Statistical Methods to Detect Variations in Provider Performance: Issues re. optimal methodology

Thérèse A. Stukel

Institute for Clinical Evaluative Sciences, Toronto
The Dartmouth Institute for Health Policy & Clinical Practice, Lebanon, NH
Institute of Health Policy, Management and Evaluation, University of Toronto

Provider Profiling

- Estimate provider-level performance & document extent of variation in outcomes
- Quality improvement, public reporting, identifying providers with poor outcomes
- Statistical methods: Fixed effects models (FEM), random effects models (REM), indirectly standardized rates (ISR)
- CMS Hospital Compare uses REM to estimate hospital mortality and readmission rates, and detect outliers; National Dialysis Facilities monitoring use FEM
- Ongoing debates in health policy and statistical circles regarding optimal methodology and the appropriate comparator: population average or one's peer group?

The rates displayed in this table are from data reported for discharges July 2006 through June 2007.

Heart Attack Death (Mortality) Rates tell you how the 30-day death rates from Heart Attack at the hospitals you selected compare to the U.S. National Heart Attack death (mortality) rate. These comparisons take into account how sick patients were before they were admitted to the hospital and differences in death rates that might be due to chance.

Hospital 30-Day Risk Adjusted Death (Mortality) from Heart Attack Compared to U.S. National Rate.

The U.S. National 30-day Death Rate from Heart Attack = 16.1%

HOSPITAL NAME	Better Than U.S. National Rate (Adjusted mortality is lower than U.S. rate)	No Different Than U.S. National Rate (Adjusted mortality is about the same as U.S. rate or difference is uncertain)	Worse Than U.S. National Rate (Adjusted mortality is higher than U.S. Rate)
HOSPITAL OF UNIV OF PENNSYLVANIA		✔	
MAIN LINE HOSPITAL LANKENAU		✔	
MILLCREEK COMMUNITY HOSPITAL		✔	

The "total number" of hospitals in the table below may differ from the total number of hospitals that voluntarily submitted process of care quality measure data. See Data Details for additional information about the data collection for the mortality measures.

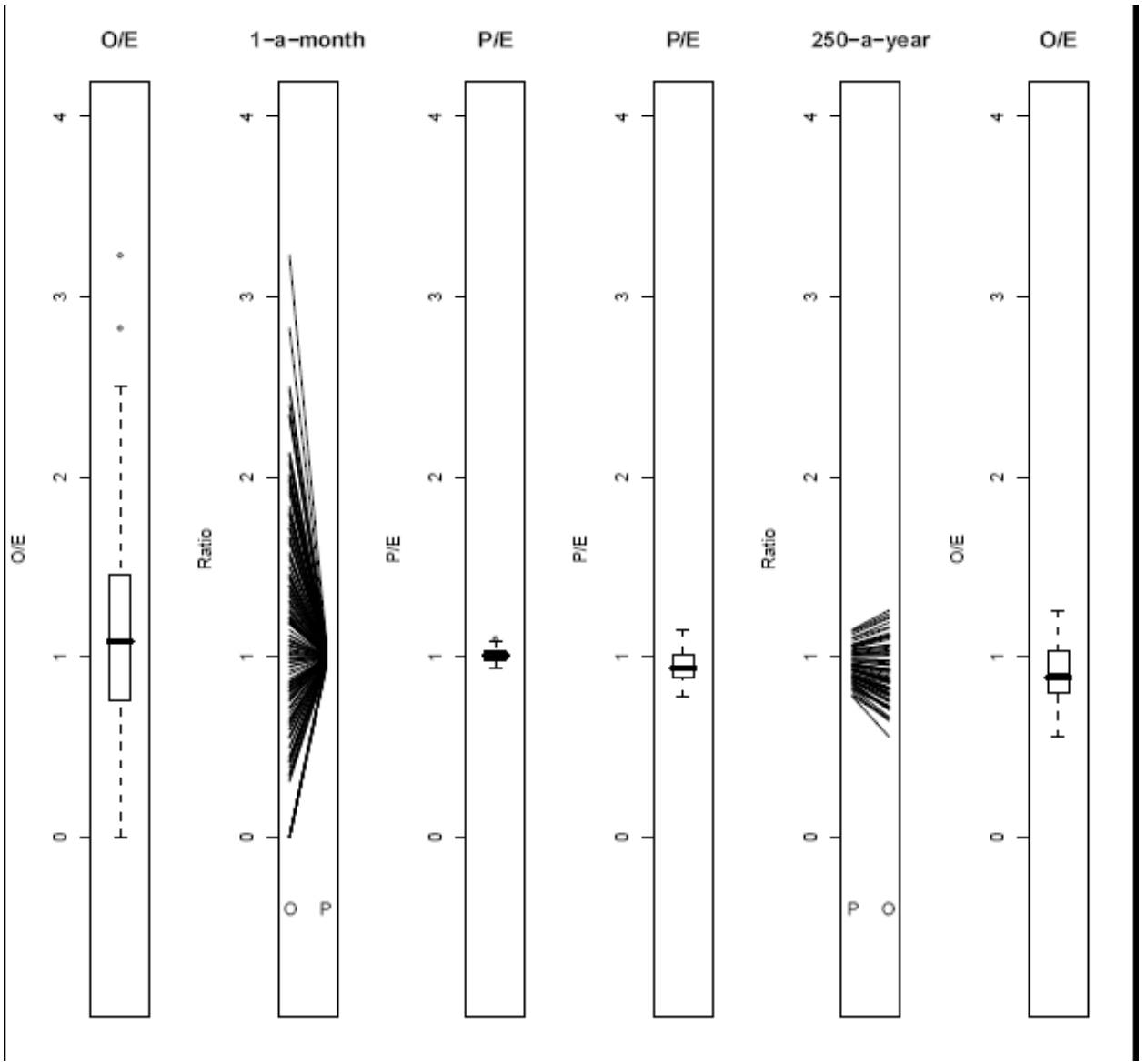
Out of 4311 in the United States →	**9** hospitals in the United States Better than U.S. National Rate	**4302** hospitals in the United States No different than U.S. National Rate	**0** hospitals in the United States Worse than U.S. National Rate
Out of 160 in Pennsylvania →	**1** hospitals Better than U.S. National Rate	**159** hospitals No different than U.S. National Rate	**0** hospitals Worse than U.S. National Rate

Note: Medicare derived the 30-Day Risk-Adjusted Death (Mortality) measures from its own data about patients on Original Medicare and the hospitals that treat them. The information in this table reflects care given only to patients who are on Original Medicare. All data are risk-adjusted.

Public Reporting

- Report Observed/Expected=O/E (FEM, ISR)
  or Predicted/Expected = P/E (REM)
- Expected (E) based on FEM or REM or ISR, accounting for patient risk factors
- Flag providers with unusually high outcome rates
- Which statistical methods are optimal?

Magnitude of shrinkage in ISR vs. RE models: O/E vs P/E



Silber, 2010

Linear Models

Linear regression model

$$Y_{ij} = (\mu + \alpha_i) + Z_{ij} \beta + \varepsilon_{ij}$$

$i = 1,\ldots, N$ indexes providers; $j = 1,\ldots, mi$ indexes patients

- Outcome, Y_{ij}
- Mean overall outcome μ
- Patient characteristics, Z_{ij}
- **Focus: Provider effects α_i**: difference between overall mean μ and provider-specific mean μ_i
- Unexplained variation $\varepsilon_{ij} \sim N(0, \sigma^2\varepsilon)$

➔ **Identify providers with poor outcomes (large $\alpha_i > 0$)**

Profiling Methods

Fixed Effects Models (FEM)

- Joint models for provider and patient effects, LS regression

Indirect Standardization (ISR)

- Models for patient characteristics; compute expected (E)

Hierarchical Random Effects Models (REM)

- Hierarchical Empirical Bayes models and shrinkage estimates to estimate provider effects, controlling for patient risk factors
- Addresses instability of small providers

All methods

- Least Squares regression, Cox model, logistic regression
- Compare O/E or P/E across providers

REM and FEM Assumptions

- Random effects model (REM) requires normal distribution of αi, and independence between αi and patient case mix

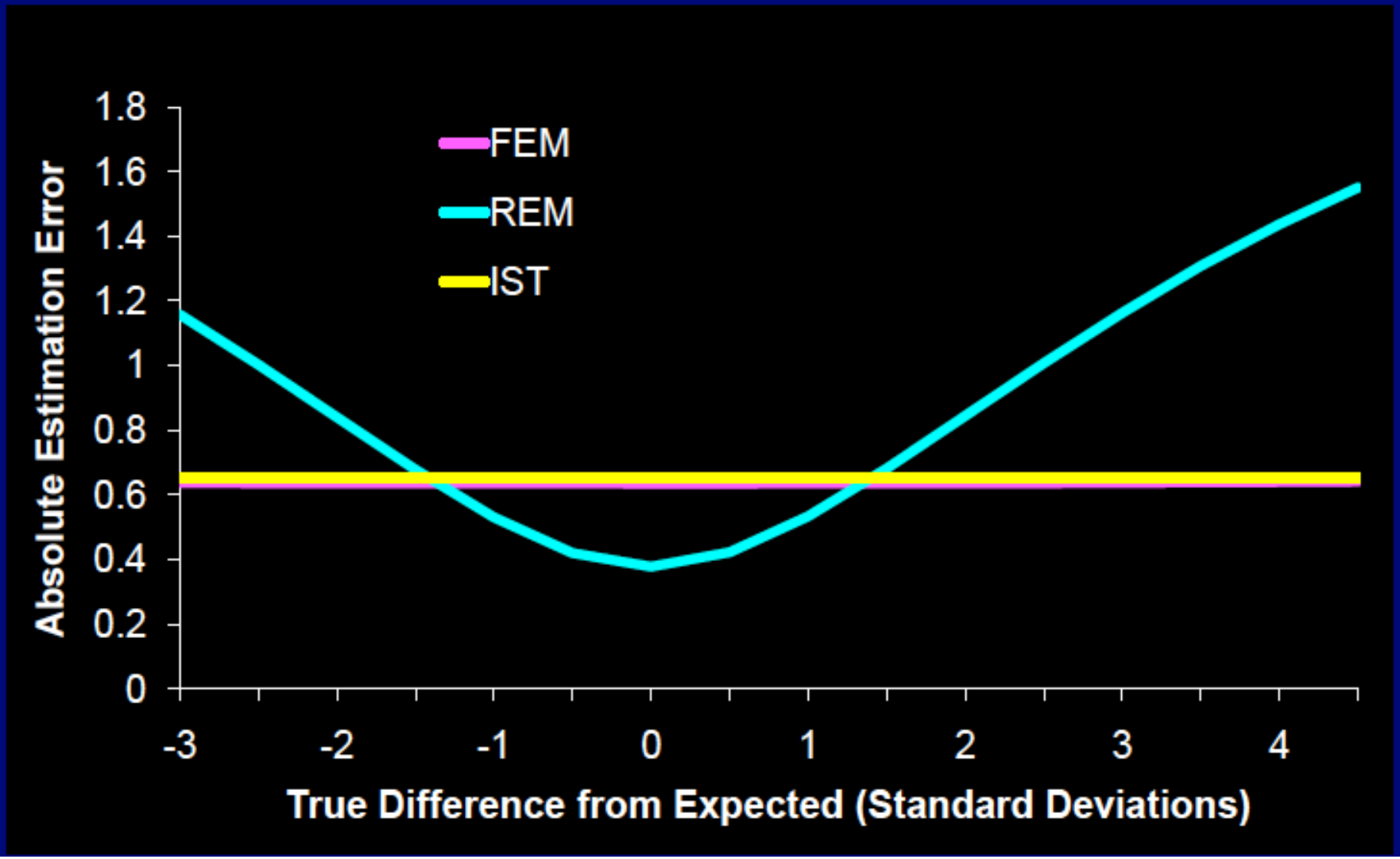
- Fixed Effects Model (FEM) assumes $\Sigma \alpha i = 0$ but not other REM assumptions; requires large *mi* (# patients per provider) for consistent results

When REM Conditions Satisfied

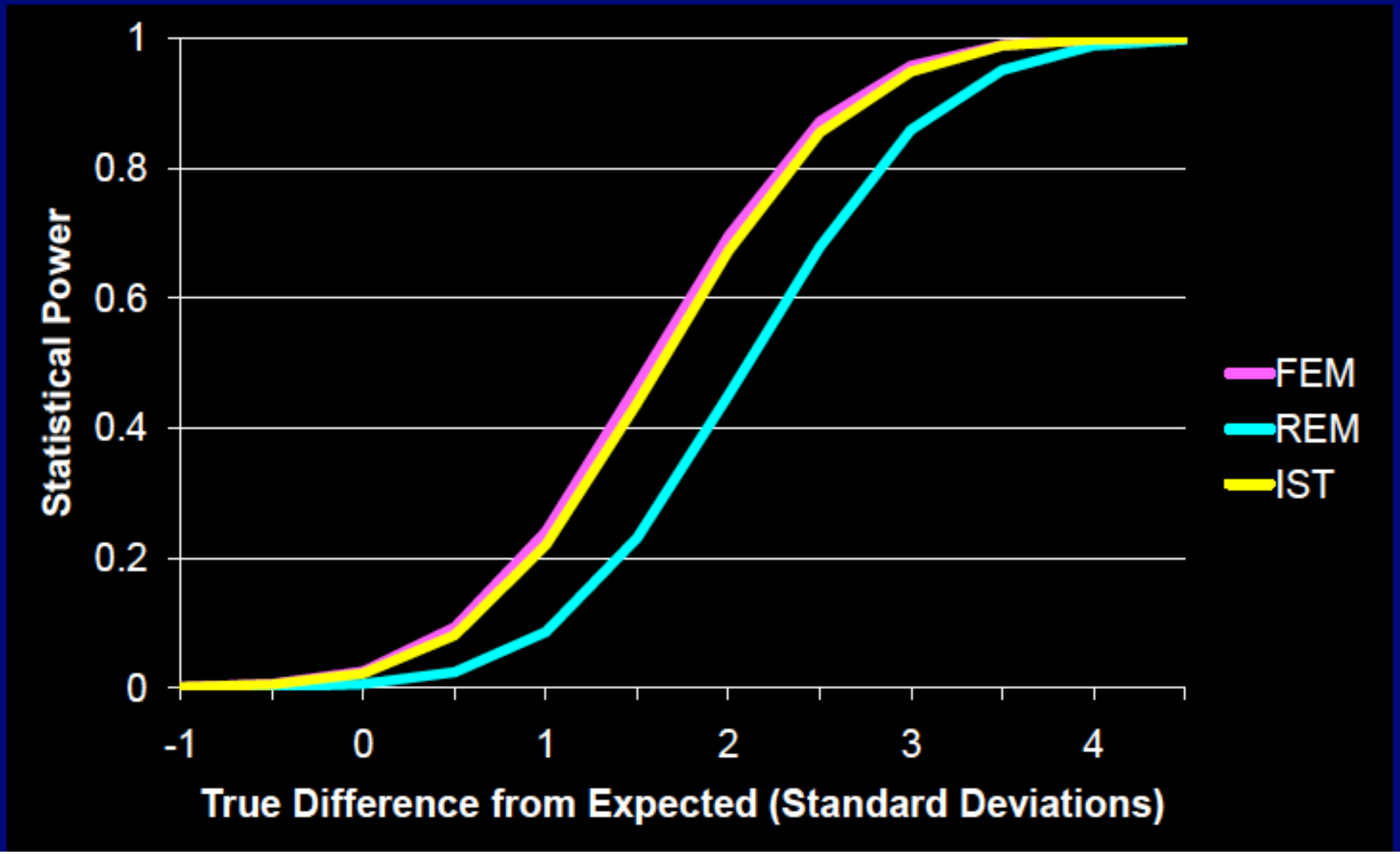
Kalbfleisch & Wolfe (2013) simulation results:

- REM estimates (***αi***) are biased; observed ("predicted") values are shrunk towards 0. Less bias for small provider effects; more bias for large provider effects
- FEM and ISR estimate accuracy is independent of provider effects (***αi****)*
- REM estimates have lower MSE = mean^2 + variance
- FEM has highest power to detect outlying providers
- Huge provider effects or large # patients per provider ➔ all methods reliably detect outliers

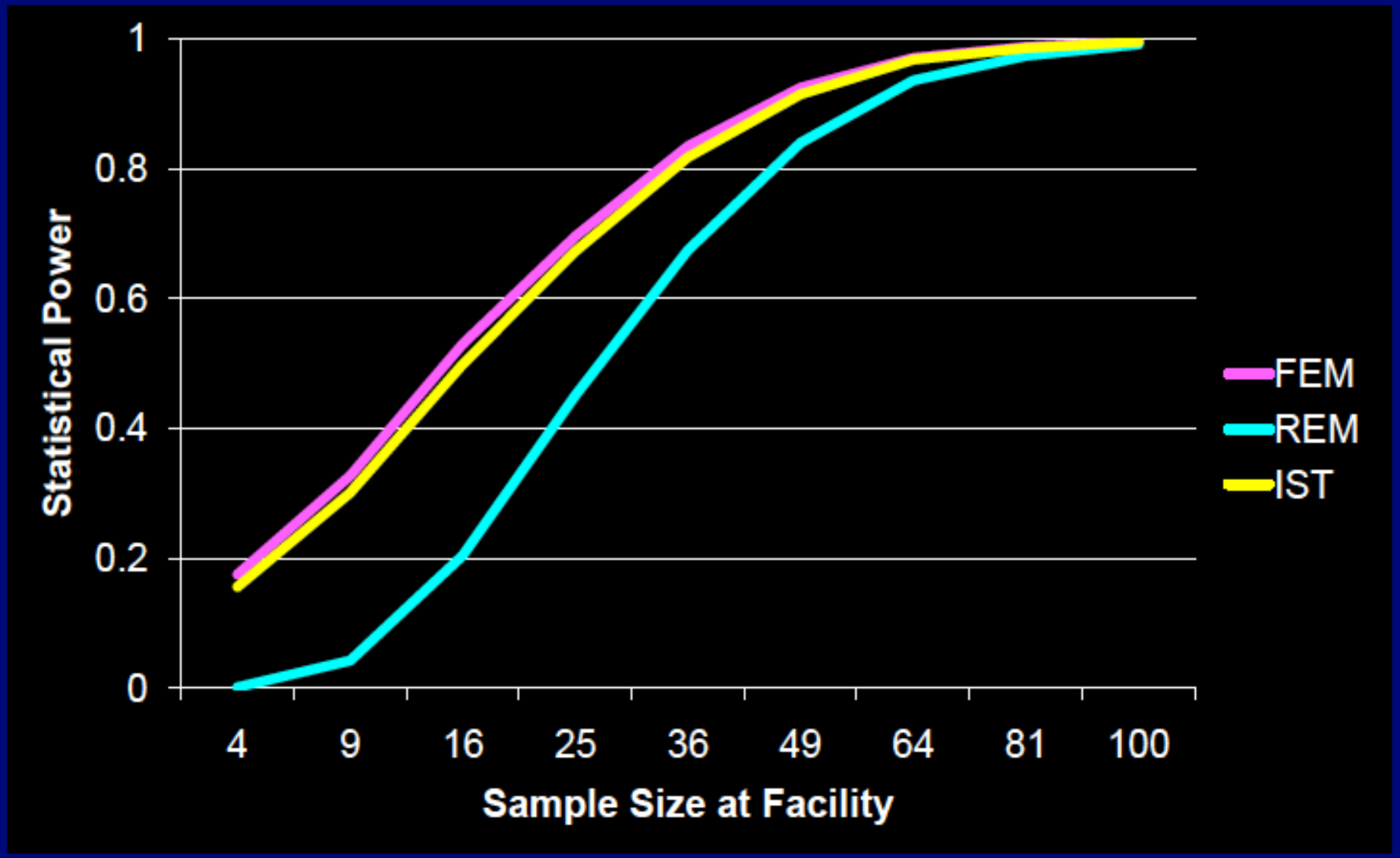
Estimation Error depends on α_1 (m_1=25)



Fraction of Facilities Flagged (m_1=25)



Fraction Flagged Among Facilities That Are 2 Standard Deviations Worse Than Expected

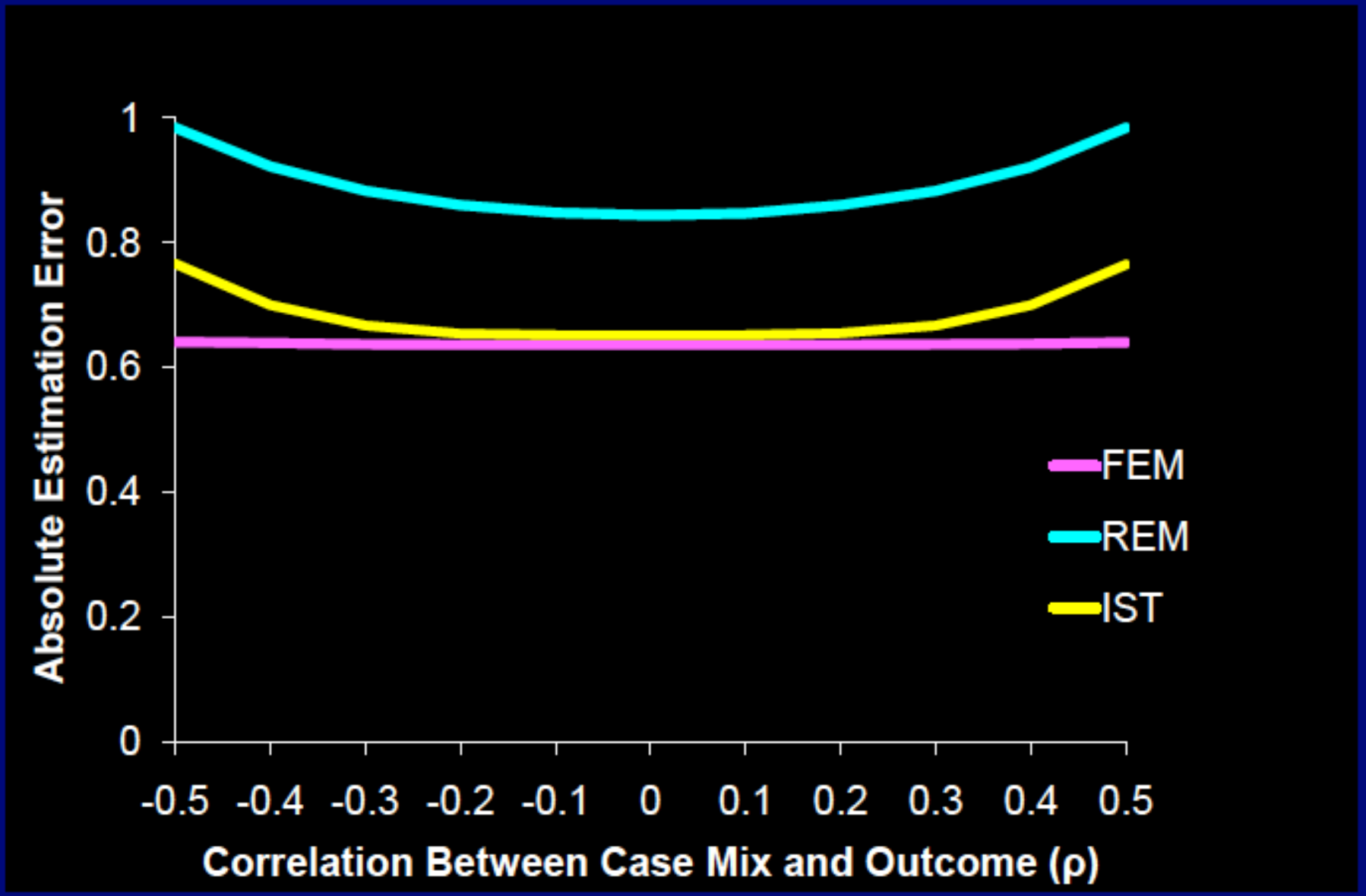


When REM Conditions not Satisfied: Correlations between Providers & Case Mix

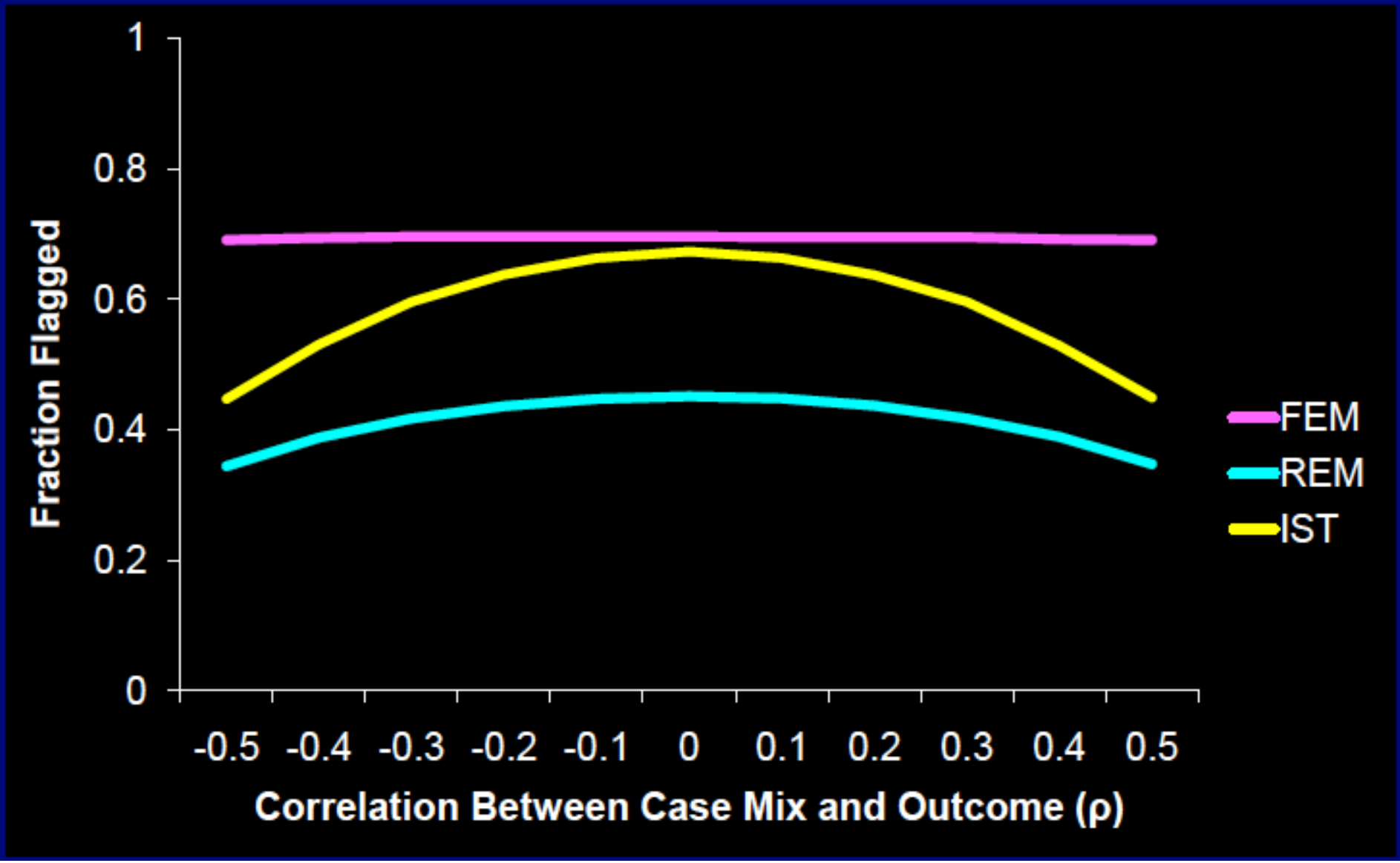
Correlation implies sicker patients referred to providers with better (negative correlation) or worse treatment strategies (positive correlation) in the presence of unmeasured confounding.

- REM estimates are more accurate for providers with outcomes close to expected ($-1.5 \leq \alpha i \leq 1.5$) (figure not shown)
- FEM and ISR estimates are more accurate for providers with extreme outcomes ($|\alpha i| > 1.5$)
- REM has lower power to detect outlying providers

$E(|\alpha_1 - \alpha_{hat\ 1}|)$ when $\alpha_1 = 2$ and ρ varies ($m_1 = 25$)



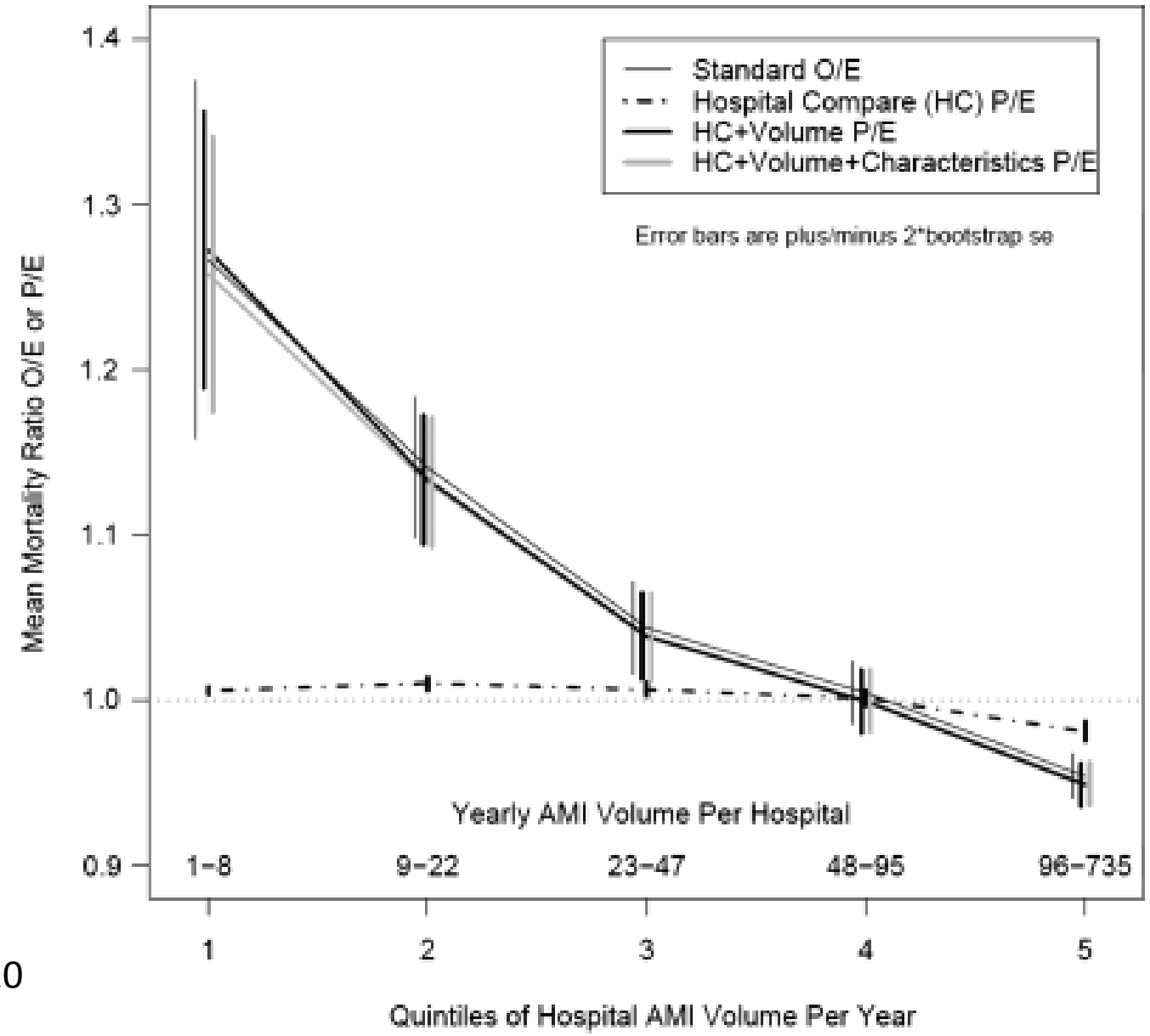
Power (or flagging rate) when α_1 =2 and ρ varies (m_1=25)



What is the proper comparator group?

- SMR = 1.0 implies provider performs at the national average for its case mix
- RE shrinks provider outcomes to the national average
- Hospital mortality and readmissions are higher in low volume hospitals
- What is the appropriate comparator group: national average or one's peer group?

Comparison of Standard Logistic Regression O/E vs. 3 Versions of Hospital Compare P/E Mortality Ratios



Silber, 2010

Profiling Methods: Comments

- REM and ISR are biased in the presence of correlation between provider and patient effects. FEM is (nearly) unbiased regardless of correlation
- FEM seem better for identify outlying centers; however, FE may flag too many large centers for review
- Debate as to whether proper comparator group is the national average or the average of one's peers; this is a policy, not a statistical question.