Choosing the Sample and Sample Size for Medical Home Evaluations: How to Ensure that Studies Can Answer the Key Research Questions

March 14, 2011

Presentation to the Medical Home Summit, Philadelphia, PA

Deborah Peikes • Stacy Dale • Eric Lundquist

MATHEMATICA
Policy Research, Inc.

Research funded by AHRQ and Mathematica Policy Research

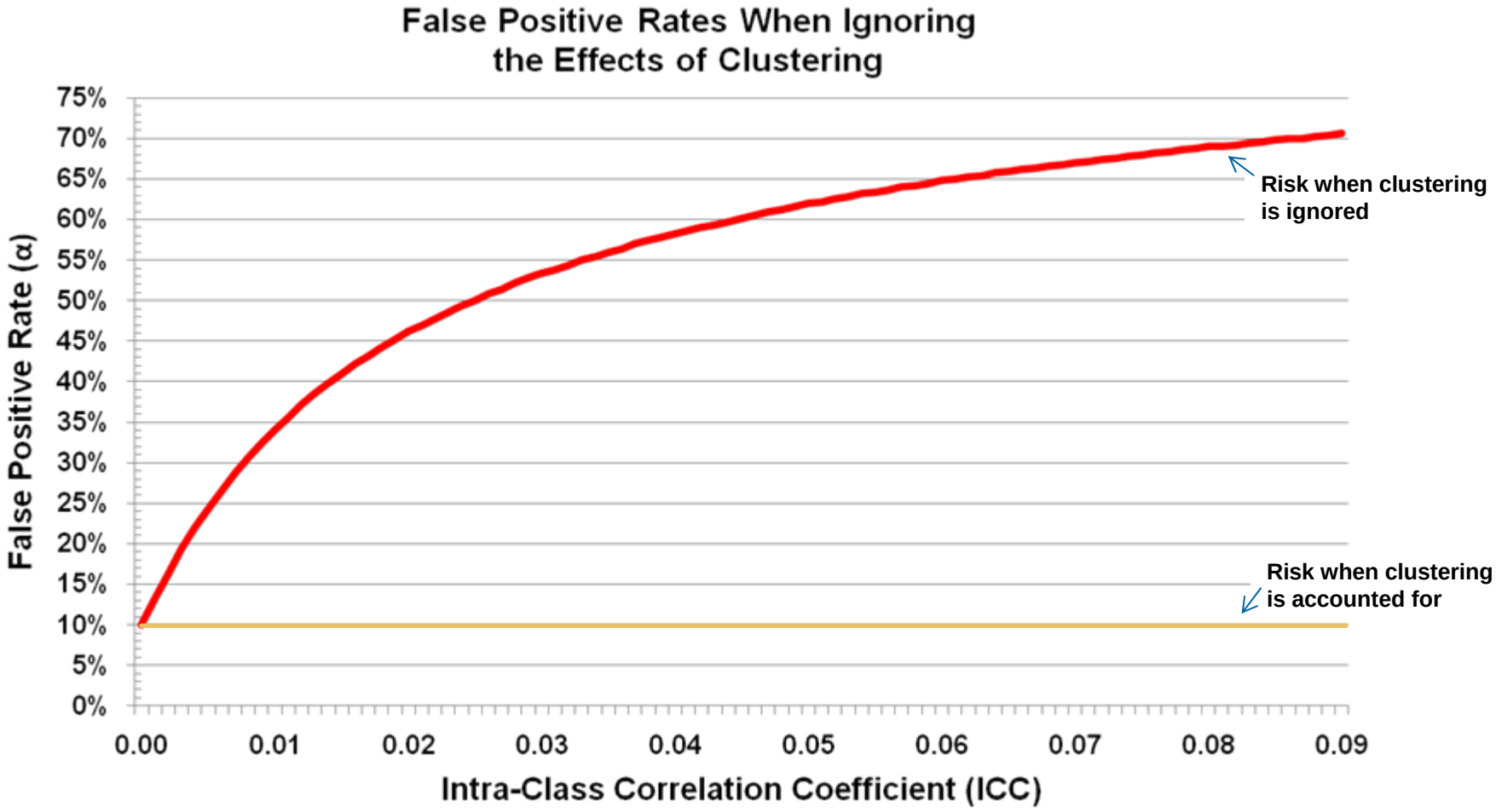
Insurers Are Testing Whether Practice-Level Interventions, Such as the Medical Home, Reduce Costs and Improve Quality

But It Is Difficult for Studies to Generate Credible Evidence

- Limited number of practices in each study
- High variation in costs and health care service use
- Correlation of outcomes within practices ("clustering")
- Interventions may only affect the costs of a small fraction of the practices' patients (chronically ill, high-risk patients)
- If studies don't take clustering into account, there is a large risk of a false positive (that is, concluding that an intervention is effective, when it is not)

Why Does Research Need to Account for Clustering?

- **Failing to account for clustering in the analysis will lead to false positives because effects will (mistakenly) appear to be statistically significant.**



Our Goals

1. **To help decision makers avoid making wrong decisions about practice-level interventions such as the medical home based on flawed studies.**

2. **To identify the approximate number of patients and practices required to detect policy-relevant yet achievable effects.**

What Is a Minimum Detectable Effect (MDE)?

- **The smallest true intervention effect that can be detected using the study sample.**

What Are Plausibly Sized (but Probably Best-Case Scenario) Intervention Effects?

	Costs	Hospitalizations	Satisfaction	Quality of Care
High-risk patients	15%	20%	20%	20%
All patients	5%	5%	20%	20%

How Large a Sample Do We Need to Detect Plausible Effects?

To answer, we primarily need to know:

- **Intra-cluster correlation (ICC)—a measure of how much outcomes vary from one practice (the cluster) to another. Variation reflects different types of patients, as well as different practice patterns.**
  - **A lower ICC makes it easier to detect effects.**
  - **Under many typical scenarios, the MDE is twice as large when the data are clustered (and accounted for correctly).**
- **Coefficient of Variation (CV) of the outcome variable—the standard deviation divided by the mean. A lower CV makes it easier to detect effects.**

Step 1: Review Literature

- **Some pilots don't have access to data**
- **We compiled CVs and ICCs for key outcomes from more than 20 published and unpublished studies**

		Satisfaction/ Quality of Care	
	Costs	Mean=50%	Mean=70%
High-Risk Patients			
CV	2.0	1.0	0.655
ICC	0.001-0.03	0.03-0.09	0.03-0.09
All Patients			
CV	3.8	1.0	0.655
ICC	0.001-0.03	0.03-0.09	0.03-0.09

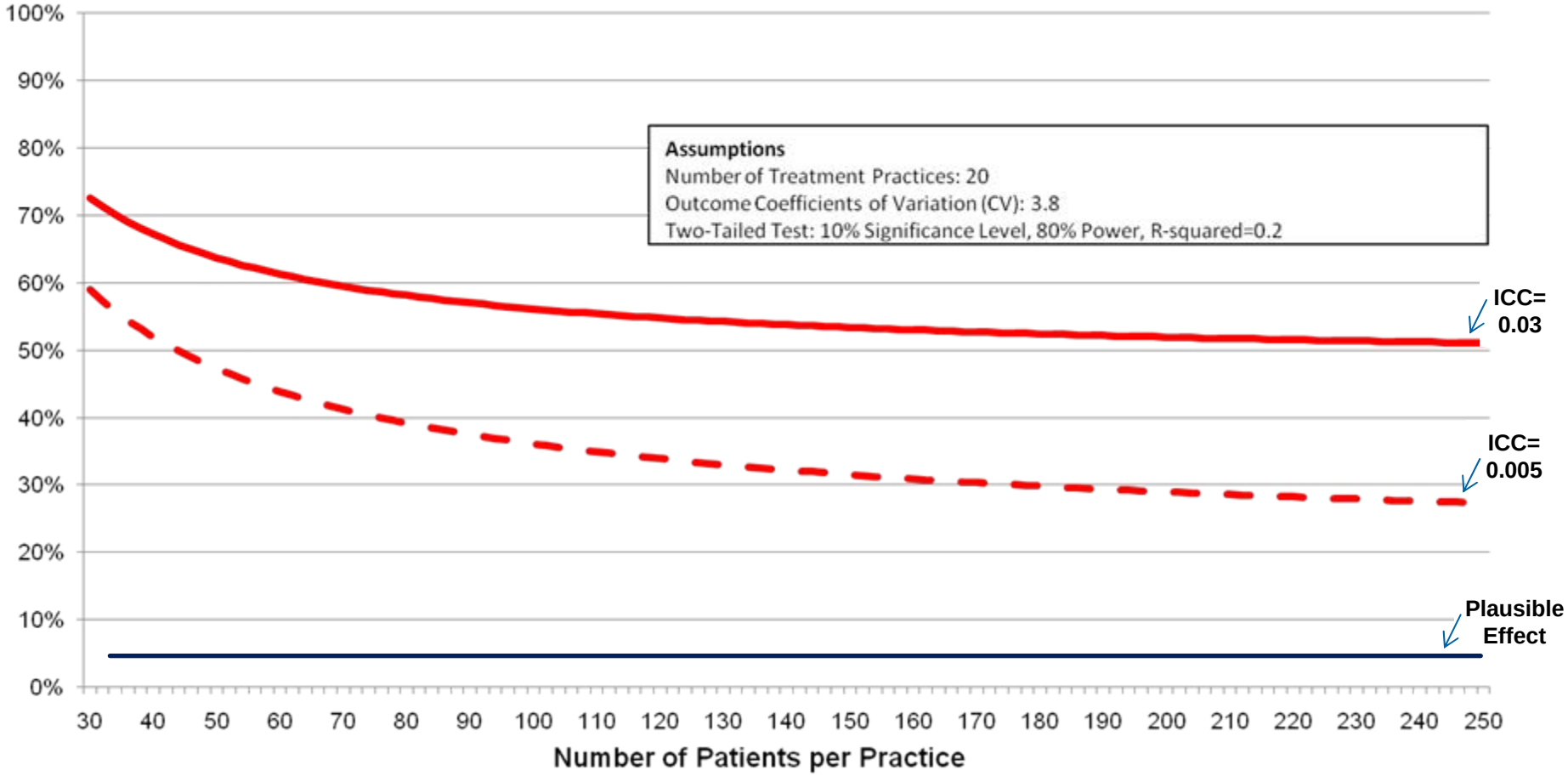
Step 2: Calculate MDEs

- **Graph MDEs using plausible values of ICCs and CVs**
  - **Assume 10% significance level (that is, a 10% chance of a false positive), 80% power (that is, a 20% chance of a false negative), and regression R-squared of 0.2**
- **Determine how many practices and patients per practice are needed to detect plausible effects**
- **Researchers should tailor MDE graphs using study-specific values of ICC and CV**

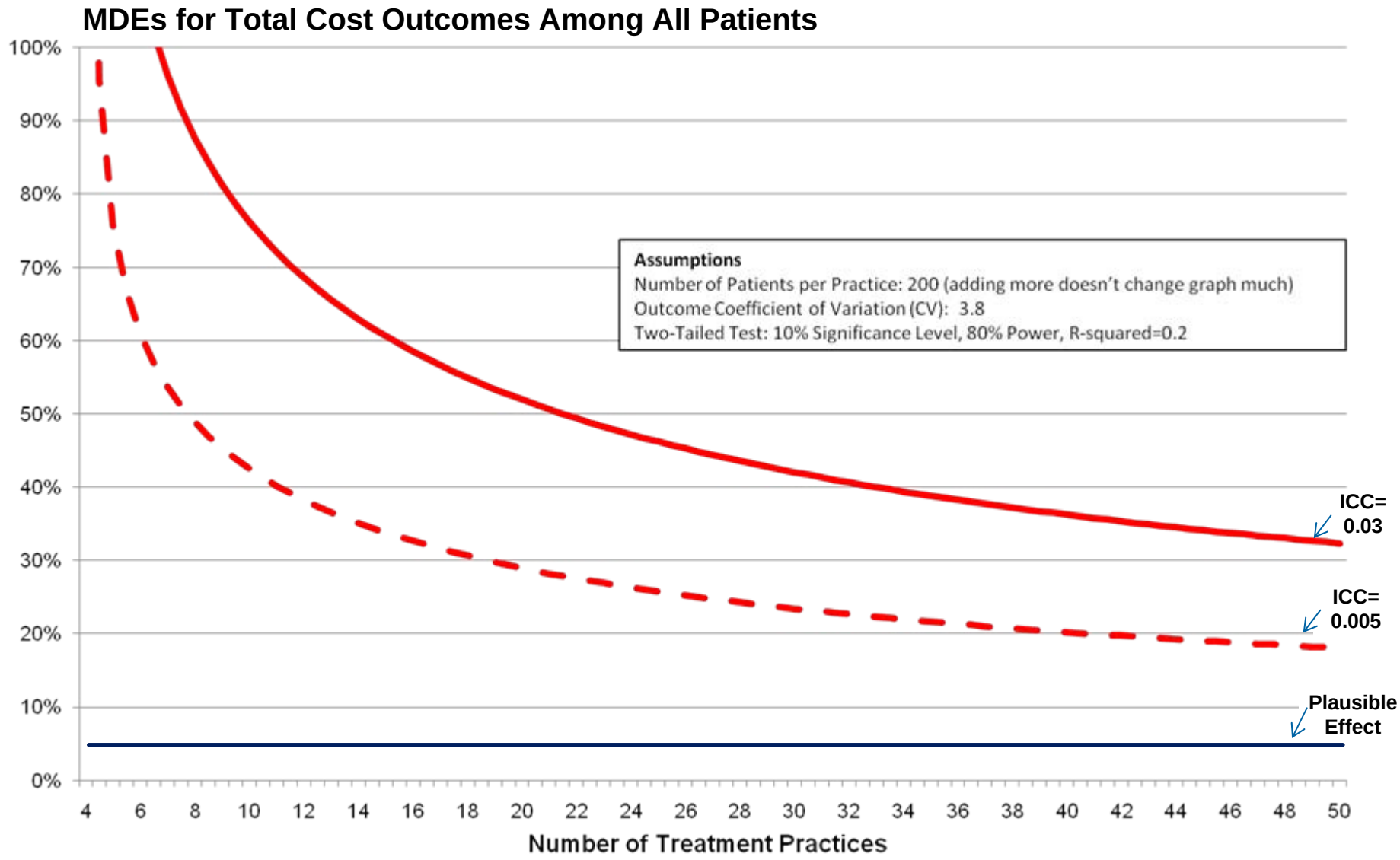
Findings:
Including More Patients Only Slightly Improves MDEs

- **If enlarging the sample increases costs, it may be worth sampling only a fraction of the patients, as the improvement in MDEs from more patients is small.**

MDEs for Total Cost Outcomes Among All Patients



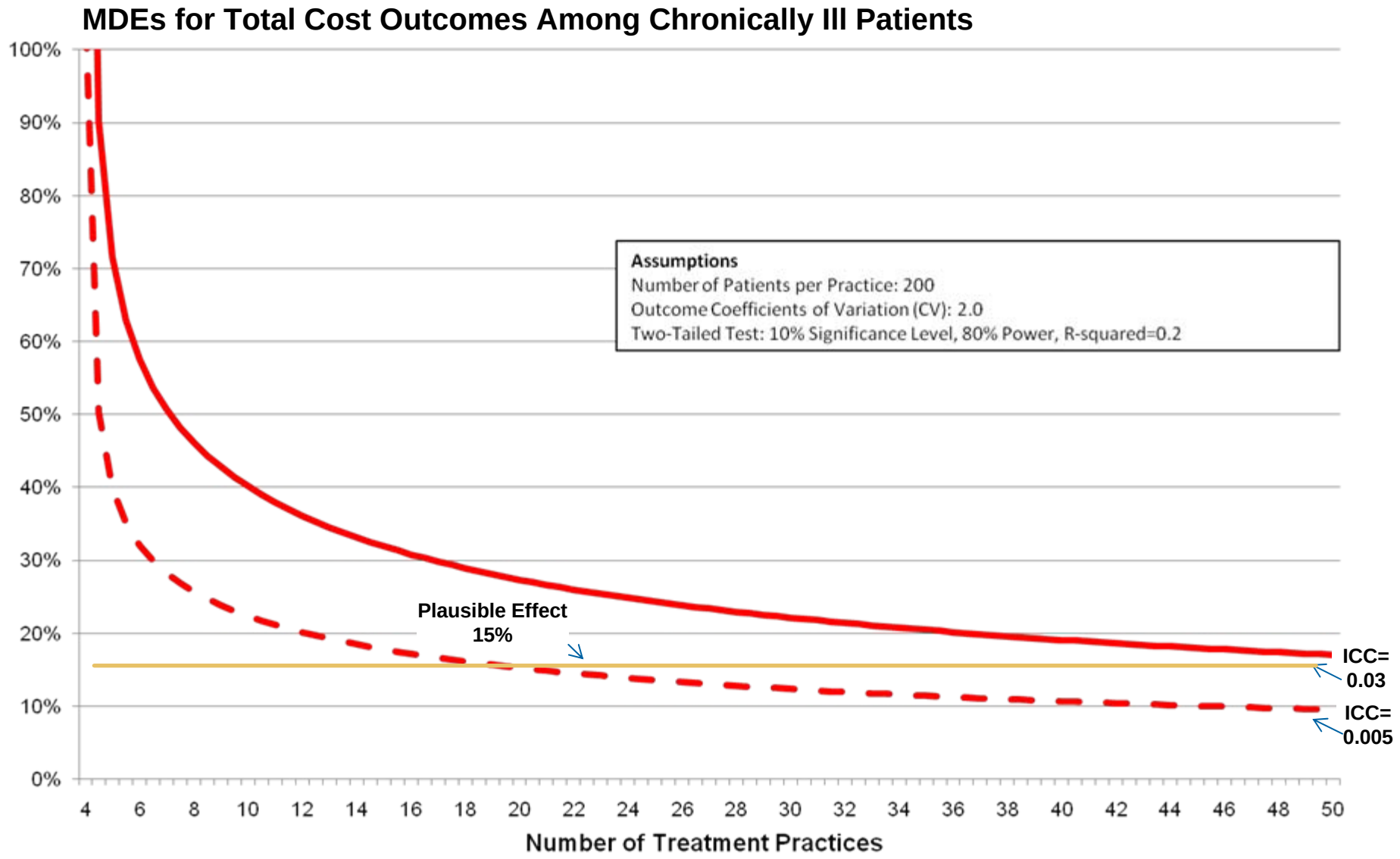
Critical to Include as Many Practices as Possible



Difficult to Measure Impacts on Costs and Service Use Among All Patients

- **Interventions are very unlikely to generate large enough cost reductions among all patients for studies to detect them**
  - **Because there is so much variation in costs, it is hard to distinguish effects of programs from noise**
  - **A 5% reduction in costs across all patients is plausible**
  - **But, even with 50 treatment practices, cost reductions would need to be 18-32% over the full patient panel to be detectable**
  - **With 20 treatment practices, this is larger: 29-52%**
  - **With 10 treatment practices, this is still larger: 43-76%**
  - **MDEs are similar for hospitalizations and even worse for bed days**

Treat All Patients, but Measure Costs Among the Chronically Ill



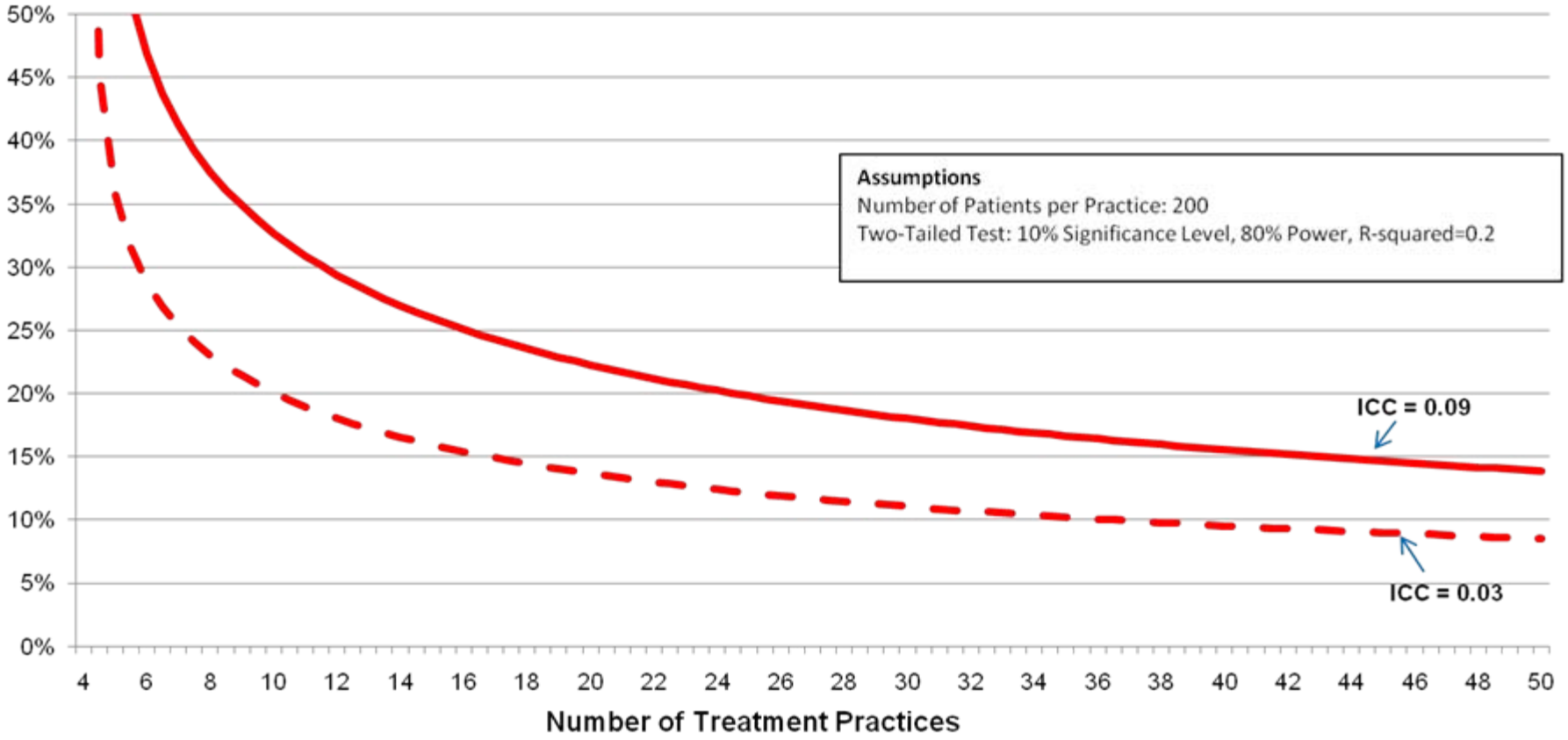
Studies Should Measure Costs Among the Chronically Ill/High-Risk Patients

- A 15% reduction in costs among chronically ill is feasible
- With 50 treatment practices, cost reductions need to be 9-17% (depending on the ICC) to detect them
- With 20 treatment practices, cost reductions need to be 15-27%
- With 10 treatment practices, cost reductions need to be 22-40%
- Savings are both possible (because there are opportunities to alter care) and detectable for chronically ill

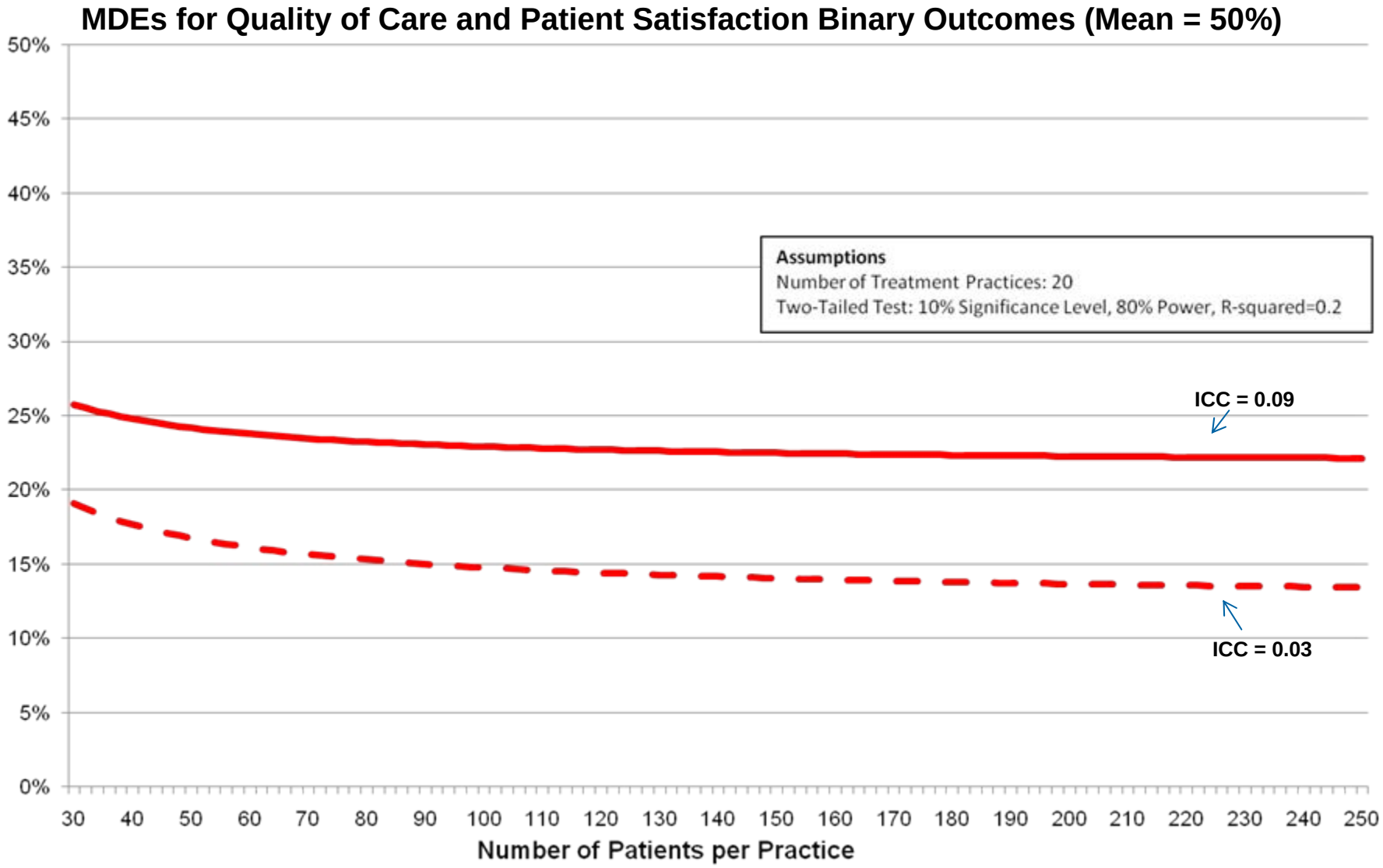
It's Much Easier to Detect Effects for All Patients on Proportions (e.g., Satisfaction, Quality of Care)

- Many interventions have effects of 20% or larger on these outcomes, and CVs are lower.
- With 10 practices, we can detect a roughly 25% change, equivalent to moving the mean from 50% to about 63%. With 20 practices, we can detect effects as small as 20% percent, from 50% to about 60%.

MDEs for Quality of Care and Patient Satisfaction Binary Outcomes (Mean = 50%)



Evaluations Only Need to Follow a Fraction of the Patients in a Practice



Conclusions

- **For costs and hospitalizations:**
  - **Studies will be unlikely to detect effects among all patients, which will likely lead to false negatives.**
  - **Even with 50 treatment practices, the intervention would need to generate effects of 18-32% to detect them.**
    - Effects this large are unlikely
  - **When limited to chronically ill patients, this drops to 9-17%. With 10 practices, the effect is a bit larger (22-40%).**
    - Effects of this size may be feasible because there are opportunities for reductions in costs for chronically ill patients
- **For quality of care and satisfaction outcomes:**
  - **Can measure these outcomes for all patients, but only need to include a fraction of patients at each practice for evaluation**
- **Include more practices rather than more patients**

Points for Future Research

- More important to include more practices than more patients (but including more practices can be operationally difficult and costly)
- Choosing outcome measures with low variances and restricting study populations for outcomes that naturally have high variances (such as costs and hospitalizations) can improve studies by dramatically reducing MDEs
- Adjusting for control variables in regression improves MDEs (we did this here)
- Medical homes can treat all patients, but use high-risk patients to measure cost and use outcomes
  - Payers may want to use risk adjustment to incentivize practices to focus resources on high-risk patients.

Points for Decision Makers

- **The lack of significant findings from underpowered studies (with too few practices) does not imply that the medical home model does not work. These findings may be false negatives.**
- **Investments should not be approved based on analyses that show significant results but are not adjusted for clustering, as such findings are likely to be false positives.**

For More Information

- **Deborah Peikes**
  - **dpeikes@mathematica-mpr.com**
- **Stacy Dale**
  - **sdale@mathematica-mpr.com**
- **Eric Lundquist**
  - **elundquist@mathematica-mpr.com**