Challenge Ⅰ
The summary of
Understanding and misunderstanding randomized controlled trials

Jingfang Liu (15320171151900)
&
Xingyue Yu (15320171151888)

Department of Economics, School of Economics,
Xiamen University

Randomized controlled trials (RCTs) are regarded as the ideal methodology for causal inference; RCTs are perceived to yield causal inferences and estimates of average treatment effects (ATEs) that are more reliable and more credible than those from any other empirical method.

They are taken to be largely exempt from the myriad problems that characterize observational studies, to 1)require minimal substantive assumptions, 2)little or no prior information, 3)and to be largely independent of "expert" knowledge. 4) They are also sometimes felt to be more resistant to researcher and publisher degrees of freedom.

This paper argue that any special status for RCTs is unwarranted. Which method is most yield a good causal inference depends on what we are trying to discover as well as on what is already know. **When little prior knowledge is available, no method is likely to yield well supported conclusions**.

This paper presents two sets of arguments:

1) An enquiry into the idea that ATEs estimated from RCTs are likely to be closer to the truth than those estimated in other ways.
2) Exploring how to use the results of RCTs once we have them.

Part Ⅰ

Give the question: RCTs give the good estimate of ATE?

1. The first argument:

1.1 Do RCTs give good estimates of average treatment effects?

This paper use a schematic linear causal model to think about the estimation of treatment effects:

$$Y_i = \beta_i T_i + \sum_{j=1}^{J} \gamma_j x_{ij} \qquad (1)$$

Where, Y_i is the outcome for unit i, Y_i is a dichotomous (1, 0) treatment dummy indicating whether or not i is treated, and β_i is the individual treatment effect of the treatment on i. The x's are observed or unobserved other linear causes of the outcome, and we suppose that (1) captures a minimal set of causes of Yi sufficient to fix its value. J may be (very) large.

By inspection of (1), the differences between the two outcomes, $Y_{i1} - Y_{i0}$, are the individual treatment effects, β_i which are typically different for different units. while we cannot observe the individual treatment effects, we can observe their mean. The difference in means is an unbiased estimator of the mean treatment effect., the equation as follows

$$\bar{Y}_1 - \bar{Y}_0 = \overline{\beta_1} + \sum_{j=1}^{J} \gamma_j \left(\bar{x}_{1ij} - \bar{x}_{0ij}\right) = \bar{\beta}_1 + (\bar{S}_1 - \bar{S}_0) \qquad (2)$$

Where, the second term or error term, which is the sum of the net average balance of other causes across the two groups.

When the sum of their net differences $\bar{S}_1 - \bar{S}_0$ is zero; this is the case of perfect balance.

1.2 How do we get balance, or something close to it?

1) When there is much prior knowledge of the other causes, the experimenter has a good chance to control
2) Failing such knowledge and control, an alternative is matching,

1.3What does randomization do?

Suppose there is an assumption of **post-randomization**, randomization provides orthogonality of the treatment to the other causes, the average of the estimated ATEs will be the true ATE in the trial sample. Conditional on the caveat that no post-randomization correlation with covariates occurs, randomization provides the basis for calculating the size of the error.

1.4 conclusion about first argument

Here are some questions is talked about:

1) We want to make the second term as close to 0 as possible, and we do not know the size of this error term, and there is nothing in randomization that limits its size though, it will tend to be smaller in larger samples. And conditional on the caveat that no post-randomization correlation with covariates occurs, randomization provides the basis for calculating the size of the error.
2) Some misunderstanding is often seen that it is common to treat the ATE from an RCT as if it were the truth, not just in the trial sample but more generally. The false belief in automatic precision suggests that we need pay no attention to the other causes in (1) or (2). In reality, the causality that is being attributed to the treatment might, in act, be coming from an imbalance in some other cause in our particular trial; limiting this requires serious thought about possible covariates
3) Supposing that no post-randomization correlations are introduced, unbiasedness is guaranteed by the randomization, whatever the test shows, and the test is not informative about the balance that would lead to precision. Accepting some bias in exchange for greater precision will often make sense, though it certainly undermines any gold standard argument that relies on unbiasedness without consideration of precision.
4) The randomization is required. several people that with different priors may be involved in an investigation and individual priors may be unreliable because of "vagueness and temptation to self-deception," defects that randomization may alleviate, or at least evade.
5) Credible inference is essential in any argument for RCTs to guarantee the quality of these estimates.
6) It is of great importance to note that randomization, by itself, is not sufficient to guarantee unbiasedness if post-randomization differences are permitted to affect the two groups.

2. The second argument:

2.1 How to use the findings?

Establishing the applicability of results is as important as constructing the trials. The argument of the primacy of internal validity is sometimes incorrectly taken to imply that the results of RCT will automatically apply elsewhere or correctly expected to be

invariant across settings. The approach to use RCT results is based on the purpose ,the way and the hypotheses designed before the trial.

The purpose that RCT results can serve them is distinguished as: simple extrapolation and generalization, drawing lessons about the population enrolled in the trial, extrapolation with adjustment, estimating what happens if we scale up, predicting the results of treatment on the individual and building and testing theory.

2.2 different purposes in which RCT served

1) Simple extrapolation and generalization

The results from a well-conducted, informative, and potentially useful RCT will not apply else where in any simple way. In some cases RCT may useless in simple generalization or extrapolation, however its results can be incorporated into a network of evidence and hypotheses that tell us more valuable information.

Establishing causality does not guarantee that the causal relation will hold in some new case, let along in general.

However, that is not to say that extrapolation is never reasonable, the point is that evidence from RCTs is not automatically simply generalizable, and its internal validity does not provide it with invariance across context.

2) Support factors in system design

Support factors is defined as factors unnecessary but sufficient for a contribution to the ATE outcome. The value of the ATE depends on the distribution of the values of the 'support factors'

$$Y_i = \beta_i T_i + \sum_{j=1}^{J} \gamma_j x_{ij} = \theta(w_i) T_i + \sum_{j=1}^{J} \gamma_j x_{ij}$$

Where the function $\theta(.)$ controls how a k-vector w_i of k 'support factors' affects individual i's treatment effect β_i.

Two populations will have the same ATE if and only if they have the same average for the net effect of the support factors necessary for the treatment to work.

In different situations, support factors will operate differently. The point is not only to find out the different distribution of support factors but also to search for what those support factors differ which needs knowledge of casual structure.

System design which will generate causal relations depend on understanding of interacting parts that support causal process. It is important to understand the support factors behind the system design.

3) No extrapolation or generalization required

Whether we require extrapolation or generalization is depend on what we want to learn. If RCT provides evidence for the general proposition or the consequence of the proposition, there is no need to require the generalization, nor is extrapolation required when an RCT is used for evaluation.

When the parameter of interest is the ATE in a well-defined population, the trial sample is itself a random sample, in this case the sample average treatment effect is an unbiased estimator of the population average treatment effect

4) Reweighting and stratifying

We should understand RCTs results under specified circumstances and different factors works in different circumstances.

In systematic way, this is done by having multiple treatment within the same trial so that it is possible to link outcomes to various combinations of treatments. In statistical way, reweighting approaches are designed to deal with treatments effects vary systematically with variations in the support factors, though these methods are often not applicable because reweighting needs some necessary requirements.

Subgroup analysis may provide information in generalization or extrapolation. To

use techniques for reweighting and stratifying. We will need to know not only the results itself but also external information to settle on characterization of the population.

5) Using RCTs to build and test

Theory can often allow us to reclassify new or unknown situations as analogous to situations where we already have background knowledge, the trial helps consolidate previous knowledge and contributes to an evolving body of theory and empirical

Thus, we can choose, based on our willingness to make assumptions and on the data that we have, a suitable combination of theoretical assumptions and observational data in order to adapt and use trial results.

When the validity and credibility of the theory is good enough, we can incorporate the trial results into previous knowledge, and extend the trial results.

6) Using RCT results when scaling up

Many RCTs are small_ scale. Predicting the same results at scale are unlikely to be as in the trial. That is not to say that RCTs is uneffective, but researchers should be careful of the different results and conclusions when scaling up.

7) Using RCT results for individuals

A well-conducted RCT delivers an ATE for the trial population but, in general, that average does not apply to each individual. In extreme cases, an RCT with an average treatment effect shown that the treatment works for no one.

2.2 Conclusion about second argument

Having an understanding of causal structure is important in RCT. It can help us find out the support factors behind casual relationship so that we can simply generalize RCT results and reweighting without blindness. As a matter of fact, RCT may not effective in scaling up and applying for each individual. There is no other serious alternative to trying in practice.

3 Conclusions about RCT

From what have been discussed above, we have two challenges, one is in medical are, the other is from social science.

3.1 The medical challenge in RCT

The medical challenge is that we have known the ATE results may not apply to specific person, can we still trust results? To be more specific, in medical area, should we take the medicine which is verified effective in RCT? As a matter of fact, many new drugs are administered in the absence of an RCT. For patients, the effect was tested is from you and your physician. This principle is applicable for those medicine which have been tested through RCTs.

3.2 The social science challenge in RCT

The other challenge is that RCTs have problems to apply, however, other methods have all of those problems. How to balance between RCT and other trails? This question is depend on our purposes.

The chief advantage of the RCT is that it can, if well-conducted, give an unbiased estimate of an ATE in a trial sample and thus provide evidence that the treatment caused the outcome in some individuals in that sample. If an unbiased estimate of the ATE is what we want and there's little background knowledge available, then an RCT may be the best choice. Though, in some cases, RCT may do better than observational study, it is a long step from that to saying that an RCT can solve the problem.

1) Invalidity to generalize RCT results

Besides the large cost, RCT can be run only in enrolled unrepresentative study sample not to full qualified population. An observational study with corrections and

much large sample may do better.

2) Inference with other methods

RCTs should be combined with methods widely used in social and economic sciences such as IV.

3) Consideration of factors besides context

RCTs are the ultimate in non-parametric estimation of average treatment effects in trial samples because they make so few assumptions about heterogeneity, causal structure, choice of variables, and functional form.

But the credibility of the results can be undermined by covariates and by excessive heterogeneity in responses. In RCT we can only recover the mean of the distribution of treatment effects, and that only for the trial sample. Yet, in the presence of outliers in treatment effects or in covariates, reliable inference on means is difficult.

4) Randomization in RCT

Randomization does nothing unless the details are right; purposive selection into the experimental population undermines inference in just the same way as does selection in observational studies.

5) The importance of causal structure

There is no option but to commit to some causal structure if we are to know how to use RCT evidence out of the original context. The lack of structure can be seriously disabling when we try to use RCT results outside of a few contexts. If the aim is to use empirical evidence, any credibility advantage that RCTs have in estimation is no longer operative.

Criticism:

Just like author said that the key point for science is discover not “what works”, but “why things work”. We can not deny the contribution of RCT that it has been applied in many field of science. But it should be emphasized that randomization does not mean everything, though it require minimal assumptions and less prior knowledge. Under some circumstance, just like we summarized above, it is meaningless. I think, the optimal way to find the truth of science is that find the mechanism behind the story and choose the suitable tool to solve it. As an old says goes “Do not put the cart before the horse”