

Calibrating effect size

Philip J. Cook

Duke University

Jerry Lee Symposium 4/24/12

“big” or “little”?

- **Hypothesis test** provides very limited info
- For policy evaluation, should report the **estimated effect (point and CI)**
- But how do we judge whether that effect is “big” enough to be of interest?
- One answer: Is the effect large enough to suggest that the associated policy innovation is **worthwhile** (serves the public interest)?
- **Sometimes a cost-benefit analysis provides guidance**

Alternative calibrations in education research

Outcome often in form of **standardized test score**

Estimated effect ÷ SD of population

Cohen (1988) sets **arbitrary guidelines** for small, medium, large

Bloom et al. (2009):

Compare effect size with av. annual growth, or with black-white gap

Easier in crime research?: **Outcomes in meaningful natural units** (eg, robberies prevented)

Still useful to have some **standard of comparison**

Cost-effectiveness analysis

C-E has **broad application**

(1) Education: **STAR treatment effect** =

.056 / \$1,000/ pupil (by grade 3)

(2) Health: Compare prevention to treatment for HIV/AIDS based on **\$\$/DALY** saved (Canning 2006)

In general, C-E allows **policy-relevant comparisons** among different interventions in **same domain**

C-E for Crime Interventions

1. **DNA processing** for burglary investigations
(Roman et al. 2009)

- Arrest rate increases: 10% to 22%
- More useful: **\$14,000/added arrest**

2. **LA Business Improvement Districts**

(Cook & MacDonald 2011)

- Reduce crime count by 28 per reporting district
- More useful: **3.4 crimes/ \$10,000 on security**

C-E for Crime (cont.)

3. CBT for juvenile detention facility

(Univ. of Chicago Crime Lab, 2012)

Reduces 1-year return rate by 9 % (66 to 57)

Yawn.

But for just \$150/kid (or ~ \$1,600/ success)

Lesson: “small” effects may emerge as worthwhile

C-E when multiple outcomes?

C-E allows comparison of interventions that share a single outcome measure

1. Health: use index (DALY, QALY) for combining deaths and spells of illness
2. Crime: simple count or weighted index?
3. But what about Perry Pre-school? (crime, education, earnings, etc.)
4. And what if intervention reduces cost? (Boot camp: Bierie 2009)

CBA as one answer

CBA: requires monetizing both costs and benefits

Intervention is worthwhile (from public interest perspective) if

sum of benefits > costs

Note: Often true that monetizing benefits is controversial and adds uncertainty.

Power calculations and CBA

Ideally, the desired MDE determines n

For a low-cost intervention, a “small” effect is worthwhile by CBA standard.

Then: either large n , or high risk of Type II error

So careful reporting is crucial

Uncertainty about effect sizes

Common practice: Limit benefit estimate to **point estimate** of treatment effect

Better: Monetize 95% CI limits as well

But: **What if treatment effect is n.s., but
EV (benefit) > cost?**

Then intervention is arguably worthwhile

For your consideration

What if we are able to confidently rule out **perverse** effects of the intervention?

Example: **Brady Law's effect on suicides**
(Ludwig and Cook, 2000)

Then **truncate** the probability distribution for the estimate of effect at zero?

Odd Result:

almost certainly deemed worthwhile

Concluding thoughts

CE and CBA provide policy-oriented calibration for judging effect sizes

“Small” effects may be worthwhile, and should not be automatically ignored

CBA framework should inform experimental design

CBA framework privileges Expected Value even if effect is n.s.