« **Foundations of Randomization » Workshop**

**June 8th and 9th, 2021**

**Schedule**

June 8th

1.30 pm - 2.20. Stephen Senn (consultant statistician, Edinburgh): Fisher’s Gambit. Understanding Randomisation

2.20 - 2.55. Evangelos Koumparoudis (Sofia University): Randomized control trials limitations and epistemological aspects

break

3.10 - 4.00. Jonathan Fuller (University of Pittsburgh): Statistics, Balance and Bias: Randomization in Clinical Epidemiology and EBM

4.00 - 4.35. Aydin Mohseni and Daniel Alexander Herrmann (University of California, Irvine): Why randomize, really?

break

4.50 - 5.40. Fabienne El Khoury (INSERM / Sorbonne Université): Randomised trials: merits, limitations, and field developments

5.40 - 6.15. Elselijn Kingma & Oliver Galgut (King’s College London): Better than randomisation? A philosophical defense of ‘dynamically allocated controlled trials’

June 9th

2 pm - 2.50. Isabelle Guérin and François Roubaud (French National Research Institute for Sustainable Development): RCTs in the field of development: a critical perspective with a focus on microcredit sector

2.50 - 3.05. Julio Michael Stern (University of Sao Paulo): Why and how to randomize and audit in legal sortition and clinical trials

break

3.20 - 4.10. Maximilian Kasy (Oxford University): Statistical decision theory cannot justify randomization or pre-registration for experiments.

4.10 - 4.45. Michel Shamy (Ottawa Hospital & Department of Medicine, University of Ottawa & Ottawa Hospital Research Institute): Evidentiary standards and the justification of randomized clinical trials: The example of hydroxychloroquine trials for COVID-19

4.45 - 5.20. Konstantin Genin (University of Tubingen) and Conor Mayo-Wilson (University of Washington, Seattle): Randomization, identifiability, and estimation of causal effects

**Abstracts**

**Fisher’s Gambit. Understanding Randomisation**

Stephen Senn, Consultant Statistician, Edinburgh

stephen@senns.uk | <http://www.senns.uk/Blogs.html> | Twitter: @stephensenn

If one starts from the position that the goal of causal inferences from experiments is to produce perfect estimates, the game is lost. There are indefinitely many confounding factors and it is impossible to balance or adjust for them all. The great statistician RA Fisher saw that this goal was hopeless and proposed an alternative: we must give up the goal of perfect estimation and accept instead, that the task is to quantify their imperfection. This is Fisher’s gambit.

His programme for dealing with this involved the following steps 1) Concentrating on probabilistic statements of effect rather than point estimates 2) Avoiding obsessions about bias 3) Using outcome variation rather than confounding to judge accuracy and precision 4) Resolving variation into between and within group components 5) Using randomisation to ensure that marginal inferences were valid.

This is not without difficulties and in particular if one believes, as Fisher did himself, that conditional inference trumps marginal inference, one may question the relevance of the marginal guarantee that randomisation provides. Nevertheless, reasonable discussions of Fisher’s approach need to discuss it for what it is, not for what it is imagined to be. In this respect the literature in the philosophy of science on this subject has from the statistical point of view been disappointing and largely irrelevant.

I shall attempt to show what Fisher achieved and what might remain to be done. I shall conclude that a basic qualification for discussing randomisation ought to be the ability to carry out and understand an analysis of variance.

**Randomized Control Trials Limitations and Epistemological Aspects**

Evangelos Koumparoudis

PhD candidate, Department of Logic, Ethics and Aesthetics, Sofia University

evangeloskoumparoudis@gmail.com/

After the first Randomized Controlled Trial (RCT) in 1948 examining the effects of streptomycin in pulmonary tuberculosis, a long way has been run placing RCTs as the ‘’gold standard’’ of evidence-based-medicine (EBM), as providing the ideal methodology for causal inference. This contribution first aims in providing a brief introductory outline of the various phases and classification of RCTs. Additionally, although EBM is defined as: ‘’the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients’’ (Sackett, 1996) and has a great reliance on RCTs there are serious limitations both concerning the trials as themselves and the medical practice. Consequently, our second aim is to reveal these limitations, not from the standpoint of total nihilism and rethinking but from a critical view which can lead us in reducing biases and provide rigorous tools to examine cause-effect relationships between medical interventions and best outcomes. As to the clinical trials, an important issue is that of the *unrepresentativeness of trial subjects*. Well-structured clinical trials, in most cases, define their preconditions of participation (Lee et al., 2001). Secondly, *long-term therapy* can also affect the reliability of a trial (Rossello et al., 2015). *Comorbidity* also plays an important role. Comorbid patients usually are excluded from the clinical trials (Van Spall et al., 2007). Furthermore, there are cases in which physicians try to balance between *statistical* and *clinical* significance (MIAMI Trial Research Group, 2001). Finally, there is the issue of misleading results in clinical trials. Bias can occur from improper statistical analysis concerning the methods used for the trial and non-publication of important statistical results, mainly those referring to negative effects of a protocol or drug (Cartwright and Deaton, 2018, Cartwright, 2008, 2010, Deaton, 2010). In medical practice, there are two main biases on the role of clinical trials; the first is the bias of *the easily measurable,* the second is the bias in *commissioned research* (Desmond and Desmond, 2001). Considering the above limitations with more emphasis on medical practice, the third point of our intervention will be the examination of mechanistic reasoning as being sufficient for medical causation in lower and high-quality controlled studies revolving around the problem of masking (Howick, 2011, Wilde, 2021). Finally, we will strive towards the problem of undetermination and if auxiliary hypotheses are needed so as we can better substantiate our findings from RCTs and Placebo-Controlled-Trials (PCTs), through the prism of Duhem-Quine thesis. The latter will help us, to question even deeper the boundaries of EBM and if is capable to understand the mechanisms of diseases from a holistic approach and thus generate new hypotheses, develop new knowledge and treatments, and if at last there is a ''gap'' between evidence and what we truly know when we are called to pass from statistics to the therapeutics of a single patient (Anderson, 2006, Chin-Yee, 2014, Norton 2003).

**Statistics, Balance and Bias: Randomization in Clinical Epidemiology and EBM**

Jonathan Fuller, History and Philosophy of science, University of Pittsburgh

About twenty years have passed since John Worrall’s (2002) classic paper, “What evidence in evidence-based medicine?” In it, Worrall considered several epistemic justifications for randomization in evidence-based medicine (EBM) independently, including considerations of statistics, balance and bias. His conclusion, further defended elsewhere, was largely negative: that generally these defenses either were not central to the evidence-based defense of randomization, failed to hold water, or were often exaggerated. Over time, Worrall’s arguments have been accepted by some and challenged by others. They remain influential in examining the status of randomization in EBM. But they are also a starting point for investigating various accounts of the role of randomization and the extent to which these roles are complementary or competing. Drawing on older as well as more recent responses to Worrall, I consider the extent to which his arguments succeed in undermining the epistemic rationale for randomization in EBM. I suggest that the three justifications he considers independently – statistics, balance, bias – are best integrated through a unified framework for thinking about causal inference. Though there are multiple such frameworks available in epidemiology, they too have their interconnections. This discussion thus bears on the question of whether there are several theoretically distinct justifications for randomization in the sciences, and if so, how they are related.

**Why Randomize, Really?**

Aydin Mohseni & Daniel Alexander Herrmann

University of California, Irvine

We present a Bayesian analysis of randomization in the context of randomized controlled trials. This analysis reveals what an experimenter must believe for her to expect randomization to be a successful strategy, addresses a puzzle regarding the nature of randomization from a Bayesian perspective, and provides normative counsel regarding adequate sample sizes.

The function of randomization in randomized controlled trials is typically ex- plained in terms of addressing hidden variable bias, where hidden variable bias is understood as the confounding effect of variables unaccounted for by the exper- imenter and which co-vary with both the outcome of interest and the treatment group.

A common explanation for how randomization is supposed to address hidden variable bias invokes the law of large numbers. As the size of the treatment and control groups grows large, the sample value of a potentially confounding variable approximates the population value of that variable. Given this, one can expect to obtain roughly equal values of the variable in the control and treatment groups, and so confounding is avoided.

We present a formal model of an experimental setup that demonstrates this explanation is not quite right. Our model involves a rich set of possible con- founders in order to evaluate the claim that randomization addresses both known and unknown confounders. This model allows us to consider different allocation strategies and evaluate their success. We consider a number of different possible formal notions of what a confounder is in this rich context, and discuss the degree to which our results are robust across these definitions.

The core moral of our arguments is that whether (and how effectively) hid- den variable bias is addressed by randomization depends crucially on background assumptions not normally accounted for—namely, the count and character of po- tential confounders. Further, the nature of the sort of randomization required can be made explicit. We elucidate this from a Bayesian perspective.

One corollary is that the experimenter’s beliefs about the empirical situation will determine the sort of randomization (if any) she expects to be successful. A precise characterization of the kinds of beliefs needed vindicates conventional randomization strategies as one (among many) strategies that are helpful in many real-world contexts. This is precisely because the experimenter’s background be- liefs in these contexts will tend to satisfy the appropriate conditions.

A longstanding puzzle addressed by our results is that, from a Bayesian perspective, a rational agent cannot expect a randomized strategy to be strictly better than all alternative deterministic strategies. This argument goes back at least as far as Savage, but makes the central role of randomization in experimental method appear puzzling. Our results affirm this argument: a (classically rational) experimenter cannot prefer randomization to all other strategies; but randomized strategies can (and often will) be amongst her set of optimal strategies. What characterizes an experimenter’s set of optimal strategies for allocating individuals to groups is not anything to do with randomization per se. Rather, the experimenter’s optimal strategies are simply those that she expects will be uncorrelated with the set of unknown confounders. This clarifies the sorts of randomized strategies an experimenter can find at least as good as (though no better than) any other optimal strategy.

Further, our results suggest that, in practice, the sample size required for vari- ous strengths of inference should be sensitive to the experimenter’s beliefs about the count and character of potential confounders in her experimental setup. Fol- lowing this line of reasoning reveals that, in a wide range of real-world cases, we have been generically underestimating the sample sizes required to license the strength of inference we take ourselves to be making

**Randomised trials: merits, limitations, and field developments**

Fabienne El Khoury, Equipe de Recherche en Epidémiologie Sociale, INSERM-Sorbonne Université

The randomised controlled trial is considered as the “gold standard” of clinical research, offering the highest quality of evidence. However, it often suffers from limitations such as ethical concerns, cost, and feasibility. We’ll discuss those merits and limitations, as well as methodological improvements and developments which have been added to the randomised design, especially in the field of behavioural interventions.

**Better than Randomisation? A philosophical defence of ‘dynamically allocated controlled trials’.**

Elselijn Kingma & Oliver Galgut

King’s College London

Randomised Controlled Trials [RCT’s] continue to be seen as the epistemically superior trial design. But it has significant shortcomings; critics have argued convincingly that RCT-based evidence has limited external validity (e.g. Cartwright, 2007), and that random allocation cannot, in individual trials, control for ‘all known and unknown confounders’ (e.g. Worrall, 2002). These shortcoming do little to undermine the RCT’s ‘gold standard’ status, however. One main reason for this may be the lack of a suitable alternative. This paper introduces and critically appraises one such plausible, and feasible, alternative: Dynamically Controlled Trials.

Dynamic allocation methods – as described in more detail in the paper – do not randomly allocate patients to intervention and control groups. Rather, they allocate patients based on a comparison of certain known characteristics of the patients with the existing distribution of those characteristics between the treatment and control groups. The aim – and result – of dynamic allocation is that treatment and control groups are more closely matched for known characteristics than is often achieved in random allocation.

This paper aims to do two things: it (1) systematically compares dynamic allocation with random allocation, and; (2) identifies the epistemic strengths and weaknesses of the former. To do so, we first critically review the arguments that support the epistemic superiority of random allocation. This identifies the main epistemic benefit of random allocation as its ability to facilitate masking. We will also demonstrate that, contrary to what is often claimed, randomisation cannot ‘control for all known and unknown confounders’. Nor is this just a theoretical problem: differences between treatment and intervention groups are well-described in the epidemiological literature as the problem of ‘baseline imbalances’.

Our review concludes that alternatives to random allocation would be epistemically superior to randomisation if (1) they are equally good or better than random allocation at facilitating masking, and (2) are better at controlling confounders/preventing the problem of baseline imbalances.

We then introduce and critically appraise the group of methods collectively known as ‘dynamic allocation’ methods. We argue, first, that these methods outperform random allocation in terms of controlling for known confounders and preventing the problem of ‘baseline imbalances’. For unknown confounders they perform no worse, and possibly better, than random allocation. Second, we contend that the more sophisticated dynamic allocation algorithms facilitate masking at least as well as random allocation. We identify and systematically define the two criteria that such ‘sophisticated’ dynamic allocation methods must meet: being probabilistic and opaque.

We conclude that probabilistic and opaque dynamic allocation methods are epistemically superior to random allocation in all but a few circumstances. All else being equal, we recommend that dynamic allocation should replace random allocation as the epistemically superior trial design.

**RCTs in the field of development: a critical perspective with a focus on microcredit sector**

Isabelle Guérin and François Roubaud

French National Research Institute for Sustainable Development

In October 2019, the *51st Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel* was awarded to three of the main proponents of Randomized Control Trials (RCTs). Yet there is reason to question the validity and repercussions of the elevation of this design to the so-called “gold standard” for impact evaluation. Have RCTs really “dramatically improved our ability to fight poverty in practice”? Which sorts of questions are they able to address and which do they fail to answer? Is causal explanation the only way to understand poverty, and do RCTs systematically manage to provide causal explanations? Are RCTs really a gold standard? What are the dangers in their misuse? Is the supremacy of experimentation in development economics, as commended by the Nobel jury, scientifically legitimate and politically desirable? Bringing together 26 leading specialists in the field (including two Nobel Prize winning economists) from a range of backgrounds and disciplines (economics, econometrics, mathematics, statistics, political economy, socioeconomics, anthropology, philosophy, global health, epidemiology, and medicine), this authoritative book presents a full and coherent picture of the main strengths and weaknesses of RCTs in the field of development - how they work, what they can achieve, why they sometimes fail, how they can be improved, and why other methods are both useful and necessary.

Based on their recent edited book, Isabelle Guérin et François, two of the editors will present the volume, its main conclusions, with a special focus on microcredit sector.

<https://global.oup.com/academic/product/randomized-control-trials-in-the-field-of-development-9780198865360?cc=us&lang=en&>

**Why and How to Randomize and Audit in Legal Sortition and Clinical Trials**

Julio Michael Stern (with Marcelo de Souza Lauretto, Rafael Bassi Stern and Carlos Alberto de Braganca Pereira)

University of Sao Paulo | jstern@ime.usp.br | https://www.ime.usp.br/~jmstern/

Our talk surveys a few contributions concerning randomization made by the Bayesian Research Group of IME-USP -- the Institute of Mathematics and Statistics of the University of Sao Paulo. This research group includes a dozen scientists working mainly at USP, UNICAMP and UFSCar -- all public universities located in the State of Sao Paulo, Brazil. Our collaborators, located in Brazil, other Latin American countries, the USA and Western Europe, include specialists in Bayesian Statistics, Computer Science, Law, Medicine, and Bioinformatics. Our work seeks integrated solutions in Randomization Techniques that encompass the following aspects:

(a) Why to Randomize, that is, what are the theoretical and philosophical justifications to delegate to chance important decisions that could otherwise be made deterministically. This item includes a variety of specific research topics ranging across pure and applied science, like Foundations of Bayesian statistics; Philosophy of Law; Decoupling of stochastic processes; Parallelization and sparse factorization techniques in computational statistics; etc.

(b) How to Randomize, that is, what are the acceptable and the best ways to implement randomization procedures. This item includes research topics concerned with integration of pure randomization with numerical optimization techniques for optimal sampling. Following a (never before fully developed) intuition of Prof. Debabrata Basu\*, we carefully introduce random noise in the formulation of mathematical programming problems aiming at optimal sampling, and do it in a way that ensures goal efficient (close to optimal) and haphazard effective (stochastically nearly decoupled) solutions.

"The conterquestion `How can you justify purposive sampling?' has a lot of force in it. The choice of a purposive plan will make a scientist vulnerable to all kinds of open and veiled criticisms. A way out of the dilemma is to make the plan very purposive, but to leave a tiny bit of randomization in the plan." Why to Randomize? - Basu (1988,ch.XIV, p.257)\*

(c) How to Audit, that is, how to make randomization procedures easily verifiable. This item includes research topics concerned with the development and implementation of GNU license or ubiquitously available software solutions enabling fair, transparent, safe, traceable and publicly auditable randomization procedures.

(d) Applications to Clinical Trials. This item includes topics related to the ethics of randomized clinical trials, including what these trials have in common and also how it differs from the ethical commitments of standard medical practice.

(e) Applications to Legal Procedures. Recent events in Brazilian politics highlighted the importance of randomization (a.k.a. sortition) procedures used by the Judicial system to assign legal cases to a specific judge or court. At the request of the National Justice Counsil (CNS), we reviewed randomization procedures used by the Brazilian Supreme Federal Court (STF), and made a series of recommendations that include the adoption of technologies already available developed at item (d).

**Statistical decision theory cannot justify randomization or pre-registration for experiments**

Maximilian Kasy

Department of Economics, Oxford University

Slides: <https://maxkasy.github.io/home/files/slides/randomization_preregistration_slides.pdf>

**Evidentiary Standards and the Justification of Randomized Clinical Trials: The Example of Hydroxychloroquine Trials for COVID-19**

Michel Shamy (with Brian Dewar, Mark Fedyk, Vignan Yogendrakumar)

Ottawa Hospital & Department of Medicine, University of Ottawa & Ottawa Hospital Research Institute

We argue that some of the reasons justifying randomization in clinical research are ethical in nature. This stance is consistent with widely held view that (a) randomization in clinical research *needs* to have ethical justification, and (b) the required justification comes from the principle of clinical equipoise. We wish to argue in support of (a) and against (b).

To wit: Physicians have a fiduciary duty to provide their individual patients with the best possible care. Since randomization involves the potential of allocating a patient to a group that will receive inferior care, it is typically justified through the invocation of the principle of ‘equipoise’ – as this is thought to ensure that individual patients are not treated as “mere” means, since this principle says that it is permissible to randomize patients when there is “uncertainty” about the efficacy of the treatment being tested. The uncertainty is independent of the patient – for a treatment may work in, say, 60% of the people but not the patient, given their unique pathophysiology. Instead, equipoise is typically defined as uncertainty within the physician community, relative to an implied standard of “works as effectively as a reasonable clinician would expect a treatment to work for a disease like the one we are concerned about.” And when this standard gets operationalized, the assessment of uncertainty boils down to the best judgment of an investigator or attending physician at the time the judgment needs to be made (which can, in certain circumstances, be emergency situations).

Against this line of argument, we have argued that the most fair and transparent method to demonstrate sufficient uncertainty for randomization is through a systematic review of the literature to determine whether there is, in fact, uncertainty surrounding a given scientific question.

But we also contend that a more granular standard is needed. The problem with most uses of the principle of equipoise is that it requires a certain kind and degree of uncertainty regardless of what is being compared. We propose that this is neither necessary nor correct. Instead, we argue that the epistemic standard necessary to provide ethical license for randomization should vary according to the nature of the treatments under comparison.

Consider four scenarios:

1. Two approved treatments are compared for relative efficacy
2. An approved treatment and a novel treatment are compared for relative efficacy
3. A novel treatment is compared against placebo
4. A standard of care that has been adopted despite an absence evidence is compared against its absence

Consider scenario A: When two approved treatments are compared, there will exist some extant literature on the efficacy of the treatment – this is a requirement for approval. As such, informed physicians can review the evidence, summarize it via a systematic review, and ultimately conclude whether uncertainty can be demonstrated within the literature. For example, a literature review can show that a proposed trial is likely to be redundant, or that a more efficient route to evidence strong enough to inform clinical judgment is a careful meta-analysis. As the scale and complexity of published clinical research grows, is it increasingly unlikely that even the most well-informed clinicians are aware of more than a fraction of the research relevant to typical clinical trials that fall into category A.

So, already we can see how the structure of the trial impacts our assessment of the ethical conditions governing the acceptability of its randomization. Now consider scenarios B and C, where the evidence supporting each of these arms of a trial varies. It is unreasonable to hold a novel treatment to the same evidentiary standards as a treatment that has been approved for therapeutic use. For the novel treatment, the trials necessary to demonstrate superiority or inferiority have not yet been performed, so there can be no informed opinion about the relative efficacy of the comparators. Contrarily, it would be equally unreasonable to use the same evidentiary standard to justify randomization of approved treatments as that used to justify experiments involving novel treatments for which there is no extant evidence base – and therefore no informed opinion as to their relative efficacy. Here, the relevant standard could be – and likely already is – biological or physiological plausibility.

Finally related to scenario (d), any practicing physician soon becomes familiar with how beliefs and practices persist can persist despite there being no trials demonstrating benefit (surgery for appendicitis) or there being existing evidence that actively shows harm (EC/IC bypass). In this case (scenario D), an epistemic standard of uncertainty is not met – because practitioners are certain the proposed treatment works – but a trial to produce evidence of non-efficacy should still be done and patients randomized would ultimately be receiving a superior treatment. Yet, equipoise would seem to rule out these trials: equipoise is inherently biased against trials designed to falsify existing clinical standards.

The history of tests of hydroxychloroquine as a treatment for COVID-19 provides examples that support our argument. Because of the pandemic’s novelty, uncertainty with the medical community about the efficacy of interventions against COVID-19 was to be expected. Hundreds of trials were launched, many on hydroxychloroquine, a known agent being applied in new circumstances. When no evidence about the efficacy of hydroxychloroquine against COVID-19 existed, randomization was justified based on biological plausibility and clinical urgency. Eventually, it became accepted as a standard of care intervention despite a lack of RCT data, and was then tested against its own withdrawal. However, as data from RCTs became available, hydroxychloroquine generally fell out of favour and was deprived of its standard of care designation.

And yet, more trials were launched, and many continue to be ongoing in 2021. The case of hydroxychloroquine can be read as a failure to adapt evidentiary standards to an evolving evidentiary base, which led to unnecessary randomization and potentially patient harm. Our four-way distinction is, we believe, an important and simple first step away from this practice.

**Randomization, Identifiability, and Estimation of Causal Effects**

Konstantin Genin and Conor Mayo-Wilson

University of Tubingen | Department of philosophy, University of Washington, Seattle

Although nearly 90 years have passed since Fisher [1935]’s famous defense of randomization, there is still no consensus about the value (if any) of randomized controlled trials (RCTs) when compared to other study designs.1 In this paper, we

(1) summarize several asymptotic reliability arguments that are traditionally used to defend RCTs,

 (2) argue that, in some settings, non-randomized experimental designs achieve those same reliability guarantees, and (3) distinguish several other small to mid-sample reliability criteria - some well-known and others novel - that might be used to defend the value of randomization. Our main contributions lie in the second and third tasks, as there is already an extensive literature discussing certain arguments for randomization.

Randomization – the random allocation of treatment to subjects – is a special type of intervention. It is well-known that, on the basis of observational data alone, neither the direction nor strength of a causal relationship is guaranteed to be identifiable [Spirtes et al., 2000]. Luckily, when two variables are associated, intervening on one variable can help discern which is a cause of the other, or whether they share some common cause.

Yet interventions come in many forms [Eberhardt and Scheines, 2007], and the types of interventions performed in RCTs are not typically necessary to identify direction of causal influence. The interventions performed in RCTs are typically special in two ways [Hernán and Robins, 2020, p. 26]. First, in RCTs, the hypothesized cause is often manipulated directly, rather than through an intermediary. For example, in drug trials, the treatment itself is typically both the hypothesized cause and the variable determined by randomization. But in other types of medical research, the hypothesized cause might be a biological variable like cholesterol that can manipulated only indirectly, e.g., via drugs or dietary changes. In theoretical work on causal discovery, researchers often investigated “ideal” interventions, in which the value of the putative cause is determined exclusively by a randomizing device (e.g., a clinical trial with perfect compliance) [Pearl, 2000]. Second, the intervention is “random” in the sense that the treatment variable must be ancillary, i.e., a variable whose distribution does not vary with the parameter of interest (here, the causal effect).

Together, these features of RCTs are sufficient to guarantee that not only that the direction of causal influence is discoverable but also that the causal effect (i.e., the strength of treatment) is identifiable [Hernán and Robins, 2020]. Here, identifiability means that if two distinct causal models disagree about the causal effect that X has on some variable Y , then the models assign different probabilities to observable data.

Identifiability is necessary for estimating a causal effect: if identifiability fails, there will be distinct models that will remain indistinguishable no matter how much data is collected in the designed experiment, but the two models will postulate different causal effects (and so they predict different results from certain policy interventions). But identifiability is far from sufficient, either in a theoretical sense or practical one. When a set of statistical models is identifiable, there may be no statistically consistent estimator [Gabrielsen, 1978], let alone an estimator that achieves a reasonable error bound at some calculable sample size. Still worse for the proponents of RCTs, consistency is attainable given other background assumptions about the data-generating process, and some such assumptions are no stronger than those researchers make when extolling the virtues of RCTs. For instance, it is well-known that, when treatment cannot be determined by a randomizing device (e.g., when non-compliance in a clinical trial is high), use of an instrumental variable can be used to estimate a causal effect [Pearl, 2011]. We explain how such standard results show that an intervention that inserts even the smallest bit of noise into an otherwise natural system (e.g., randomly providing access to treatment to one in ten million patients) can allow one to estimate causal effects consistently under some of the same assumption made by proponents of RCTs.

Similarly, recent theoretical work in causal discovery shows that even in- tervening may be unnecessary. For example, [Genin and Mayo-Wilson, 2020] show that, if one assumes the causal relations of interest are linear and error terms are non-Gaussian (i.e., the underyling model is what is called LiNGAM [Shimizu et al., 2006]), then there is a statistical decision procedure for learning the direction of causal influence from purely observational data, assuming no causal relations are confounded. In this paper, we summarize some unpublished work that extends Genin and Mayo-Wilson’s results to confounded models and show how the same arguments allow one to consistently estimate the strength of causal effects, again from purely observational data.

The general philosophical point we hope to stress is clear: identifiability and asymptotic reliability guarantees are typically insufficient to justify the use of RCTs, even from a purely “scientific” perspective that ignores ethical considerations. In the final paper of the paper, we therefore explore other reliability criteria that might be used to justify the use of randomization. Some such criteria – uniform consistency and efficiency – are already well-known among statisticians, but others – such as statistical decidability [Genin and Kelly, 2017] – deserve more attention, as least so we shall argue.