Can unequal be more fair? Ethics, subject allocation, and randomised clinical trials

Andrew L Avins  Veterans Affairs Medical Center, San Francisco and University of California, San Francisco, USA

Abstract
Randomised clinical trials provide the most valid means of establishing the efficacy of clinical therapeutics. Ethical standards dictate that patients and clinicians should not consent to randomisation unless there is uncertainty about whether any of the treatment options is superior to the others ("equipoise"). However, true equipoise is rarely present; most randomised trials, therefore, present challenging ethical dilemmas. Minimising the tension between science and ethics is an obligation of investigators and clinicians. This article briefly reviews several techniques for addressing this issue and suggests that unbalanced randomisation, a technique rarely employed in current clinical trial practice, may be useful for enhancing the ethical design of human experimentation.

Keywords: Medical ethics; clinical trials; research design; random allocation; decision making; human experimentation

In theory, a randomised clinical trial (RCT) of a new intervention should be initiated only if there is uncertainty as to whether any of the alternative treatments under study is superior.1,2 Under traditional ethical paradigms, if clinicians have any preference for one of the experimental options, they are obliged to inform patients of their preferred approach and reject randomisation.2,3

In reality, such true indifference, or "equipoise," is unusual.4,5 Presumably, investigators would not initiate a clinical trial if they didn’t have some reason to believe that the new intervention had advantages over the current standard of care. Even if the probable outcome were quite uncertain, clinicians and patients might still perceive an advantage to using “cutting edge” therapy for the emotional benefits alone.3 Furthermore, confirmatory trials (those intended to validate the results of previous experiments) cannot begin with an assumption of equipoise, since prior information about efficacy and safety exists.

When clinicians cannot sincerely claim a state of indifference, serious problems may develop. Previous clinical trials have suffered when clinicians failed to enrol all eligible patients, introducing problems of generalisability and reduced statistical power. In one example of inadequate subject recruitment, clinicians identified problems with claiming true uncertainty as an important reason for refusing to enrol many eligible subjects.6 Such situations also raise ethical concerns, since subjects are exposed to untested interventions with a diminished chance of yielding meaningful information. Zelen has proposed a set of “randomised consent” designs in which subjects are randomised prior to being asked to enrol in a trial.7 These designs may substantially aid recruitment in difficult situations, but carry their own set of methodologic and ethical concerns.8,9

If we are to continue performing clinical trials in the absence of true equipoise, how can we reconcile the ethical mandate and the practical reality? This article presents a brief overview of several proposals for addressing the dilemma and offers an appeal for more widespread application of a little-used randomisation variant for this purpose. Throughout this discussion, the type of clinical trial addressed is the “parallel-comparison-of-treatment” design, the most common type of clinical trial.10 Alternative designs, such as the cross-over trial, expose all subjects to all treatment alternatives, and do not carry the same ethical burdens as the parallel-comparison method.

A philosophical approach
Freedman has offered a unique approach to the dilemma of equipoise and randomised trials. He has argued that the absence of expert clinical consensus provides adequate grounds for resolving a research question through human experimentation. That is, the absence of true uncertainty regarding the relative benefits of alternative therapies (theoretical equipoise) is less relevant than the presence of disagreement among knowledgeable clinicians and investigators (clinical equipoise).11 There remains, however, the conflict for those clinicians who cannot accept the inconsistency among their peers as a substitute for their own
clinical intuition. There is also the problem of widespread consensus among experts who nevertheless acknowledge great uncertainty about their typical approach. Still, this argument provides a persuasive ethical foundation for many clinical trials.

An intriguing implementation of this proposal has been suggested by Korn and Baumrind. In this schema, patients are randomised not simply to alternative therapies, but to alternative physicians who have a preference for one of the treatment options. That is, when there exists substantial disagreement among treating physicians as to the preferred therapy for a patient, the patient is randomised to a physician who then treats the patient with the therapy that he or she believes is best. In this way, the fiduciary relationship between the doctor and the patient is preserved, while treatment assignment is random. The major problems with this technique are that double-blind studies are not possible and bias may be introduced if physicians who tend to provide generally superior clinical care also tend to prefer one of the treatment alternatives.

Methodologic approaches
While philosophical approaches have sought to justify the concept of randomisation, methodologic approaches take a different perspective: given that one treatment arm of a clinical trial may prove inferior, the methodologic techniques attempt to minimise the number of subjects who ultimately receive the inferior therapy. These techniques, employed singly or in combination, are not "solutions" to the ethical problem; their object is to reduce the tension, not eliminate it. Experimentation with people will always be difficult and uncomfortable; these techniques help to diminish the problems but cannot resolve them entirely.

Method 1: Adjusting the overall significance level
In most clinical trials, the criterion for statistical significance is a two-sided alpha level of 0.05. Despite the ignoble origin of the 0.05 cutpoint, few seriously question the assertion that we must be 95% confident that an outcome at least as extreme as that observed in the trial could not have occurred by chance if the null hypothesis were true (in the frequentist statistical perspective). However, are there not some research questions for which we could comfortably be 90% or 85% confident? For example, highly fatal diseases for which no adequate therapy exists may not require the same criterion as other conditions.

The number of subjects spared randomisation by altering the significance level may be substantial. Table 1 shows the reduction in the number of subjects needed to achieve statistical significance in a hypothetical clinical trial if the criterion for significance is increased from 0.01 to 0.15. For example, if the baseline success rate is 30%, increasing the significance level from 0.05 to 0.10 requires 20% fewer subjects; increasing it to 0.15 requires 34% fewer subjects.

Unfortunately, the arbitrary 0.05 cutpoint is so deeply ingrained in research practice that there seems little hope that logic will find its way into this aspect of clinical investigation.

Method 2: Group sequential designs
The concept underlying group sequential designs is that interim analyses of accumulating data in a clinical trial may provide sufficient evidence of an effect. Recruitment may be stopped early and subjects in the inferior treatment arm may be provided the superior therapy more rapidly.

To contend with the problem of inflating the experiment-wise type I error (i.e., the probability of incorrectly rejecting the null hypothesis of "no effect" when it is true), statisticians have developed a set of techniques, called group sequential designs, to permit interim analyses while constraining the overall probability of a type I error. In general, this requires setting more stringent significance levels for the interim analyses relative to the final analysis. Unlike adjustments to the overall significance level, these easily implemented techniques have become a requisite part of clinical trial monitoring and have resulted in the early termination of many RCTs.

Table 1 Sample size estimates for a hypothetical clinical trial as a function of different significance levels. Success rate in the control group is assumed to be 30%. Sample size ranges are shown as a function of the success rate in the experimental group and the statistical power.

<table>
<thead>
<tr>
<th>Success rate</th>
<th>Number of subjects needed if significance level is:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Treatment</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td>Power=80%</td>
<td></td>
</tr>
<tr>
<td>30%</td>
<td>40%</td>
</tr>
<tr>
<td>30%</td>
<td>45%</td>
</tr>
<tr>
<td>30%</td>
<td>50%</td>
</tr>
<tr>
<td>30%</td>
<td>55%</td>
</tr>
<tr>
<td>30%</td>
<td>60%</td>
</tr>
<tr>
<td>Power=90%</td>
<td></td>
</tr>
<tr>
<td>30%</td>
<td>40%</td>
</tr>
<tr>
<td>30%</td>
<td>45%</td>
</tr>
<tr>
<td>30%</td>
<td>50%</td>
</tr>
<tr>
<td>30%</td>
<td>55%</td>
</tr>
<tr>
<td>30%</td>
<td>60%</td>
</tr>
</tbody>
</table>
Method 3: Adaptive randomisation
Adaptive randomisation techniques involve the continuous adjustment of the chance of receiving each of the therapeutic options as the trial progresses: as one arm shows emerging evidence of superiority, the probability of subsequent patients receiving that therapy is increased.\textsuperscript{16,23,24} While the choice of therapy is still ultimately left to chance (thereby preserving the bias reduction of randomised trials), such techniques may minimise the number of subjects ultimately assigned to the inferior treatment. These techniques are somewhat complex to implement and it is unfortunate that this complexity has slowed wider application.\textsuperscript{25} An even greater problem, however, is the practical limitations of the methods: many research questions simply do not lend themselves to rapid assessments of efficacy and, therefore, will not benefit from adaptive randomisation designs. Furthermore, patients entered early in the trial present problems identical to those in the traditional clinical trial design.

Method 4: Bayesian techniques
Bayesian methods are slowly gaining acceptance into clinical trial methodology. These techniques seek to ascertain posterior probabilities by revising subjective prior probability estimates using the relevant likelihood function.\textsuperscript{16,26} Such methods are appealing in that they can explicitly incorporate all relevant information into the design, monitoring, analysis, and interpretation of clinical trials. This strength is also the source of most controversy: since prior estimates are subjective, it is difficult to provide a generally convincing interpretation if there is great uncertainty or controversy about the prior probability of success (the condition of "clinical equipoise" under which clinical trials are supposed to proceed). Bayesian methods can also be challenging and time-consuming to implement.\textsuperscript{27} Nevertheless, these techniques are of great current interest and are undergoing rapid development.

An additional method: unbalanced randomisation
Most clinical trials randomise subjects in equal proportions to the various treatment options (balanced randomisation). Pocock has suggested, however, that unbalanced randomisation (with more subjects in the intervention than the control arm) may have ethical advantages over balanced designs, since more subjects would be randomised to what was thought to be the superior therapy.\textsuperscript{28} This suggestion, while generally ignored in current research practice, deserves greater attention. In particular, this simple proposal may be refined to provide a useful technique for enhancing the ethical construct of randomised trials.

In usual clinical practice, patients choose whether to receive a given therapy. In the research context, however, patients choose whether to receive a probability of receiving a given therapy. Unlike the truly dichotomous decisions of clinical practice, the clinical trial offers only a chance of receiving a therapy and there is no a priori reason why that chance must be exactly 50%.

Balanced randomisation implies a state of true equipoise. Suppose an individual has a very slight preference for one of the treatment alternatives. In this case, balanced randomisation is unlikely to cause much uneasiness. As the preference for one alternative grows, however, increasing discomfort will probably develop to the point where enrolling in the trial becomes untenable. Thus, there is a continuous, not simply a dichotomous, departure from equipoise. The degree of uncertainty about the superiority of one option (and, therefore, the degree of deviation from equipoise) will vary according to several factors: the likelihood of success with the new intervention, the probable magnitude of superiority of the new intervention, the known efficacy of the current standard treatment, the severity of the particular disease process, and the risk tolerance of the potential subjects. For example, there would likely be little discomfort regardless of the study design for illnesses that involve little mortality or long-term morbidity risk (for example, a clinical trial of therapies for the common cold). For more serious conditions, however, potential subjects may have strong preferences for a treatment option, particularly in view of the other aspects noted above.

To the degree that informed individuals generally have preferences or concerns about one treatment option, adjusting the randomisation ratio to reflect those attitudes should result in a more ethically acceptable trial. If the likelihood of success with the new intervention is great and subjects would generally prefer receiving this intervention, randomising a greater proportion of subjects to it would be desirable. This tendency, however, should be mitigated by the known efficacy of the control therapy: if standard treatment is known to be highly effective, one should be cautious about randomising even half the subjects to any untested intervention, particularly for diseases with great morbidity and mortality.

In mathematical terms, we wish to maximise the expected utility of the experiment, based on the prior probabilities of success with each therapy and the estimated utilities of the possible outcomes. In
more concrete terms, we should approach the clinical trial just as the investor approaches an investment. In each case, the goal is to maximise benefit under conditions of uncertainty. Unlike the typical clinical trial, however, a wise investor does not arbitrarily place equal portions of money into alternative investments with differing risk and yields; rather, one would generally “hedge one’s bets” and allocate money according to realistic estimates of success and the implications of suffering a loss.

Balanced randomisation is generally regarded as the preferred allocation scheme in clinical trials since it optimises statistical power. Unbalanced randomisation schemes are rarely employed in randomised trials, though there may be surprisingly little loss of power with even moderate imbalances. Pocock has shown that, in many cases, there is only a minor loss of power using randomisation ratios as unbalanced as 3:1 (that is, randomising three times as many subjects to one arm as the other), though power declines rapidly with ratios greater than this. Figure 1 shows power curves for a hypothetical clinical trial. The top graph shows the decrease in power associated with different allocation ratios, assuming a constant number of subjects; the bottom graph shows the total number and distribution of subjects required under different allocation ratios, assuming a constant statistical power. In general, the loss of power or the increase in the total number of required subjects is small with ratios less than 2:1, and only slightly more with ratios as large as 3:1. Of course, explicit power analyses for

Figure 1: Sample size and power curves for a hypothetical clinical trial with a dichotomous outcome. Both graphs assume a two-sided a level of 0.05; a 50% probability of success in the experimental group and 30% in the control group. In Figure 1A, a constant sample size of 248 is assumed; the curve depicts the reduction in power associated with increasing proportion of subjects allocated to the experimental group (ie, increasing imbalance in subject allocation from 1:1 [50% in the experimental group] to 9:1 [90% in the experimental group]). In Figure 1B, a constant power of 0.9 (beta = 0.1) is assumed; the curves represent the sample sizes of the experimental group (long dashes), control group (short dashes), and both combined (solid line), associated with an increasing proportion of subjects allocated to the experimental group.
each trial should be performed before implementing any experimental design.

Whose point of view?
A critical question arises: from whose perspective should the allocation decision be made? Most discussions of equipoise focus on clinicians and investigators, since these individuals have the sophistication to make truly informed judgments about the biomedical issues. However, since it is ultimately the subjects who must bear the risks of the experiment, opportunities for involving participants in the design of the study must be aggressively explored. Establishing an allocation ratio provides such an opportunity. Qualitative decisions such as how to weight increased efficacy against increased side effects can only be made realistically by those who will experience the trade-off. A working partnership between potential subjects and investigators is needed to define the optimum solution; this is, admittedly, not typical in current research practice.

Input from those who understand the scientific issues is essential, however, as previous work with surrogate patients has shown. MacKillop et al surveyed expert physicians about their willingness to participate in specific clinical trials and gave this information to healthy surrogate “patients”. The investigators found that the willingness of the surrogate patients to enter a clinical trial of chemotherapy (though, interestingly, not of surgery) for lung cancer was heavily influenced by the opinions of the physician experts. It should be noted that, to further the interests of subjects, “expert” opinion must be realistic, unfettered, and unbiased; explicit acknowledgment of the degree of prior uncertainty is essential. When investigators have unrealistic expectations of a new and untested intervention, or when elements of trial design are hindered by imprudent regulatory mandates, the goal of designing more participant-oriented trials may be seriously compromised.

Clearly, only truly informed individuals can make independent decisions; in general, a partnership among all relevant parties is required. For some conditions, this can be particularly difficult. Research with vulnerable populations (for example, children and patients with dementia or severe psychiatric illness) requires special care, especially when incorporating the values and preferences of potential subjects. In these situations, investigators have a singular duty to ensure that the interests of subjects are respected. Methods that may be employed to help protect these subjects include close cooperation with guardians, special oversight from institutional review boards, and involvement of advocacy groups for patients with the particular illnesses.

Serious misunderstandings of the biomedical issues on the part of subjects may lead to poor choices; similarly, inadequate appreciation of the personal issues on the part of investigators may produce inappropriate decisions. When clinicians, investigators and potential subjects all have opportunities to share in the design of clinical experiments, research ethics may be greatly enhanced.

Choosing an allocation ratio
The final challenge, then, is to identify a method for choosing the allocation ratio that best reflects the state of clinical uncertainty and with which potential subjects (and their health care providers) would feel comfortable. Ideally, the procedure should force individuals explicitly to acknowledge the degree of uncertainty and to consider the factors noted above (the probability and magnitude of superiority of the new intervention, the efficacy of standard care, and the severity of the illness under study). In the case of confirmatory trials, one could conceive of this technique as a form of adaptive randomisation using previous trials, rather than patients, as the unit of analysis: as more trials suggest the superiority of one treatment arm, a greater proportion of subjects could be assigned to that arm in subsequent trials. Techniques such as cumulative meta-analysis could be modified for this situation, by providing an evolving assessment of the relative benefits of alternative treatments.

Other possibilities include the use of consensus-development techniques (such as the Delphi method) or surveying clinicians and identifying the proportions who prefer one of the possible treatment options. Modifications of preference-elicitation techniques commonly used in decision analysis may also prove useful. The basic reference gamble, for example, could be adapted to help quantify the utilities and trade-offs perceived by potential subjects. Clearly, any procedure would require empirical evaluation before it could be advocated for general use.

Since individuals’ values and preferences vary, it is unlikely that any single allocation ratio would be universally accepted as optimum. While variation in choice of an allocation ratio is expected, investigators should consider the nature of this variability carefully, rather than dismiss the process entirely. For example, if the problem is that some potential participants would prefer a very high likelihood of receiving the new intervention (for example, a 3:1 allocation ratio), while others would prefer a more modest imbalance (for
example, a 2:1 ratio), certainly adopting a 2:1 randomisation scheme would be more generally preferable to the almost universal convention of balanced randomisation. To the extent that subjects enter a clinical trial in order to gain access to new therapeutics, acknowledging this desire in setting an allocation ratio shows greater respect for their preferences. Some degree of imbalance, when introduced to further participants' interests, should be considered in trial design, even if universal agreement on the degree of imbalance is not easily achieved.

Discussion
The modern randomised trial has become the gold standard for tests of therapeutic efficacy while generating concerns about the ethics of clinical research. Despite extensive debate, there is still no consensus about the absolute need for equipoise, and how equipoise should be measured and expressed. The wider use of unbalanced randomisation may serve as one mechanism for diminishing the ethical tension.

Currently, unbalanced randomisation designs are rarely employed. The MEDLINE database contains 1,367 studies identified as randomised trials for the five-year period 1991-1995 in the five general medical journals with the highest science citation index impact factor (New England Journal of Medicine, Lancet, Annals of Internal Medicine, Journal of the American Medical Association, and the British Medical Journal). In only 31 of these studies were subjects randomised with unbalanced allocation (list available from author on request) and in all cases, the greater proportion of subjects were randomised to the intervention arm(s). Of the 31 trials using unbalanced randomisation, 13 gave no clear reason for its use and 15 cited pragmatic design concerns (for example, increased power to estimate side effects of a new intervention, lower study costs, or anticipation of a large number of differential drop-outs or cross-overs from one treatment arm to the other). One study used unbalanced randomisation to "encourage patient enrolment and compliance", another at the insistence of a local institutional review board, and one "for ethical reasons". Thus, unbalanced randomisation is rarely used to further the interests of study subjects.

As with any other research design, there are potential problems with this method. The major objection is that subjects may become uneasy about unbalanced randomisation, since it may appear that the choice between the treatment arms is not neutral. Proper explanation of the reasons for unbalanced allocation should alleviate concerns and may actually improve recruitment.

Including potential subjects in the decision may also diminish this problem and strengthen enrolment. In fact, available evidence suggests that this fear is probably unfounded. Contact was made with the investigators of several recent clinical trials which employed unbalanced randomisation and all reported that there were no substantial problems with acceptance or compliance by subjects (M Saag, M Cairo, S Wiesner, M Gottlieb, D Knopman, K Stanley, personal communications). All of these trials were for serious illnesses, including opportunistic infections, primary biliary cirrhosis, Alzheimer's disease, and neonatal sepsis. One recent RCT of prophylaxis for an HIV-related infection employed 2:1 randomisation on the recommendation of the community representative to the trial; this example demonstrates that truly informed subjects will clearly prefer unbalanced randomisation in some cases (T Louis, personal communication). How potential subjects respond to unequal treatment assignment will require further study, but the use of unbalanced randomisation should not be rejected simply because researchers presume that it may complicate and reduce subject recruitment.

Placebo-controlled trials (or trials with a no-treatment arm) present particularly difficult problems. Despite the demonstrated possibility that a new therapy can be worse than no therapy, it is likely that subjects willing to enrol in many clinical trials would prefer receiving the active treatment. In this setting, subjects and investigators might see an advantage to using unbalanced randomisation with as large a ratio as resources permit. Since there may be a substantial loss of statistical power with allocation ratios greater than 3:1, ratios larger than this should generally be avoided unless the cost of enrolling more subjects is easily absorbed. Allowance should also be made for potential unblinding and greater drop-out rates in the smaller group, since the 3:1 ratio is at the "shoulder" of the power curve and greater imbalances may severely diminish statistical power.

Another interesting situation occurs with trials in which unbalanced randomisation favours standard therapy over the new intervention. While allocating a majority of subjects to the control condition may seem like a strange proposition, there are settings in which such allocation should be considered. Future trials of antiarrhythmic drug therapy, for example, might randomise more subjects to placebo than active therapy, given the experience of the Cardiac Arrhythmia Suppression Trials (in which treatment of post-myocardial infarction patients with several antiarrhythmic...
agents resulted in increased mortality compared to placebo-treated subjects). 21 49 50

Though the ethical problems of experimenting on people can never be entirely resolved, there are several techniques that may help to minimise the dilemmas. In particular, unbalanced randomisation may enhance the ethical construct and acceptability of clinical trials, and the loss of statistical power with even a moderate imbalance may be minor. As researchers, clinicians, and patients strive to learn the best means of curing disease and reducing disability, creative approaches will be needed to lessen the tension between medical needs and ethical mandates.

Acknowledgments

The author gratefully acknowledges the advice and insightful comments of Drs Miriam Kupfermann, Andrew Bindman, Molly Cooke, Robert Wachter, Esther Hudes, and the anonymous reviewers. The author also thanks the cited investigators for their willingness to share their experiences with unbalanced randomisation designs.

Andrew L Avins, MD, MPH, is Assistant Professor in the General Internal Medicine Section, Veterans Affairs Medical Center, San Francisco, and Assistant Professor Department of Epidemiology and Biostatistics, University of California, San Francisco, USA. Please address all correspondence to: Andrew L Avins, MD, Veterans Affairs Medical Center (111-A1), 4150 Clement Street, San Francisco, California 94121.

References

5 Baum M. A new approach for recruitment into randomised controlled trials. Lancet 1993;341:812-3.
36 High DM. Advancing research with Alzheimer disease subjects: investigators' perceptions and ethical issues. Alzheimer Disease and Associated Disorders 1994;8, suppl 4:97-104.

News and notes

New website for Linacre Centre

The Linacre Centre for Health Care Ethics, the only Catholic bioethics centre in the UK and Ireland, has recently launched its new website. As well as information about the centre and its publications, including full contents and extensive reviews, the website also contains papers on topical issues in medical ethics such as human cloning, euthanasia and advance directives.

The site address is: www.linacre.org.

News and notes

Epistemology and Medicine

The 13th annual conference of the European Society for Philosophy of Medicine and Health Care, *Epistemology and Medicine*, will be held from August 18 - 21 1999 in Linkoping, Sweden.

Abstracts and registration deadline by January 1 1999.

For more information please contact: Professor Dr Henk ten Have, secretariat ESPMH, Department of Ethics, Philosophy and History of Medicine, Faculty of Medical Sciences, Catholic University of Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands. Fax: 024 - 3540254; from abroad: ++ 31 - 24 - 3540254.