Examining the Experimental Designs and Statistical Power of Group Randomized Trials Funded by the Institute of Education Sciences

Jessaca K. Spybrook

A Presentation for the Evaluation Café at Western Michigan University

February 19, 2008

Background

- Evidence-based education
- Randomized trials
- Group randomized trials / Cluster randomized trials

Background

- Institute of Education Sciences (IES)
  - National Center for Education Research (NCER)
  - National Center for Education Evaluation and Regional Assistance (NCEE)
- Produce research that provides reliable evidence on which to base education policy and practice

Background

- NCER
  - Goal 3 Projects – Efficacy and Replication
    - Test effectiveness of intervention under specific conditions
    - ~ $250,000 - $700,000 per year
  - Goal 4 Projects – Effectiveness Evaluations
    - Test effectiveness of intervention under more typical conditions
    - Up to $1.2 million per year

Background

- NCEE
  - Conduct rigorous evaluations of federal programs
  - Contracts not grants
  - At least $1 million per year

Background

- Group randomized trial ≠ Reliable, scientific evidence
- Strong design
- Large enough sample size to conclusively determine whether or not an intervention can improve student outcomes by a specified margin (adequate power)
- Power of 0.80 is usually considered acceptable in social sciences

Background - Terms

- Minimum detectable effect size (MDES) – Smallest effect size that can be detected with power = 0.80
  - Sample size at all levels
  - Intra-class correlation
  - Covariate-outcome correlation
  - Presence and strength of blocking variable

Central Goal of this Study

- Examine the designs and power analyses for the group randomized trials funded by the National Center for Education Research (NCER) and the National Center for Education Evaluation and Regional Assistance (NCEE)

Key Questions

1. What designs do these studies use?

Key Questions

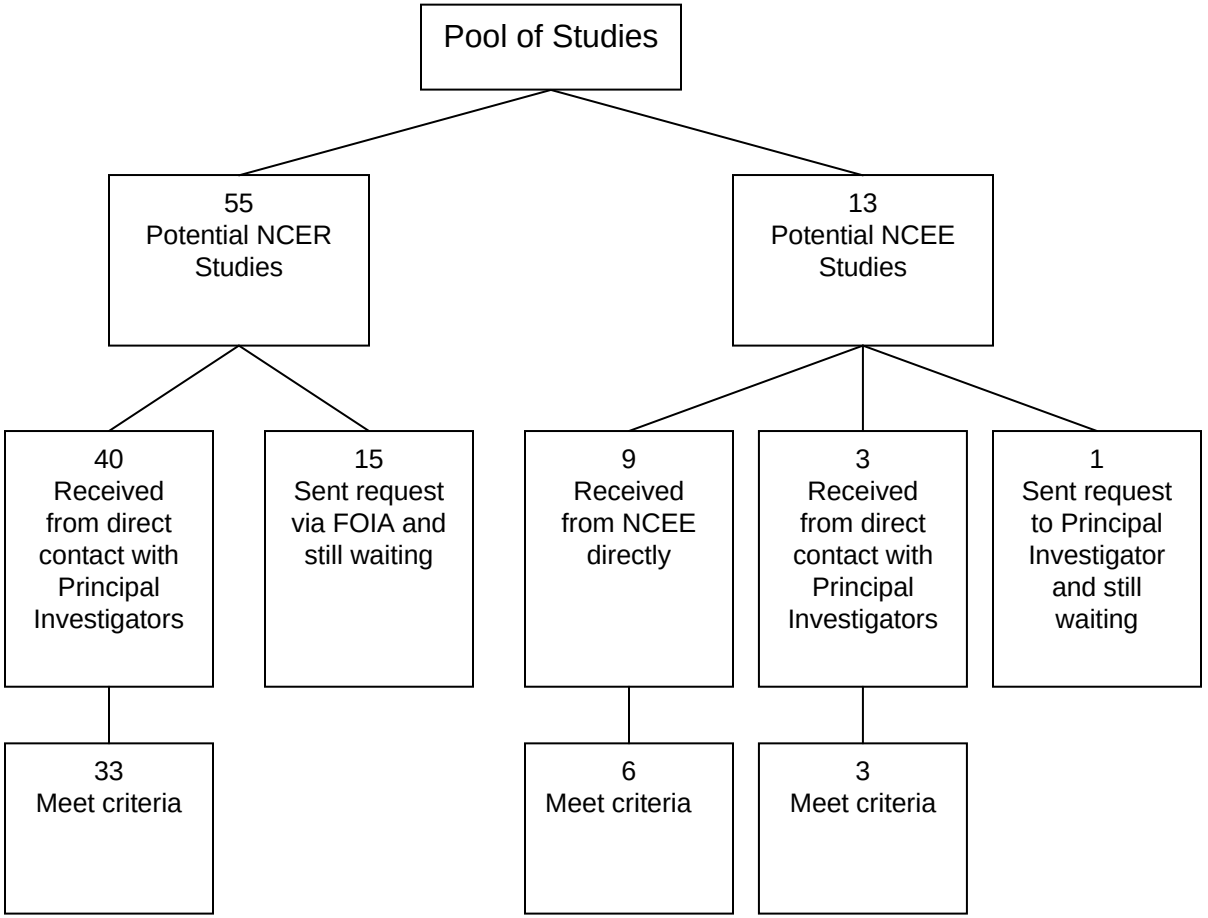
2. Under plausible assumptions about intra-class correlations, covariate-outcome correlations, and explanatory effects of blocking, what are the minimum detectable effect sizes's (MDES) of the studies in the sample?

Key Questions

3. What is the relationship between the MDES stated in the proposal and the MDES under plausible assumptions regarding the design parameters? To the extent that there are discrepancies between the two values, what are the possible sources of the inconsistencies?

   - Is there a power analysis? Is it documented? Does it correspond to the study description?
   - Are the intra-class correlations documented? If so, what are the estimated values?
   - Are covariates included in the power analysis? If so, are the covariate-outcome correlations documented? If so, what are the values?
   - Is blocking included in the description of the study? If so, is blocking included in the power analysis and are the explanatory effects of blocking documented? Is the treatment of the blocks (ie. fixed or random) stated, and if so, is it justified?

Sample


Pool of Studies
55
Potential NCER
Studies
13
Potential NCEE
Studies
40
Received
from direct
contact with
Principal
Investigators
15
Sent request
via FOIA and
still waiting
9
Received
from NCEE
directly
3
Received
from direct
contact with
Principal
Investigators
1
Sent request
to Principal
Investigator
and still
waiting
33
Meet criteria
6
Meet criteria
3
Meet criteria

Sample

	Number of Studies
National Center for Education Research	33
Goal 3 Study	25
Goal 4 Study	8
National Center for Education Evaluation and Regional Assistance	9

Methods

- Classify the study design
- Determine plausible values for design parameters – intra-class correlations, covariate-outcome correlations, explanatory power of blocking
- Calculate the recomputed MDES
- Compare recomputed MDES to stated MDES

Results – Experimental Designs

	Two-Level Cluster Randomized Trial	Three-level Cluster Randomized Trial	Three-level Multi-site cluster randomized trial[a]	Four-Level Multi-site cluster randomized trial
Number of Levels	2	3	3	4
Level of Randomization	2	3	2	3
Blocking?	No	No	Yes	Yes
Number of Studies	5	5	20	11
Example of Nesting	Students, Schools	Students, Classrooms, Schools	Students, Classrooms, Schools	Students, Classroom, Schools, Districts

Results – Experimental Design

Experimental Design	Number of NCER Proposals	Number of NCEE Proposals
Two-Level Cluster Randomized Trial	5	0
Three-Level Cluster Randomized Trial	5	0
Three-Level Multi-site cluster randomized trial	13	7
Four-Level Multi-site cluster randomized trial	9	2

Results - The Recomputed MDES

- Plausible values for ICCs
  - Bloom et al., 1999
  - Schochet, 2005
  - Hedges & Hedberg, 2007
  - Bloom, Richburg-Hayes, & Black, 2007

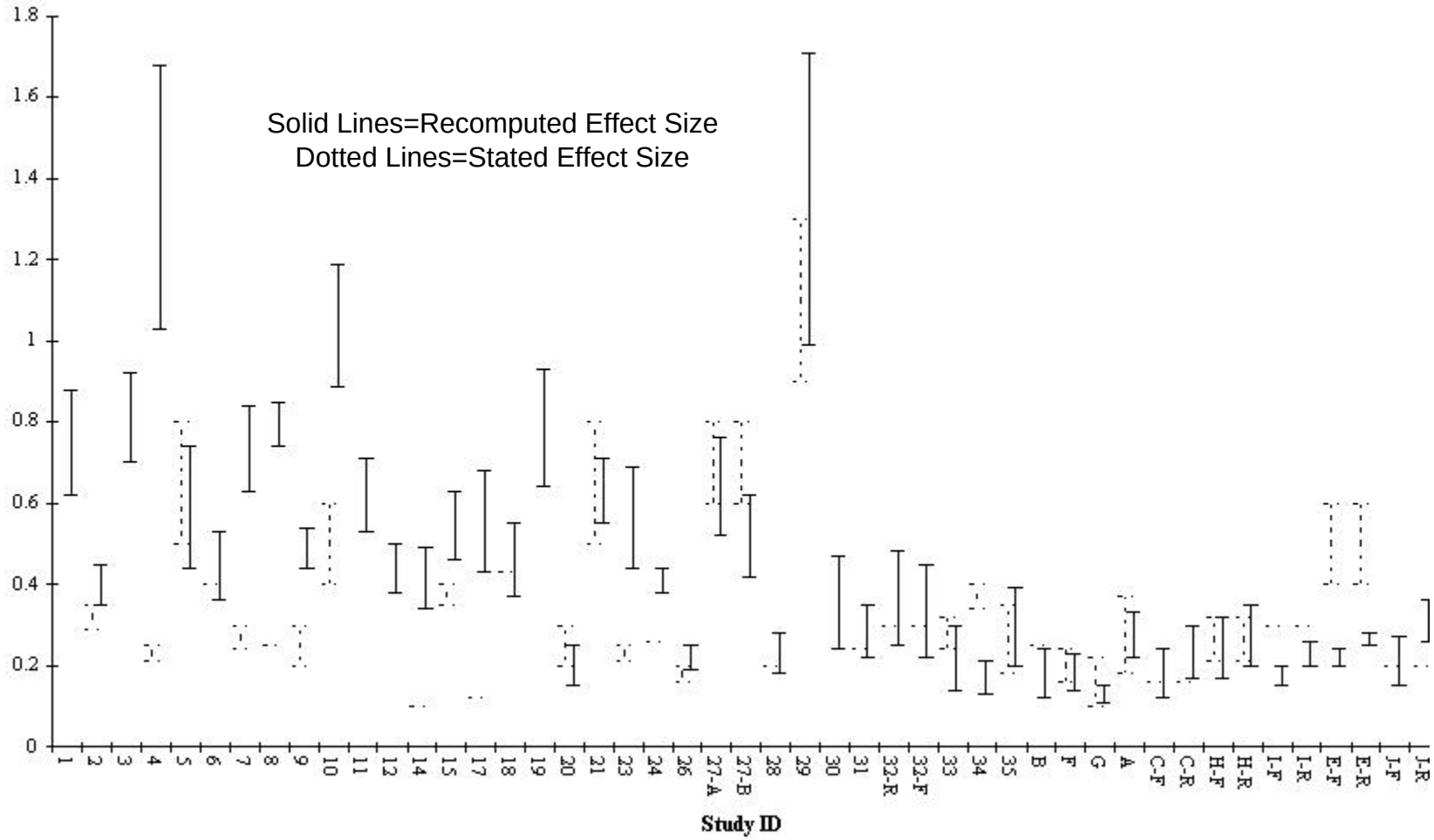
  - Murray & Blitstein, 2003

Results – The Recomputed MDES

- Plausible values for covariate-correlations
  - Bloom, Richburg-Hayes, & Black, 2007

- Plausible values for variance explained by blocking
  - Hedges & Hedberg, 2007

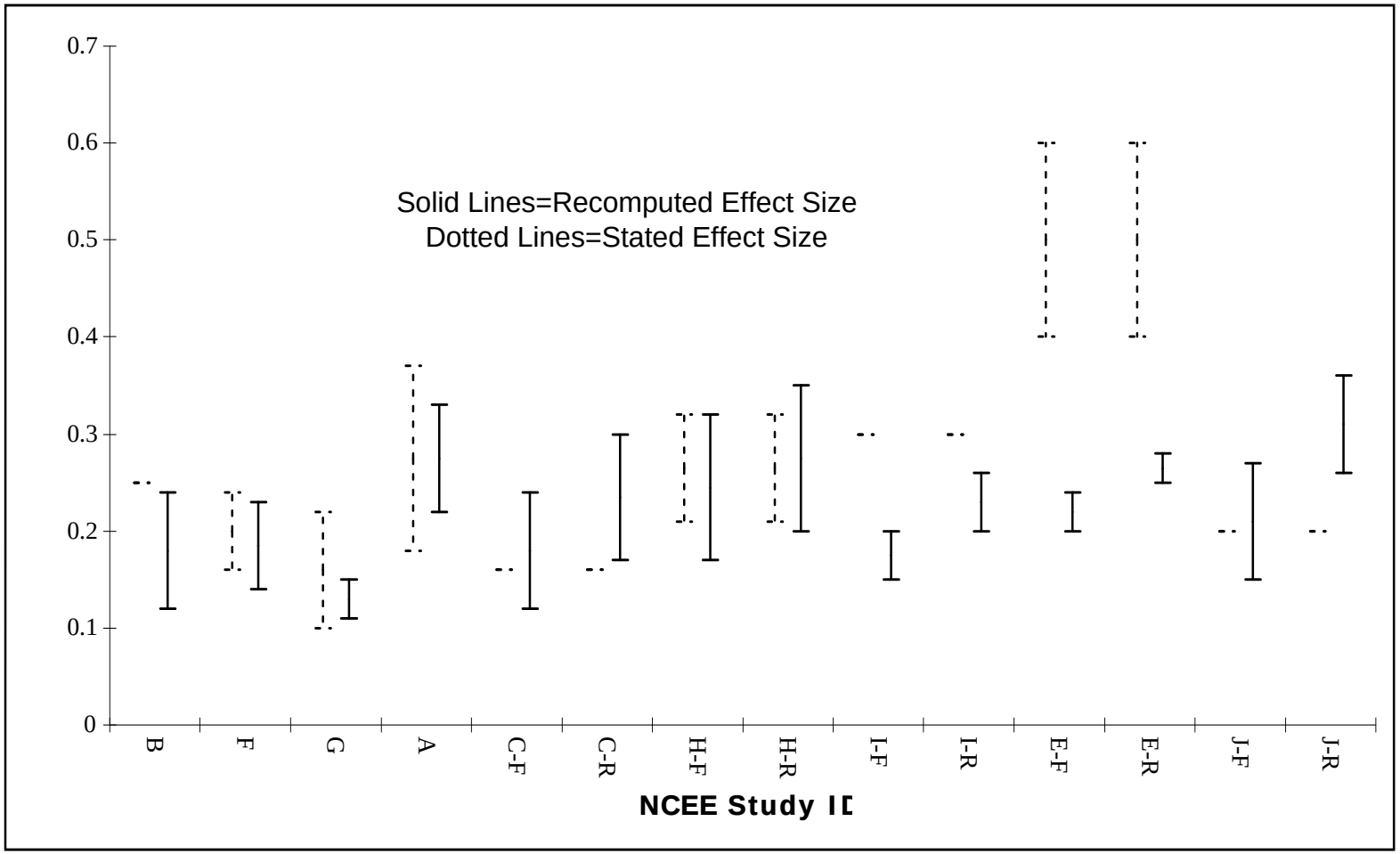
Results – Recomputed and Stated MDES



Results

- Studies 1-24, MDES ranges from 0.40-0.90
  - NCER studies funded in 2002, 2003, 2004
  - Less likely to use a covariate
- Studies 26-J, MDES ranges from 0.18-0.40
  - NCER studies funded in 2005, 2006
  - NCEE studies
  - More likely to use a covariate

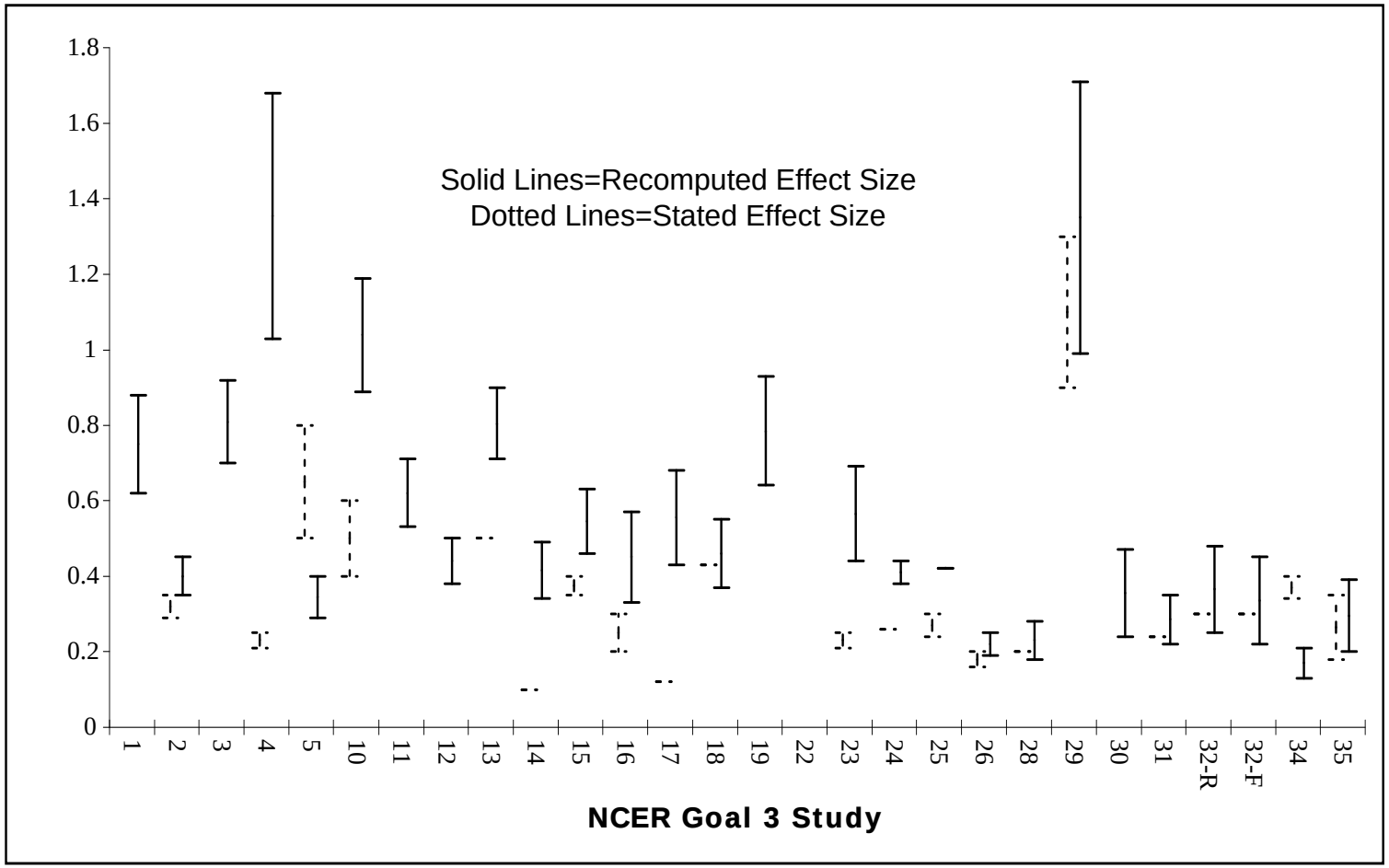
Results - NCEE



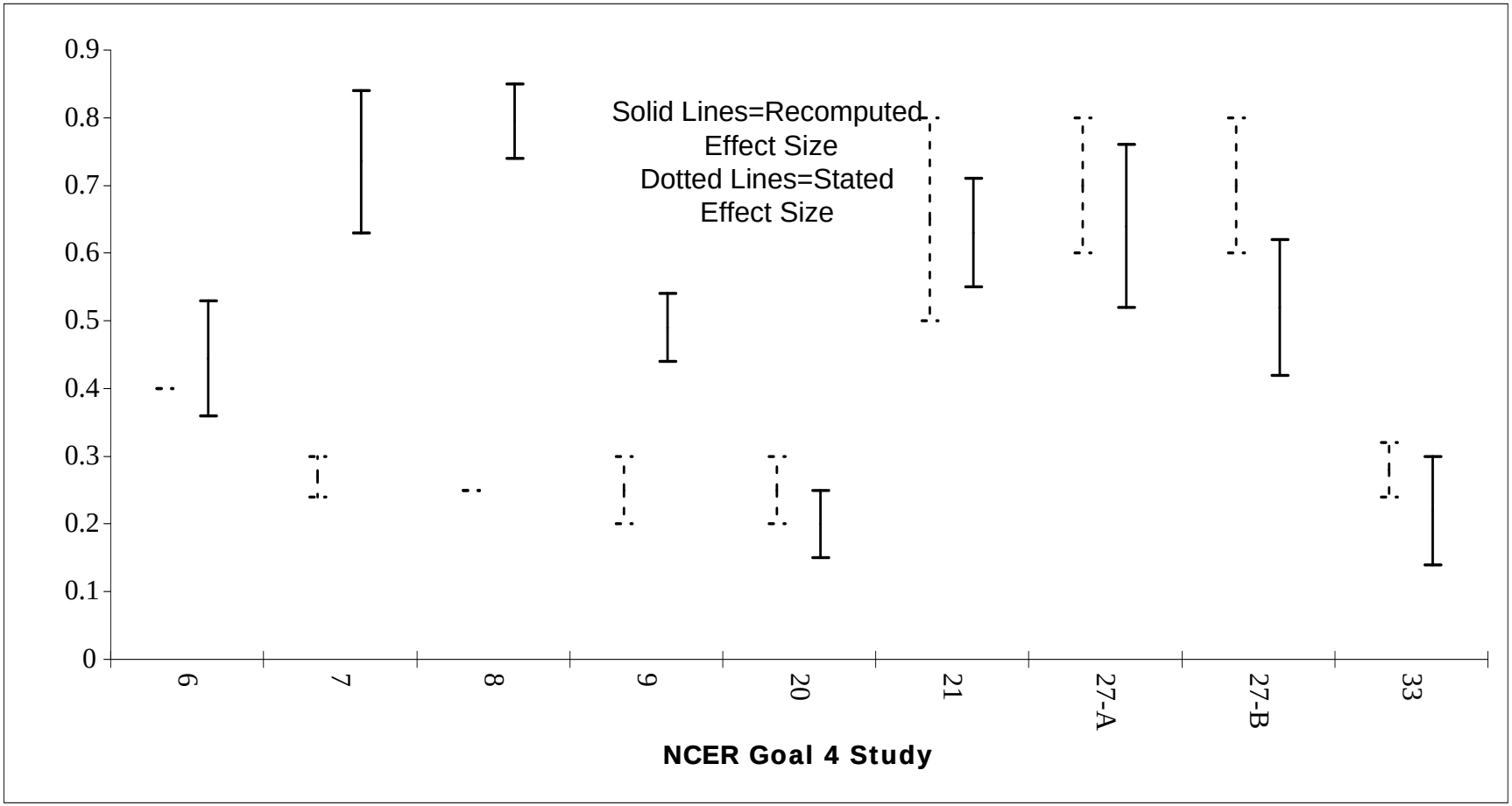
Results - NCEE

- Recomputed MDES ranges from 0.10 – 0.40
- Majority of recomputed and stated MDES are in the same range

Results - NCER



Results - NCER



Results - NCER

- Similar for goal 3 and 4 studies
- Recomputed MDES ranges from 0.18 – 1.70
- Approximately half of the studies have recomputed and stated MDES in the same range

Results – Relationship between stated and expected MDES

	Number of NCER Proposals	Number of NCEE Proposals
MDES within the same range	14	7
Stated MDES < Expected MDES	12	0
Expected MDES < Stated MDES	1	2

The 6 NCER studies without a power analysis are not included.

Results – Details of Power Analyses

	Number of NCER Proposals			Number of NCEE Proposals	
	Same (n=14)	Stated<Recomputed (n=12)	Recomputed<Stated (n=1)	Same (n=7)	Recomputed<Stated (n=2)
Simple statement of power with/without brief citation	6	11	0	0	0
Detailed power analysis with software or documented calculations	8	1	1	7	2
Optimal Design	7	1	1	0	2
Other	1	0	0	7	0

Results – Details of Power Analyses

	Number of NCER Proposals			Number of NCEE Proposals	
	Same (n=15)	Stated<Recomputed (n=11)	Recomputed<Stated (n=1)	Same (n=7)	Recomputed<Stated (n=2)
ICC estimate not included in proposal	4	7	0	2	0
ICC estimate included in proposal	11	4	1	5	2
Academic ICCs					
Within 0.10 to 0.20	7	1	1	4	0
Not within 0.10 to 0.20	3	1	0	1	2
Social or health ICCs					
Within 0.01 to 0.05	0	1	0	0	0
Not within 0.01 to 0.05	1	0	0	0	0

Results – Details of Power Analyses

	Number of NCER Proposals			Number of NCEE Proposals	
	Same (n=15)	Stated<Recomputed (n=11)	Recomputed<Stated (n=1)	Same (n=7)	Recomputed<Stated (n=2)
No covariate	6	6	0	1	0
Covariate mentioned not documented	5	3	1	2	1
Covariate documented	4	2	0	4	1
0.01-0.30	0	1	0	0	0
0.31-0.50	0	0	0	1	0
0.51-0.70	4	1	0	1	1
0.71-0.99	0	0	0	2	0

Results – Details of Power Analyses

	Number of NCER Proposals			Number of NCEE Proposals	
	Same (n=14)	Stated<Recomputed (n=7)	Recomputed<Stated (n=1)	Same (n=7)	Recomputed<Stated (n=2)
Blocking included in the description	14	7	1	7	2
Blocking included in the power analysis	0	0	1	3	2
Include explanatory power of blocking	0	0	0	3	0
Explicitly treat blocks as fixed effects	0	0	0	1	0
Explicitly treat blocks as random effects	0	0	1	0	0
Specify the effect size variability	0	0	1	0	1

Conclusions

- Blocked designs are most common
  - Good for precision

- NCEE studies tend to have smaller MDES
  - Differences in funding
  - Differences in methodological guidelines

Conclusions

- NCEE studies tend to be more accurate
  - Training
- Growth is evident in accuracy and precision of NCER studies
  - More precise over time (use of covariates, blocked designs)
  - More accurate over time

Limitations

- Study proposals as data
- Use of original funded proposal