Are Outcome-Adaptive Allocation Trials Ethical?

Spencer Phillips Hey, Ph.D [Postdoctoral Fellow] and

Studies of Translation, Ethics and Medicine (STREAM), Biomedical Ethics Unit, McGill University,
3647 Peel Street, Montréal QC H3A 1X1

Jonathan Kimmelman, Ph.D [Associate Professor]

Studies of Translation, Ethics and Medicine (STREAM), Biomedical Ethics Unit, McGill University,
3647 Peel Street, Montréal QC H3A 1X1

Abstract

Randomization is firmly established as a cornerstone of clinical trial methodology. Yet, the ethics of randomization continues to generate controversy. The default, and most efficient, allocation scheme randomizes patients equally (1:1) across all arms of study. However, many randomized trials are using outcome-adaptive allocation schemes, which dynamically adjust the allocation ratio in favor of the better-performing treatment arm. Advocates of outcome-adaptive allocation contend that it better accommodates clinical equipoise and promotes informed consent, since such trials limit patient-subject exposure to sub-optimal care. In this essay, we argue that this purported ethical advantage of outcome-adaptive allocation does not stand up to careful scrutiny in the setting of two armed studies and/or early phase research.

Keywords

adaptive randomization; ethics; equipoise; therapeutic misconception

Randomization is firmly established as a cornerstone of clinical trial methodology. Yet, the ethics of randomization continues to generate controversy. On its face, random allocation appears to conflict with both a patient's best interest and the ethics of clinical practice. Why would a patient ever agree to receive a randomly chosen treatment?

The concept of clinical equipoise provides an answer and helps to resolve this fundamental tension. Clinical equipoise demands that there exists a state of community uncertainty about the relative therapeutic merits across all arms of a trial. Insofar as this condition holds, all patient-subjects enrolled in a trial can be assured of receiving nothing less than competent medical care. Generally speaking, the randomized controlled trials (RCT) for serious illnesses are ethically acceptable only where conditions of clinical equipoise hold [1] (for a survey of debates over clinical equipoise, see [2]).

However, clinical equipoise is silent on the nature of the random allocation scheme. The default, and most efficient, allocation scheme randomizes patients equally (1:1) across all

arms of study. But many RCTs use unequal allocation schemes (e.g., 2:1 or 3:1), which assign more patients to the experimental intervention [3], or outcome-adaptive allocation schemes, which dynamically adjust the allocation ratio in favor of the better-performing treatment arm [4]. These alternative allocation schemes are often justified by appeals to increasing patient-subject benefit [5–6]. That is, by weighting the allocation ratio in favor of the experimental intervention or “better performing” arm, the trial increases the number of patients who receive the presumed superior treatment. Therefore, advocates contend that unequal allocation schemes can better accommodate clinical equipoise, since such trials limit patient-subject exposure to sub-optimal care [7].

While this argument is appealing, it deserves careful scrutiny. Elsewhere, we argued that unequal allocation ratios in confirmatory trials are ethically problematic, since they fail to minimize patient burden, leverage patients’ therapeutic misestimations, and potentially introduce new internal validity threats [8]. What about trials that use outcome-adaptive allocation?

In what follows, we consider whether criticisms of unequal allocation apply to two armed studies using adaptive allocation (for examples see [9–10]). These studies—which are the most contentious adaptive allocation study types—generally begin by randomizing patients on a 1:1 basis to treatment or control. As preliminary patient-response data emerge, the allocation ratio is adjusted in favor of the better-performing arm. Once the probability of accepting or rejecting the null hypothesis has dropped below a pre-specified threshold, the trial is closed. After reviewing the ethics of two armed adaptive allocation trials, we briefly consider the ethics of a less contentious study design: multi-arm adaptive allocation (for examples see [11–12]).

1 Minimizing Patient Burden

A foundational ethical principle is that trials should minimize research burdens. Proponents of outcome-adaptive allocation designs argue that they advance this charge, since they maximize the number of patients receiving superior treatments (and consequently minimize the number receiving inferior ones) [5–6].

This argument holds little water when considered against the background of drug development. First, the primary objective of early phase clinical testing is to establish the evidentiary grounds for conducting confirmatory trials. Whereas confirmatory trials employ clinical endpoints (e.g. survival) and large sample sizes, early phase trials typically use surrogate endpoints (e.g. tumor response) that can be collected in a shorter time frame, and smaller sample sizes. That surrogate benefit and underpowered studies are fickle guides for clinical benefit is indicated by the frequent discordance between effect sizes in phase 3 studies and those in phase 2 [13]. Indeed, the average probability that an intervention reaching phase 3 testing will advance to regulatory licensure—a good proxy for clinical utility—is approximately 50%. This probability is considerably lower in realms like neurology and oncology, where adaptive allocation is frequently employed¹⁶; not infrequently, agents that look very promising in early phase trials actually make patients worse off against comparators in late phase trials. This means that, in a best case scenario

where one drug outperforms another in an outcome-adaptive study, patients receiving the better performing drug still only have a 50% probability of receiving a drug that is competitive with or better than standard of care [14].

Second, new treatments tend to deliver only small improvements over standard ones. Several systematic reviews of randomized controlled trials in cancer, for example, show that the odds ratios cluster close to unity [15]. However, adaptive randomization tends to offer advantages only when effect differences between two interventions are large [5]. As a consequence, the purported therapeutic advantages of adaptive randomization will rarely be realized, yet their disadvantages in terms of requiring larger sample sizes will often be encountered.

These arguments take on greater force in the context of phase 3 trials. In favor of adaptive allocation in this context is the fact that allocation adjustment will be based not on surrogate, but rather clinical endpoints. This means that better performing arms are much more likely to be clinically superior than they would be in phase 2 studies. This advantage, however, will often- though not always- be nullified by the lengthy period between enrollment and observation of clinical outcomes. These trials must therefore be designed to either recruit very slowly, in order to gain the benefits of adaptive allocation, or else they may have already completed enrollment by the time sufficient outcome evidence is available to adjust the allocation ratio.

Advocates of adaptive allocation might offer two rejoinders. First, they might argue that inefficiencies described above work to the advantage of patients, since larger sample sizes afford more opportunities for patients to enter trials and access superior treatment (e.g., Berry (2011) seems to suggest this point [16]). Leaving aside the fact that this advantage only holds if standard of care proves inferior, this reasoning neglects the economies that arise in dividing labor. Research systems are set up to resolve scientific uncertainties; care systems are established to deliver therapies based on that resolution. Running larger trials because they offer patients better care options asks research systems to shoulder tasks that care systems are designed to carry. It is far better to resolve scientific questions as efficiently as possible, so that on the one hand research resources can be channeled towards resolving other uncertainties, while on the other, the therapeutic “baton” can be expeditiously passed to systems that are designed to deliver care.

Concerning the efficiency of the research enterprise as a whole, all commentators agree that outcome-adaptive trials are more complicated and expensive to plan and coordinate. While some research centers [4] may be able to implement outcome-adaptive allocation as a rule, the requisite funding and stakeholder support cannot be assumed to hold across the research enterprise. To be sure, the current unavailability of support does not mitigate the force of proponents who claim that trials ought to be done this way, but it does undermine the claim that we can just “slot in” adaptive allocation into the usual course of research. Absent this support, and in light of the FDA’s ambivalence toward the evidence produced in adaptive trials [17], outcome-adaptive trials seem more likely to make research less efficient on the whole.

A second argument in defense of two-armed adaptive randomization might be made for trials employing placebo comparators in spite of the existence of standard care options. In such cases, adaptive randomization will indeed decrease the number of patients receiving placebo, should the novel agent show activity. However- we regard such departures from clinical equipoise as unethical- at least in the context of serious illnesses. The appropriate way to resolve the ethics here is to use placebo- or to advise patients against entering such trials.

In the end, it seems doubtful that adaptive allocation generally improves risk/benefit for patients. To the contrary, the larger sample sizes translate to more patients enduring more research procedures and extra clinic visits. Since costs scale with sample sizes, it also means more resources are consumed in answering a single research question than would have been the case with a 1:1 design.

2 Informed Consent

Another cornerstone principle in medical research is that trials should safeguard the autonomy of research subjects. This can pose challenges in trials involving life-threatening illness, because patients can lack a realistic understanding of risk/benefit (therapeutic misestimation [18]) or they can fail to understand the ways that study protocols antagonize treatment objectives (“therapeutic misconception” [19]). Random allocation with fixed proportion is a good example of a protocol-driven element that often frustrates a patient’s treatment objectives. Some have suggested that, by using all available information to adjust a protocol, adaptive allocation offers a “partial remedy” for the therapeutic misconception [6].

For reasons canvassed in the previous section, this argument is not convincing in settings where allocation adjustment is driven by surrogate and short-term outcomes. Indeed, given the marginality of advantage and the clinical uncertainty that should exist after a phase 2 study is ended, adaptive allocation seems instead to invite confusion about risk/benefit rather than dispel it.

But there are also other problems. Studies show that many subjects struggle with the concept of randomization, believing erroneously that investigators will allocate them to the treatment arm they believe is most favorable [19]. Imagine that all patients enter trials incorrectly believing that study teams will allocate them to the best possible treatment. That is, they enter the trial with 100% confidence that they will receive the drug believed to work best by the study team, a highly plausible misconception in the context of a trial with adaptive randomization. Imagine further that as the trial progresses, the study drug trends toward superiority against standard of care and the allocation ratio is adjusted to 4:1. Now, all patients harbor less wrong beliefs on trial entry, since indeed their probability of being allocated to the arm that is believed to be superior is 80% rather than 50%. This is the remedy for therapeutic misconception that proponents of adaptive allocation point out. However, valid adaptive allocation designs dictate allocation of some patients to the flagging arm—in our example, every fifth patient on average. Those patients, who also entered the study believing they will be allocated to the better arm, are now allocated to a treatment that is actually believed to be inferior. Contrary to dissolving the research-treatment distinction,

outcome-adaptive trials simply redistribute its tensions, concentrating misunderstandings for patient-subjects in the flagging arm.

Previously, we argued that using uneven, fixed allocation ratios works against informed consent by leveraging unrealistic patient therapeutic expectations to the advantage of clinical investigators [8]. It also provides no incentive for investigators to work with patients to dispel unrealistic expectations. These arguments almost certainly hold for early phase trials using adaptive allocation. They probably also hold for most late phase trials, where the advantages of outcome-adaptive allocation are even less compelling.

3 Validity

The overarching aim of trials is to deliver evidence that can support clinical decision-making. First and foremost, this requires that studies support valid inferences about the causal relationship between an intervention and a clinical effect. The very purpose of randomization is to limit the confounding effects of population variability, since randomization all but ensures that patients in all arms of a study are drawn from the same population. It is therefore critical to consider whether adaptive allocation introduces validity threats that are avoidable using other means.

Both critics and proponents alike acknowledge that adaptive allocation introduces some threats to internal validity. In the first place, trials are dynamic entities. Populations and treatments change over the course of testing; new research sites drawing from different populations might be added or protocols might be modified; investigators become more skilled with delivery techniques. The trial environment is dynamic as well: care standards can change over the course of a study, as can supportive therapies and reagent suppliers. These factors all have the potential to introduce bias into the treatment comparisons when allocation proportions are changed. Moreover, adaptive allocation should have an influence on physician-investigator behavior: Given that the odds of receiving the better treatment will continually improve over the course of the trial, it is in the best interests of patient-subjects (and the physicians advocating on their behalf) to wait and enroll as late as possible [20].

Of course, not all eligible patient-subjects can afford to wait. For some potential patient-subjects—particularly those who are more seriously ill or treatment-refractory—enrollment in research may be their only chance for an effective treatment. This introduces statistical challenges due to the systematic difference in the early- versus late-enrolling patient populations [21]. The later-enrolling population—more likely to receive the better treatment—is likely to be healthier. As explicitly acknowledged by the FDA in their guidelines concerning adaptive methods, this predictable time-trend in the study population increases the risk of a biased effect estimate [17]. Block design methods can reduce some of these validity threats. However, these stymie the very advantages of using adaptive randomization [22].

4 The Multi-Arm Context

Let us now briefly consider outcome-adaptive allocation in the context of multi-arm trials. These studies evaluate many different research questions at once. For example, they might

compare the efficacy or toxicity of three or more different drug-dosage regimens; or they might evaluate response rates of three or more biomarker populations. Again, as patient-subject data becomes available, the probability of being allocated to a particular treatment arm is adjusted in accord with the response rate (or sometimes a poorly performing arm may be dropped altogether). In contrast to the above two-arm trial, whose goal is often to reject the null hypothesis and demonstrate efficacy, the goal in these trials is often to identify the most promising regimen or patient population for further investigation.

Given their ability to evaluate many different hypotheses at once, outcome-adaptive multi-arm trials may have advantages over rigid allocation schemes [16,22]. Gains in efficiency are ethically attractive, and by eliminating flagging arms relatively quickly, multi-armed adaptive studies may also diminish total subject burden associated with testing multiple hypotheses. However, for the same reasons canvassed in the previous sections, multi-armed designs in early phase settings seem no more likely to enhance therapeutic benefit than two armed adaptive allocation schemes. Nor are they likely to alleviate moral concerns about therapeutic misconception.

5 Conclusion

We are dubious of the suggestion that adaptive allocation studies offer generic ethical advantages. At least in the two-arm setting, they seem to worsen total burden by increasing patient exposure to research procedures and to drugs that—even at the end of testing—remain unproven. We are also skeptical that adaptive allocation alleviates the therapeutic misconception. Instead, telling patients that allocation will be adjusted according to evolving (but still highly fallible) evidence seems to invite therapeutic overestimation, while leaving unresolved tensions between the scientific and care objectives in research. Adaptive allocation introduces new sorts of validity threats—though these will vary from study to study. Finally, at least in the two-arm setting, adaptive allocation asks research systems to shoulder some of the weight that care systems are designed to carry. So can we say that adaptive allocation is unethical?

We can contemplate several scenarios where adaptive allocation might have a compelling ethical basis. First, we remain open to the application of multi-armed adaptive allocation studies—not because they increase therapeutic advantage or ameliorate consent problems, but rather because they may enable more efficient resolution of uncertainty where many hypotheses are entertained. Second and related, validity threats need not deter adaptive allocation where outcomes are collected over short periods, or where there are sound reasons to anticipate stability in the trial environment. Third, consent discussions should at least explain that adaptive allocation involves allocating some patients to treatments that accumulating evidence disfavors, and that at least in the case of studies using surrogate endpoints and small sample sizes, even if one arm performs better than another, there remains substantial uncertainty about its actual clinical utility. Indeed, were this disclosed to patients, one suspects purported recruitment advantages for adaptive allocation studies [23] would diminish.

In the end, the principal question that should guide evaluation of novel trial designs is whether they can resolve medical uncertainties with the least patient burden and fewest resources. We favour a rebuttable presumption against adaptive randomization in the setting of two armed trials.

Acknowledgments

The authors thank referees for helpful comments. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors

References

1. Weijer C, Miller PB. When are research risks reasonable in relation to anticipated benefits? *Nat Med*. 2004; 10:570–573. [PubMed: 15170195]
2. London, AJ. Clinical equipoise: Foundational requirement or fundamental error?. In: Steinbock, B., editor. *The Oxford Handbook of Bioethics*. New York: Oxford University Press; 2007. p. 571–596.
3. Dumville JC, Hahn S, Miles JNV, Torgerson DJ. The use of unequal randomisation ratios in clinical trials: a review. *Contemp Clin Trials*. 2006; 27:1–12. [PubMed: 16236557]
4. Biswas S, Liu DD, Lee JJ, Berry DA. Bayesian clinical trials at the university of texas md anderson cancer center. *Clin Trials*. 2009; 6:205–216. [PubMed: 19528130]
5. Lee JJ, Chen N, Yin G. Worth adapting? Revisiting the usefulness of outcome-adaptive randomization. *Clin Cancer Res*. 2012; 18:4498–4507. [PubMed: 22753588]
6. Meurer WJ, Lewis RJ, Berry DA. Adaptive clinical trials: A partial remedy for the therapeutic misconception? *JAMA*. 2012; 307:2377–2378. [PubMed: 22692168]
7. Collins SP, Lindsell CJ, Pang PS, et al. Bayesian adaptive trial design in acute heart failure syndromes: moving beyond the mega trial. *Am Heart J*. 2012; 164:138–145. [PubMed: 22877798]
8. Hey SP, Kimmelman J. The questionable use of unequal allocation in confirmatory trials. *Neurology*. 2014; 82:77–79. [PubMed: 24306005]
9. Maki RG, Wathen JK, Patel SR, et al. Randomized phase ii study of gemcitabine and docetaxel compared with gemcitabine alone in patients with metastatic soft tissue sarcomas: results of sarcoma alliance for research through collaboration study 002. *J Clin Oncol*. 2007; 25:2755–2763. [PubMed: 17602081]
10. Garcia-Manero G, Jabbour E, Borthakur G, et al. Randomized open-label phase ii study of decitabine in patients with low-or intermediate-risk myelodysplastic syndromes. *J Clin Oncol*. 2013; 31:2548–2553. [PubMed: 23733767]
11. Giles FJ, Kantarjian HM, Cortes JE, et al. Adaptive randomized study of idarubicin and cytarabine versus troxacitabine and cytarabine versus troxacitabine and idarubicin in untreated patients 50 years or older with adverse karyotype acute myeloid leukemia. *J Clin Oncol*. 2003; 21:1722–1727. [PubMed: 12721247]
12. Kim ES, Herbst RS, Wistuba II, et al. The battle trial: Personalizing therapy for lung cancer. *Cancer Discov*. 2011; 1:44–53. [PubMed: 22586319]
13. Zia MI, Siu LL, Pond GR, Chen EX. Comparison of outcomes of phase ii studies and subsequent randomized control studies using identical chemotherapeutic regimens. *J Clin Oncol*. 2005; 23:6982–6991. [PubMed: 16192585]
14. Hay M, Thomas DW, Craighead JL, et al. Clinical development success rates for investigational drugs. *Nature Biotechnol*. 2014; 32:40–51. [PubMed: 24406927]
15. Djulbegovic B, Kumar A, Glasziou P, et al. Trial unpredictability yields predictable therapy gains. *Nature*. 2013; 500:395–396. [PubMed: 23969443]
16. Berry DA. Adaptive clinical trials: the promise and the caution. *J Clin Oncol*. 2011; 29:606–609. [PubMed: 21172875]
17. FDA, et al. Guidance for industry: adaptive design clinical trials for drugs and biologics. 2010.
18. Horng S, Grady C. Misunderstanding in clinical research: Distinguishing therapeutic misconception, therapeutic misestimation, & therapeutic optimism. *IRB*. 2003; 1:11–16.

19. Appelbaum PS, Roth LH, Lidz CW, et al. False hopes and best data: Consent to research and the therapeutic misconception. *Hastings Cent Rep Report*. 1987; 17:20–24.
20. Saxman SB. Ethical considerations for outcome-adaptive trial designs: A clinical researcher's perspective. *Bioethics*. Epub ahead of print 19 Feb 2014.
21. Der Graaf R, Roes KCB, Van Delden JJM. Adaptive trials in clinical research: Scientific and ethical issues to consider. *JAMA*. 2012; 307:2379–2380. [PubMed: 22692169]
22. Korn EL, Freidlin B. Outcome-adaptive randomization: Is it useful? *J Clin Oncol*. 2011; 29:771–776. [PubMed: 21172882]
23. Tehranisa JS, Meurer WJ. Can response-adaptive randomization increase participation in acute stroke trials? *Stroke*. 2014; 45:2131–2133. [PubMed: 24916909]