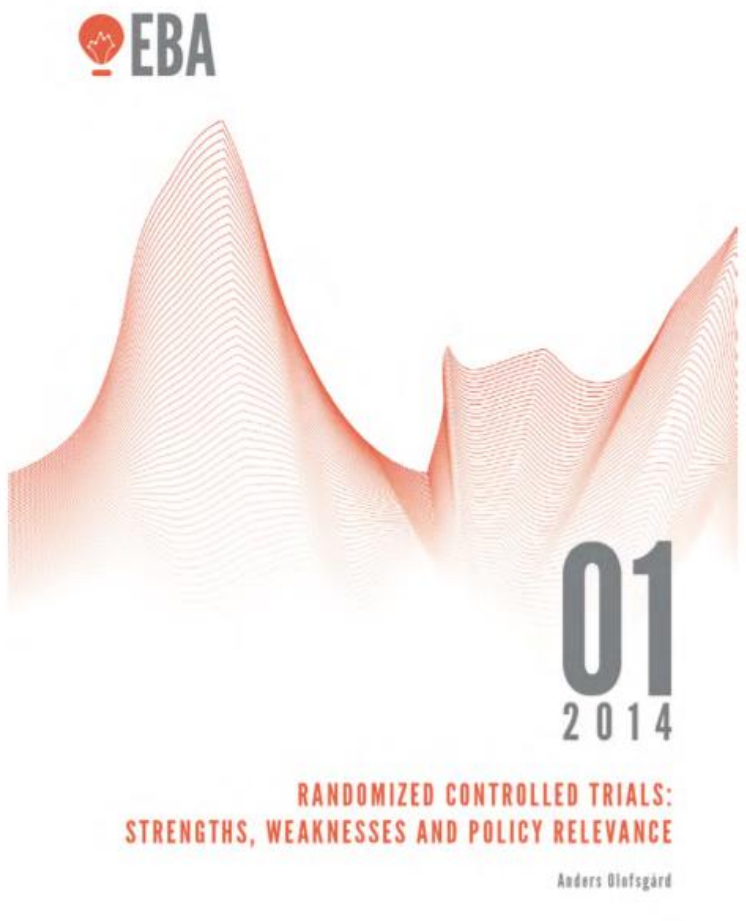
Randomized Controlled Trials

STRENGTHS, WEAKNESSES AND POLICY RELEVANCE

EBA Report 01:2014

3 parts

1. Short introduction to method with some examples.
2. Discussion of strengths and weaknesses.
3. Discussion of the method's application in Swedish aid policy.



Ambition today

1. A brief and simple explanation of the main logic of the method.
   I. Long history, dating back to at least 19th century.
   II. Often associated with drug tests in medical science.
   III. Increasingly important among social scientists, not least development economists.
2. Discuss selective strengths and limitations.
   I. Scientific
   II. Practical/ethical
   III. Resources/Finances

Example: School buildings -> student learning

Q. Why can't we just compare national test results of students in villages with proper schools with those in villages without proper schools?

A. Test results depend on more than buildings, and villages with and without proper school buildings may differ in systematic ways along these other dimensions.

I. Better off villages may be more likely to have proper school buildings for both economic and political reasons, and children in better of households typically do better in school (positive selection/bias).
II. Schools with proper buildings may be able to attract better teachers and more motivated students if mobility is possible (positive selection/bias).
III. Donors/governments may target particularly poor villages with school building investments (negative selection/bias)

Q. But what if we could compare before and after between these two types of schools (diff-in-diff)?

A. Helps quite a bit, but time trends may also vary systematically with income and effect of buildings may differ depending on income (heterogeneous treatment effects).

A typical RCT on school buildings

1. Select a sample of roughly similar villages (stratification) without proper school buildings. Then randomly assign villages as treatment (t) or control (c) villages, and build schools in treatment villages only.
2. Prior to actually building schools, make sure to have a baseline of student achievement in both t and c villages, and information on potential confounding variables. Create a balance table to test whether randomization has "worked" (first and sometimes second moments). Also helps identify variables to control for in a multivariate regression format.
3. After appropriate time, conduct an end line of student achievement and confounding variables. In-between monitoring to make sure the intervention is on track also often needed.
4. The average treatment effect (ATE) is estimated as the difference between the average improvement in the treatment schools and the average improvement in the control schools. The effect is interpreted as causal and applies to the underlying population.

S & W: ”Scientific”

STRENGTHS

1. Internal validity: We can feel relatively sure we are capturing a causal effect of the intervention.
2. Control and discipline: requires early involvement, instruments can be designed to make sure relevant information is collected, measurable outcome variables clearly defined, objectives clearly defined.

WEAKNESSES

1. External validity (place, provider, scale).
2. Case selection (not necessary, not unique to RCT).
3. Narrow focus: Often focused on one outcome parameter, policy makers may be interested in many additional factors.

S & W: Practical/ethical

STRENGTHS

1. Logic easy to understand and procedures relatively standardized.

2. Ethics: reduced risk of political capture, helps policy makers make informed decisions (less “experimental”, not more).

WEAKNESSES

1. Many types of interventions cannot reasonably be randomized.

2. Typically requires early involvement, difficult as an afterthought.

3. Ethics: if budget is limited and there are good reasons to believe the intervention has particularly good effects for certain groups.

S & W: Resources/Finances

STRENGTHS

1. The big cost is rarely evaluation, but the intervention, so if a better method of evaluation can lead to better interventions then it can save money.

2. It's cheaper than the Large Hadron Collider….

WEAKNESSES

1. Usually takes long time to get results.

2. Often quite expensive.

3. Certain quantitative skills are necessary.

Bottom-line: Use it, but selectively

1. Few credible previous evaluations are available so the value added of new information is high (be wary of external validity though).

2. The intervention can be randomized at reasonable cost.

3. The intervention is planned to be scaled up (terminated) if the randomized trial shows satisfactory (un-satisfactory) impact.

BUT!!! It can only be one of several tools in the toolbox.