At what level of collective equipoise does a clinical trial become ethical?

Nicholas Johnson, Richard J Lilford, Wayne Brazier  St James’s University Hospital, Leeds

Authors’ abstract

It has often been argued that if a clinician cannot decide which of two treatments to offer, a trial may be ethical, but it is unethical if she/he has a preference. Since individual clinicians usually have a preference, most trials could be judged unethical according to this line of argument. A recent important article in the New England Journal of Medicine argued that individual preferences are not as important as the collective uncertainty of informed clinicians. If clinicians are equally divided, there is a state of collective equipoise and a trial is ethical. However, clinicians will seldom be equally divided. We conducted an ethometric study to find out how much collective equipoise can be disturbed before the potential subjects in a trial think that it is unethical. Half of our subjects perceived a trial as unethical when equipoise was disturbed beyond 70:30. In other words, when 70 per cent of experts favour one treatment, 50 per cent of subjects would prefer that treatment to be administered rather than subjected to critical assessment. When equipoise is disturbed beyond 80:20, less than 3 per cent of subjects would consider human trials morally justifiable.

Introduction

Clearly if doubt exists about the efficacy of a particular treatment or the superiority of one treatment over another, it is desirable to find an answer. Solutions to scientific problems have their foundations in the deduction of first principles and these ideas are supported by observations and then confirmed by experimentation. In the case of medicine the most reliable experiment is a randomised controlled clinical trial.

Before a clinical trial can be considered ethical both the procedural and substantive ethics must be acceptable. Procedural ethics in a controlled clinical trial largely revolves around the question of consent and this is central to the question of how to perform a trial, not whether a trial is ethical in itself (1).

Substantive ethics questions whether the trial should be conducted at all. It is not enough to rely totally on consent. In some cases it might be wrong to offer patients entry into a trial – to tempt them to make a choice which most people would argue is strongly against their interests. If this were not so, then the function of the ethics committee would simply be to ensure that fact disclosure was obtained before a patient agreed to enter a trial. But ethics committees consider whether the trial should be done at all – whether the trial is substantially ethical. Indeed, the very existence of ethics committees engenders a certain level of dependency, because a potential subject can assume that the trial was seen as reasonable by a separate, informed and authoritative panel charged with protecting the subject from exposure to a study that would be judged unreasonable. Such committees must balance two countervailing attitudes. On one side of the coin are society’s objections to controlled clinical trials and emotional reactions against the very concept of human experimentation (2,3). On the other side, history has taught us the dangers of foregoing clinical trials (for example, Stilboestrol and progesterone for miscarriage prophylaxis, intermittent positive pressure ventilation with oxygen for chronic bronchitis and steroids for head injuries (4)) and some of the dilemmas clinicians now face (such as whether to administer digitalis therapy for congestive cardiac failure or do routine ultrasound examinations for all low-risk pregnancies) are due to the failure to initiate a controlled clinical trial at the birth of a new therapy. As a general rule, critical experimentation should precede the routine use of a new treatment and such trials are ethically acceptable because doctors will not have preconceived preferences for one new treatment over the other. Trials testing established therapies are more difficult to justify because a preference will have developed for a particular treatment, perhaps because both treatments have been used for some time, or because previous evidence is available from observational studies or previous inconclusive randomised trials.

‘Equipoise’ is the point where we are equally poised in our beliefs between the benefits and disadvantages of a certain treatment modality, (or the preference of treatment A over treatment B) (5,6). At this point we are agnostic or resting on the fulcrum of a preference.
Individual equipoise (also known as theoretical equipoise) applies to an individual clinician, whilst collective equipoise (also known as clinical equipoise) refers to the profession as a whole. Individual equipoise, demands an opinion from an individual that the evidence is equally split (50 per cent:50 per cent) between the benefits and disadvantages of one plan of action over another. In this situation most people would think the clinician justified in offering the patient entry into a controlled trial. However, individual clinicians are rarely in a state of pure equipoise (6). Usually there is a preference for one or other treatment option. The strength of belief in the superiority of one treatment over the other is variable – the clinician may have a strong or weak preference. In theory this strength of belief can be expressed numerically; for example, as the odds the clinician would accept if betting on the outcome of a trial. When the clinician believes that the odds of one treatment being superior are high, many people will intuitively feel that a trial is unethical. This is usually the case and this has led some people to state that most clinical trials are unethical – the doctor will have a preference (albeit not necessarily a very strong one) and should therefore give the patient his/her preferred treatment. This argument follows the Papworth principle – sometimes called the gold standard – that one should treat a patient as one would wish oneself or one’s mother to be treated. A crucial contribution to this debate came from Freeman (1987) (6) who argued that doctors are not bound by this principle but they are bound by the confines of expert opinion. The criteria for entry into a specialty and for censure in negligence cases are based on collective beliefs not on our own preference. We acknowledge the validity of this directive although we do not judge that adherence to collective norms completely removes any obligation based on personal preference. We think there must come a point where a clinician’s personal preference becomes so strong as to compel him/her to act in a particular way, irrespective of collective doctrine. However, we do not wish to deflate Freeman’s argument in this paper. Here we wish to begin where Freeman left off and to accept for the moment that it is ethical to offer patients entry into a randomised trial provided that each treatment modality in the study is a ‘recognised’ method falling in collective norms. We wish to extend Freeman’s analysis by asking another question: how broadly can we define these collective norms. In the same way that individual equipoise will seldom be complete, so collective equipoise will also seldom be equally divided. Usually there will be an imbalance between the number of experts favouring one treatment versus another. We decided that it would be informative to ask subjects in a hypothetical trial to judge the degree of collective equipoise necessary for a specific type of trial to be ethical – we therefore did an ethmometric study.

Subjects and methods
(MEASURING THE LEVEL OF EQUIPOISE)

To evaluate the level of equipoise that people think should be satisfied if a clinical trial is ethical, subjects were asked to specify the level of collective doubt between two treatment modalities they would accept if casting a vote on an ethical committee. This level of doubt was expressed to subjects on a series of cards showing varying numbers. Most subjects could comprehend pictograms containing various numbers of experts favouring each treatment. A series of scenarios (appendix I) were presented and the order selected at random. A sample of convenience of one hundred and thirteen subjects was recruited. The subjects were selected from a broad range of society and geography (table 1). Our intention was to select lay people rather than patients to represent those who may sit on ethical committees and who would therefore take both sides of the argument into account. This fulfils the Rawlsian criteria that our subjects could imagine themselves as future patients who could benefit from the knowledge gained by trials but who would also wish for protection in trials where one treatment was already favoured over another by a certain proportion of experts. Subjects were therefore given propaganda about the importance of proper clinical trials in protecting the public from poorly tested and potentially harmful treatments.

Our sample size was calculated using data from a pilot study and our intention was to reject with 99 per cent confidence the null hypothesis ‘a trial is accepted.

<table>
<thead>
<tr>
<th>Cohort definition</th>
<th>Location</th>
<th>Age</th>
<th>Number</th>
<th>No excluded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patrons of public house</td>
<td>Leeds</td>
<td>16-65</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>Bridge players</td>
<td>Cornwall</td>
<td>50-70</td>
<td>25</td>
<td>2</td>
</tr>
<tr>
<td>Italians (students)</td>
<td>London</td>
<td>16-20</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>Tax officers</td>
<td>London</td>
<td>20-30</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Nurses</td>
<td>Leeds</td>
<td>18-30</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>Neighbours</td>
<td>Cornwall</td>
<td>40-70</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Medical students</td>
<td>Leeds</td>
<td>20-34</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>
32 At what level of collective equipoise does a clinical trial become ethical?

even when equipoise approaches 0.95'. Twenty subjects were excluded either because they could not understand the concept, or because the intellectual effort required was too demanding.

Results
The responses to three of the scenarios is shown in figure 1. The human trial most likely to be tolerated was the comparison of headache analgesics (scenario A). In this trial, the median level of equipoise was 67 per cent (67:33), the interquartile range was 60 per cent – 75 per cent (60:40–75:35) but even in this case only 3 per cent of subjects would declare the trial substantively ethical if equipoise was disturbed above 80 per cent (80:20). The data from the other scenarios is presented in table 2. Individuals responded very differently, with responses ranging from 50 per cent to 99 per cent. The trend seems to favour a comparison of established treatments rather than an evaluation of a new one and a greater acceptance of a clinical trial if the likely risks of the condition are less, reversible and not limiting (for example, chemotherapeutic trials demand a high level of equipoise yet a trial involving a headache cure is more readily tolerated). A high level of collective equipoise is demanded if the issues are emotive, for example a clinical trial involving infants is accepted only if collective equipoise is high (close to 50:50), whereas lower levels of equipoise are required of trials involving old people and experiments on rats require less stringent standards than those on beagles. There was no statistically significant difference in response between any of the subgroups described in table 1. Doctors were likely to consider conducting a clinical trial at a higher level of equipoise (the amount of uncertainty is less) than they would accept if they themselves were the subjects but this is not statistically significant.

Discussion
In general, these findings suggest that trials are perceived as unethical when equipoise is disturbed

---

![Figure 1](image_url)

**Figure 1**

Maximum number of experts who may favour one treatment over another before the trial becomes acceptable (level of equipoise expressed as a percentage) for each scenario.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Mean</th>
<th>SD</th>
<th>Mode</th>
<th>Median</th>
<th>Third Quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario A</td>
<td>67%</td>
<td>7</td>
<td>72%</td>
<td>70%</td>
<td>76%</td>
</tr>
<tr>
<td>Analgesia trial for headaches</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario B</td>
<td>63%</td>
<td>11</td>
<td>65%</td>
<td>60%</td>
<td>73%</td>
</tr>
<tr>
<td>Trial of cytotoxics in leukaemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario C (i)</td>
<td>65%</td>
<td>13</td>
<td>65%</td>
<td>65%</td>
<td>75%</td>
</tr>
<tr>
<td>Pneumonia trial in neonates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario C (ii)</td>
<td>69%</td>
<td>15</td>
<td>65%</td>
<td>70%</td>
<td>75%</td>
</tr>
<tr>
<td>Pneumonia trial in old people</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario E (i)</td>
<td>68%</td>
<td>19</td>
<td>60%</td>
<td>75%</td>
<td>80%</td>
</tr>
<tr>
<td>Trials in beagles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario E (ii)</td>
<td>73%</td>
<td>23</td>
<td>75%</td>
<td>75%</td>
<td>85%</td>
</tr>
<tr>
<td>Trials in rats</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*A, B and E (ii) are plotted on figure 1
experts beyond 70:30. In other words, when 70 per cent of experts favour treatment A, then 50 per cent of subjects would prefer that treatment A be administered rather than subjected to critical assessment. This is a narrow scale of uncertainty and the use of such limits to practice would preclude many important trials and this effect might be compounded in many cases by the lack of individual equipoise of practising clinicians. Admittedly a different population would have arrived at a slightly different cut-off point and there is no doubt that the exact framing of the scenarios will affect the precise conclusion (7) but when equipoise is disturbed to nearly 95:05 then the trial clearly and substantially exceeds the limits that almost all people would consider acceptable. Even when equipoise is disturbed beyond 80:20, less than 3 per cent of subjects would consider human trials morally justifiable. We suggest that if the level of equipoise is less than 80:20 then, as a ground rule, trials will be substantially ethical and should be permitted. All these observations presuppose that the average strength of belief of clinicians on both sides of the debate is the same. Notwithstanding this concept, we think that a survey of expert opinion would be very useful when discussing the ethics of a proposed randomised trial, for example, the multivitamin trial to prevent recurrence of neural tube defect. If this issue was under discussion, say, at a Medical Research Council committee meeting, a survey of the preferences of experts (based on their individual interpretation of existing data) would be illuminating. If they were split 70:30, then their level of collective equipoise would be very close to the median equipoise limits of the population at large, but if nine experts thought that treatment A was superior and two favoured treatment B, then the level of equipoise would lie outside the limits which most people would tolerate. This would be an argument against declaring the trial substantially ethical. If opinion of doctors or other people, in possession of all existing facts, was split 10 for treatment A, 40 for treatment B and 50 with no preference at all (completely agnostic or individually equipoised), then the results should not be interpreted 20:80, (and therefore unethical), since this disenfranchises those who are individually equipoised. Perhaps their vote should be added equally to each side giving a ratio of 35 v 65, thereby making the trial ethical – at least on this criterion. The most sophisticated ethometric study of all would take into account the strength of preference on each side of the argument and weight the analysis by this factor. If voting was 85 v 15 for treatments A and B, respectively, but the mean strength of opinion for treatment B was \( \frac{9}{10} \) and for treatment A \( \frac{3}{10} \), then the weighted collective equipoise would be 65 v 35 and therefore within the boundaries of the median cut-off value for collective equipoise.

The argument we wish to present here, is not that the substantive ethics trials should be subjected to routine ethometric studies although we do think that these might be illuminating in certain (particularly contentious) cases. Our intention is to demonstrate that the concept of collective ethometrics, while an important contribution to the debate, is not an all or none phenomenon and can (at least theoretically) be quantified. Thus, Freeman, while contributing greatly to the debate by explaining the difference between individual and collective equipoise and while explaining the importance of the latter, has not resolved this issue to the degree suggested in the New England Journal of Medicine article (6). This underlines a still more important facet of the ethics of randomised trials which is that they are subject to gradualism not absolutism. The ethical undesirability of a trial grows as, for example, the strength of certainty that a particular treatment is superior increases. If this view is accepted, then the substantive ethics of trials is not unlike that which can be advanced for termination of pregnancy where the medical and ethical undesirability of the procedure can be perceived to rise as gestational age increases. The point in this gradual increment where a procedure becomes substantially unethical can be submitted to ethometric evaluation. It should also be noted that the limits of equipoise vary with the subject under discussion, although by a smaller amount than we had anticipated. This suggests that our results may be influenced by an element of mid-range anchoring and this, together with various framing effects, explains why we do not think that our results are accurate and binding. They indicate an order of magnitude, rather than a precise cut-off limit. It is nevertheless clear that the majority of our subjects are less likely to tolerate trials involving issues with which the subject can personally identify or those which cover emotive issues. This might partly be related to aesthetics and an emotional, unsubstantiated 'gut feeling' but it is also related to the perceived quality of the future life of those who have been selected for the trial (neonates are more aesthetically attractive and have a greater life expectancy than old people, and beagles have a greater perceived 'quality' of life than rats).

It should be noted that these limits of equipoise obtained by consumer trade-off surveys cannot be clearly translated into the probability (p value) obtained by statistical analysis of results. Science is about belief; clinical trials like legal trials are designed to influence belief. Strength of belief (individual equipoise) or the number of experts believing in the superiority of one treatment or another (collective equipoise) will not vary in parallel with the p value. Further experiments are, therefore, required to determine at what p value 70 per cent of doctors, (having started in complete and equal disagreement) will favour one treatment. Nevertheless, continuing a trial until \( p<0.05 \) per cent may be too extreme because both individual and collective equipoise will be strongly disturbed by this stage. The p value that trials seek is an issue of procedural ethics that has been discussed elsewhere (8,9) but our overall argument is...
applicable not only to the point when a trial becomes unethical but to the precision of a proposed trial.

Conclusions
Collective equipoise is an important (albeit not overriding) consideration in the substantive ethics of randomised trials. In general, when more than 2/3 of experts favour one treatment over another, more than half of people will feel that a human trial is substantively unethical. Our subjects seem to tolerate slightly more relaxed levels of equipoise as the subject of research becomes less emotive. They demonstrate speciessism, preferring experiments on rats to beagles and beagles to humans. The substantive ethics of randomised trials depend on collective and individual beliefs about the superiority of various treatments. Utilitarians will usually favour randomised trials until equipoise is very strongly unbalanced, because more people outside the trial stand to lose from acceptance of an inferior treatment than do people inside the trial who are allocated to an inferior arm. Therefore, unless the lifespan of a treatment is likely to be very short, the greatest good comes to the greatest number by doing a large number of trials to a high level of precision. However, if overriding importance is given to the individual patient who cannot be asked to make any sacrifice for the common good, then trials are only ethical in the face of complete collective (and perhaps even individual) equipoise. We cannot get away from the truth that, rare instances of complete equipoise apart, trials represent a trade-off between maximising our obligations to an individual patient and those to future patients. Our study shows that most subjects accept a small trade-off of the individual good for the collective good.

Nicholas Johnson MRCoG FRCS is a Lecturer in the Department of Obstetrics and Gynaecology at St James's University Hospital. Richard J Lilford MRCoG MRCP PhD is Professor and Head of Department of Obstetrics and Gynaecology at St James's and Wayne Brazier is a student at St James's University Hospital.

Appendix I

SCENARIO A
Doctors are uncertain as to which treatment is best for headaches. Two pain-killing drugs (A and B) have been extensively tested and are known to be safe.

SCENARIO B
At the moment death can be prevented in half of those patients with leukaemia by prescribing the established drug. An alternative drug is available but doctors are uncertain as to which of these two treatments is the most effective. If either drug fails the patient will surely die. There is no possibility of trying the other treatment because by that time it will be too late.

SCENARIO C (i)
Some patients are at risk of developing pneumonia following an operation. Doctors are uncertain as to which of two antibiotics (A and B) to use to decrease the risk. Only one of these drugs may be used because they interact. If the antibiotic fails then the sick patient may develop pneumonia and this could be fatal. The trial is to be performed in geriatric patients.

SCENARIO C (ii)
Doctors also want to perform the same trial in newborn babies recovering from an operation. Assume that pneumonia is equally dangerous in geriatric patients and babies.

SCENARIO D
You are to be a subject in clinical trial A, B and C.

SCENARIO E
Two anti-leukaemia drugs for use in humans have been developed. It is planned to test the drug in animals with leukaemia.

i) in beagles?
ii) in rats?

References