



13

+

Sham Surgery and Reasonable Risks

Alex John London

One of the most fundamental tensions within clinical research arises from the need to balance the goal of advancing the frontiers of science in order to improve the standard of care available to future persons with the goal of responding with diligence and compassion to the important interests and health needs of the research participants who make such progress possible (Jonas 1969; see also London 2003). Although this tension pervades clinical research, it is particularly salient in the debate over the use of placebo controls in clinical trials.¹ Recently however, the already contentious debate about the conditions under which it is permissible to include a placebo arm in a clinical trial has been further complicated by several clinical trials in which participants have been randomized to a sham-surgery control.

To many critics, sham-surgery controls differ in morally significant ways from traditional placebo controls (Macklin 1999; see also London and Kadane 2002). Generally speaking, traditional placebo controls are inert substances that are chosen precisely because of their causal inefficacy. The primary worry associated with the use of such substances are the opportunity costs that participants may incur from being randomized to them. These opportunity costs are cause for ethical concern, for example, when randomizing participants with a particular medical condition to a placebo deprives them of the opportunity to access an alternative treatment modality that is otherwise available for their condition. While opportunity costs remain a concern in the case of sham-surgery controls, the latter also raise additional concerns because the so-called placebo or sham surgeries often involve actual surgical interventions that carry their own special risks and burdens. Unlike the relatively benign profile of the traditional, inert placebo, sham-surgery controls may require recipients to undergo invasive and

burdensome surgical procedures whose purpose is not to treat the recipient but simply to maintain the methodological rigor of the study. The affirmative risks and burdens that may be associated with such procedures have marked out sham-surgery controls for special scrutiny and concern (Dekkers and Boer 2001; Clark 2002; London and Kadane 2002; Weijer 2002).

However, recent interest in utilizing sham-surgery controls has been motivated, in large part, by an increased desire to subject surgical procedures to what is viewed as the gold-standard for clinical research, the randomized, double-blind, placebo-controlled trial. Unlike pharmaceuticals or implantable medical devices, surgical procedures are largely free from regulatory oversight (Reitsma and Moreno 2002; Bower 2003). Those who defend the use of sham-surgery controls, therefore, emphasize the desirability of requiring rigorous evidence of efficacy for surgical procedures (Freeman et al. 1999; Horng and Miller 2002; Miller 2003). In particular, they claim that a sham-surgery control is necessary in order to effectuate a blinded study in which the first-person experience of participants in each trial arm is comparable. Moreover, proponents of sham-surgical controls are quick to point out that there are many elements of clinical trials that subject participants to risks or burdens without the prospect of direct personal benefit. That is, proponents argue that from an analytical point of view a sham-surgery control is no different from the extra blood draws, spinal taps, or other diagnostic procedures to which subjects are routinely subjected within the context of a well-designed clinical trial but which would not be administered in the context of routine clinical practice.

Perhaps somewhat ironically, therefore, one of the central issues in the debate about the ethics of sham-surgery controls is whether these practices even raise special concerns over and above those that routinely arise in the evaluation of clinical research. While some view them as largely contiguous with existing methods and practices in clinical research, others see them as practices that require special justification or which should be prohibited outright. In either case, these disagreements reflect the significant lack of consensus within research ethics about the moral status of sham-surgery controls.

In the discussion that follows, I argue that even if the ethical issues that are associated with these practices are not qualitatively different from standard ethical issues that arise in the course of clinical research, the ethical issues that these practices do raise are nevertheless particularly important and warrant placing the burden of proof on researchers to show that the use of such a design in a particular case is ethically permissible. On a deeper level, however, I argue that the lack of consensus about the moral status of sham-surgery controls reflects a more profound conflict within the research ethics community about the nature and the extent of the risks to which it is permissible to subject research participants. In order to clarify the ethical issues that are raised by the use of sham-surgery controls, as well and to illustrate

this deeper conflict within the research ethics community, I begin with a brief portrait of three sham-surgery-controlled clinical trials. I then examine several proposed standards for evaluating trials of this kind and argue that none is entirely adequate. Finally, I conclude with a proposal that clinical research must conform to a particular principle of equal respect and argue that sham-surgery controls should be permitted only in cases where they are consistent with such a principle.

A TALE OF THREE SHAM SURGERIES

Case 1. Arthroscopic Surgery for Osteoarthritis of the Knee

Each year roughly 650,000 arthroscopic débridement or lavage procedures are performed as a treatment for osteoarthritis of the knee (Moseley et al. 2002). At roughly \$5,000 per procedure, the annual cost of these procedures is about \$3.25 billion. In 2002 Moseley and colleagues reported the results of a sham-surgery-controlled, double-blind, randomized clinical trial of arthroscopic surgery for osteoarthritis of the knee. In this study, a total of 180 patients with osteoarthritis of the knee were randomized to receive either arthroscopic débridement, arthroscopic lavage, or “placebo surgery.” Those who were randomized to the sham-surgery control were given a short-acting intravenous tranquilizer and an opioid while members of the surgical team simulated a standard arthroscopic débridement procedure. In other words, subjects were conscious in the operating theater, where their knees were draped and prepped and the surgical team manipulated both the subject’s knee and the medical instruments as though standard operations were being performed. In actuality, subjects received only a one-centimeter incision in their skin; no instruments were inserted into the opening. Each group received comparable postoperative care consisting of walking aids, a graduated exercise program, and analgesics. Each subject was followed for two years and the primary end point of the study was pain, although functionality was a secondary efficacy endpoint.

Case 2. Fetal Nigral Cell Transplants for Parkinson’s Disease

Parkinson’s disease is a degenerative neurological disorder characterized by a loss of dopaminergic neurons in the basal ganglia of the brain, producing tremors, muscle rigidity, and abnormal movements. The standard treatment, oral doses of the dopamine precursor Levodopa, reverses these symptoms in most patients, but over time its effects tend to wear off and its side-effect profile increases. In 1999 Freeman and colleagues reported the results of a double-blind, randomized, sham-surgery-controlled trial in which 36 subjects were randomized to one of three arms, two receiving bilateral fetal

nigral transplantation and one receiving bilateral placebo surgery. Throughout the study, all subjects continued to receive standard medical therapy. Just as members of the first two arms underwent two surgical procedures, the control group received two placebo surgical procedures that were designed “to provide an equivalent experience for the subjects and their family members” (Freeman et al. 1999). Each placebo procedure involved the placement of a stereotactic frame—a frame attached to the cranium with surgical screws, which allows for accurate location of targeted areas in the brain—a magnetic resonance imaging (MRI) scan, a positron-emission tomography (PET) scan, the administration of general anesthesia, and the drilling of two dime-sized burr holes into the skull through scalp incisions. In the control group, the burr holes did not penetrate the dura and no material was injected into the brain. All subjects, however, received intravenous antibiotics and cyclosporine for six months after surgery.

Case 3. Trial of Glial Cell Line–Derived Neurotrophic Factors (GDNF) for Parkinson’s Disease

In 2003, Nutt and colleagues reported the results of a randomized, double-blind, placebo-controlled trial of GDNF for Parkinson’s disease. This phase 1–2 trial was designed to assess the safety, tolerability, and biological activity of GDNF—a peptide that had been shown to promote the survival of dopamine neurons in animal models—in subjects with moderately advanced Parkinson’s disease. Its primary end points for safety and tolerability were adverse events, vital signs, and various laboratory measures, and the trial was not powered to detect specific changes in any efficacy measure. In this trial, each of the 80 enrolled subjects received an implanted intracerebroventricular (ICV) cannula connected to an access port that was implanted under their scalp. In other words, each subject was placed into a stereotactic frame and had an opening drilled through their skull into which a small flexible tube was inserted so that either the study material or a placebo could be delivered to the subject’s brain. Subjects were then randomized to receive either escalating doses of ICV GDNF or an ICV placebo for a period of six to eight months.

CONFLICTING EVALUATIONS

The recent debate over the use of sham-surgery controls has focused, in large part, on central ethical issues that are spelled out in the U.S. code of federal regulations which institutional review boards (IRBs) in the United States are required to use in evaluating particular research initiatives. These regulations require IRBs to ensure that:

1. Risks to subjects are minimized: (i) by using procedures which are consistent with sound research design and which do not unnecessarily expose subjects to risk, and (ii) whenever appropriate, by using procedures already being performed on the subjects for diagnostic or treatment purposes.
2. Risks to subjects are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result. In evaluating risks and benefits, the IRB should consider only those risks and benefits that may result from the research (as distinguished from risks and benefits of therapies subjects would receive even if not participating in the research). The IRB should not consider possible long-range effects of applying knowledge gained in the research (for example, possible effects of the research on public policy) as among those research risks that fall within the purview of its responsibility.²

Although each side of this debate couches its ethical analysis in the language that is laid out in the above guidelines, each arrives at very different views about some of the above cases. The reason for this conflict in assessments lies in the fact that each side ultimately embraces significantly different views about the larger social role of clinical research and the scope and limits of what it is permissible to ask of research subjects.

For example, the arguments offered by Freeman and colleagues to defend their use of a sham-surgery control in case 2 are emblematic of the approach of those who defend the use of sham-surgery controls more generally. First, Freeman and colleagues argue for the methodological necessity of including a sham-surgery control in their trial. In particular, they claim that it is an indispensable component of a trial that is designed to answer their chosen research question: is fetal-tissue transplantation a safe and effective treatment for Parkinson's disease and, if so, are the observed benefits the result of the fetal-tissue transplant or of some associated placebo effect? Second, they then argue that, against this methodological background, "the risks to participants are reasonable and have been minimized as far as possible" (Freeman et al. 1999, 991). In particular, members of the control group continue to receive standard medical therapy for Parkinson's disease, a partial burr hole is used instead of penetrating the dura, no material is inserted in the brains of subjects, and renal function is monitored to detect adverse reactions to cyclosporine.

The claim of Freeman and colleagues that the risks to the control group have been minimized in their trial takes the risks that are associated with the active arm as the proper baseline of evaluation. Whereas the holes in the heads of subjects in the active arm penetrate the dura, those of control arm do not. Whereas subjects in the active arm have material inserted into their brains, members of the control group do not. Relative to this baseline, the risks to members of the control group have been minimized as far as is consistent with maintaining the integrity of a sound clinical trial design. Although sham-surgery controls may subject participants to risks and burdens

that are not compensated for by any prospect of direct therapeutic benefit, this approach claims that their use does not *unnecessarily* expose subjects to risk as long as those risks and burdens have been minimized as far as is consistent with a sound trial design.

In contrast, critics of sham-surgery controls are reluctant to use the risks that are associated with the intervention in the active arm as the baseline for determining whether the risks to members of the control group have been minimized. For example, in her evaluation of case 2, Ruth Macklin argues that “the question of how great the risks of sham surgery are in any particular trial is distinct from the question of whether a surgical intervention carries risks of harm that are greater than those associated with no surgical intervention” (1999, 993). Here, Macklin appears to be asserting that the proper baseline against which the risks to members of the control group should be evaluated is the situation in which they are not subjected to any surgical procedure at all. In other words, the proper baseline in this particular case would be the provision of their standard medical therapy and a more traditional, inert placebo. It is against this background assumption that Macklin asserts that “it is undeniable that performing surgery in research subjects that has no potential therapeutic benefit fails to minimize the risks of harm” (993).

On this particular issue, therefore, the difference between critics and proponents of sham-surgery controls boils down to the more fundamental question of how to set the proper baseline against which the risks to subjects in the control group are evaluated. Each of the above proposals has one distinct virtue: operational clarity. That is, Macklin and Freeman and colleagues each provide clear, though very different, standards for determining the limits of the risks to which members of the sham-surgery arm of a trial may permissibly be exposed. Moreover, each of their proposed standards yields a determinate evaluation of each of the three cases described above. Despite this singular virtue, however, each position appears to err in opposite directions. Whereas Macklin’s position is overly conservative, the approach of Freeman and colleagues is overly permissive.

The conservative nature of Macklin’s position can be illustrated by consideration of case 1. Proponents have been quick to argue that the sort of sweeping condemnation of sham-surgery controls that results from Macklin’s position would rule out as unethical the sham-surgery control that Moseley and colleagues employed in this study (Miller 2003, 45–46). After all, subjects in the control arm received a tranquilizer, an opioid, and one-centimeter incisions in the skin of their knees. These risks are greater than the baseline situation of not receiving any surgery at all and would therefore be ruled out as ethically impermissible in Macklin’s view. In fact, it looks like Macklin’s view would rule out a similar trial design in which subjects were spared the risks associated with the tranquilizer and the opioid but were still subjected to one-centimeter skin incisions in their knees. To critics, this result reduces such a rigid and conservative position to absurdity.

In contrast, the approach endorsed by Freeman and colleagues would justify not only each of three cases described above but also the use of even the most invasive and burdensome sham-surgery controls. The permissiveness of this approach results from two factors (London and Kadane 2002). First, because it evaluates the risks to subjects in the control arm against the baseline of the risks to which subjects are exposed in the active arm or arms, this approach would permit extremely invasive and burdensome sham-surgery controls so long as those controls had fewer risks and burdens than the intervention in the active arm. If, for instance, the active arm were a new coronary surgery, this approach would permit a sham-surgery control in which the body cavity of subjects is opened but no additional intervention is performed.

Such a control would have to be justified, of course, as having reduced the risks as far as possible consistent with the integrity of the trial design. Notice, however, that the integrity of a trial design relates to its ability to generate the information necessary to answer the particular question that the trial is designed to answer. The second feature of this approach that makes it overly permissive is that it provides few resources, if any, for assessments of whether the question that the trial is designed to answer is an appropriate or acceptable question to pursue at that time. The level to which the risks to subjects in the sham-surgery arm must be reduced is a direct product of the specific question that the researchers have chosen to pursue. If, as in case 3, researchers want to distinguish the clinical effects of the implanted material from the effects of the general surgical procedure that accompanies that intervention, they would be largely free to subject members of the control group to highly invasive and burdensome sham-surgical procedures in order to do so. Although such information may always be of interest from a purely scientific point of view, such fastidiousness may be unnecessary from a more pragmatic or clinically oriented point of view. The approach under question, however, lacks the internal resources to draw such lines in a clear and principled way.

Worries of the latter sort are particularly salient in case 3 above. Although intracerebroventricular administration of GDNF had shown promising results in rodent and monkey models of Parkinson's disease, its tolerability, safety, and effectiveness in humans had not been established. The purpose of the study by Nutt and colleagues was to assess the safety, tolerability, and biological activity of ICV GDNF in patients with advanced Parkinson's. The placebo-controlled design was used to maintain the double-blinded standard and to hold constant changes in the condition of recipients that might be due to the ICV catheter and the administration of a substance into the brain, thereby more accurately isolating effects specifically related to the GDNF itself. This particularly intrusive placebo control therefore enabled the researchers to isolate adverse events that were associated with the administration of GDNF from those associated with the surgical elements of the procedure itself.

In this case, serious questions can be raised about the appropriateness of trying to answer this explanatory question rather than assessing the safety and tolerability of ICV administration of GDNF and *all that it entails* relative to the baseline condition of subjects with advanced Parkinson's disease receiving a more benign placebo—one that involved more sham and less surgery. After all, the various components of this intervention cannot be separated in practice; one cannot administer GDNF to the brain of a patient without creating a pathway of access through the skull. From a clinical standpoint, therefore, concerns about safety and tolerability do not apply simply to the GDNF and its effects on the brain. Rather they encompass all necessary elements of the procedure, including the ICV cannula and access port. Even if we grant the claim that Nutt and colleagues reduced the risks to study participants as far as possible, consistent with the integrity of their preferred trial design, it is questionable whether it was appropriate at the time to ask the particular research question that required such a fastidious and burdensome trial design.

Prospective worries of this sort appear to be borne out by the results of this particular study. The trial was terminated after a total of 50 participants, 12 of whom were randomized to the placebo, completed the double-blind portion of the study. With respect to effects on measures relating to parkinsonism, the placebo weakly dominated the active intervention, meaning that there were either no significant differences between the placebo and the active agent or the measured difference favored the placebo. Similar results were obtained on measures of safety and tolerability. Whereas 92 percent of subjects randomized to placebo suffered treatment-associated adverse effects, adverse effects were reported by 100 percent of subjects who received the active agent. Adverse effects that related to the implanted cannula and access port included headache (25 percent of the placebo group, 71 percent of the active group) and nausea (25 percent of the placebo group, 87 percent of the active group), with serious adverse events including an extended hospitalization of a patient due to difficulties removing the device and a bacterial colonization of the access port in another patient whose port therefore had to be removed and reimplanted.

Although the nature of the sham-surgery in case 3 is significantly more burdensome than the one employed in case 2, similar reservations have been articulated in the latter case as well (London and Kadane 2002). In both cases, legitimate questions arise about the permissibility of employing the burdensome methods necessary to explain which effects measured in the trial are associated with which elements of the experimental intervention before one has addressed the more pragmatic issue of whether or not the experimental procedure as a whole has effects that make it attractive from a clinical point of view.

UNCERTAINTY ABOUT REASONABLE RISKS

Not all proponents of sham-surgery controls are wedded to the framework proposed by Freeman and colleagues. Miller, for example, has argued that

Macklin's position is overly restrictive largely on the grounds that it misconstrues the requirements of the federal regulations (Miller 2003, 45–46). After all, although condition 2 in the above-cited regulations requires that risks must be reasonable in relation to the benefits subjects may receive from participating in the research, it is also clear that subjects need not themselves receive any such benefit for the research to be acceptable. What condition 2 actually requires in this regard is an evaluation of the reasonableness of the risks posed to participants in relation to the importance of the knowledge that may result from the research (Weijer 2000).

Miller argues that in case 1, the risks associated with the sham-surgery control were justified in light of the methodological rigor of the trial design and the importance of the research question. After all, the resources that are expended each year on the procedures being studied are far from inconsequential and the surgical procedure itself is far from risk free. Moreover, although a research initiative must be judged as ethical or unethical at its inception, Miller claims that the merits of their position are borne out by the actual results of the trial. At the conclusion of the study, Moseley and colleagues reported that there was no point at which recipients of either of the active interventions reported less pain or better functionality than recipients of the sham-surgery control. In fact, recipients of actual débridement had poorer objective measures of walking and stair climbing at two weeks and one year, and they showed a trend toward worse functioning at two years than did recipients of the sham-surgery control. The research that generated these significant findings, they emphasize, would have been prohibited under Macklin's guidelines.

Although Miller's approach to assessing the risks to which it is permissible to subject research participants is more permissive than the one articulated by Macklin, it lacks the operational clarity that is a hallmark of her approach as well as the one endorsed by Freeman and colleagues. In particular, neither Miller's approach nor the federal regulations from which it is derived provide an account of (1) what constitutes "the importance of the knowledge that may reasonably be expected to result" from a particular research initiative or (2) how this might be measured. The practical result of this absence of operational clarity is that the boundaries of what different deliberators might accept as a measure of the importance of such knowledge is set only by the limits of the imaginations of the various deliberating agents (London and Kadane 2003). Moreover, even if a set of agents share the same view on this issue, neither Miller nor the federal regulations provide a clear standard for determining the permissible limits of the risks to which research subjects may be subjected in *exchange for such gains in knowledge*—however they are understood.

On this point, at least, Miller is clear: he holds that "the ultimate question of risk-assessment" is whether the risks of sham surgery are justified by the anticipated scientific value of the study, and says that, ultimately, "we lack any objective tools for measuring research risk–benefit ratios" (Miller 2003,

46). However, if we remain content with such a state of affairs, then we must also be prepared to accept the reality that the discrepancies between proponents such as Miller and critics like Macklin simply boil down to different intuitions about whether the risks of sham-surgery controls are in some inchoate sense “worth it” (Kim 2003). Such an approach provides little guidance about how to bring reasoned resolution to the significant range of unresolved conflicts that exists when deliberators who apply a shared set of standards are incapable of reaching a consensus on the moral standing of a particular research initiative. In the worst case, the field would simply be divided by conflicting intuitions about how to proceed here as well. In Macklin’s opinion, for example, when reasonable people disagree about the risks that subjects bear in sham-surgery-controlled clinical trials, the default position should be a conservative approach to the use of such practices. In contrast, proponents of sham-surgery controls may incline more toward a position that allows methodologically rigorous clinical research to go forward unless there is a social consensus that it would be unethical to do so.

While Miller seems to think that his approach would be less permissive than the one I am attributing to Freeman and colleagues, it is unclear whether this would necessarily be so. Because the judgments that Miller’s view requires deliberators to make are largely intuitive, and because the framework provides little operational guidance about how to make such judgments, it is not clear that it would help well-intentioned deliberators form a considered opinion about more controversial cases such as case 2 or case 3. Nor is it clear that Miller’s approach is inconsistent with that of Freeman and colleagues. In fact, it seems clear that *both* Macklin and Freeman and colleagues could reasonably argue that their positions should be understood as presuming Miller’s general framework in which risks to subjects are balanced against gains in knowledge. That is, both could argue that they are presenting operationally clear, concrete proposals regarding *how to determine* whether the risks to subjects are permissible in light of the potential benefits to science.³

The fact that views as disparate as those of Macklin and Freeman and colleagues can be presented under the rubric of ensuring the reasonability of risks provides a powerful illustration of the uncertainty surrounding the operational content of this general requirement. It also illustrates how relatively local skirmishes over the ethics of sham-surgery controls reflect these more fundamental uncertainties. This connection helps to explain my contention that sham-surgery controls raise issues of special ethical significance, even if these issues are not qualitatively different from those that are faced in trials that are not sham-surgery controlled. Persistent uncertainties about sham-surgery controls are symptomatic of a larger uncertainty in the field concerning the limits of the risks to which trial participants may permissibly be subjected. Although sham-surgery controls highlight in a particularly dramatic fashion the extent to which the demands of science may be at odds

with the interests of trial participants, other elements of clinical trials frequently raise similar issues. This general problem is of fundamental moral importance because it deals with the extent to which the interests of trial participants can be compromised or sacrificed for the good of future persons. Any effort to forge a consensus about the ethical use of sham-surgery controls, therefore, will have to confront ambiguities surrounding these larger issues explicitly and in a principled manner.

FAIRNESS TO EQUALITY: A TENTATIVE PROPOSAL

I suggested above that the central difference between Macklin and Freeman and colleagues lies in the different baselines they use to evaluate the risks in a clinical trial. Although the difficulties associated with each of their approaches stem from limitations in the specific baselines they adopt, the fact that each articulates such a clear baseline for making such judgments gives their views an operational clarity that is missing in the Common Rule. Any view that hopes to overcome these limitations in a way that preserves this kind of operational clarity will have to provide a clear answer to each of the following questions (London 2005).

1. How should the concept of “reasonable risk” be understood?
2. What are the criteria or practical markers that can be used in order to delineate in an operationally useful way the parameters or boundaries that separate reasonable from excessive risks?
3. What are the tests or mechanisms that deliberators can use in order to determine whether or not these operational criteria have been met in any particular case?

In this section I will outline tentative answers to each of these questions and show how the resulting framework would discriminate between the cases described above.

To begin with, it is not sufficient to link the reasonableness of a risk simply to the importance of the prospective benefits of research. One must go further and specify whether or not, for the individual trial subject, there is a threshold beyond which risks cannot be outweighed by benefits in science. As a starting point, therefore, we should consider what reasonable limitations there should be on the risks that it is permissible for society to allow its clinical researchers to offer to prospective trial participants. Put slightly differently, if the institution of clinical research is going to function as part of a social division of labor that is justifiable to the members of the community whose interests it is supposed to serve, then what are the limits of the risks that it may offer to prospective participants while serving this function?

Each member of society can recognize that all have a fundamental interest in ensuring that the basic social structures of their community function to safeguard their most basic human interests. In this context, “basic interests” refers to the set of interests that each individual community member shares with all other community members in being able to cultivate and exercise their rudimentary intellectual, affective, and social capacities in the pursuit of a meaningful life plan. Basic interests should be distinguished from what I will call “personal interests,” in that the latter are interests that individuals form as the result of pursuing a particular set of goals or plans and as the result of exercising the set of basic interests that make these ends and activities possible (London 2003). In liberal democratic communities, individuals may differ widely in their personal interests and may reasonably disagree about the value or significance of the personal interests of their fellow citizens. Nevertheless, each is capable of recognizing that all share a set of basic interests in being able to cultivate and exercise the rudimentary intellectual, affective, and social capacities that make it possible for them to form, pursue, and revise their respective personal interests.⁴

Because basic interests can be profoundly restricted or defeated by sickness and disease, each can recognize a reason to support medical research as a social institution insofar as it strives to advance the state of medical science and therefore the standard of care that is available to community members. As one element within a larger social division of labor that must be justifiable to the members of the community whose basic interests it is supposed to serve, clinical research must pursue its goal of advancing the interests of future patients in a way that is consistent with an equal regard for the basic interests of the present persons whose participation makes those results possible.

The requirement to respect the basic interests of both present participants and future beneficiaries supports the following proposal for a definition of *reasonable risk*.

Concept of Reasonable Risk: Reasonable risks are those that are necessary in order to generate important scientific information and that are consistent with an equal regard for the basic interests of study participants and the members of the larger community whose interests that research is intended to serve.

The requirement that risks be consistent with an equal regard for the basic interests of study participants and members of the larger community is intended to reflect the idea that even if the beneficiaries of the research enterprise can cite a moral imperative to carry out research as part of an effort to help to safeguard their basic interests, such an imperative cannot legitimate requiring others to sacrifice or to forfeit their basic interests in the process.⁵ In other words, the same concern to advance the interests of future patients

that underwrites the research enterprise as a social institution cannot be withheld from present, prospective research participants.

Operational criteria can be generated for this conception of reasonable risk by considering how the basic interests of community members are safeguarded and advanced by the larger social division of labor. For the kinds of cases we have been considering here, where the basic interests of persons are threatened or restricted by sickness, injury, or disease, this job falls, in large part, to the health care system. This suggests adopting the following operational criterion for determining whether efforts to advance the standard of care for future patients show an equal regard for research participants and nonparticipants.

Operational Criterion: Within the research context, equal regard for the basic interests of participants and nonparticipants requires that the basic interests of participants be protected and advanced in a way that does not fall below the threshold of competent medical care.

It is important to note several features of this operational criterion. First, its scope is limited to the basic interests of participants for two primary reasons: (1) It is supposed to delineate the level of risk that it is permissible to *offer* to prospective participants; participants are then free to decide for themselves whether the risks that remain in a trial that meets this standard are acceptable in light of their particular personal interests. (2) The focus on basic interests reflects the normative claim that it is permissible to ask individual community members to alter, risk, or even to sacrifice some of their *personal interests* as part of an effort to advance or secure the *basic interests* of others. It is not, however, permissible to ask community members to sacrifice their basic interests in order to advance or safeguard the personal interests of others.

Another significant feature of the stated operational criterion is that several reasons underwrite using the threshold of competent medical care as a practical standard for determining whether the level of protection and care that a research initiative provides is consistent with an equal regard for the basic interests of participants and nonparticipants: (1) Competent medical care represents the socially enforceable standard of professional knowledge, skill, and ability that community members have a legitimate claim to receive when they access the medical system; although some clinicians may rise above the rest in terms of various professional excellences, competent medical care denotes the level of care that the medical profession is accountable for providing on a uniform basis. (2) Competent medical care refers to the use of practices, procedures, and methods that have a reasonable likelihood of success; in this respect, it serves as an indicator of what expert medical professionals believe is a causally efficacious means of effectuating desired clinical goals.

It is often the case that competent medical care does not provide a single, well-ordered standard of care for dealing with particular medical conditions. The boundaries of competent care often include a variety of alternative approaches for dealing with a particular condition. This may be the result of different traditions of practice that adopt different methods for dealing with a particular condition. It can also result when there is uncertainty within the expert medical community about what constitutes the optimal method of dealing with a condition generally or for treating specific individual patients with that particular medical condition.

The fact that there is often no single, well-ordered standard of competent medical care can thus be used to create a practical test for assessing whether particular research initiatives are acceptable in light of the operational criterion articulated above.

Practical Test: For each individual within a particular clinical trial, the care and protection that is provided to that individual's basic interests falls within the threshold of competent medical care when it represents an admissible intervention in light of either uncertainty in the form of agnosticism or conflict in the expert medical community about the relative net therapeutic advantage of that package of care in comparison to alternative packages that are available either within the trial itself or within the context of clinical care.

This practical test is similar to what Freedman (1987) referred to as "clinical equipoise," although there are some important differences.⁶ For example, clinical equipoise is almost universally viewed as deriving its moral force from norms that are internal to the doctor-patient relationship. In particular, the equipoise requirement is traditionally supported as a means of reconciling the demands of sound scientific practice with the physician's therapeutic obligation. The moral force of the practical test outlined here, however, is grounded in a different source, namely, in the claim that it represents a requirement that is necessary to justify the conduct of scientific research as one element of a larger social division of labor that must be justifiable to each of the individual members of the community whose interests that division of labor is intended to serve.

Similarly, the equipoise requirement is sometimes applied to entire trial populations, whereas the above practical test is to be applied to each prospective trial participant individually. The reason for this is simply that conflict or uncertainty about the relative therapeutic merits of a set of interventions may exist for some individuals and not for others, depending on their particular clinical characteristics.

The above test is also explicit in distinguishing between uncertainty that arises from a state of agnosticism and uncertainty that arises from a state of conflict. Briefly, clinical agnosticism refers to the situation in which the ex-

pert medical community as a whole is in agreement that there are not sufficient grounds to make a definitive judgment as to the relative therapeutic merits of a set of interventions for a particular patient. Such a situation might arise, for example, in the case of a new, investigational intervention for a condition that is currently untreatable. Clinical conflict exists, however, when expert clinicians have definite opinions about the superiority of one intervention over another, for example, but their opinions are in conflict, with one physician (or set of physicians) preferring intervention A over B for patient P and another preferring B over A for P. According to the standard elaborated above, it would be permissible to offer to P the option of participating in a clinical trial in which she would be randomized to either A or B because both interventions are regarded as admissible in this scenario.

Although this very brief sketch requires significant additional elaboration, it is nonetheless sufficient to discriminate between the cases of sham surgery that were described above. In particular, it highlights as salient two significant differences between case 1, on the one hand, and cases 2 and 3 on the other.

First, arthroscopic surgery for osteoarthritis of the knee is the subject of genuine conflict in the clinical community, with a significant portion of practitioners offering this intervention to their patients and a significant portion of the expert medical community either uncertain about or affirmatively skeptical of the therapeutic merits of this intervention. In such a case, it may be permissible to perform what is referred to as an "explanatory" trial that is capable of passing the practical test articulated above.⁷ In such trials, the goal is to identify which specific components of a procedure are responsible for its causal efficacy. Such a trial might pass the above test if a reasonable minority of reputable medical experts perceive some therapeutic merits to the actual arthroscopic procedure but disagree about whether these benefits result from actual débridement and lavage or from the "experience of surgery." If this were the case, then this particular sham-surgery control might be admissible on the grounds that some experts believe that it offers the prospect of a benefit that would not be received in a no-treatment arm with fewer risks than those that are associated with the actual procedure. Randomization to this arm would not provide subjects with a level of care or protection for their basic interests that falls below what is shown for subjects who seek care directly.⁸

Such a justification does not seem to be available in cases 2 and 3. In these cases, the operative clinical question is whether the new interventions being proposed provide a net therapeutic advantage over the unaugmented existing standard of care. Before it would be permissible to answer the explanatory question of which elements of these interventions are responsible for their causal efficacy, therefore, it must first be established that these interventions are efficacious in a way that makes them attractive as clinical interventions.

Second, there is an alternative prima facie case for the permissibility of the sham-surgery control in case 1. This approach has several components: (1) The state of conflict in the clinical community in case 1 makes both the procedure and no-treatment admissible interventions. (2) The operational criterion and the practical test articulated above each relate to the basic interests of trial participants and not to their broader set of personal interests; in this case, the risks associated with the sham-surgery control are largely limited to harms that pose the most credible or material threat to the personal and not the basic interests of trial participants. (3) Finally, the risks to the personal interests of subjects that are posed by the sham-surgery control in this case have been minimized as far as is consistent with preserving the blind of the study. Together, these considerations support the prima facie claim that it would be permissible to offer a trial with this risk profile to subjects on the grounds that it provides equal regard for their basic interests and for the basic interests of other community members while allowing participants to decide whether the risks that the trial poses to their personal interests are acceptable in light of its scientific goals.

This alternative prima facie case is more difficult to make in support of case 3. In particular, it is difficult to see the risks associated with this sham surgery as limited to the personal interests of participants. In part, this is due to the much higher ratio of surgery to sham in this control. That is to say, unlike case 1 where the sham surgery involved much more sham or theater, the sham surgery in this case is actually very invasive; holes are drilled through the skull, a catheter attached to an access port is inserted into the subject's head, and at regular intervals for six months saline injections are delivered to the subject's brain. In large part, however, this is due to the fact that the anticipated adverse effects associated with this control constitute more significant impediments to the functionality of subjects whose abilities to pursue their particular life projects is already being restricted by a degenerative illness.

It may be the case, however, that this sort of alternate prima case could be used to support something similar to the trial that was conducted in case 2. To build such a case, one would have to show that the risks associated with the sham surgery have been limited to harms that pose the most credible or material threat to the personal interests of subjects and that the remaining risks to the personal interests of participants had been reduced as far as possible, consistent with effectuating a blinded study. This might be done not only by utilizing partial burr holes that do not go completely through the skull but also by substituting placebo substances for any of the antibiotics or other medications whose provision cannot be justified by their therapeutic merits for the recipient. Generally speaking, the clearer it becomes that the risks associated with the sham-surgery control are limited to the personal interests of participants and that they have been reduced as far as is consistent with the integrity of the trial design, the greater the prima facie case that can

be made in support of the trial. In such cases, offering the option of participating in such a trial is consistent with a regard for the basic interests of participants that is equal to that which is being shown for the interests of future beneficiaries that motivates the research in the first place.

As it actually stands, however, the trial in case 2 does not meet the conditions necessary for justification under either of the approaches sketched above. Clearly, the framework that has been articulated here requires both further refinement and clarification as well as a more substantial philosophical defense. In particular, significantly greater attention will have to be paid to the distinction between basic and personal interests. Nevertheless, even this relatively rudimentary sketch is sufficient to highlight some of the features of sham-surgery controls that can make them morally problematic. In particular, the affirmative risks that are associated with sham surgeries can endanger the basic interests of trial participants in a way that is not consistent with the same kind of respect and regard that motivates the very quest to advance the boundaries of scientific understanding for the benefit of future persons.

NOTES

1. For example, compare Freedman 1990 and Rothman and Michels 1994 with Miller and Brody 2002.
2. 56 Federal Register 28012, 45 CFR 46.
3. For a general elaboration of this criticism against Miller's proposed standard and a defense of equipoise against some of his recent criticisms, see London 2006.
4. For a defense of this claim, see Rawls 1982.
5. For further discussion, see London 2003, 2006.
6. For a fuller elaboration of the differences between the view articulated here and Freedman's position, and for a general defense of this particular understanding of equipoise, see London 2006a.
7. On the difference between pragmatic and explanatory trials, see Schwartz, Flamant, and Lellouch 1980. On the relevance of this distinction to the evaluation of trials involving a sham surgery control, see London and Kadane 2002.
8. This argument is elaborated more fully in London and Kadane 2003. There, we emphasize that although an argument of this form may be sufficient to justify the conduct of this trial, we remain skeptical about whether the disagreement in the medical community over the therapeutic merits of this trial was actually grounded in an assessment of the available data.

REFERENCES

- Bower, Vicki. 2003. The ethics of innovation: Should innovative surgery be exempt from clinical trials and regulations? *EMBO Reports* 4 (4): 338–40.
- Clark, Peter A. 2002. Placebo surgery for Parkinson's disease: Do the benefits outweigh the risks? *Journal of Law, Medicine & Ethics* 30 (1): 58–68.
- Dekkers, Wim, and Gerard Boer. 2001. Sham neurosurgery in patients with Parkinson's disease: Is it morally acceptable? *Journal of Medical Ethics* 27: 151–56.

- Freedman, Benjamin. 1987. Equipoise and the ethics of clinical research. *New England Journal of Medicine* 317: 141–45.
- . 1990. Placebo-controlled trials and the logic of clinical purpose. *IRB* 12 (6): 1–6.
- Freeman, Thomas B., Dorothy E. Vawter, Paul E. Leaverton, James H. Godbold, Robert A. Hauser, Christopher G. Goetz, and C. Warren Olanow. 1999. Use of placebo surgery in controlled trials of a cellular-based therapy for Parkinson's disease. *New England Journal of Medicine* 341 (13): 988–92.
- Hornig, Sam, and Franklin G. Miller. 2002. Is placebo surgery unethical? *New England Journal of Medicine* 347 (2): 137–39.
- Jonas, Hans. 1969. Philosophical reflections on experimenting with human subjects. *Daedalus* 98 (2): 219–47.
- Kim, Scott Y. H. 2003. The sham surgery debate and the moral complexity of risk-benefit analysis. *American Journal of Bioethics* 3 (4): 68–70.
- London, Alex John. 2003. Threats to the common good: Biochemical weapons and human subjects research. *Hastings Center Report* 33 (5): 17–25.
- . 2005. Does research ethics rest on a mistake? The common good, reasonable risk and social justice. *American Journal of Bioethics* 5 (1): 37–39.
- . 2006. Clinical equipoise: Foundational requirement or fundamental error. In *Oxford Handbook of Bioethics*, ed. Bonnie Steinbock. New York: Oxford University Press.
- London, Alex John, and Joseph B. Kadane. 2002. Placebos that harm: Sham surgery controls in clinical trials. *Statistical Methods in Medical Research* 11 (5): 413–27.
- . 2003. Sham surgery and genuine standards of care: Can the two be reconciled? *American Journal of Bioethics* 3 (4): 61–64.
- Macklin, Ruth. 1999. The ethical problems with sham surgery in clinical research. *New England Journal of Medicine* 341 (13): 992–96.
- Miller, Franklin G. 2003. Sham surgery: An ethical analysis. *American Journal of Bioethics* 3 (4): 41–48.
- Miller, Franklin G., and Howard Brody. 2002. What makes placebo-controlled trials unethical? *American Journal of Bioethics* 2 (2): 3–9.
- Moseley, J. Bruce, Kimberly O'Malley, Nancy J. Petersen, Terri J. Menke, Baruch A. Brody, David H. Kuykendall, John C. Hollingsworth, Carol M. Ashton, and Nelda P. Wray. 2002. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *New England Journal of Medicine* 347 (2): 81–88.
- Nutt, J. G., K. J. Burchiel, C. L. Comella, J. Jankovic, A. E. Lang, E. R. Laws Jr., A. M. Lozano, R. D. Penn, R. K. Simpson Jr., M. Stacy, and G. F. Wooten. 2003. Randomized, double-blind trial of glial cell line-derived neurotrophic factor (GDNF) in PD. *Neurology* 60: 69–73.
- Rawls, John. 1982. Social unity and primary goods. In *Utilitarianism and beyond*, ed. Amartya Sen and Bernard Williams, 159–85. Cambridge: Cambridge University Press.
- Reitsma, Angelique M., and Jonathan D. Moreno. 2002. Ethical regulations for innovative surgery: The last frontier? *Journal of the American College of Surgeons* 194 (6): 792–801.
- Rothman, Kenneth J., and Karen B. Michels. 1994. The continuing unethical use of placebo controls. *New England Journal of Medicine* 331: 394–98.
- Schwartz, D., R. Flamant, and J. Lellouch. 1980. *Clinical trials*, trans. M. J. R. Healy. London: Academic Press.
- Weijer, Charles. 2000. The ethical analysis of risk. *Journal of Law, Medicine & Ethics* 28: 344–61.
- . 2002. I need a placebo like I need a hole in the head. *Journal of Law, Medicine & Ethics* 30 (1): 69–72.