Evidence and Ethics in Medicine

John Worrall

ABSTRACT Ethics and epistemology in medicine are more closely and more interestingly intertwined than is usually recognized. To explore this relationship, I present a case study, clinical trials of extracorporeal membrane oxygenation (ECMO; an intervention for persistent pulmonary hypertension of the newborn). Three separate ethical issues that arise from this case study—whether or not it is ethical to perform a certain trial at all, whether stopping rules for trials are ethically mandated, and the issue of informed consent—are all shown to be intimately related to epistemological judgments about the weight of evidence. Although ethical issues cannot, of course, be resolved by consideration of epistemological findings, I argue that no informed view of the ethical issues that are raised can be adopted without first taking an informed view of the evidential-epistemological ones.

Because so much more work has been done on biomedical ethics than on any other aspect of philosophy of medicine, the latter is often, in effect, identified with the former. But there are issues even in the heartland of biomedical ethics that cannot properly be addressed unless due attention is paid to epistemological questions using insights from the philosophy of science. The central reason for this is that ethical and epistemological issues are often closely interrelated, and interrelated in more intricate and more contentious ways than is recognized. This claim is argued here in two stages. First, an historical case study of a series of randomized clinical trials is sketched; and second, the ethico-epistemological issues raised by that case study are examined to show how it pro-

Department of Philosophy, Logic and Scientific Method, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom.
E-mail: j.worrall@lse.ac.uk.

Perspectives in Biology and Medicine, volume 51, number 3 (summer 2008):418–31

vides—in three related but separate ways—vivid and precise confirmation of the thesis that ethics is importantly dependent on epistemology.

The Story of Extracorporeal Membrane Oxygenation

In 1980, pulmonary hypertension of the newborn had a fairly well-established mortality rate of around 80% in hospitals across the United States and Europe. In the 1980s, Bartlett and colleagues at the University of Michigan started to treat babies suffering from this condition using a technique called extracorporeal membrane oxygenation (ECMO). The idea behind this treatment is simple: venous blood is taken from the baby, and pumped round a circuit that includes a membrane where the blood is oxygenated, reheated to body temperature, and passed back into one of the baby's carotid arteries—thus bypassing the baby's lungs, the immaturity of which is implicated in the persistent hypertension.

Bartlett and colleagues achieved an approximately 80% survival rate for neonates suffering from persistent pulmonary hypertension using ECMO (Bartlett et al. 1982). Nonetheless, they "felt compelled to conduct a prospective randomized study," despite the fact that they "anticipated that [in such a trial] most ECMO patients would survive and most control patients would die" (Bartlett et al. 1985, p. 479). As this makes clear, Bartlett and his colleagues were already convinced by the evidence of their success with ECMO, as compared to the high mortality rate when treating apparently similar patients with the conventional treatment, that ECMO was a (greatly) superior treatment for this condition. Their switch to using ECMO can be considered to have effected an "historically controlled trial," in which the controls are provided not by a contemporaneously studied group given the established conventional treatment, but rather by earlier (though recent) patients who had been treated for the same condition under the old regime.

These investigators believed, however, that if ECMO's superiority was to be generally accepted by the medical community and hence was to become the accepted treatment worldwide, the superiority of ECMO had to be demonstrated in a prospective randomized trial. Such a trial involves two sets of patients studied contemporaneously—one given the new treatment (constituting what is usually called the "experimental" arm), here ECMO, and the other given the conventional treatment (CT, for "control" arm), where the division between the experimental and control arms is made in some way or another that involves a random process (usually whether successive entries in a table of random numbers are even or odd).

As the remark quoted above shows, Bartlett and his colleagues clearly had ethical qualms about assigning babies involved in such a trial to the CT arm when they could have treated those babies with ECMO; on the other hand, they felt driven to perform a trial that at least had a random element in order to convince

the medical community of the superior efficacy of ECMO. They therefore adopted a modified protocol for their trial—one that is usually called "randomized play the winner" (or "randomized play the leader"; Truog 1999). This involves a randomization of the first patient—say by drawing a ball from an urn with one red (ECMO say) and one white (CT) ball. Suppose—as in fact turned out to be case—the first patient is assigned to ECMO. The idea of "randomized play the winner" is that if that first patient survives, then before the second patient is randomized, another red (ECMO) ball is added to the urn, while if the first patient does not survive, a white (CT) ball is added to the urn. Thus as each new patient is entered into the trial, the odds of being assigned to one treatment or the other are progressively stacked in favor of the treatment (if there is one) that seems to be doing better on the basis of the results so far. In the more usual randomized trial, of course, the decision on whether or not any given patient goes into the experimental or control group is made by the same random process in each case—no attention being paid to the results so far.

As it turned out, the first baby in this "randomized play the winner" trial was assigned to ECMO and lived, while the second was assigned to CT and died. This produced a quite heavily biased urn, and the urn became ever more biased as successive babies drew an ECMO ball and survived on ECMO treatment. The overall result of this trial was in fact that a total of 11 babies were assigned to ECMO, all of whom lived; while one baby was assigned to CT and died (Bartlett et al. 1985).

In view of the evidence from the historical trial, should this trial have been performed at all, since it involved babies—as it turned out, only one—being assigned to CT? A strong body of opinion, led by orthodox frequentist statisticians, held that, on the contrary, the historically controlled trial was scientifically non-telling and so also was this first trial: since it was not a properly randomized study, it had produced no scientifically telling evidence of the superior effectiveness of ECMO (Wade and Epstein 1985). After all, many people in medicine believed (and continue to believe) that "the only source of reliable evidence . . . is that obtained from . . . carefully conducted randomized trials" (Tukey 1977, p. 679). This strong body of opinion recommended a second, and this time properly randomized trial, which was duly performed.

The second trial involved orthodox randomization with the standard significance level of 5%; but also a "stopping rule" that specified that the trial was to be halted once four deaths had occurred on either of the treatment arms (O'Rourke et al. 1989). The outcome of this trial was that the nine babies allocated to ECMO all survived, while 10 babies were assigned to CT, four of whom died—the fourth death triggering the stopping rule.

Anyone having qualms about the ethics of the first trial would surely have still more misgivings about this second one—and in particular about the deaths on the control arm. Again, however, a strong body of opinion held that, not only was this second trial justified because before it was performed there was no *really*

telling, *really scientific* evidence of the superiority of ECMO, but also that *even this second trial* failed to provide proper scientific evidence (Pocock 1993). Having a stopping rule at all is a major problem for any orthodox statistician (it automatically means that no power calculation can be performed), and in this case it meant that the trial was stopped without the result achieving statistical significance—that is, without the null hypothesis (of no real difference between ECMO and CT) being refuted according to the standard statistical canon (even assuming, statistically illicitly, that the trial was always intended to have the number of patients that it finished up having courtesy of the stopping rule). In fact, the observed outcome has a probability of occurring of $p = 0.054$ if there is no difference between the two treatments; thus the observed outcome failed, albeit narrowly, to achieve the "magic" $p = 0.05$ level. Some statisticians argued, therefore, that a still *further* trial was necessary to obtain proper scientific evidence of the superiority of ECMO and hence good reason to recommend ECMO's general introduction into medical practice.

The eminent medical statistician Stuart Pocock (1993) was among those who argued exactly this point. He wrote:

> a decision was taken to halt randomization [in this second trial] when the data disclosed four deaths among ten infants receiving conventional medical treatment compared with none among nine infants having ECMO (p= 0.054). . . . randomization was [thus] stopped early on the basis of a fairly small amount of data, all subsequent patients being allocated to ECMO [actually 20, 19 of whom survived]. The investigators were sensitive to the individual ethics of seeking parental consent and randomization for the next newborn infant. . . . However, with only 19 patients this does not represent strong evidence of the superiority of ECMO and provides little scope for making reliable judgments on the benefits of this treatment for universal use in such newborn infants in the future. (p. 1459–60)

Pocock acknowledged but gave no weight to the fact that this second trial had already recruited a further 20 patients when the stopping rule dictated an end to the trial proper. Because the trial (or at least "phase 1" of the trial) had now officially ended, these further patients were not "randomized" but were instead all assigned to ECMO, and 19 of the 20 survived. Statistical orthodoxy dictates that since these babies were not officially part of the randomized trial, this extra information (from what on some accounts was regarded as "phase 2" of the trial) cannot count as scientific evidence. But obviously had the babies been randomized and been, as a group, (enormously) lucky—by being randomly assigned to ECMO—then this trial result would have counted as (very) significant support for ECMO in even the most orthodox statistician's book. Of course the Bayesian approach to statistics places no bar on considering these further patients as evidence, and so, as (almost) always, seems to be in line with scientific commonsense.

But leaving this issue of the extra patients aside, Pocock's analysis clearly leads

to the implicit recommendation of still a third trial; and such a trial was duly performed (UK Collaborative ECMO Trial Group 1996). This time however, the trial was aimed at assessing ECMO's effectiveness for a wider range of neonatal conditions than simply pulmonary hypertension, looking instead at neonatal respiratory failure—as caused in a variety of ways. Again the trial used orthodox randomization, and on this occasion no explicit stopping rule.

Although ECMO scored considerably less well when compared with conventional ventilation for respiratory failure in general than it did when the issue was restricted to treatment for pulmonary hypertension, the U.K. trial, which finished up involving 185 neonates, nonetheless was stopped early because of what were judged to be too many deaths on the conventional treatment arm. (Even trials that are not governed by explicit stopping rules generally make provision for an oversight committee—or "wise woman or man"—that is allowed to see the decoded results (decoding not in fact being necessary in this case since the trial could not be blinded) as they come in and that may stop the trial if in the committee's "considered judgment" the evidence gathered so far already clearly tells in favor of one or the other treatment.)

ECMO became the conventionally accepted treatment around the world for persistent pulmonary hypertension of the newborn, and indeed for a range of other causes of respiratory failure in neonates (and more generally).

Ethical and Evidential Issues Closely Intertwined

Three separate ethical issues that arise from this case study all involve in clear-cut ways questions about what does or does not count as telling evidence for the effectiveness of a treatment—questions that many in medicine seem to regard as settled or uncontroversial, but that are in fact very much open. No one should claim that the ethical issues to be discussed here are settled by informed consideration of the epistemological issues about evidence, but the claim that will be argued is that no informed decision can be made about those ethical issues without informed consideration of the epistemological-evidential ones. The three ethical issues concerned are (1) under what conditions is it ethical to perform a clinical trial; (2) is a stopping rule (or, equivalently, some sort of oversight mechanism that can end a trial early under certain circumstances) ethically mandated; and (3) what constitutes genuinely "informed consent"? The first and third of these issues have already been discussed in relation to the ECMO story in two articles by Robert Truog (1993, 1999).

Under What Conditions Is a Clinical Trial Ethical?

Any physician who contemplates entering a patient into a randomized clinical trial (RCT) faces a prima facie ethical difficulty (Botros 1990). After all, such

a physician must—at least in many cases—be giving that patient a half-chance of receiving what is in fact a less effective treatment. Indeed, in the case of placebo-controlled trials, it is hoped and generally expected that the experimental treatment will outperform the placebo. How is this compatible with laudable sentiments that are taken to govern the practice of all physicians, such as "the health of my patient will be my first consideration" (Physician's Oath, Declaration of Geneva)?

The usual response is, of course, that the question of what is objectively the case, and the question of what the physician knows to be the case are quite separate. It is perfectly acceptable for a physician to enter a patient into a trial so long as she or he is in "equipoise"—interpreted initially as requiring at least that the physician does not know which of the two treatments is superior. Note that this moral issue overlaps with the realm of epistemology: we need to think more clearly about what it takes to "know" that one or the other treatment is superior. If, for example, we were to interpret *know* in the traditional Cartesian sense, then the condition of being in equipoise would surely be achieved too easily. Objectively speaking, we know nothing in this ultra-demanding sense, and so, if this were the underlying sense of "know," any physician would always be in (objective) equipoise with respect to any treatment both before any therapeutic trial and, indeed, after any such trial. We would surely want to say, however, that it was ethically questionable for a clinician to enter a patient into a trial, one of whose treatments he or she did not know for sure to be—but still had "good reason to believe" ahead of the trial was—inferior. But what does it take to have "good reason" ahead of a trial to think that one or the other treatment is inferior?

If we understand having "good reason to believe" in a subjective sense, then this condition seems seldom to hold for clinicians involved in RCTs. Meta-level sociological studies have indicated that clinicians are usually very positive about the treatment they are testing ahead of the RCT. (And this optimistic view will have been reinforced by positive results in all the preliminary investigations that the United States, United Kingdom, and other regulatory authorities require to be undertaken before "phase 3 studies"—the RCTs—are licensed.) Of course, this can produce an uncomfortable situation for those clinicians, since if they are convinced that the new treatment will prove superior, they are automatically also convinced that those patients assigned to the control arm in their trial will receive an inferior treatment.

Consolation here is usually sought within the notion of "collective equipoise." This idea was introduced by Benjamin Freedman (1987), who took the view that "It is not necessary for a physician to be in personal equipoise to ethically enroll a patient [in a trial] . . . so long as there is genuine uncertainty within the medical community" (p. 141). This is clearly the sort of idea to which Bartlett and colleagues were appealing when organizing their first trial, with the underlying justification that, although that trial might well involve condemning babies

on the control arm to a likely death, many more deaths would be caused worldwide if ECMO were not generally accepted as a superior treatment, and general acceptance would come about only via a successful RCT.

I would argue, however, that an alternative and stronger reaction can be found in the likely response that the "hard-liners" will make to all this. This response is that it does not matter at all what anyone subjectively believes—instead the question is one of objective evidence. So long as there is no objective evidence that one or the other treatment is inferior, a researcher's personal convictions are irrelevant. Or rather, those personal convictions ought to be critically examined and replaced by views that are in line with the objective evidential situation. It may prove difficult psychologically, but surely clinicians ought to ask themselves where any convictions about the greater effectiveness of a treatment come from, and, if the answer is "not from properly analyzed evidence," then they ought to modify those convictions. This critical process might reveal that—despite their initial position with its associated but ill-founded convictions—they were objectively in equipoise because they had no really telling evidence that the control treatment was inferior; and hence that no ethical issues in fact arise about the trial (even though the clinicians running it initially believed that such issues did arise).

This hard-line defense is—perhaps paradoxically—strengthened by the fact that some (meta-level) studies have indicated that subjective confidence tends to evaporate rather significantly once "properly controlled" RCTs are performed. As early as 1977, for example, a study looking at trials of new therapies in surgery and anesthesia found that only 49% of new therapies when tested in RCTs in fact "proved" superior to standard treatments (Gilbert, McPeek and Mosteller 1977). This was true even though the authors categorized as "superior" treatments that only seemed equally good in treating the "target disorder" but that seemed to exhibit fewer unfortunate side-effects. It is, by the way, especially notable that they found such a low percentage of "successful" trials, given that they were looking only at published studies. As is widely recognized—it constitutes a significant problem for meta-analyses and systematic reviews more generally—medical journal editors and even researchers themselves exercise a strong selection bias against publishing "negative" results (ones in which the "null hypothesis" of no difference between the two treatments involved fails to be refuted). So it seems entirely reasonable to conclude (as Gilbert, McPeek, and Mosteller do) that

> when assessed by randomized clinical trials, innovations in surgery and anaesthesia are successful about half the time. Since innovations brought to the stage of randomized trials are usually expected by the innovators to be sure winners, we see that . . . the evidence is strong that the value of the innovation needs careful empirical checking. (p. 693)

Surely we should endorse the "hardliner" claim that what matters is objective evidence, rather than any purely subjective opinions. But then the next question,

of course, is what exactly counts as "objective" evidence. One already noted and very strong view, which is nonetheless advocated by many in the medical profession, is that objective evidence is garnered *only* from RCTs, and that therefore ahead of a properly performed RCT, any "evidence" that we might appear to have consists simply of "guesswork" (Baum 1990) or of "hunches" (Schafer 1982). Herbert (1977) put this view succinctly: "uncontrolled studies may point in a direction, but cannot be *evidence* as lawyers use the term evidence to mean something probative . . . [that is] tending to prove or actually proving" (p. 690). Of course if this were correct, then returning to the "objective equipoise" issue (and assuming that there is no ethical issue about performing a trial if the researchers are indeed in "equipoise"), there would never be any ethical issue about at least the first "proper" RCT performed on some proposed new treatment, no matter what had gone on before that RCT by way of historically controlled trials or anything else. Nothing short of an RCT can provide real scientific evidence, and so, whatever they may subjectively believe, no one can have objective scientific evidence for the superiority (or inferiority) of a new treatment ahead of the first RCT performed to assess it. Before the first RCT, any researcher is *objectively* in "equipoise," whether she or he likes it or not.

However, if there should be no question but that what counts is objective evidence rather than subjective opinion, the claim that *only* RCTs supply such objective evidence can certainly be questioned. Indeed, for all that it is often asserted, it seems altogether too polarized a view to be even remotely plausible (Worrall 2007). No RCT has ever been performed (nor ever surely will be) on appendectomy as a treatment for acute appendicitis, for cholecystectomy as a treatment for gallstone disease, or aortic aneurysm repair, or for pretty well any major surgical intervention. The same holds, I believe, for a wide range of medical interventions including (famously) penicillin for pneumonia and rabies vaccine following a bite from a rabid dog, but also diuretics in heart failure and even aspirin for mild headache relief. For all the fact that evidence-based medicine advocates can point to *some* treatments (grommets for glue ear, and suppression of ventricular ectopic beats are favorite examples) that had been generally accepted in medical practice but then "proved" to be ineffective when subjected to trials, this surely cannot be true for *all* accepted but non–RCT-tested treatments. Surely there is strong evidence in favor of the effectiveness of some such treatments (such as appendectomy for acute appendicitis)—strong evidence that cannot, then, have been delivered by an RCT.

Advocates of EBM, having foresworn the view that only RCTs really count as evidence, now defend a much more nuanced view involving an "evidence hierarchy." This position still gives the leading role to RCTs (and to meta-analyses thereof)—for reasons it is sometimes hard to discern—and in general the rankings embodied in the hierarchy often seem short on justification. The hierarchy also seems to embody the rule that one properly performed RCT "trumps" any number of historically controlled trials with contrary results, and

so seems to entail that there is always a sound evidential reason to perform an RCT on any intervention, no matter how strong the evidence for the effectiveness of that treatment from historically controlled trials may appear to be.

The principal question to ask of any piece of evidence that appears to be positive for some theory and in particular for the effectiveness of some treatment is whether or not there are plausible ways in which that evidence might have arisen even if the theory under test were false. Returning to the ECMO case, the question to ask is whether there is a plausible way in which the change from around 80% or so mortality to around 80% or so survival could have arisen from other causes without ECMO itself being an effective treatment. The first issue, judged from this commonsensical scientific perspective, is whether the babies given ECMO were being specially (and substantially) selected in some way so that the treated group was systematically different from (presumably a priori more likely to recover than) the group earlier given the then-conventional treatment. The answer to this is, I believe, no: basically *all* the babies with this condition were now treated by Bartlett and colleagues with ECMO rather than CT. We do need to take into account the fact that the researchers, of course, knew that every baby that they were actively treating in the "historically controlled trial" was being given ECMO. There is clear evidence of quite exceptional efforts being made by the researchers and ancillary medical staff on behalf of the babies being given the new treatment (Bartlett et al. 1982), but then—more scientific commonsense—although there does seem to have been some "treatment bias" (as it might be called), it needs to be asked whether it is likely that this would account for the change from 80% mortality to 80% survival or, indeed, any significant portion of it. From this perspective, then, it looks as if the historically controlled evidence was already telling, and that the qualms of the Michigan researchers about the babies on the CT arm of any trial were not based merely on subjective conviction, but rather on a justifiable view of the evidence. They were not in "objective equipoise" ahead of the trial, but instead had good reason to think that the control treatment was inferior. It seems that the better course of action might have been to try to convince the medical community that it was in the grip of an overly simple view of what counts as real scientific evidence.

I need not argue any such claim in detail here. My thesis maintains only that this is a precisely delineated area in which ethical issues are closely intertwined with epistemological ones. And whether or not the historical evidence should be regarded as objectively telling, it is surely clear that there is such intertwining. If the only telling evidence is that garnered from properly performed RCTs, then none of the three trials (randomized play the winner trial, first "properly randomized" but stopped early trial, or the third U.K. trial) can be regarded as unethical. (I am of course assuming here that it is unethical to perform a trial only if there is good reason to think ahead of the trial that one of the arms, usually the control arm, involves an inferior treatment. It could be argued that this is not the whole ethical story—but consideration of further issues is beyond the scope

of the current paper.) Whereas, if one holds that "proper evidential reason" may be supplied by sophisticated properly controlled historical trials, then there may well have been good objective reasons ahead of any of the three trials to hold that ECMO was superior, and hence good reason to think that any baby assigned to the control arm in any of these trials was being given an inferior treatment—and hence that all three trials were unethical. So epistemology affects ethics and, despite some overblown claims on behalf of RCTs, since neither of the two epistemological antecedents in these conditionals is obviously correct, this example illustrates how a wider moral context may play an adjudicating role in judging the epistemological basis of medical decision making. In short, an informed view of the ethical situation coupled to an informed view of the epistemological evidence is required to derive a rational decision.

The Ethics of Stopping Rules

The second ECMO trial involved a stopping rule dictating that the trial be ended once four deaths had occurred in either arm. Both in this case and more generally, the decision to specify such a stopping rule in advance (or to introduce a less formal arrangement, such as an oversight committee also able to halt the trial on the basis of interim results) is, it would seem, aimed at avoiding the following ethically troubling situation: the clinicians involved in the trial have already become convinced on the basis of the results so far that one or the other treatment is clearly superior and hence the other is clearly inferior; and yet they are forced to continue the trial and hence randomize some patients to a treatment that they already firmly believe is inferior. Ian Kennedy (1988) puts the problem this way: "As a trial progresses, a trend in the evidence may appear, suggesting that it would be harmful to continue with [the treatment on one "arm"]. . . . Such a trend may appear, of course, before any statistically valid conclusion can be drawn," and he advocates a definite solution:

> In my view . . . the trial must [in such a situation] cease, despite the unfortunate consequences this may have for medical science. To argue otherwise . . . is hard to defend in any ethical system which takes seriously the principle of respect for autonomy. It would put the interests of others, at present unknown, and the interests of science . . . above the interests of the patient whom the doctor has undertaken to treat. (pp. 219–20)

This again raises epistemological issues. Indeed, Kennedy makes an obvious epistemological mistake: trends do not "appear"—the claim that the evidence exhibits a "trend" cannot suffice for a (sophisticated) evidential judgment based on argument. It is not at all clear, as Kennedy suggests, that "a trend may appear . . . before any statistically valid conclusion can be drawn." On the contrary, a frequentist statistician will deny that we can have real evidence for a trend (a definite and constant tendency for one of the treatments to perform better) except from the results of properly designed statistical tests. Everyone must acknowledge

that it is at least logically possible that the first set of results from a trial *may* be significantly unrepresentative.

Again, as in the issue of beginning the trial in the first place, the statisticians' view here starts from the correct position that what matters is not anyone's personal convictions but rather the objective evidence. The hard-line frequentist statistician would go on to argue that there is no ethically troubling situation for the stopping rule to address. Whether or not the researchers involved in the second ECMO trial, for example, would have become convinced that the control treatment was inferior once four babies assigned to it had died is a purely a subjective matter, likely based on a prejudice about earlier historical "evidence" that in fact carries no weight. The researchers in that trial, whatever personal convictions they might have happened to have, were in fact in "objective equipoise," even after four deaths had occurred in the control arm. Stuart Pocock (1993) takes this line. In continuing his critique of the early stopping of the second ECMO trial, he writes: "Thus collective ethics may have been compromised by such early stopping. . . . if ECMO really is effective the prolonged uncertainties maintained by lack of really substantial evidence may well have led to fewer newborn infants worldwide receiving it than would have been the case had the trial continued longer" (p p. 1459–60). The obvious question is why the "investigators were sensitive to the individual ethics of seeking parental consent and randomization for the next newborn infant." After the fourth death on CT, they clearly believed that ECMO was the superior treatment, and therefore that randomizing further babies to the control arm would definitely be condemning them to receive an inferior treatment. The crucial issue is whether this was merely a matter of subjective conviction, or whether the objective evidence at this point (or even earlier) was telling for the inferiority of CT.

As before, statistical orthodoxy is correct that it is objective evidence, not subjective conviction, that counts—but again, as before, statistical orthodoxy is *not* clearly correct in taking its hard-line view about what counts as real evidence. The fact that the epistemological view embodied in statistical orthodoxy is debatable suffices for showing that arriving at an informed judgment about the ethics requires making an informed judgment about the epistemological-evidential issues. If the epistemological view that earlier "evidence" plus the four deaths on CT was not really telling is correct, then the stopping rule in this trial was not justified. What was needed was a properly powered trial producing a statistically significant result, and the stopping rule automatically scotches the first and, as it turned out, the second also. Indeed, if this epistemological view is correct, then, as Pocock hints, there is an argument that including the stopping rule made the trial unethical—*not* because there was already evidence that ECMO was the superior treatment, but rather because the stopping rule introduced the possibility that patients would be involved in a trial that might be terminated without producing compelling evidence one way or the other for ECMO's effectiveness. If, on the other hand, this view is wrong and the researchers' belief

that the historical evidence—*plus* the outcome of the first trial, *plus* the zero out of nine deaths on the ECMO arm compared to four out of 10 on the CT arm in their own trial—did not simply de facto lead to a subjective conviction on their part that ECMO was superior (but de jure amounted to telling objective evidence that ECMO is the superior treatment), then stopping the trial is ethically to be applauded as having very likely prevented deaths that would have served no extra epistemic purpose. Once again ethics and epistemology are intertwined: no informed view can be taken about the ethics without having an informed view on the epistemology.

Ethics and Informed Consent

Standard practice requires that patients have given their informed consent before being definitively recruited to a trial. Again the interplay between ethics and epistemology readily appears in our case example. *Informed* is a success word, and different views concerning the available evidence may entail different views about whether a particular trial participant has truly been informed (or has been given the opportunity to be truly informed), and hence about whether that trial is ethical. (It is taken as given here that a trial is ethical only if the participants have given their informed consent.) However another feature of the ECMO story provides a particularly sharp version of this perhaps obvious general lesson.

The second trial on ECMO (the first "properly randomized" one) used what is often called the "Zelen method" of obtaining "informed consent" (Zelen 1979). This involves randomizing a patient (in this case a neonate) in advance and without their knowledge, and asking for agreement to be in the trial *only* if they have been randomized to the experimental treatment. (In this case, of course, it was the baby's parents who were asked for consent, and it was ECMO that was the "experimental" treatment.) The justification for using this method relies on the assessment that CT was by definition the treatment that a baby suffering from persistent pulmonary hypertension could then expect to receive in routine practice, and therefore that there was no need to inform patients' parents that a trial was going on at all if their baby was to be assigned to CT. So the outcome was that none of the parents of the 10 babies assigned to CT (of whom four died) were told that their baby was in a trial, or that other babies in the same trial were being given ECMO.

Can these parents be regarded as having given their informed consent, and hence can the trial be regarded as ethical in this respect? The motivation for using the Zelen method in this trial was, presumably, the fear that if parents had been told of the involvement of ECMO then they might have inquired about the earlier evidence (or, depending on one's epistemological point of view, *alleged* evidence) of ECMO's effectiveness and consequently demanded that their child be given ECMO—a demand that might have been difficult to resist, whatever one's views about the true effectiveness of the earlier "evidence." This quandary might in effect then have made it very difficult to conduct the trial.

Again, one's attitude towards this ethical issue will surely depend on one's epistemological view about what does or does not count as genuinely telling evidence. *If* statistical orthodoxy is correct that neither the historically controlled trial nor the first "randomized play the winner" investigation provided substantive scientific evidence for the superiority of ECMO, it might be argued that the investigators were under no obligation to inform patients that, unbeknown to them, they were involved in a trial that included an alternative treatment to the one to which they had been allocated, and that claims had been made about the superior effectiveness of this alternative. (Even then this seems to involve an unacceptable degree of paternalism, in that it presumes that the parents are not likely to understand that the so-called previous evidence for ECMO is nothing of the kind—as those advocating the use of the Zelen method are committed to believe.) If, on the other hand, statistical orthodoxy is incorrect, and these earlier trials do supply reasonably telling evidence of the superiority of ECMO, then it clearly seems ethically unacceptable not to inform parents of the involvement of this alternative treatment.

Whether or not clinicians would have been under an ethical obligation to accede to any demands from such further "informed" parents to switch their child to ECMO again depends on what view is taken of the evidence. Surely not, if statistical orthodoxy is correct: no matter how much priority is given to individual liberty, it cannot be countenanced that patients are entitled to demand from a state- or private insurance–funded system whatever "treatment" meets their fancy on the basis of any view that they care to take about evidence on what does and does not work. On the other hand, surely so, if statistical orthodoxy is incorrect, and the earlier results really do constitute good evidence for ECMO's superior effectiveness. Indeed, as indicated, the trial was already unethical on this latter supposition, independently of any considerations about informed consent.

As in the case of the issues discussed above, there is no need, for the purposes of the present paper, to argue for one view or the other on the epistemological-evidential question—the fact that an argument is needed establishes that, in this respect as in the earlier two, no informed ethical judgment can be made without an earlier informed epistemological judgment about evidence.

REFERENCES

Bartlett, R. H., et al. 1982. Extracorporeal membrane oxygenation for newborn respiratory failure: 45 cases. *Surgery* 92:425–32.

Bartlett, R. H., et al. 1985. Extracorporeal circulation in neonatal respiratory failure: A prospective randomized study. *Pediatrics* 76:479–87.

Baum, M. 1990. The ethics of clinical research. In *Ethics and law in health care and research*, ed. P. Byrne, 1–8. London: John Wiley.

Botros, S. 1990. Equipoise, consent and the ethics of randomised controlled trials. In *Ethics and law in health care and research*, ed. P. Byrne, 9–24. London: John Wiley.

Freedman, B. 1987. Equipoise and the ethics of clinical research. *N Eng J Med* 317:141–45.

Gilbert, J. P., B. McPeek, and F. Mosteller. 1977. Statistics and ethics in surgery and anaesthesia. *Science* 198:684–89.

Herbert, V. 1977. Acquiring new information while retaining old ethics. *Science* 198:690–93.

Kennedy, I. 1988. Consent and randomized controlled trials. In *Treat me right*. Oxford: Oxford Univ. Press.

O'Rourke P. P., et al. 1989. Extracorporeal oxygenation and conventional medical therapy in neonates with persistent pulmonary hyptertension of the newborn: A prospective randomized study. *Pediatrics* 84:957–63.

Pocock, S. 1993. Statistical and ethical issues in monitoring clinical trials. *Stat Med* 12: 1459–69.

Schafer, A. 1982. The ethics of the randomized controlled trial. *N Eng J Med* 307:720.

Truog, R. D. 1993. Randomized controlled trials: Lessons from ECMO. *Clin Res* 40:519–27.

Truog, R. D. 1999. Informed consent and research design in critical care medicine. *Crit Care* 3:R29–R33.

Tukey, J. W. 1977. Some thoughts on clinical trials, especially problems of multiplicity. *Science* 198:679–84.

UK Collaborative ECMO Trial Group. 1996. UK Collaborative randomized trial of neonatal extracoroporeal membrane oxygenation. *Lancet* 348(9020):75–82.

Ware, J. H., and M. D. Epstein. 1985. Comments on extracorporeal circulation in neonatal respiratory failure: A prospective randomized study, R. H. Bartlett et al. *Pediatrics* 76:849–51.

Worrall, J. 2007. Evidence in medicine and evidence-based medicine. *Philos Compass* 2(6): 981–1022.

Zelen, M. 1979. A new design for randomized clinical trials. *N Eng J Med* 300:1242–45.