Social Science & Medicine 210 (2018) 26–28

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Commentary

The trials of randomized control: Probability, intuition and the dinosaur risk. A commentary on Deaton and Cartwright

Kaushik Basu

Department of Economics and SC Johnson College of Business, Cornell University, Ithaca, NY, United States

1. Introduction

The arrival of randomized control trials (RCTs) in economics, once the mainstay of epidemiology and certain branches of medicine, for uncovering patterns and causality in economic and social phenomena, has been the big story of our time in the social sciences. Its need is beyond dispute. In designing economic policy interventions, almost every day, we have to form judgements about what intervention X may do to variable Y or Z, and what side effects this may have on seemingly unrelated variables, which can have implications for human well-being. Getting a handle on this important question seems worth every effort. Without doubt RCTs have given us several important insights. However, given the need and some successes, it is also easy to go overboard, making claims that may not stand up to scrutiny and bull-dozing other methodologies, which may have an important function but cannot withstand the celebration of RCTs.

The need is not to decide on the acceptance or rejection of the method of RCTs, but to take stock of exactly what it does or does not do, to have an understanding of what this has achieved and what it has not done and, more importantly, cannot do. This is because of the risk of over-reach, which can transport science to superstition and ultimately do harm. This is what the paper by Angus Deaton and Nancy Cartwright attempts to do (Deaton and Cartwright, 2018). The paper is, at the same time, a comprehensive survey and an impartial (though occasionally, one suspects based on some of their earlier writings, through clenched teeth) evaluation. The paper has two parts. The first argues that, when it comes to average treatment effects (the adjective, average, is important), RCTs cannot be bettered. The second part takes on the larger question of when to use the RCT results and how. I have relatively little to add, and nothing to contest, to their first part and so the next section, which simply elaborates on it, is relatively straightforward. The two sections after that develop the argument in Deaton and Cartwright and propose an open research agenda stemming from their paper and also the many recent works on the methodology of randomized control trials.[1]

2. RCTs as description

If we draw a sufficiently large random sample from a population and then administer some intervention to that sample and find that, in terms of some particular response following the intervention the average of the sample differs from the average of the people who were not subjected to the intervention, then that difference in response can be treated as a consequence of the intervention. In other words, we get to find out the average treatment effect by this method. Occasionally, such random trials have to be designed and executed by researchers and at times, we are lucky to find these occur naturally because of some government intervention. In either case the upshot is that RCTs of this kind can reveal important properties of a population, for instance, how it will respond to some policy intervention.

An iconic example of an RCT that uses an intervention which occurred without researcher intervention but yielded some deep insights is the paper by Chattopadhyay and Duflo (2004). The Indian Government in 1992 made a constitutional amendment reserving one-third of the *panchayats*, or village councils, so that they could only have women as leaders; and the one-third earmarked for this would have to be chosen by lottery. This created a perfect setting to find out if the selection of women made a difference to the provision of public goods in the village. Chattopadhyay and Duflo (2004) studied this with data for Bengal and Rajasthan and showed that the selection of women leaders indeed had a positive impact. Such villages were more progressive in many ways and in particular in the provision of the most important public good, to wit, water. The randomization ensured that the standard critique that researchers often pick up the reverse causality, namely, that it is the more progressive villages that elected women would not apply.

A good example of randomization designed by researchers in search of a regularity is the celebrated paper of Miguel and Kremer (2004). Can deworming lead to greater fitness and improve school attendance? By designing an RCT, whereby they administered deworming medicine to children in a bunch of randomly selected schools in Kenya, they showed that participation rates of students improved greatly by virtue of deworming.

Deaton and Cartwright (2018) rightly caution us against the error

E-mail address: kb40@cornell.edu.

[1] See Banerjee and Duflo (2009), Cartwright (2010), Deaton (2010), and Banerjee et al. (2016).

https://doi.org/10.1016/j.socscimed.2018.04.036

Available online 27 April 2018

that many practitioners of RCTs fall into, namely, that of assuming that these kinds of results "provide strongest evidence for causality and for effectiveness" of policy interventions (see also Cartwright, 2011). This brief observation however deserves dissection.

Let me begin by assuming that the world runs on causality or, what is the same, is causally determined. This means that if two worlds were identical prior to time t, they would end up identical *at* time t. Most of us work on the presumption that the world is indeed causally determined and that is the assumption that I will make here. I would hasten to add, as David Hume had reminded us, this cannot but be an axiom. There is no way to prove determinism. With this assumption in the background, what much of science tries to do, and so does economics, is to try and discover some of the causal links in the world.

My claim is that RCTs do establish a limited causality that I refer to as "circumstantial causality" but they do not, and in fact cannot, establish "universal causality" and for this reason they do not provide evidence for "effectiveness" (Basu, 2014, 2018). Most significantly, RCTs by themselves cannot tell us what is the right policy intervention.

To understand this, consider the Chattopadhyay and Duflo study. For the population on which the RCT was performed the result does show causality. For Bengal of the time period studied by them, it is reasonable to believe that it was the election of women which caused the provision of public good to improve. But "Bengal of the time period" had many traits—its weather, its history, its eating habits, various government interventions that were going on and so on. Let me call this bundle of traits T. So what the study shows is that given T, an intervention which selects a woman leader is on average likely to improve the provision of public goods. Since no one can describe exactly what the T is except in tautological terms, namely, "the traits that characterized Bengal of that time", there is no way of knowing if the election of women leaders in another society or another time will have the same effect.

Based on the above findings, we cannot assert that electing a woman leader in a Chiapas village will improve water provision there, and deworming kids in a school in Kenya in 2019 will improve school attendance. All these two RCTs do—and RCTs are ideal for this—is to establish historical, circumstantial causal links. They give the impression of having pinned down a universal causality—an intervention that would work always; but that is not the case. In short RCTs are very good at describing the past, and good descriptions are not to be belittled (Sen, 1980). But they are not in themselves sufficient to guide policy interventions in the future.

3. Description and commonsense

Are RCTs then of no use in designing actual policy interventions? Fortunately, that is not so. They can be useful when we combine them with intuition that we often possess. The problem is this intuition will at times suggest that we abandon the results of RCTs and what I would argue is that it will be folly not to do so as a rule.

This can be explained with the 'Noah's Ark' construction (Basu, 2018). Suppose you live in a village where everybody is afflicted with flu. Researchers come and draw a random sample and inject them with a green liquid and discover that those injected with this strange-looking liquid recover immediately, whereas the average person without such an injection takes extremely long to recover. This RCT seems to make a good case for you to take the green injection.

Now add some details to this story which are statistically irrelevant. Suppose this village is called Noah's Ark, where you (and may be one more creature) is human and all others are reptiles. Next suppose in a neighboring village, where the inhabitants are all humans, there is an outbreak of the flu, and the same injection is administered to some inhabitants (not randomly selected but those wearing bright-colored shirts because they were easier to spot) and this results in some deaths and very few getting cured.

All your intuition will tell you to reject the findings of the RCT conducted basically on reptiles, and to take the poorly-conducted study in the neighboring village seriously.[2] What I am arguing is that it will be foolish for you not to give in to your intuition. By its very nature, intuition is difficult to define but we have to, nevertheless, use it. There are reasons for this. Our intuition itself may have an informal statistical basis. When a doctor prescribes a medicine and you feel that is not right for you, there is reason not to dismiss this gut feeling out of hand. Fortunately, some doctors recognize this, as is described wonderfully in Groopman and Hartzband (2012). This is because, while doctors may be using large amounts of statistics from other human beings, you have evidence from your own past experience. And evolution suggest that the intuition by which we amass this past evidence informally cannot be all wrong. Our intuition has an evolutionary basis and we will be wrong to dismiss it out of hand. Unfortunately, both users of RCTs and evidence-based medicine have a propensity to de-emphasize intuition (see Sackett et al., 1996 and Worrall, 2007).

Turning to the two above examples, most of us would intuitively feel that the benefits of deworming demonstrated by Miguel and Kremer for Kenya would carry over to other societies and other times, such as Bengal today or Kenya in 2018, whereas the provision of public goods that resulted from electing women leaders in Bengal, as shown by Duflo and Chattopadhyay, may not carry over to Kenya or the Amazon region. The first study has a biological element and our intuition suggests that is more common to human beings than psychological and social propensities, and so, once it is discovered in one society, it will carry over to others. So while both these studies meet the gold standard for describing what happened in particular populations, their wider applicability may be quite different. To reach that conclusion we have to use reasoned intuition. This is not a foolproof method but arguably the most reasonable one for establishing external validity.

This argument about intuition has bearings on the large debate that has taken place on external validity, not just in contemporary economics but also in medicine (see Freiden, 2017). Some have taken the line that for external validity, that is, validity beyond the population on which the experiment was conducted, we have to do more experiments on diverse populations and collate the results. This is, in reality, a non-starter for solving the problem, because for all populations studied there will be more beyond, if nothing else, the people of tomorrow. For external validity, we have to use the kind of intuition discussed above. As Banerjee et al. (2016) point out, external validity is inherently subjective.

4. Reasoned intuition, evolution and the dinosaur risk

Though intuition is subjective, there is some reason to rely on it because the intuitive sense that we have is likely a trait acquired through long evolutionary processes. Hunter gatherers who did not have the right intuition, perished, and so the ones who survived, are likely to have the right kind of intuition. Hence, there is good reason to utilize the intuition that we have. Unfortunately, we cannot leave it at this. There is a case for bringing reason to bear on intuition and, if need be, over-rule our natural intuitions in some cases.

To understand this note that the dinosaur in its last days also had traits which had survived past evolution but were no longer ideal for survival in the future. This could be because of the slow drift of the world's environment, the fauna and flora around us or even the dinosaur's own traits. Hence, the fact that we have a certain trait (say, a

[2] One can find many examples from real-life, non-randomized studies that have given us real insights. Thus Freiden (2017), in discussing RCTs in medicine, talks about the high rate of sudden infant death syndrome (SIDS) in New Zealand, which led to a study of the conditions under which 128 infants who died from SIDS led to major a break-through, including an understanding of the critical role of sleeping positions (see Mitchell et al., 1991). Furthermore, as he rightly points out, this is also an area where the ethics of doing an RCT may be called into question, something that Deaton and Cartwright (2018) also refer to in the context of the social sciences.

certain kind of intuition), does not mean that holding on to that or giving in to that will necessarily help us with our survival. We could be dinosaurs who are on the verge of going extinct.

Note now that human beings are arguably the first creatures on earth who are, like all other creature, subject to the laws of evolution, but are, in addition, aware of their own predicament. Hence, they are capable of analyzing whether their current traits create the risk of them being the dinosaurs of today, that is, in line for extinction. If so, they can analyze and try to change their own traits. Herein lies the scope for exciting new research that combines theory with empirical studies.

We can use evolutionary theory to study behavior traits that are evolutionarily stable, in the sense of Maynard-Smith and Price (1973). An interesting recent example is the finding by Alger and Weibull (2013), that a minimal sense of morality, is an essential human trait for long-run survivability. What they actually show is that it is a convex combination of selfishness and morality, the latter in the Kantian sense of doing what would lead to maximal payoff if everybody did so, that is robust in an evolutionary sense.

What I would like to add here and is not a part of the paper by Alger and Weibull, is that their result does not mean that human beings who are around today *have* this trait of a *homo moralis*. It is entirely possible that, through the rough and tumble of market competition or through our reading and misunderstanding of Adam Smith we have become totally selfish. In other words, we now run the 'dinosaur risk'—that of waiting in line to go extinct. What this means is that the Alger and Weibull paper has, whether or not it was intended by the authors, an implicit prescriptive element. It says that, given that most of us want human beings to survive in the long-run, we must nurture our sense of morality. Pure selfishness will not take us there.

To return to the topic on hand, if we want to take that next step from RCTs and to decide what kinds of policies we should implement tomorrow, we have to make room for our intuition, but a 'reasoned intuition'. We have to use reasoning to decide which parts of our intuition are evolutionarily robust. The answer to this will have to come from theory, in conjunction with what we know are reasonable axioms describing the world we inhabit. The result of discovering the robustness of morality was a surprise. We may yet be able to discover rules by which we can make that leap from the results of randomized control trials to actual policy that is not purely a leap of faith.

References

Alger, I., Weibull, J., 2013. Homo moralis: preference evolution under incomplete information and assortative mating. Econometrica 81.
Banerjee, A., Chassang, S., Snowberg, E., 2016. 'Decision-Theoretic approaches to experiment design and external validity. NBER Work. Pap. 1 No. 22167.
Basu, K., 2014. Randomisation, causality and the role of reasoned intuition. Oxf. Dev. Stud. 42 (4).
Basu, K., 2018. The Republic of Beliefs: a New Approach to Law and Economics. Princeton University Press (forthcoming).
Banerjee, A., Duflo, E., 2009. The experimental approach to development economics. Annual Review of Economics 1.
Cartwright, N., 2010. What are randomized trials good for? Phil. Stud. 147.
Cartwright, N., 2011. The art of medicine: a Philosopher's view of the long road from RCTs to effectiveness. Lancet 377 April 23.
Chattopadhyay, R., Duflo, E., 2004. Women as policy makers: evidence from a randomized policy experiment in India. Econometrica 72 (5).
Deaton, A., 2010. Instruments, randomization, and learning about development. J. Econ. Lit. 48.
Deaton, A., Cartwright, N., 2018. Understanding and misunderstanding randomized control trials. Soc. Sci. Med. 210C, 2–21.
Freiden, T., 2017. Evidence for health decision making: beyond randomized control trials. N. Engl. J. Med. 377 (5).
Groopman, J., Hartzband, P., 2012. Your Medical Mind: How to Decide what Is Right for You? Penguin, London.
Maynard-Smith, J., Price, G.R., 1973. The logic of animal conflict. Nature 246.
Miguel, E., Kremer, M., 2004. Worms: identifying impacts on education and health in the presence of treatment externalities. Econometrica 72.
Mitchell, E.A., Scragg, R., Stewart, A.W., et al., 1991. Results from the first year of the New Zealand cot death study. New Zealand Medical Journal 104.
Sackett, D.L., et al., 1996. Evidence-based medicine: what it is and what it isn't. Br. Med. J. 312.
Sen, Amartya, 1980. Description as choice. Oxf. Econ. Pap. 32.
Worrall, J., 2007. Evidence in medicine and evidence-based medicine. Philos. Compass 2/6.