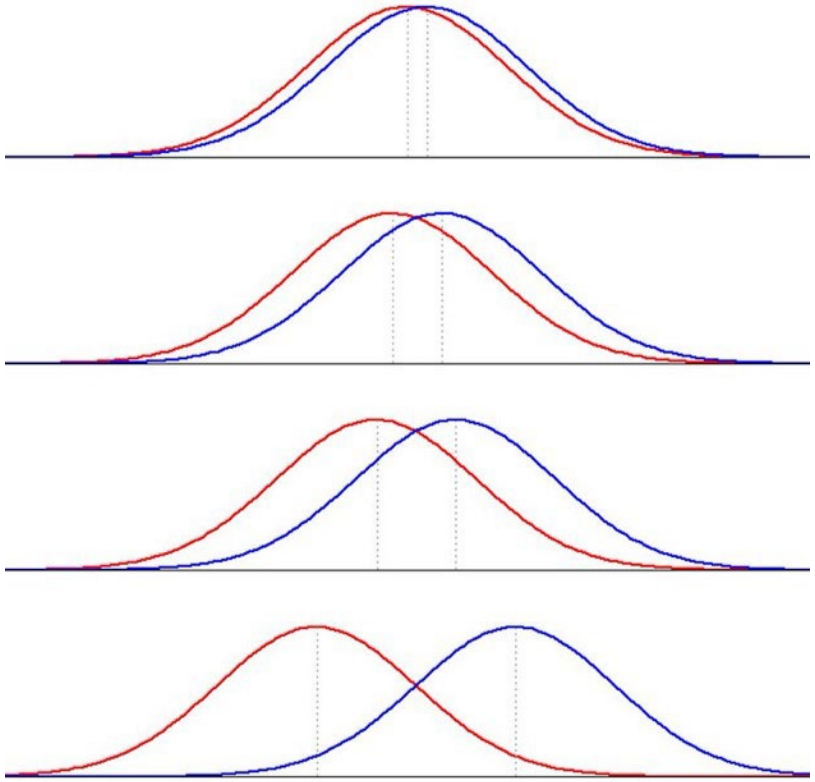
BEHAVIORAL SCIENCE & POLICY INSTITUTE

Interpreting Effect Sizes in Psychology:

Application to Social Media and Youth Mental Health



Presentation for the National Academies of Sciences, Engineering and Medicine

DR. DAVID S. YEAGER

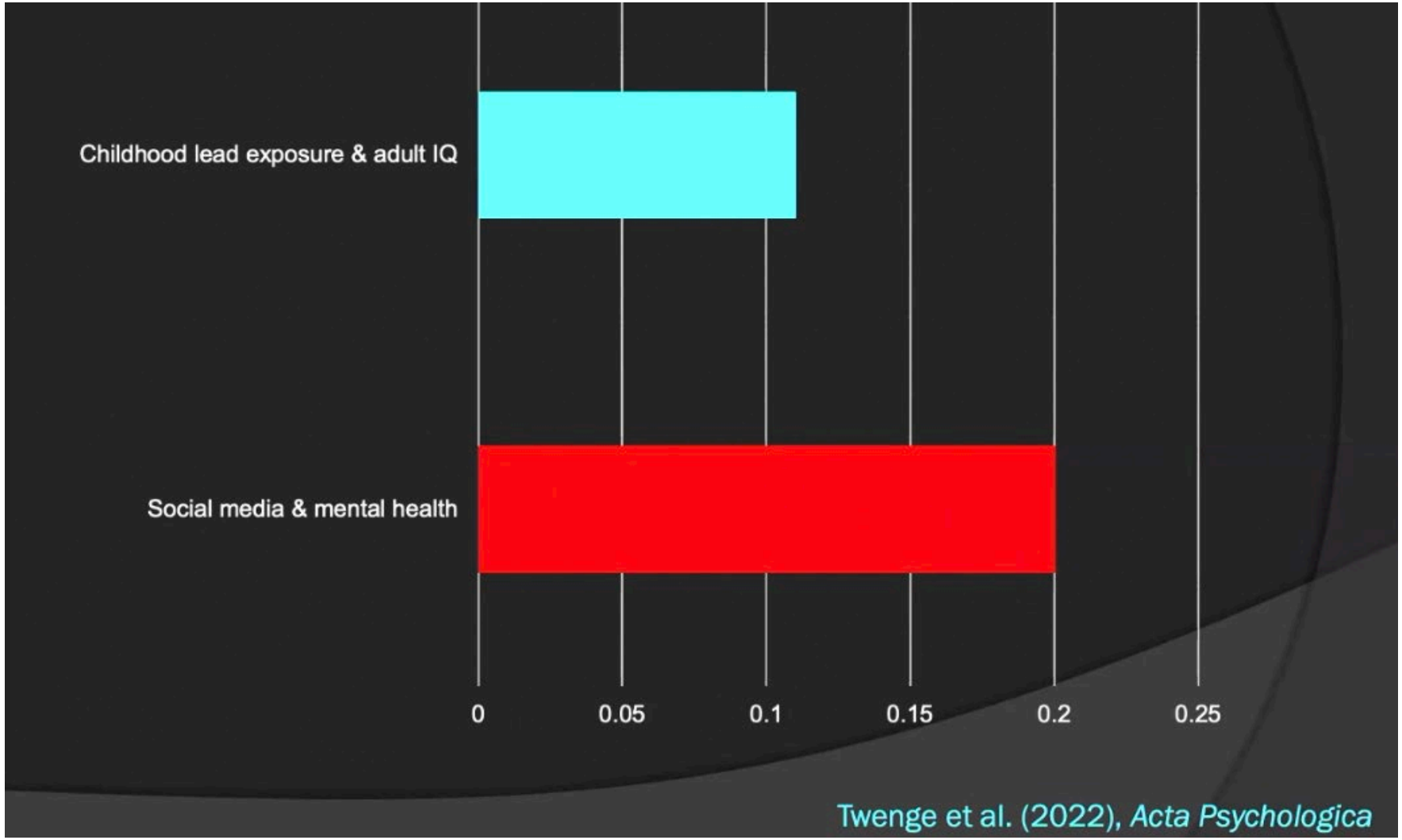
Associate Professor of Psychology and Co-Founder,
Texas Behavioral Science and Policy Institute,
The University of Texas at Austin

TEXAS The University of Texas at Austin

NSF (1761179 and 2201928; PI: Yeager), NICHD (R01HD084772, PI: Yeager; P2CHD042849); William and Melinda Gates Foundation (INV-047751 and INV-004519); All opinions are my own, not funding agencies'.

BEHAVIORAL SCIENCE & POLICY INSTITUTE

What's Wrong With this Picture?



BEHAVIORAL SCIENCE & POLICY INSTITUTE

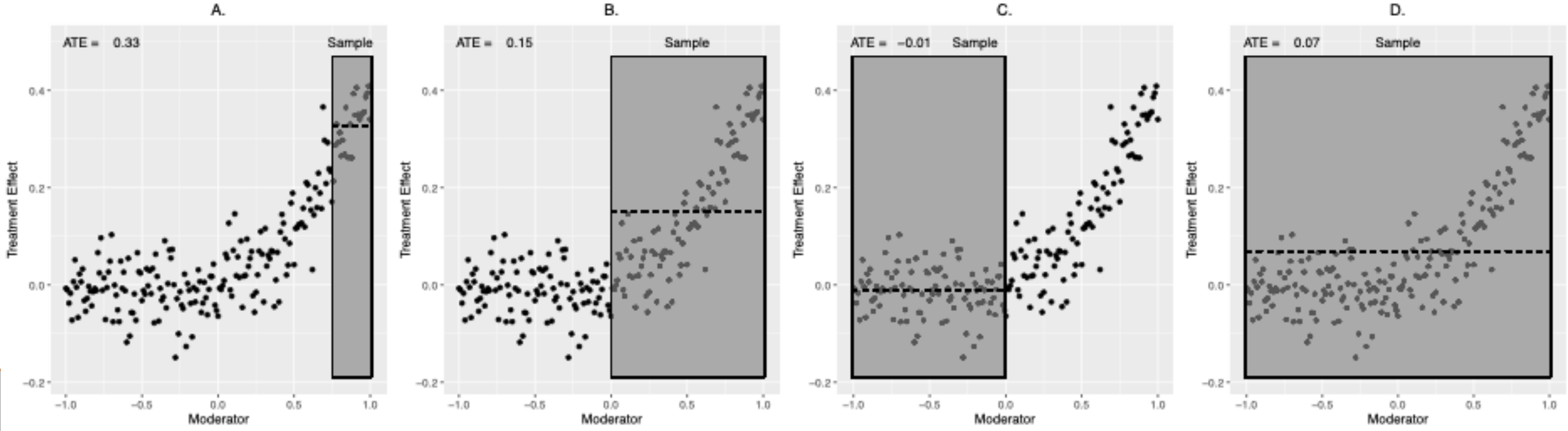
"The Effect" is Misleading

Tipton, … Yeager (in press), *Psychological Bulletin;* Bryan et al. (2021), *NHB*

Elizabeth Tipton,
Northwestern University

Individual studies and meta-analyses often describe "the effect" of X on Y

- But effects are heterogeneous across groups and dependent on many methodological choices



BEHAVIORAL SCIENCE & POLICY INSTITUTE

Three Key Considerations

Kraft (2020), *Educational Researcher*

1. Results From Correlational Studies Presented as Effect Sizes Are Not Causal Effects

2. The Magnitude of Effect Sizes Depends on What, When, and How Outcomes Are Measured

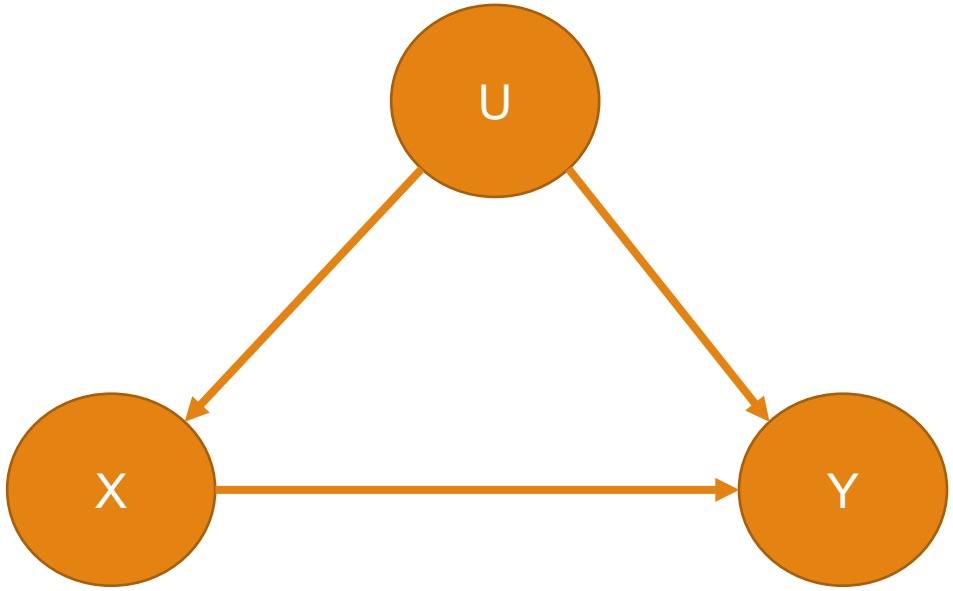
3. Subjective Decisions About Research Design and Analyses Influence Effect Sizes

4. *Costs Matter for Evaluating the Policy Relevance of Effect Sizes*

5. *Scalability Matters for Evaluating the Policy Relevance of Effect Sizes*

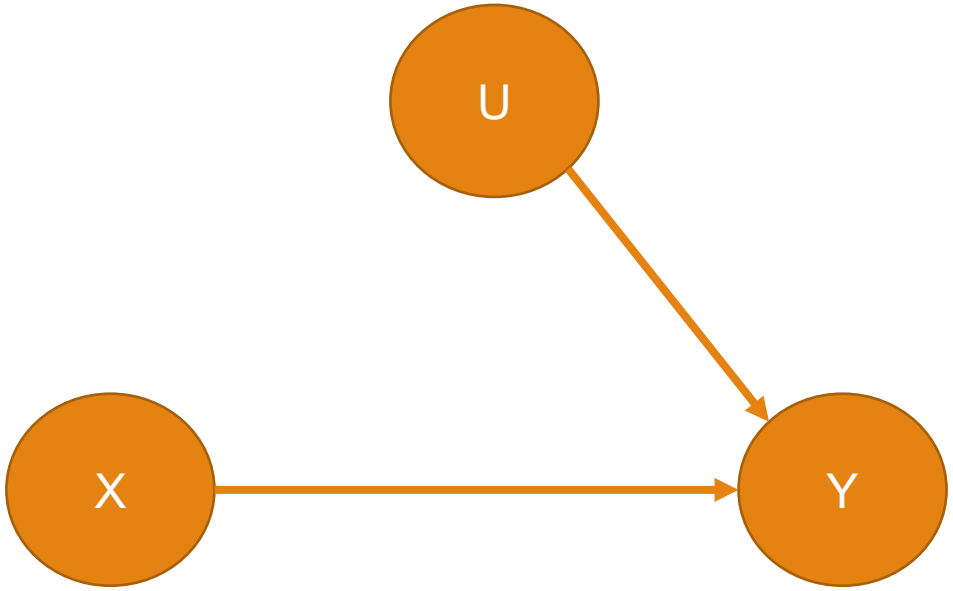
1. Correlational vs. Causal

Observational Study



$ES = r_{xy} + r_{xu} * r_{yu}$

Experimental Study



$ES = r_{xy}$

Bottom line: Correlational studies usually over-estimate effects, but meta-analyses often average the two together or compare to causal ESs

1. Correlational vs. Causal (pt2)

The “treatment” is not comparable across study designs

- Correlational: Examining the entire range of social media use
  - E.g. 1 hr vs. 8 hrs
  - Or “thresholds” – e.g. <4hrs vs. >4 hrs (each with a wide range)
- Experimental: Examining only the achievable change in behavior,
  - E.g. 10 min, 20 min
  - Probably only for a subgroup at a certain part of the distribution
  - Absent strong treatment or policy, probably won’t change someone at 8 hrs into someone using 1hr

Bottom line: Even in the absence of confounding correlational effects are not representing the realistic causal effects of an exogeneous shock

2. ESs Depend on Outcomes

1. "Interim" or "proxy" outcomes > end-of-the-line, accumulated outcomes
   - Self-esteem vs. clinical diagnoses

2. Immediate post-test > lagged / delayed outcomes
   - Fadeout (Bailey et al., 2021, *PSPI*)

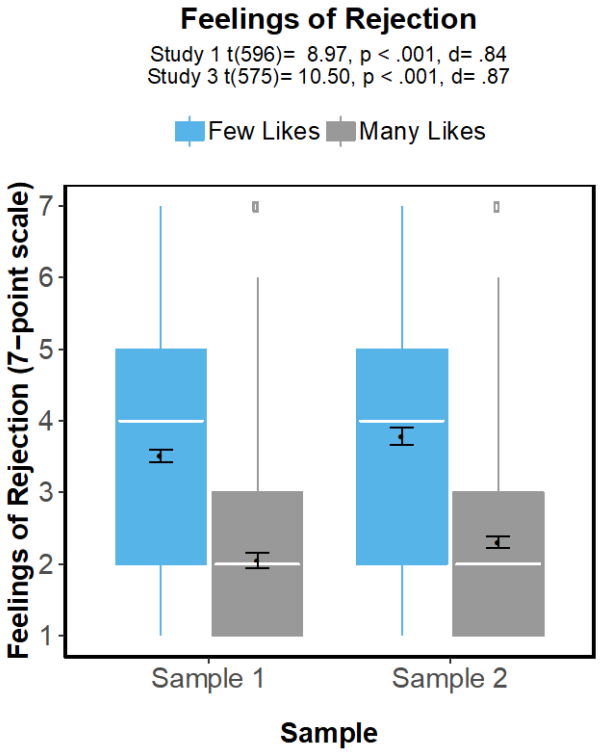
3. Reliable outcomes > Unreliable outcomes
   - Key point: reliability of total test scores > reliability of change scores, but change scores are usually the source of our effect sizes

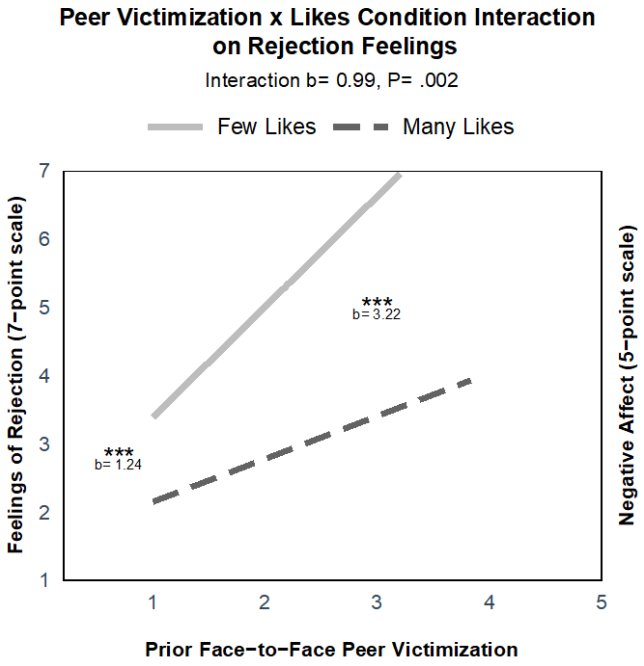
4. The "multifinality problem" (treatments with diffuse, modest effects)

2. ESs Depend on Outcomes

ES of "likes" on well-being: d = .84

Do we believe that?





BEHAVIORAL SCIENCE & POLICY INSTITUTE

3. Study Design

1. Sample (e.g. average effects, subgroup effects)
   - Targeted vs. universal and Rose's paradox
   - Depends critically on sample representativeness (and recruitment)

2. Standard deviations of outcomes
   - Measurement and sample considerations (e.g. targeted/indicated)

3. Treatment vs. control contrasts (counter-factual)
   - Schizophrenia "decline effects;" Kemp et al. (2010)

4. Type of treatment (ITT, TOT) and take-up rates of treatment

BEHAVIORAL SCIENCE & POLICY INSTITUTE

Three Key Considerations

Kraft (2020), *Educational Researcher*

1. Results From Correlational Studies Presented as Effect Sizes Are Not Causal Effects

2. The Magnitude of Effect Sizes Depends on What, When, and How Outcomes Are Measured

3. Subjective Decisions About Research Design and Analyses Influence Effect Sizes

4. *Costs Matter for Evaluating the Policy Relevance of Effect Sizes*

5. *Scalability Matters for Evaluating the Policy Relevance of Effect Sizes*

BEHAVIORAL SCIENCE & POLICY INSTITUTE

Recommendations

Kraft (2020), *Educational Researcher*

1. Conduct heterogeneity-attuned meta-analyses with *best evidence synthesis*

2. When possible, compare effect sizes to “benchmark” studies
   - E.g. in education, class size reduction

3. Locate effect sizes on the distribution of effects that have been observed in similarly-designed studies