Low and Middle Income Households) Need Us to Slay Quickly?



Infeasibility of Consensus-Based MD Measurements



→ Let consumer leaders set measurement validity standards



Attributional Paralysis



→ Use attribution to multiple MDs



Measurement Burden Resistance



→ Strategically expand required billing data



Adaptation to Darkness



→ Cease tolerance of performance blindness now



Weighing Due Care ("MD Measurement Science is New") And Due Speed ("the Hour is Late")



- **Mainstreet pays the heaviest price**
- **Mainstreet is personally familiar with imperfect performance measurements**
- **Let consumer leaders set minimum validity standards for MD measurements**