Why Randomised Interventional Studies*

Adam La Caze
School of Pharmacy
The University of Queensland
a.lacaze@uq.edu.au

9 November 2010

Abstract

A number of arguments have shown that randomisation is not *essential* in experimental design. Scientific conclusions can be drawn on data from experimental designs that do not involve randomisation. John Worrall has recently taken proponents of randomised studies to task for suggesting otherwise. In doing so, however, Worrall makes an additional claim: randomised interventional studies are epistemologically equivalent to observational studies providing the experimental groups are comparable according to background knowledge. I argue against this claim. In the context of testing the efficacy of drug therapies, well-designed interventional studies are epistemologically superior to well-designed observational studies because they have the capacity avoid an additional source of bias. While arguments for interventional studies are present in the medical literature, these arguments are too often presented as an argument for randomisation. Randomisation in interventional studies is defended on Bayesian grounds.

Keywords: Randomised controlled trials; Evidence in medicine; Epidemiology; Observational studies; Evidence-based medicine

*This is a preprint of a forthcoming paper to be published in *Journal of Medicine and Philosophy*. Please reference published paper.

1 Introduction

Worrall (2002, 2007a,b) argues that there is no *unique* justification for randomisation in experiments. In this argument he follows the work of Peter Urbach (1985; 1993) and Colin Howson and Urbach (2006) in suggesting that randomisation is not essential, or *sine qua non*, in science. This is right. Randomisation is not essential to science (or medicine), and to suggest otherwise is far too strong a claim. Worrall, however, goes one step further. He suggests that experimental designs that randomly allocate subjects to treatment and control provide *no* epistemological benefit in addition to that which can be achieved through alternative means. On Worrall's view, if experimental groups are comparable according to background knowledge, then whether the groups were formed by random allocation in an interventional study, or deliberate matching in an observational study, is of no epistemological import.[1] Randomised trials are not necessarily bad, Worrall suggests, but they provide no benefits over observational studies. This view is in stark contrast with scientific practice in clinical medicine.

If Worrall is correct, marked changes could be made to how medical research is conducted. Currently, prior to permitting a new drug on the market, large scale randomised interventional studies are required to show the therapy is efficacious. An 'interventional' study is a prospective study in which the investigators or a random process decide which participants receive active treatment and which participants receive control. Randomised interventional studies are commonly called 'randomised controlled trials' or 'RCTs'. I use 'randomised interventional studies' because this terminology more clearly separates two concepts, that is, whether or not a study is an interventional study, and whether or not an interventional study is randomised. Interventional studies can be contrasted with 'observational' studies. In an observational study, participants select (actively or passively) whether or not to take the treatment under investigation (or expose themselves to a risk factor) in their day-to-day life. Considerable time and expense could be saved if equally reliable evidence regarding the efficacy of therapies could be gained from observational or historically controlled studies.[2] Some study designs, especially case-control and historically controlled studies, can be conducted much more quickly than randomised interventional studies. And, while prospective observational cohort studies are conducted on similar timelines, recruitment is much easier in cohort studies compared to interventional studies.

But Worrall is not correct; the evidence provided by an observational study (or a historically controlled trial) regarding a treatment's efficacy is

not equivalent to the evidence provided by a well-conducted randomised interventional study, even when each of the studies involve experimental groups that are comparable according to background knowledge. There are good reasons for wanting to establish claims of a treatment's efficacy in a randomised interventional study rather than an observational study. In Section 2, I outline Worrall's claim that there is no epistemological distinction between randomised interventional studies and observational studies. In Section 3, I provide a positive argument for testing claims of efficacy in randomised interventional studies rather than observational studies.

2 Worrall on randomised (interventional) studies

According to Worrall, providing the groups under comparison are suitably controlled according to background knowledge, non-randomised non-interventional (that is, observational) studies are epistemologically equivalent to randomised studies.

> The best we can do (as ever) is test our theories against rivals that seem plausible in the light of background knowledge. Once we have eliminated other explanations that we know are possible (by suitable deliberate, or *post hoc*, control) we have done as much as we can epistemologically. (Worrall, 2007b, p. 486)

While the claim that randomised interventional studies are essential for science (or medicine) is so strong that it is no surprise that it is false, Worrall's counter-claim that there is *no* epistemological distinction between randomised interventional studies and observational studies (when groups appear matched according to background knowledge) is also quite strong, just in the opposing direction.[3]

It is important to be clear on the study designs under comparison. Randomised interventional studies are more commonly referred to as 'randomised controlled trials' or 'RCTs'. Unfortunately this terminology fails to distinguish randomisation from whether or not a study is an interventional study. While a randomised controlled trial is always an interventional study, an interventional study need not be randomised. Clinical medicine (in particular, evidence-based medicine) focusses on the distinction between randomised interventional studies and observational studies, as opposed to the distinction between randomised and non-randomised interventional studies.[4] But this is sometimes obscured by the terminology that is employed.

Studies are 'observational' when it is the participant, in consultation with their regular health carers, who select which treatments they take—no

intervention by study investigators is made on the participant's treatment. Observational study designs rely on deliberate matching, or *post hoc* adjustment, to compare the groups under investigation. Historically controlled studies, which compare patients taking a 'new' treatment with a historical cohort of patients who were treated with the conventional treatment, rely on similar matching and *post hoc* adjustment to analyse clinical data.[5] Interventional studies, by contrast, *allocate* patients to active treatment or control. In randomised interventional studies, allocation is achieved using a random process. Providing the study is large relative to the number of potential confounders, randomised allocation ensures that the experimental groups are roughly equally matched for characteristics that may influence a patient's response to treatment. To the extent this rough matching for confounders is achieved, deliberate matching and *post hoc* adjustment is not required.

In non-randomised interventional studies some other method is used to allocate patients to treatment or control. Perhaps experimental groups are deliberately matched by individuals blinded to which group will eventually receive the treatment. Blinding allocation may allay concerns about selection bias—more specifically, the selection bias called 'allocation bias'. A selection bias is a bias that comes about due to differences between the experimental groups other than the treatment under investigation. Allocation bias refers to investigators influencing the make-up of the experimental groups by preferentially selecting patients for one group over another due to patient characteristics (a process which may be subconscious). Observational studies do not suffer from allocation bias (because they do not 'allocate' patients), but they do suffer from other forms of selection bias. Selection biases are a particular problem for observational studies because it can be extremely difficult to identify, isolate and adjust for the multitude of factors that may influence participants to take, or not take, the treatment under investigation.

The distinction between interventional studies and observational studies raises more questions of practical and ethical importance than the distinction between randomised and non-randomised interventional studies. Interventional studies are considerably harder to run, and cost more than observational studies. The ethical questions raised by interventional studies are more troubling than the questions raised by observational studies. In interventional studies participants must consent to be *allocated* treatment or control. Unless there is an epistemological benefit in conducting randomised interventional studies, medicine might as well conduct observational studies. By contrast, there is no practical imperative in favour of non-randomised

interventional studies over randomised interventional studies.

3 Why randomised interventional studies

Worrall is right to criticise those who claim too much on behalf of randomisation. Randomisation does not *guarantee* the results of a clinical trial. When judging evidence for therapeutic decisions, the design of the study providing the evidence is but one of many important considerations. Rather than all of science, or all of medicine, well-conducted randomised interventional studies are epistemologically superior to well-conducted observational studies in a much narrower domain. My aim is to show that the randomised interventional study design has some epistemological benefits over observational study designs in testing the efficacy of drugs.

Efficacy refers to whether the drug works as expected under experimental conditions. Efficacy can be contrasted with *effectiveness*: the effects of the drug when used in routine care. Most clinical trials are set up to test the question of whether the drug is efficacious. Questions of efficacy can be made precise. Does the drug benefit the selected patients in the primary clinical outcome examined in the study? The quality of the evidence that the trial provides depends on whether the trial employed appropriate methods to rule out, or minimise, as many sources of systematic error (bias) as possible.

The argument for preferring randomised interventional studies in tests of a treatment's efficacy proceeds in two parts. Part one, provided in Section 3.1, is an argument for *interventional* studies over observational studies. Interventional studies have the capacity to rule out a form of bias that observational studies are particularly prone to. This argument has been provided in the epidemiological literature, but in the epidemiological literature, (i) interventional study designs and randomisation are taken together and too often conflated and (ii) the primary argument provided for randomisation is commonly the argument that Worrall has rather successfully criticised. The second part of the argument is for randomisation in interventional studies. I discuss the Bayesian rationale for randomised allocation in interventional studies in Section 3.2.

3.1 Why *interventional* studies

When it comes to the clinical sciences—inferences regarding how a patient or group of patients will respond to a therapy—there is often more unknown, or uncertain, than known. Causal knowledge in clinical medicine is provided by the basic medical sciences, such as physiology and pharmacology. Clinical

science both tests and applies this knowledge to patients—it is a translational science. Findings in basic science often raise new research questions in clinical science.

Basic and clinical sciences have a very different focus. Pharmacology and pathophysiology are extensions of biology; focus is given to the biological effects of drugs and diseases respectively. To gain an understanding of these biological effects, systems and mechanisms *within* the body are isolated. The individual, as a single complex whole, is the first thing that is abstracted away. This abstraction is necessary to develop, test, and improve the causal explanations provided by these basic sciences. Such causal explanations play an important role in the clinical sciences, but focus is given to a entirely different question.[6]

Clinical science works at the level of individuals (or groups of individuals with similar characteristics), and provides tools to predict, or understand, what effect treatment will elicit. Two difficulties present themselves at this level. Sometimes there is a lack of sufficient biological knowledge to be able to predict the outcome in an individual. Other times there is a wealth of biological knowledge, but insufficient detail to differentiate how the systems and mechanisms at play will interrelate in a given individual; that is, despite a comprehensive understanding of the underlying causal processes, it is impossible to predict what the overall effect of treatment will be. It is at this level that the empirical information provided by randomised interventional studies is invaluable, and superior to that provided by observational studies.

The set-up and analysis of observational studies requires investigators to estimate the effects of causal processes that influence the effects of a treatment in a patient—more so than is necessary in randomised interventional studies. Observational studies follow patients undergoing routine care. It is the patients who have made the choice whether or not to take the treatment under investigation, often sometime prior to the study. There are many possible factors that may influence whether or not patients choose to take a particular treatment, and any one (or a combination) of these may also influence the patient's outcomes. Due to the uncertainty in causal knowledge, it is usually impossible to isolate which of the many *possible* causal factors differentially distributed in the experimental group *may* influence the observed effects of treatment.[7] The investigators of observational studies, by deliberate matching or *post hoc* adjustment, attempt to minimise the effects of such possible confounding factors. This either requires estimation of the effects of different patient characteristics on their response to treatment, which in turn requires an extensive understanding of the causal processes at play, or relies on adjusting the observed data based solely on statistical

considerations. Both of these approaches rely on substantial assumptions, and the correct analysis of observational studies is dependent on the veracity of these assumptions.

Interventional studies that employ methods to avoid allocation bias don't have this problem. There is no need to identify causal factors that may influence the outcomes under investigation from the myriad of causes that may have led a patient to be on, or not on, the treatment under investigation. In interventional studies, participants are *allocated* treatment or control. And by allocating patients in an appropriate way, interventional studies can avoid this form of selection bias—whether allocation is via randomisation (with a post-randomisation check to ensure groups are adequately matched) or by some other method. The analysis of interventional studies does not rely on assumptions made about the possible influences of factors that have led participants to choose whether or not to take the therapy.

Of course, interventional studies rely on many other assumptions and methods to ensure they provide reliable results. Bias, including other forms of selection bias, can occur in interventional studies. For instance, there can be important differences in experimental groups at baseline in randomised interventional studies. While the random allocation can be checked to ensure that known confounders are adequately distributed, no checks are possible for unknown confounders. And even if the allocation appears appropriate at baseline, the possibility of differences arising after the allocation still needs to be minimised. This is achieved by maintaining allocation concealment (from participants and investigators), and by ensuring that the analysis does not rely on any post-allocation characteristics. Considerations such as these emphasise the care needed in conducting and interpreting any type of clinical reasearch. But, importantly, none of this undermines the benefit that interventional studies possess over observational studies. One specific source of bias is ruled out; bias due to patients *choosing* whether or not to take the treatment under investigation. And because this source of bias is eliminated, well-conducted randomised interventional studies, when compared to well-conducted observational studies, do not rely to the same extent on assumptions about the causal processes at play.

The importance of the decisions made on the basis of tests of efficacy also plays a role in the argument for interventional studies. Tests of efficacy are used by regulatory agencies to decide whether a new drug should be put on the market. Prior to marketing, data on the effects of a drug in the general population are limited. This makes it difficult (and, at times, impossible) to isolate and adjust for all of the causal factors that could influence the effects of a therapy (and be differentially distributed in groups of patients

who would select to take the therapy versus those patients who would elect not to). Given that results of tests of efficacy are used to decide an issue of such importance to public health, it is important that the test is rigourous. In this context, it is important to rule out or minimise as many potential sources of error as possible, and conducting interventional studies rather than observational studies as tests of efficacy help to achieve this rigour.

The argument for interventional studies rather than observational studies in tests of a treatment's efficacy can be seen in the medical literature. However, the argument that is presented is often less than clear, in part, due to the terminology used ('randomised trials' versus 'non-randomised studies'), and, in part, due to ambiguity in use of the term 'bias'. The following two quotes provide good examples of the argument for randomised interventional studies as it is presented in the medical literature.

> ... [R]andomisation minimises systematic errors (i.e. biases) in the estimates of treatment effects, allowing any moderate effects that exist to be detected unbiasedly in studies of appropriately large size. By contrast, observational studies—such as cohort studies and case-control studies—involve comparisons of outcome among patients who have been exposed to the treatment of interest, typically as part of their medical care, with outcome among others who were not exposed (or comparisons between those with different amounts of exposure). The reasons why certain patients received a particular treatment while others did not are often difficult to account for fully, and, largely as a consequence, observational studies are more prone to bias than are randomised trials. (Collins and MacMahon, 2007, p. 16).[8]

> The placement of RCTs at the top of the therapeutic research hierarchy has occurred due to the realisation that RCTs are superior to observational studies in evaluating treatment because RCTs *eliminate* bias in the choice of treatment assignments and provide the only means to control for unknown prognostic factors. (Devereaux and Yusuf, 2003, p. 107) [emphasis added]

The argument for 'randomisation' provided by Collins and MacMahon (2007, p. 16) is the same as the argument for interventional studies provided above. Interventional studies have the capacity to rule out bias that can arise due to participants choosing whether or not they take the treatment. Unfortunately, the contrast Collins and MacMahon draw is between

'randomisation' and observational studies. It is not a particularly good argument for randomisation, because methods other than randomisation can be used in interventional studies to eliminate this form of bias, but it is a good argument for interventional studies over observational studies. A charitable reading (put into the terminology I have been using) would suggest that Collins and MacMahon accept this, and intend their argument to be for randomised *interventional* studies over observational studies, rather than randomisation *per se*. Note that the sense of bias employed in Collins and MacMahon's argument is informal. Here bias refers to "a process at any stage of inference tending to produce results that depart systematically from the true values" (Fletcher et al., 1996, p. 7). This sense of bias is commonly seen in the medical literature. But there are other senses of 'bias'. In particular, the technical sense of 'bias' that is used in parameteric statistics. I will refer to this as 'statistical bias'.

Statistical bias refers to the expectation of a estimator of an unknown parameter. An estimator is unbiased if the mean of the estimator over the entire sample space equals the true value of the unknown parameter.[9] A test-statistic is an estimator of the unknown parameter under investigation in a clinical trial. If the test-statistic provides an unbiased estimator of the unknown parameter, then, were the trial repeated indefinitely, the mean value of the test-statistic would equal the true value of the unknown parameter. Of course, trials are never repeated indefinitely.[10] Nevertheless statistical bias is put to use in the epidemiological literature to argue for randomisation in interventional studies.

Devereaux and Yusuf are referring to statistical bias when they suggest that randomised interventional studies eliminate bias in the choice of treatment assignments. Part of the argument is sound, providing 'bias' is interpreted as statistical bias. Randomisation does eliminate statistical bias, and does control (in a sense to be articulated) for known and unknown prognostic factors. But as Worrall (2007a; 2002; 2007b) has convincingly argued, the elimination of statistical bias does not provide an adequate justification for placing randomised interventional studies at the top of the therapeutic research hierarchy of evidence. Statistical bias refers to the influence of confounding factors in *expectation*—that is, the expected influence of prognostic factors over the entire sample space. This means that were the trial to be repeated indefinitely, with a new random allocation of trial participants performed on each occasion, then, in the indefinite sequence of trials, the imbalance of any confounder on a particular random allocation will be counteracted by the distribution of that confounder on another allocation. The overall effect of known and unknown confounders in the indefinite sequence

of trials will be zero. However, on any *particular* randomised allocation of experimental groups there is a possibility that confounding factors will be unevenly distributed between the groups—and hence, the more informal sense of bias can still occur. While the distribution of known confounders can be checked after any single random allocation, this is clearly not possible for unknown confounders. Indeed, as noted by both Worrall (2002, p. S324) and Howson and Urbach (2006, pp. 195–6), the probability that any single confounder is unevenly distributed on any particular random allocation ranges from zero to one.

Proponents of the argument that randomisation is essential in medicine—because it eliminates statistical bias from uneven distribution in confounding factors at baseline—all too often neglect to explicate what eliminating 'statistical bias' actually amounts to. Often any reference to expectation, or the indefinite repetition of trials, is obscured or absent (as is this case in the quote provided from Devereaux and Yusuf). The argument is made as if randomisation ensures the even distribution of known and unknown confounders in a particular random allocation. Were randomisation to achieve this, then it may be considered *sine qua non*. Randomisation provides categorical assurances on the the distribution of known and unknown confounders, but only in the indefinite sequence of trials.

Randomisation in interventional studies does not ensure the even distribution of all possible confounders at baseline. But, interventional studies do rule out bias (of the first, more informal sense) originating from the differential distribution of confounders linked to whether or not a patient chooses to take a particular therapy. Observational studies can certainly provide important scientific information. And, when we have a rich understanding of the causal processes at play, the epistemological benefits of interventional studies will be minor. But none of this undermines the importance of the distinction between interventional studies and observational studies in the context of testing the efficacy of drugs.

Worrall (2007a, pp. 1017–1018) acknowledges that observational studies are more prone to certain types of bias than randomised interventional studies, but gives this very little emphasis. Instead, Worrall makes the point that providing the treatment effects are sufficiently large, the effects of selection bias in observational studies are likely to be too small to entirely swamp the beneficial effects of the treatment. This is true, but does not resolve the problem for decision makers. Estimating the magnitude of any treatment effect is vital for decisions (a point that Worrall 2007a, p. 1017 accepts). Selection bias may lead to either an over- or under-estimation of the effect of treatment witnessed in an observational study. Even if the potential bias

is smaller than the size of the treatment effect, it will often be impossible to estimate the magnitude of the bias—or for that matter its direction—with any degree of confidence. The difficulty this poses for therapeutic decision makers underlines the importance of interventional studies when testing the efficacy of therapies.

3.2 Why *randomised* interventional studies

It is easier to argue for randomisation in interventional studies than it is to argue for randomisation *simpliciter*. Once the need for interventional studies for testing the efficacy of treatments is established, the question moves to how participants should be allocated. In the context of large drug trials, randomised allocation has a number of practical and epistemic benefits. Bayesian considerations provide a rationale for randomisation as the preferred method of ruling out allocation bias in interventional studies. A Bayesian argument for randomisation in interventional studies has been provided by Lindley (1982) and Suppes (1982) (and extended upon by Berry and Kadane (1997) and Kadane and Seidenfeld (1990)).[11] An outline of the argument is provided below.

Lindley recognises the importance of the experimental groups being evenly matched for known confounding factors. And also that whether the experimental groups are considered well-matched is ultimately a subjective judgment. He refers to an allocation in which the investigator judges the experimental groups evenly matched as 'haphazard'. The need for the allocated groups to be haphazard is more important than randomisation (Lindley, 1982, p. 439). But, while it is possible for the Bayesian to deliberately allocate the experimental groups, and hence ensure a haphazard design, the complexity accrues quickly. The expected utility of each possible allocation needs to be calculated. To calculate this utility the effect of each possible confounder needs to be estimated. This is no easy task when many of the confounding factors are merely plausible rather than well understood, which is precisely the situation in the clinical sciences. In response to this problem, Lindley suggests a "reasonable approximation to the optimum design" is to randomise and then check to ensure the groups are haphazard (Lindley, 1982, p. 439).

> Consequently the two, apparently conflicting, views of the randomiser and the Bayesian have been brought into agreement. It is the haphazard nature of the allocations, not the random element, that is important; and the use of a haphazard design saves

> the Bayesian a lot of trouble, with small chance of any appreciable gain, by producing a situation relatively easy to analyse. A further point is that a detailed, Bayesian consideration of possible covariates would almost certainly not be robust in that the analysis might be sensitive to small changes in the judgements about covariates.

The final sentence of this quote recognises the importance of context. If the number, and overall effect, of possible confounders is uncertain, then the Bayesian calculation needed for deliberate matching becomes difficult. This is certainly the case for the large medical trials conducted to test the efficacy of a new drug treatment. In much smaller trials, with few possible confounders, deliberate matching of the experimental groups may be more convenient than randomising and checking whether the groups are well-matched (and re-randomising if necessary).

Suppes (1982) also provides Bayesian reasons to randomise. In situations such as in the clinical sciences, where knowledge of causal processes is available, but unable to accurately predict the response to treatment, randomisation both simplifies computation, and aids communication to (and acceptance by) a sceptical scientific audience. These two reasons for the Bayesian to randomise are linked. Randomising simplifies the Bayesian's computation for the reasons noticed by Lindley. There is no shortage of *plausible* causal processes in the clinical sciences, rather, what the clinical sciences lack is knowledge of how these plausible causal processes will interrelate in any particular therapeutic situation. Thus, in addition to a small number of well known, and reasonably well understood confounders—which can be checked to ensure they are evenly distributed in the experimental groups—there is a potentially limitless combination of additional causal processes that may effect the outcome of the treatment. The Bayesian can take (a least some of) these additional plausible causal processes into consideration in forming a prior based on a particular allocation of experimental groups, but the resulting prior will have a high variance, and be dependent on personal judgements. Randomising (with a check to ensure the allocation is haphazard) and *assuming* that randomisation has either resulted in the even distribution of additional plausible (but uncertain, or yet to be fully elucidated) causal factors, or that such factors are likely have a minimal effect on the intervention, provides a rationale for adopting a much simpler prior distribution, as well as helping narrow the many ways that the experimental results could be incorporated into the likelihood function (see Suppes, 1982, pp. 464–466).[12]

This simplification of the Bayesian computation also aids communicating the experimental results. Simplification directly assists communication, a point that should not be ignored. But perhaps more persuasively, randomising can provide some common ground for agreement on both the Bayesian prior and the analysis of the experimental results (that is, randomisation may aid agreement on the experimental distribution—whether a Bayesian or frequentist analysis is to be conducted). Of course, this requires the audience to grant the assumption that merely plausible confounding factors are either evenly distributed or unlikely to affect the results. (Remember, any well understood confounders will be checked to ensure they are evenly distributed). But if this assumption is granted, there is a much improved possibility of reaching agreement on how the experiment should be analysed.

The possibility of reaching agreement on the personal judgements needed by the Bayesian to justify a particular deliberate allocation is much less likely. The possibility of consensus is an important advantage of randomisation. It blocks a certain type of sceptical reply—a scepticism towards the personal judgements of the investigator. The more assumptions needed for the analysis, the more difficult it is to persuade a sceptical scientific audience. This is directly analogous to the problem encountered in observational studies—interpretation of observational studies is more difficult because the analysis relies on the assumptions made by the investigators when they deliberately match the experimental groups or make *post hoc* adjustments.

A sceptical reply can, of course, also be made against a randomised interventional study, but it is of a different character. A sceptical audience may question the analysis of a randomised trial on the basis of an unevenly distributed plausible confounding factor. Indeed, a reply of this sort is always possible given that randomisation does not ensure the even distribution of confounders on any particular allocation. In contrast to non-randomised interventional studies, however, the burden of proof is on the sceptic rather than the investigator. The sceptic needs to argue that the differential distribution of the causal factor is sufficient to explain the experimental results. In the clinical sciences the debate sometimes follows such a path. The alternative hypothesis proffered by the sceptic, if plausible, can then be tested in a new trial. Such toing-and-froing plays an important role in how knowledge accumulates in the clinical sciences.

Where does the Bayesian justification for randomising in interventional studies leave a dyed-in-the-wool frequentist? After all, contemporary clinical trials are almost exclusively analysed by frequentist methods. Let me finish with a couple of brief comments that may convince a frequentist that the Bayesian rationale for randomising should not be seen as too much of a

problem for continued frequentist analysis of clinical trials (if that is your view on how clinical trials should be analysed).

First, randomisation—done for whatever reason—provides the experimental distribution that underpins the frequentist hypothesis test. Surely, the important issue for the frequentist statistician is that the experimental groups were randomised, not why they where randomised.[13] Second, a sticking point in the Bayesian argument for randomisation for the frequentist is presumably its reliance on personal judgements. The frequentist position is often construed as rejecting *any* reliance on subjective judgement in inference. Such a position, however, is not tenable (and I doubt it is held by too many frequentist statisticians when explicitly discussed). As Howson and Urbach (2006) show (time and time again) the judgement of the investigator (or analyst) plays a central role in frequentist inferences. Instead, the frequentist might restrict their view to a rejection of subjective judgement playing a role in drawing inferences from observed data once the experimental analysis has been specified.[14] If this restriction is accepted, then perhaps the frequentist can accept the Bayesian justification for randomisation, or develop a justification along similar lines to the argument that has been provided, without being explicitly Bayesian.

4 Conclusion

Randomisation does not provide the guarantee that all possible confounders are evenly distributed in experimental groups, and therefore does not provide some irrefutable epistemic good. However, given fallible access to knowledge of causal processes in the clinical sciences, *some* epistemic good is provided by conducting randomised interventional studies rather than observational studies. Randomised interventional studies rule out a source of bias that can occur in observational studies. Worrall is right to argue for a more positive view of observational studies than that provided in much of medicine (especially in evidence-based medicine) (Worrall, 2002, p. S329). But randomised interventional studies are not epistemologically equivalent to observational studies when known confounders appear adequately balanced. There is good reason to conduct randomised interventional studies rather than observational studies when testing the efficacy of drugs. Of course, efficacy is only part of the story when it comes to informing therapeutic decisions. Once the drug is on the market, clinicians need to make judgements about the *effectiveness* of a therapy in individuals and groups. These judgements rely on the external validity of clinical research, and raise a number of additional

challenges.

Acknowledgements

To be added.

Notes

[1]See Worrall (2007b, p. 486) (quoted in Section 2), and Worrall (2002, p. S328): "It is difficult to see how we can do better than control, whether in advance or post hoc, for all plausible alternatives."

[2]In this context historically controlled trials might be considered quasi-interventional. The main reason historically controlled trials are infrequently conducted in contemporary medicine is because they are seen to be epistemologically inferior to randomised interventional studies. And due to this epistemological inferiority historically controlled trials are considered unethical for testing new, or 'experimental' treatments. If Worrall's epistemological are arguments successful, then the ethics of these trials needs to be reconsidered.

[3]In other places, Worrall (2007a, pp. 1017–1018) recognises that observational studies are exposed to additional sources of bias when compared to randomised interventional studies, but he dismisses the importance of this on the basis that any such biases are unlikely to influence the direction of the observed effects. I discuss this view below.

[4]See, for instance, hierarchies of evidence given by proponents of evidence-based medicine, such as Phillips et al. (2009).

[5]I will focus on the distinction between randomised interventional studies versus *observational* studies, rather than historically controlled trials. Observational studies, such as cohort and case-control studies are the studies more frequently used in contemporary medical research. However, the argument I provide for an epistemological distinction between randomised interventional studies and prospective cohort studies is just as relevant if historically controlled cohort studies are used for comparison.

[6]See La Caze (forthcoming) for further discussion on the links between basic science and clinical research.

[7]Note, it is the *possibility* of this specific form of selection bias that differentiates observational studies from interventional studies—a possibility that cannot be ruled out because of the fallibility of causal knowledge in clinical science.

[8]For an earlier example, see Yusuf et al. (1984)

[9]Stuart and Ord (1991, p. 609) provide this definition of bias. Stuart and Ord emphasise that this sense of bias "should not be allowed to convey overtones of a non-technical nature".

[10]Even if a trial (which was perfectly internally valid) was to be repeated an incredible number of times, the mean of the observed values of the test-statistic would not necessarily equal the true value of the unknown parameter. The assurance of unbiasedness refers to the estimation of the test-statistic over the *entire sample space*. In any long run of trials, even an infinite long run, the estimator will not necessarily equal the true value of the unknown parameter.

[11]Lindley's position is of particular interest because Worrall attributes to him a different view. Worrall suggests there is no straightforward Bayesian argument for randomisation, and enlists Lindley's support.

> As always with Bayesianism, there are a variety of positions on offer (the phrase "*the* Bayesian account" always makes me smile), but the most straightforward one articulated, for example, by Savage (who later however, for reasons it seems difficult to fully understand, decided it was 'naive') and Lindley, as we in effect noted earlier, sees no role for randomisation here at all. (Worrall, 2007b, p. 482)

Perhaps Worrall is best interpreted as suggesting there is no Bayesian argument for randomisation to be *sine qua non* (and on this we would agree), because Lindley certainly acknowledges a role for randomisation in certain contexts. (Note, in this quote Worrall is referring his earlier recognition of Lindley's suggestion that as the number of confounders increases, the probability is high that any one of these confounders ends up being unevenly distributed in the randomly allocated groups—an important point, but not Lindley's last word on the role of randomisation in experiments.)

[12]Clearly, simplifying the prior distribution and the likelihood function are benefits if a Bayesian analysis of the clinical trial is to be conducted. It should be noted that randomising also simplifies the experimental distribution for frequentist statistical analysis. I discuss some of the ramifications of adopting a Bayesian justification for randomisation for the continued frequentist analysis of clinical trials below.

[13]The arguments against randomised allocation provided by Worrall (2007a) and Howson and Urbach (2006, pp. 188–194) (and others) are arguments against randomisation (*per se*) somehow guaranteeing the results of large clinical studies, *not* against randomised allocation being a convenient method for arriving at experimental groups that are adequately (or roughly) similar for a limited number of confounders in large interventional studies.

[14]I think this is a more accurate representation of the view, but in any case, it is certainly more tenable. It is difficult to see how subjective judgement can be excluded from the specification of an experiment.

References

Berry, Scott M. and Joseph B. Kadane. 1997. Optimal bayesian randomization. *Journal of the Royal Statistical Society. Series B (Methodological)*, 59(4):813–819.

Collins, Rory and Stephen MacMahon. 2007. Reliable assesment of the effects of treatments on mortality and major morbidity. In Peter M. Rothwell, editor, *Treating Individuals: From randomised trials to personalised medicine*. Philadelphia: Elsevier.

Devereaux, P. J. and Salim Yusuf. 2003. The evolution of the randomized controlled trial and its role in evidence-based decision making. *Journal of Internal Medicine*, 254:105–113.

Fletcher, Robert H., Suzanne W. Fletcher, and Edward H. Wagner. 1996. *Clinical Epidemiology: The Essentials*. Baltimore: Lippincott Williams and Wilkins, 3rd edition.

Howson, Colin and Peter Urbach. 2006. *Scientific Reasoning: The Bayesian Approach*. Chicago: Open Court, 3rd edition.

Kadane, Joseph B. and Teddy Seidenfeld. 1990. Randomization in a bayesian perspective. *Journal of Statistical Planning and Inference*, 25(3):329–345.

La Caze, Adam. forthcoming. The role of basic science in evidence-based medicine. *Biology and Philosophy*.

Lindley, Dennis V. 1982. The role of randomization in inference. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1982:431–446.

Phillips, Bob, Chris Ball, David L. Sackett, Doug Badenoch, Sharon E. Straus, R. Brian Haynes, and Martin Dawes. 2009. Oxford Centre for Evidence-Based Medicine Levels of Evidence. Technical report.

Stuart, Alan and J. Keith Ord. 1991. *Kendall's Advanced Theory of Statistics*, volume 2. London: Edward Arnold, 5th edition.

Suppes, Patrick. 1982. Arguments for randomizing. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1982:464–475.

Urbach, Peter. 1985. Randomization and the design of experiments. *Philosophy of Science*, 52:256–273.

———. 1993. The value of randomization and control in clinical trials. *Statistics in Medicine*, 12:1421–1431.

Worrall, John. 2002. What evidence in evidence-based medicine? *Philosophy of Science*, 69:S316–S330.

———. 2007a. Evidence in medicine and evidence-based medicine. *Philosophy Compass*, 2(6):981–1022.

———. 2007b. Why there's no cause to randomize. *British Journal for the Philosophy of Science*, 58(3):451–488.

Yusuf, Salim, Rory Collins, and Richard Peto. 1984. Why do we need some large, simple randomized trials? *Statistics in Medicine*, 3:409–420.