



Published in final edited form as:

Hastings Cent Rep. 2015 ; 45(1): 30–40. doi:10.1002/hast.407.

SUPPORT and the Ethics of Study Implementation:

Lessons for Comparative Effectiveness Research from the Trial of Oxygen Therapy for Premature Babies

John D. Lantos and Chris Feudtner

The Surfactant, Positive Pressure, and Oxygenation Randomized Trial (SUPPORT) has been the focal point of many different criticisms regarding the ethics of the study ever since publication of the trial's findings in 2010 and 2012. These criticisms center on two major concerns. The first concern rests on a belief that, when the SUPPORT study was being designed, the state of medical knowledge was already sufficiently clear about the optimal amount of oxygen to give premature babies and that therefore any study of differing oxygen therapy protocols was unnecessary and unethically exposed infants to a treatment known to be inferior. In this paper, we will not address this point. Instead, we will assume, as did the Office for Human Research Protection,¹ that the study addressed an important area of uncertainty in neonatology, that there was genuine uncertainty in the expert community as to which level of oxygen was the best, and that there was widespread practice variation within the range of oxygen levels that were generally thought to be acceptable.

In this article, we focus on the second major concern, namely, that although SUPPORT may have been warranted based on the uncertainty of the best oxygen therapy strategy for premature infants, the technical design and implementation details of the study itself were ethically flawed. While the OHRP focused on the consent form, rather than on the study design and implementation, OHRP's critiques of the consent form reveal views about the study design and implementation that we believe are fundamentally flawed. These views were more fully articulated by Public Citizen and many others who criticized the study at an open meeting held by the Department of Health and Human Services (HHS) in August 2013. These criticisms about the design and implementation of SUPPORT, if generalized, become relevant concerns about these aspects of many comparative effectiveness research studies.

CER, by design, compares outcomes between patients who receive two different treatments that are both in widespread use. Some prominent voices have warned that the term "comparative effectiveness research" is used inappropriately (even disingenuously) to suggest that such research is different from research on new treatments. Such critics claim that there are no relevant ethical differences between the two types of clinical research and thus that the term "comparative effectiveness research" is a "smokescreen" meant only "to blur or eliminate the distinctions between research and therapy, scientist and physician, and subject and patient."² We disagree. We believe (along with others³) that CER can be conceptually distinguished from what we might call "innovative therapy research" in which a new treatment is compared with a placebo or a standard practice. The differences arise precisely because, in CER, both therapies are already in widespread use and, in spite of extensive clinical experience, expert physicians do not know which is better. In this situation, patients outside of studies are much more likely to be treated in ways that are very

similar to the ways that patients in CER studies are treated. The primary difference is that, outside of CER studies, the choice of a treatment is by idiosyncratic practice variation, whereas in a CER study the treatment a patient receives is determined by formal randomization. We will elucidate why we think that these circumstances and aspects of CER research are ethically distinct from the conduct of innovative therapy research.

Our analytical approach will be to use the SUPPORT study as a prime example of CER and show why this particular study of premature infants challenges some prevailing assumptions about the riskiness of research. We will address five aspects of the study design and implementation: randomization, treatment by protocol, choice of endpoints, (4) lack of a “standard-care” control group, and the use of altered oximeters. Examining these aspects will allow us to answer two specific central questions. The first is a methodological question with ethical implications: was the study designed in such a way as to answer the primary study question? The second question is whether the study design added or decreased risk to the babies enrolled in the study compared to babies who were not in the study.

Before we begin our analysis, however, we should clarify our concept of “risk” in the particular circumstances of the SUPPORT study. Premature babies born between twenty-four and twenty-seven weeks of gestation—the gestational ages of babies included in the SUPPORT study—are known to have a predictable likelihood of developing certain medical problems. In studies done prior to SUPPORT, the mortality rate for such babies was between 20 and 40 percent. Among survivors, 20 to 30 percent developed moderate to severe retinopathy of prematurity, and a similar percentage developed neurocognitive impairment. Each of these problems is independent of the others. Risk, in this context, is the population- or group-level prevalence of developing these problems. Thus, if among one group of one hundred babies, the mortality rate is 25 percent, and among another group, it is 20 percent, the second group has a lower risk of mortality. The point of the SUPPORT study was to use careful study design, data collection, and statistical analysis in order to measure the degree to which the different approaches to treatment increase or decrease risk observed in groups of patients. For the different outcomes, groups of infants could quite possibly have a higher risk of one particular outcome and a lower risk of another. Risk is, in this view, a measure that can be determined only with regard to a group of patients. Once risk has been determined in a group of patients, then knowledge of this group-level risk can be applied in choosing treatments for individual patients, but always with the background assumption—and a leap of inference—that an individual patient would experience the risk alteration observed for the group. With more data and analysis, the size of the group to which the patient is being compared may grow smaller—and thus the inference of risk for the patient more precise—but individual risk assessment remains an inference. While the idea of determining risk at the level of the individual patient is alluring, it is conceptually impossible.

Randomization and Treatment by Protocol

The first two controversial aspects of the study design—randomization and treatment by protocol—reflect very similar ethical concerns. Many critics of SUPPORT contended that physicians who randomized patients to predetermined treatment protocols compromised

their ethical obligation to do what was best for each patient. Instead, in this view, such physicians were primarily adhering to an ethical obligation to do what was best for the clinical trial, for science, and for the creation of generalizable knowledge, even if this required them to sacrifice the interests of individual patients. Such concerns were articulated forcefully by the citizen's advocacy group Public Citizen. In a letter to the Secretary of Health and Human Services, they wrote,

As part of routine care for such infants outside the research context, oxygen therapy would have been individually titrated with a goal of maintaining oxygen saturation levels somewhere within the range of 85% to 95%. Such individualized care would have been based on the parents' wishes for balancing the risks of administering lower levels of oxygen (including neurologic injury and death from hypoxemia [lower levels of oxygen hemoglobin saturation]) with the risks of administering higher levels of oxygen (including severe retinal injury, lung injury, and death from oxygen toxicity) Determining which level of oxygen to administer as part of routine care is based on what is in the best interests of that infant, as determined by the infant's parents in conjunction with the infant's physicians and other members of the health care team.⁴

This line of criticism was included in a letter from the OHRP to the University of Alabama at Birmingham, the lead site for the study. They wrote, "Ultimately, the issues come down to a fundamental difference between the obligations of clinicians and those of researchers. Doctors are required, even in the face of uncertainty, to do what they view as being best for their individual patients. Researchers do not have that same obligation."

This view was at the center of many criticisms of the study voiced by people who spoke against it at the public meeting held by HHS in August 2013 to discuss the controversy. For example, Vera Sharav of the Alliance for Human Research Protection stated, "In standard care the doctor's fiduciary responsibility is to prescribe treatments that serve each patient's best interest, adjusted in response to each patient's individual, fluctuating need. In research, treatment is predetermined by protocol that seeks to resolve uncertainty and contribute generalizable knowledge."⁵

This critique of randomization and treatment by protocol has been repeated in published papers by prominent lawyers and bioethicists. George J. Annas and Catherine L. Annas opine, "[I]n treatment a patient has a physician who is bound by a fiduciary duty to act in the patient's best interests (of course, with the patient's consent). In contrast, in research the researcher is duty-bound to follow the research protocol and may not deviate from it, even if the researcher believes a deviation is in the best medical interest of the subject."⁶ Similarly, Ruth Macklin and Lois Shepherd assert, "It is the doctors, not the researchers, who have a fiduciary obligation and long-standing ethic to pursue the patient's best interests above all other considerations."⁷ Of note, few clinical researchers agree with these broad and rigid generalizations. In fact, most clinical researchers would feel an obligation to deviate from a research protocol if they thought doing so was in the best interest of the patient. But clinical investigators seem to have a more modest view of their abilities to discern what treatment is best, in the absence of evidence, than do these lawyers.

These criticisms of SUPPORT that focus on randomization and treatment by protocol can be generalized as criticisms of any prospective randomized clinical trial. All such trials randomize patients and treat according to a predefined protocol. Such trials have always been ethically controversial. We need to review these criticisms in order to see how they have been responded to in the past, so that we can then judge whether SUPPORT in particular and CER generally raise any new issues in these regards.

Two decades ago, Samuel Hellman and Deborah Hellman wrote that randomized trials raise “the classic conflict between rights-based moral theories and utilitarian ones.”⁸ Further, their work suggests that this conflict is inherent and irreducible because it is part of the fundamental nature of clinical trials. They write, “Researchers *are required* to modify their ethical commitments to individual patients and do serious damage to the concept of the physician as a practicing, empathetic professional who is primarily concerned with each patient as an individual.” We added the emphasis to this quotation because the verb is so important. A researcher, according to this view, does not have the freedom to balance ethical obligations. The ethical commitments of research *require* a primary commitment to science.

Franklin Miller and Donald Rosenstein have a similar view.⁹ They note that randomized clinical trials differ fundamentally from standard care in their purpose, characteristic methods, and justification of potential harms. Interventions evaluated in randomized trials are allocated according to chance. Double-blind conditions and, often, placebo controls are used. For scientific reasons, protocols governing clinical trials typically restrict flexibility in the dosing of study drugs and the use of concomitant medications.

Echoing Hellman and Hellman, Miller and Rosenstein suggest that

the principles of beneficence and nonmaleficence governing medical care direct the physician to help individual patients and to avoid subjecting them to disproportionate harms. In clinical research, beneficence is primarily concerned with promoting the well-being of future patients, and nonmaleficence places limits on the potential harms to which research participants are exposed for the benefit of future patients and society. (p. 1384)

Miller and Rosenstein clearly see clinical research as designed to promote the good of future patients even if that means compromising the good of current patient-subjects. This moral stance, they believe, inevitably puts today’s research subjects at increased risk compared to patients who are not part of a research study.

These are powerful arguments. They are not, however, as applicable to CER as they are to other forms of research. Furthermore, there are particular aspects of oxygen therapy in neonatology that make these arguments particularly inapplicable to the SUPPORT study. In fact, using this line of argument to criticize SUPPORT reflects a deep and important misunderstanding of how clinical decisions about oxygen saturation targets have been and continue to be made for individual premature babies in neonatal intensive care units (NICUs).

Neonatologists generally did not—and do not—make individualized decisions about oxygen saturation targets for each patient based on that patient’s particular clinical situation and their clinical judgment about whether a higher or lower oxygen saturation should be targeted. Instead, as explained by neonatologist Keith Barrington, neonatologists always treat babies by protocol. Each doctor, or more often each NICU group, chooses an oxygen saturation target range based on their assessment of the evidence about the harms and benefits of oxygen levels that are too high or too low. They then use that protocol for all the babies in their NICU who are of a certain gestational age. Each baby, then, is always treated according to a predetermined protocol. In describing the situation prior to the SUPPORT study, Barrington noted,

Some centers felt that a saturation of 90 to 95% was the best idea, largely avoiding hyperoxia, and not wanting to risk hypoxia. So they would have a unit routine. The alarm limits for all preterm babies in the NICU would be set to, say, 89 and 96. Then the nurse would adjust the oxygen delivered, sometimes every couple of minutes, to stay within the target range.... Every preterm baby in the NICU would have the same limits, until they no longer needed oxygen. Another center examining the same data would set their oximeter alarm limits to a lower range. [Furthermore, NICUs do not] generally discuss choices about oxygen saturation levels with parents. Instead ... within each NICU doctors would agree, based on their review of the evidence, about the target saturations that they would try to achieve.¹⁰

The reasons for this state of affairs are the very reasons that the SUPPORT study was necessary. A neonatologist cannot possibly know, based only on observation and clinical judgment, just how much oxygen is best for any particular baby. The immediate benefits of giving more oxygen to a baby with respiratory distress are obvious. The baby’s oxygen saturation will increase, the baby will look healthier, and the baby will have less respiratory distress. Thus, clinical judgment, using the types of clinical factors that clinicians in other domains of medicine would use, would dictate that all babies be given high levels of oxygen because then the babies would improve, clinically, in the short term. But the special and crucial feature of neonatology is that oxygen is uniquely toxic to premature babies in a way that is not toxic for any other patient. Thus, the doctors must treat with less oxygen than their judgments based on clinical observation would suggest is appropriate. In such a situation, to rely upon clinical judgment, rather than protocol, would lead (as was the case in the past) to high and avoidable levels of blindness and chronic lung disease. To avoid confusion, we should also underscore the fact that this balancing-act approach to oxygen therapy is specific to the treatment of extremely premature babies with hyaline membrane disease and is not necessarily the best way to treat babies with other medical problems.

For these reasons, we think that the concerns about treatment by protocol are particularly irrelevant to the SUPPORT controversy. Treatment by protocol was and is the standard of care in all NICUs, and thus treatment by protocol was not a unique feature of the SUPPORT study. Instead, the question in the SUPPORT study was simply whether one commonly used treatment protocol increased or decreased the risk of complications in a group of babies for whom there was a known high risk of these problems.

Concerns about randomization are somewhat different. They turn on questions about whether doctors truly are uncertain about which treatment is best. If one treatment is known to be superior, then randomization to a treatment that was known to be inferior would be unethical. If, however, the experts are uncertain about which treatment was better and if there is evidence to suggest a complex and unpredictable balance of possible harms and possible benefits, then randomization would be ethically acceptable because the act of assigning patients to different treatments randomly would not predictably increase risk to either group. Of course, one group may turn out to fare better than the other. Discovering such a difference in outcomes is, after all, the goal of doing such a study. But if, at the outset, the clinician-investigators know which group would do better, then they should not do the study. If, furthermore, the professional uncertainty about the best treatment is so widespread that expert physicians have come to different conclusions and accordingly have decided to use different protocols in their day-to-day practice, then the risk of randomization would be comparable to the risk inherent in the variation of actual clinical practices that flowed from the disagreements within the professional community. While random treatment assignment will shift the treatments given to some patients within the range of currently accepted standards of care, given the uncertainty about which treatment is superior (which is the essential presupposition of clinical trials), no patient or physician would be sure whether this random shift in treatment would lead to benefit or harm for either group of patients and thus for any individual patient within those groups. We believe that the SUPPORT study occurred in precisely this set of circumstances.

Some have argued that, even if randomization and treatment by protocol do not add clinical risks, these procedures might still cause psychological harm. By this argument, a patient (or parent) can be psychologically harmed simply by being informed that his or her doctor does not know which treatment is best. Marilyn Morris and Robert Nelson contend that participation in a randomized trial “confronts research participants and/or their families with the inadequacy of current medical knowledge, which may be upsetting in the context of a critical illness.”¹¹ Elizabeth Robinson and colleagues studied the attitudes of healthy adults regarding such studies. They found that many study participants use psychological defenses such as denial to avoid the stress that such disclosures might cause. Robinson et al. report that, among subjects who had participated in randomized trials, “around half of the participants denied that a doctor could be completely unsure about the best treatment. A majority of participants judged it unacceptable for a doctor to suggest letting chance decide when uncertain of the best treatment.”¹²

Viewed through the lens of this concern, the ethics of randomized trials seems perched on a paradox. On the one hand, most ethical frameworks for conducting such trials insist upon informed consent. Prospective study subjects should be informed about what is or is not known about the study in question in order to decide whether to enroll. On the other hand, if prospective study subjects are not to be subjected to psychological harms that accompany their doctors’ disclosure of uncertainty, then they cannot be adequately informed of the risks and benefits of either enrolling in the study or of being treated outside of the study. The obligation to inform patients of the true state of knowledge and uncertainty is a precondition of research (and, presumably, of all clinical treatment), but providing such information can violate the ethical dictum to do no harm, including psychological harm.

Such concerns go well beyond the informed consent process for clinical research. A doctor who does not disclose uncertainty in seeking consent for clinical treatment is not giving the patient all the information that a reasonable person might want in order to make a decision. Disclosure of uncertainty in clinical practice should mirror the disclosure of uncertainty in clinical trials. In this sense, then, clinical trials do not add risk by disclosing uncertainty, except to the extent that consent for treatment outside the research context is inadequate and misleading. In that case, a more honest and comprehensive approach to consent for research might, in fact, add risk of more psychological problems in research subjects (or their parents).

If many patients do not believe, or do not want to acknowledge, that their physicians are genuinely uncertain about which treatment is best, and if patients must be informed of this in order to be enrolled in a clinical trial, then any enrollment in a clinical trial will cause harm to some subset of potential study subjects. According to this view, however, the obligation to protect research subjects from harm is in tension with the obligation to disclose the truth. This could be solved by never doing research. This is what the parent of a baby in the SUPPORT study suggested when she spoke at the HHS open meeting. Carrie Pratt criticized the researchers “for even approaching us at such a stressful time.” The implication was that there were situations in which patients or parents were so vulnerable that even being asked to make a decision was unbearably stressful.

Such considerations raise questions about the way we assess and compare different types of potential harms. They suggest an important lesson to learn from the SUPPORT controversy. Clinical researchers conducting CER studies must acknowledge their own uncertainty about which of two (or more) treatments is better and inform prospective research subjects of their uncertainty. Given that uncertainty, they may also inform potential study subjects that their outcomes will not be predictably worse if they allow their treatment to be chosen at random, rather than allowing their physician to make a clinical judgment about what is the best treatment. For some if not most potential research subjects, this information may be disturbing, yet this information reflects the (often unspoken) reality of clinical care, where the process of informed consent does not often provide a full disclosure of the disease-related hazards confronting the patient (which is to say, the risks of morbidity or mortality due to the underlying condition) and the uncertainty about the potential harms or benefits of treatment options.

In this context of too-often-incomplete disclosure of the clinical situation, institutional review boards and federal regulatory bodies must understand and respond to the tension within CER studies between full disclosure and psychological harm. If the disclosure of the uncertainty that justifies randomization is itself a harm, then IRBs need additional guidance about how they ought to balance this potential harm against the ethical obligation to be honest.

Choice of Study Endpoints

The researchers who designed the SUPPORT study chose two composite outcomes as the primary outcome measures in the study. One endpoint was a composite of either death or

retinopathy. The other was a composite of either death or neurodevelopmental impairment at two years of age.

These combined endpoints have been the focus of two very different sorts of criticisms. One suggested that, since death was a component of each combined endpoint, then the investigators anticipated that there would be a mortality difference and were ethically obligated to disclose this. The other suggested that these composite endpoints treated death and disability as ethically comparable.

While both of these concerns are understandable, they misinterpret the historical record and the analytic reason for using combined endpoints. First, in the years leading up to the SUPPORT study, strong evidence suggested that there was no mortality difference among babies whose oxygen saturation targets were in the high 80s versus the low 90s. A Cochrane meta-analysis of all available studies done on babies during the early neonatal period concluded that there was no evidence that restricted compared with liberal oxygen administration had significant independent effects on death rates or on the combined measure of death or retinopathy.¹³ Barrington recounts a planning meeting for the Canadian Oxygen Trial (a study similar to SUPPORT in randomizing babies to different oxygen saturation targets) and notes that doctors argued vehemently for and against different target saturations based on the claim that one or the other target would be safer for each of the various endpoints—mortality, retinopathy, and neurocognitive impairment.¹⁴

Researchers, as best we can tell, did not expect to find a mortality difference. Why, then, did they include death in their composite outcome measures? Why did they calculate death rates? Beyond the fact that all well-conducted studies keep track of deaths (every death must be reported to IRBs and to data safety monitoring boards [DSMBs]), the answer has to do with performing a statistically accurate and appropriate analysis. To not include death as part of the primary endpoint for the SUPPORT study would have meant that the sickest babies—those who died—would have been excluded from the final analysis. Such an analysis, confined to only those babies who survived, would potentially overestimate the impact of the different interventions on the other outcomes, since the sicker babies would be more likely to have had the other outcomes. The standard goal of a clinical trial is to measure the impact of an intervention accurately for all enrolled subjects, not just for the survivors. Tyson and colleagues explained the reasoning:

[T]he primary outcome in trials of high-risk patients is often a composite outcome that includes deaths when no effect on mortality is expected. Such patients may die before they can develop the outcome the intervention is hypothesized to prevent (e.g., severe ROP [retinopathy of prematurity] in SUPPORT). Death is then a competing outcome. Failure to account for differences in mortality would violate the intention-to-treat principle and can seriously bias the primary analysis. Had the primary outcome in SUPPORT been limited to severe ROP without including deaths, the primary outcome would have indicated incorrectly that the low saturation goal was superior.¹⁵

In SUPPORT, everybody knew that there would be deaths in both groups of patients. The babies in the study were, after all, critically ill. Past studies had shown death rates of 20 to

50 percent for babies born at the gestational ages of the research subjects.¹⁶ In this context of a relatively high baseline level of mortality, the SUPPORT composite endpoints were a way of being analytically careful by not allowing the primary endpoints of interest—retinopathy and neurodevelopmental impairment—to be artificially deemed statistically significant only because some babies who would have developed those complications did not survive.

Using combined endpoints is a standard analytic technique intended to minimize the statistical misinterpretations that might arise if not all study subjects are available to be assessed at the end of the study. Choice of endpoints is related to decisions about how large a study must be in order to be adequately powered to answer the scientific question that it sets out to answer. While SUPPORT was powered to find differences in the composite endpoint, the international collaboration in which SUPPORT was one of three multicenter studies was powered to find small differences in mortality. Choice of endpoints is also related to the interim analysis of data by a DSMB in order to decide whether to terminate a study early. The choice of composite endpoints in the SUPPORT study did not preclude or prevent the DSMB from carefully examining individual endpoints.

Critics of SUPPORT bring up both these concerns. Jon Merz and Divya Yerramilli criticize SUPPORT for not specifying mortality as an independent endpoint rather than part of a composite endpoint.¹⁷ They suggest that the researchers were “lucky” that mortality differences turned out to be significant but that the researchers were remiss in not powering the study to detect a mortality difference. This concern overlooks the fact that SUPPORT was part of an international consortium that was designed to find small differences in mortality, that mortality was carefully tracked, and that other studies in other countries performed interim analyses of their data when the SUPPORT results were reported in order to see whether they had a similar mortality difference.

The anticipation at the time the studies were designed was that any mortality difference, if it existed at all, would be so small that it would require a meta-analysis of multiple studies in multiple countries that all used the same protocol. Those other studies were done. Results from all three studies have now been published.¹⁸ Two showed a difference in mortality between the higher and lower target oxygen saturations. One did not. The pooled data from all three shows a mortality difference. Thus, knowledge about the implications of different oxygen therapy saturation targets for premature infants regarding mortality has advanced, but this was not expected.

Furthermore, the fact that mortality was not a primary endpoint of the study does not mean that mortality was not carefully monitored, as some have insinuated. In all studies, mortality is considered to be a “serious adverse event.” All serious adverse events in a study would be reported to both the IRB and DSMB. The DSMB would then be able to compare the rates of serious adverse events between the two arms of the study to see if they were statistically significant. The DSMB would then make a difficult decision if they saw trends that had not yet reached statistical significance.

In the SUPPORT study, the mortality rate prior to discharge from the hospital was 19.9 percent (130 out of 654) in one arm and 16.2 percent (107 out of 662) in the other. This just crossed the threshold of statistical significance (95 percent confidence interval: 1.01–1.60; $p = 0.045$). The mortality rates at thirty-six weeks of postconceptual age, by contrast, were 114 out of 654 (17.7 percent) and 94 out of 662 (14.2 percent). This was not statistically significant. A DSMB might have looked at both and would then have made a judgment call about whether to continue or stop the study.

When the SUPPORT study was published, the DSMB for the BOOST trial in Australia, New Zealand, and the United Kingdom examined their interim data. They reported, in a letter to the *New England Journal of Medicine*, that their interim analysis did not reveal a statistically significant difference in mortality. They thus elected to continue the study and wrote,

The independent data and safety monitoring committee in the Australian Benefits of Oxygen Saturation Targeting II (BOOST-II) trial and the New Zealand BOOST-NZ trial, whose protocols are similar to those in SUPPORT, has reviewed outcomes in 1352 patients. The committee reported no clear difference between the two oxygen targets in terms of the rate of death before hospital discharge. When considering the Australian BOOST-II and New Zealand BOOST-NZ trials in combination with SUPPORT, the committee found that there was significant uncertainty about the effects of treatment on mortality. They recommend that recruitment of study subjects continue. To our knowledge no such randomized controlled trial has reported survival free of disability in childhood, the major end point of all current oxygen-targeting trials. Until these results are known it would be premature to adopt either the higher or lower oxygen target for use in routine care.¹⁹

These analyses and decisions show that the investigators and their DSMBs were analyzing mortality data independently, even though the difference in mortality was not one of the primary endpoints of the study.

There is no completely objective way to determine whether a study ought or ought not to continue. If there were, then there would be no need for DSMBs. Instead, there would be straightforward stopping rules that could be invoked by study statisticians whenever the predetermined statistical threshold was reached. Instead, DSMBs must make difficult judgment calls when interim analyses show worrisome trends that have not yet—or have just barely—reached statistical significance. That was done in this study.

Lack of a Concurrent Standard-Care Control Arm

Some critics of SUPPORT focus on the lack of a standard-care control arm. They claim that, because the standard of care at the time was to target an oxygen saturation of 85 to 95 percent, the randomization to the low end and the high end of that range was ethically problematic. This randomization denied patients access to a standard of care that, at the time, was thought to be the best care. Thus, at the August 2013 meeting convened by HHS, Dr. Charles Natanson argued that “randomizing the critically ill to extremes of titrated

therapies creates practice misalignments which carry risks and do not represent usual care.”²⁰

Among experts in research methodology, there is a vigorous debate about whether a standard-care control arm is essential in order to evaluate the efficacy of variations in standard care. To a certain extent, the debate, as it applies to SUPPORT, turns on questions of what standard care actually was in NICUs around the world for babies born at twenty-four to twenty-seven weeks gestation. Natanson assumes that, because the recommended treatment was to target an oxygen saturation of 85 to 95 percent, this was both the agreed upon standard and the most common practice. This assumption is likely wrong. Many textbooks of neonatology recommended other target saturations. Many published papers recommended other targets. Many neonatologists chose other targets as their standard NICU protocols. Rather than imagining that the conventional treatment at the time was to target 85 to 95 percent, a more accurate statement would be that, in each NICU, there was a different target and that most of those targets were within the range of 85 to 95 percent. Some, as part of their standard practice, targeted the lower end of the range. Others targeted the higher end. To study each of these treatments and compare one against the other would have been methodologically impractical. At the very least, such a study would have been much larger, requiring, not just two arms or the addition of an elusive and illusory standard-of-care arm, but many more arms to reflect the widespread and manifold variation in standard treatment.

We are not the only ones to see problems in this line of criticism. Maureen Meade and Francis Lamontagne point out three concerns with usual-care control arms. First, there is the question of how broadly to define “usual care.” Some centers in multicenter studies may not meet established standards of care. Thus, with SUPPORT, the standards in 2005 were to treat toward a target oxygen saturation of 85 to 95 percent. Some centers allowed lower oxygen saturations. Others targeted higher saturations. Should such centers have been included in the usual-care arm? If so, should they have been instructed to change their “usual-care” practices and to treat within the target range of 85 to 95 percent? If so, that would have been a deviation from “usual care,” necessitating yet a fourth arm. If usual care was not standardized in some way, the study results would likely have been uninterpretable. The study certainly would have taken much longer to complete, as it would have required many more subjects.

Second, as we noted above, “usual care” in neonatology generally was provided by a predetermined protocol. Many critics of SUPPORT imagine that “usual care” would be to provide individualized adjustments of oxygen levels based upon the baby’s clinical condition from moment to moment. But that is not, as we have shown, usual care in neonatology.

A third problem with “usual-care control arms” is that usual care changes over time. The longer a study takes to complete, the more “usual care” might change from the beginning to the end of the study. We know, for example, that practices with regard to oxygen therapy were in flux during the SUPPORT years. The longer the study went on, the more changes would likely have occurred in usual care. To the extent that usual care changed to reflect the experiences of clinicians who were simultaneously involved in a prospective study, it likely

would have come to resemble the study treatments and been more difficult to distinguish from them.

A final argument against the need for usual-care control arms is that there is often good historical data as well as concurrent (but not randomized) data from similar patients who are not enrolled in the study. With regard to the SUPPORT study in particular, there were excellent historical data from well-established national databases that allowed comparison of study babies with babies who were treated outside of the study. Of course, there are known problems with nonrandomized comparison cohorts. But those problems don't disappear when study subjects are randomized. For subjects in a randomized trial, biases are introduced by differences between those who choose to enroll and those who do not. For nonrandomized comparison groups, similar sorts of selection biases are inevitable. Yet, if analyzed in a rigorous and thoughtful manner, each comparison group gives valuable information.

Use of Altered Oximeters to Mask Treatment Assignment

Many critics were concerned about the altered oximeters that were designed to give slightly false readings of oxygen saturations. These were used in order to conceal from the doctors and nurses which group each baby was in. Double masking, concealing evidence of group membership from both investigators and patients, is a standard technique for randomized controlled trials. Masking usually doesn't garner attention because, in most trials, the therapy is a pill or an injection. The masking takes place by making each pill or injection look the same. There is no need to change the technology.

In SUPPORT, however, where patients were randomized to different target oxygen saturation ranges, the masking had to disguise the true oxygen saturation. Most of the critiques of the altered oximeters focused on the ways in which they prevented doctors from exercising individualized clinical judgment. As we have shown above, however, this concern is misplaced in the context of the NICU, where oxygen supplementation is seldom given based upon individualized clinical judgment. The masked oximeters did not prevent doctors from treating patients as they would have done anyway—that is, by protocol. Instead, the masked oximeters simply allowed doctors to do so with the sort of masking that is well accepted in randomized trials.

Concerns about Consent

For any study to be ethically appropriate, there must be uncertainty about the optimal way to treat patients. To the extent that this criterion was met, the investigators did not know the overall balance of advantages and disadvantages of being assigned to one treatment arm or the other. As noted above, if this overall balance of potential benefits and harms were known in advance to be more favorable in one treatment group, then the study should never have been done. This basic fact allows for consideration of what should have been included in the consent form or communicated during the consent process.

Parents of babies who are eligible for clinical trials need information about the potential harms and benefits of enrolling in a study in order to decide whether to permit their babies to

participate. Informed consent for CER is different from informed consent for studies of previously untested therapies in at least three important ways.

First, in studies of new and innovative therapies, the potential harms of these therapies are, generally, truly unknown. One of the primary goals of such studies is to understand the safety and efficacy profile of a previously untested therapeutic innovation. For a CER study such as SUPPORT, by contrast, all the treatments have been in widespread use. Much is known about them. Thus, we can explain to prospective study subjects what is known, what is not known, and what we hope to learn by doing the study. We can also compare the risks of being in the study (and thus being assigned a treatment by formal random allocation) versus being outside the study (and thus being offered a treatment based on the preferences of the attending physician). Of note, in NICU care, parents generally do not choose their physician. They do not usually know the NICU protocol for oxygen saturation targeting. And, in customary practice, they are not given a choice of doctor or of oxygen levels. Thus, from the perspective of the parents, the difference between a random assignment to one of the SUPPORT study treatments and the ordinary ways of being assigned a similar treatment were small.

Second, in studies of a previously untried therapy, the “default” position for patients who choose not to enroll in the study is clear. They will get the existing standard of care. In CER, by contrast, there is no obvious “default” position. Prospective study subjects must be told that if they do not enroll in the study, they might get the exact same treatment that they would have gotten if they had enrolled in the trial. They should also be told whether and how treatment will be different if they enroll in the study. However, as noted above, knowing just how different treatment might be in the study is difficult because knowing what is the current standard treatment is difficult. This is true when there is known and widespread practice variation, as there was with regard to oxygen saturation targets in NICUs at the time that SUPPORT was designed. In each NICU, the task was more tractable, as there was likely to be an existing standard care, based on an existing oxygen saturation targeting protocol. Therefore, within each NICU, the babies in the study might have received predictably different treatment had they not enrolled in the protocol, and this information could have been provided to parents. But across all the NICUs (assuming that some had protocols similar to one arm of the study and some similar to the other arm), there was no well-established “default therapy.” That is the nature of the CER universe.

Finally, in any clinical situation, prospective study subjects or their parents must be told about the potential harms that they or their child face from their underlying disease, regardless of whether they enroll in the trial. In the NICU, parents must first be told of the potential harms that are present as a result of their baby’s early birth, regardless of whether or not the baby is enrolled in the trial. That is, parents need to understand the potential harms of death, eye disease, blindness, chronic lung disease, and neurodevelopmental impairment that are associated with extreme prematurity. This is not an easy conversation to have. There has been much discussion and debate among expert clinicians about how to explain these potential harms.²¹

The process of informed consent for a research study such as SUPPORT, then, must explain the potential harms and benefits of the treatments that are currently in use and the ways in which those potential harms and benefits might increase or decrease if the treatments are assigned randomly, rather than idiosyncratically. Such discussions are difficult to have under any circumstances. They are particularly difficult when a mother has just given birth to a critically ill baby. We know from studies of consent in other situations that parents often have trouble understanding randomization. For example, Eric Kodish and colleagues have shown that half of parents who consented for their children to be in a study of cancer chemotherapy did not understand randomization.²² And those parents were discussing randomization in a relatively less stressful situation—that is, they did not have to make a decision about enrollment in the clinical trial within minutes.

Consent for such trials could, in theory, be simplified and improved. A parent could have been given the following key information honestly and straightforwardly (and perhaps even this concisely):

Your baby was born extremely prematurely. Many babies who are born this early die. Many of those who survive have long-term complications, including eye disease, chronic lung disease, cerebral palsy, and brain damage. Most survivors, however, do not have any of these problems. We are doing a study to try to learn the best ways to prevent these things from happening. The study involves giving babies two different levels of oxygen: higher and lower. Some babies in the study may do better—and some babies worse—than other babies in the study. But we don't know which group will have better outcomes. (If we knew, we wouldn't be doing the study.) Right now, babies in NICUs across the United States receive both levels of oxygen and many levels in between. We also don't know whether babies in the study will have better or worse outcomes than babies who are not in the study. For babies in the study, we will decide what oxygen level to provide by a random choice (similar to flipping a coin.) Every baby will have a 50-50 chance of getting either low or high oxygen levels. Babies who are not in the study are treated according to our NICU protocol. You can decide whether to be in the study or not. If you decide to be in the study, this decision will determine only the level of oxygen that we use and will not affect the care that your baby gets in any other way.

Such an approach illustrates the goals of informed consent in CER. The consent process should not discuss just the potential harms of being in a study; it should also include a comparison with the potential harms of not being in the study.

Implications for the Ethics of Clinical Research

The controversy about the SUPPORT study looks very different from controversies about research ethics in the past. This controversy is decidedly not about trade-offs between the interests of current patients and the interests of future patients. Instead, because experts were genuinely uncertain of which arm would prove better, babies who were enrolled in the study were at no higher risk of complications than babies who were not in the study. Instead, the

controversy should be about understanding and communicating the comparative potential harms and benefits of enrolling in the study.

When a study is evaluating an innovative therapy, the research question is usually whether a new and untested approach to treatment is better than a widely accepted one. The potential harms and benefits of the widely accepted treatment are well-known, while those of the innovative treatment are less well characterized. One important goal of such studies, then, is to better characterize the safety and efficacy profile of the new treatment. In such a situation, the default comparison is clear—the outcomes of patients receiving the new treatment will be compared to the outcomes of those who receive standard treatment. The default treatment is also clear—if a patient does not enroll in the trial, he or she will not receive the intervention under study but instead receive the standard therapy.

With comparative effectiveness research, by contrast, both therapies are in widespread use and can be considered standard. Some doctors tout one approach, some another. If a patient chooses not to be in the study, they will receive the treatment that is favored by the doctor who is caring for them or at the institution where they are receiving care. As we have tried to show, CER raises issues that are different from those raised by studies of innovative therapies. These differences have important implications for the way we think about study design and informed consent. Consent discussions for such studies require different ways of thinking about and talking about the relative potential harms and potential benefits of being in the study compared to not being in the study.

Grappling with the ethics of the way in which the SUPPORT study was implemented advances the way that we think about the many levels of comparisons inherent in any clinical trial, especially when they occur in the larger context of clinical care. Without doubt, specific aspects of the rationale and implementation of SUPPORT are unique. Nevertheless, our ethical analyses regarding the various facets of the implementation of this study suggest that there are elements of SUPPORT that are likely to arise in other CER studies. Specifically, many studies will require treatment by protocol, randomization, choice of study endpoints, data safety monitoring, masking of treatment assignment, pitting two conventional and widely used treatments against each other, and obtaining consent when what is being studied are two established treatments. Because of this, we think that the SUPPORT study is a good paradigmatic case upon which to develop regulatory policy for similar studies in the future. Regulations that do not acknowledge the distinctiveness of such CER studies may inhibit our ability to learn whether one widely used treatment is better or worse than another. That will expose all patients, not just research subjects, to the risks of unstudied and poorly understood treatments. Careful attention to the risks and benefits for patients who choose to enroll in CER studies and for those who do not will enable us to simultaneously fulfill three goals: systematically improving medical care, protecting human research subjects, and obtaining meaningful informed consent for both treatment and research.

Acknowledgments

John Lantos's work on this paper was supported by a Clinical and Translational Science Award grant from the National Center for Advancing Translational Sciences awarded to the University of Kansas Medical Center for

Frontiers: The Heartland Institute for Clinical and Translational Research # UL1TR000001 (formerly #UL1RR033179). The contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes for Health or NCATS.

References

1. Office for Human Research Protection, letter to University of Alabama at Birmingham. <http://www.mazzaschi.com/OHRP.6.4.2013.pdf>.
2. Annas GJ. Questing for Grails: Duplicity, Betrayal and Self-Deception in Postmodern Medical Research. *Journal of Contemporary Health Law and Policy*. 1996; 12:297–324. at 297. [PubMed: 8666728]
3. Faden R, et al. Ethics and Informed Consent for Comparative Effectiveness Research with Prospective Electronic Clinical Data. *Medical Care*. 2013; 51(8, supplement 3):S53-S7. [PubMed: 23793051]
4. Carome MA, Wolfe SM. letter to Secretary of Health and Human Services in re The Surfactant, Positive Pressure, and Oxygenation Randomized Trial (SUPPORT). 2013 Apr. <http://www.citizen.org/documents/2111.pdf>.
5. Sharav V. HHS public meeting. 2013 Aug. https://www.youtube.com/watch?v=IAfBKgYmxtg&list=PLr17E8KABz1Gc_ndt-9grGg8O_jE5G1RNC&index=8.
6. Annas GJ, Annas CL. Legally Blind: The Therapeutic Illusion in the SUPPORT Study of Extremely Premature Infants. *Journal of Contemporary Health Law and Policy*. 2013; 30(1):1–36. at 4.
7. Macklin R, Shepherd L. Informed Consent and Standard of Care: What Must Be Disclosed? *American Journal of Bioethics*. 2013; 13(12):8–14. at 10. [PubMed: 23311831]
8. Hellman S, Hellman DS. Of Mice but Not Men—Problems of the Randomized Clinical Trial. *New England Journal of Medicine*. 1991; 324:1855–1859.
9. Miller FG, Rosenstein DL. The Therapeutic Orientation to Clinical Trials. *New England Journal of Medicine*. 2003; 348:1383–1386. at 1384. [PubMed: 12672867]
10. Barrington KJ. Personalized Medicine in the NICU. *American Journal of Bioethics*. 2013; 13(12): 33–35. at 34. [PubMed: 24256532]
11. Morris MC, Nelson RM. Randomized, Controlled Trials as Minimal Risk: An Ethical Analysis. *Critical Care Medicine*. 2007; 35:940–944. at 942. [PubMed: 17255879]
12. Robinson EJ, et al. Lay Public's Understanding of Equipoise and Randomisation in Randomised Controlled Trials. *Health Technology Assessment*. 2005; 9(8):1–192. [PubMed: 15763039]
13. Askie LM, Henderson DJ, Ko H. Restricted versus Liberal Oxygen Exposure for Preventing Morbidity and Mortality in Preterm or Low Birth Weight Infants. *Cochrane Database of Systematic Reviews*. 2001; 4 art. no. CD001077.
14. Barrington K. Pre-SUPPORT, What Did We Really Know? 2013 Sep. <http://neo-natalresearch.org/2013/09/07/pre-support-what-did-we-really-know>.
15. Tyson JE, Walsh M, D'Angio CT. Comparative Effectiveness Trials: Generic Misassumptions Underlying the SUPPORT Controversy. *Pediatrics*. 2014; 134:651–654. [PubMed: 25201795]
16. Lantos JD. The Weird Divergence of Ethics and Regulation with Regard to Informed Consent. *American Journal of Bioethics*. 2013; 13(12):31–33. [PubMed: 24256531]
17. Merz JF, Yarramilli D. SUPPORT Asked the Wrong Question. *American Journal of Bioethics*. 2013; 13(12):25–26. [PubMed: 24256528]
18. Carlo WA, et al. Target Ranges of Oxygen Saturation in Extremely Preterm Infants. *New England Journal of Medicine*. 2010; 362(21):1959–1969. [PubMed: 20472937] Stenson BJ, et al. Oxygen Saturation and Outcomes in Preterm Infants. *New England Journal of Medicine*. 2013; 368(22): 2094–2104. [PubMed: 23642047] Schmidt B, et al. Effects of Targeting Higher vs Lower Arterial Oxygen Saturations on Death or Disability in Extremely Preterm Infants: A Randomized Clinical Trial. *Journal of the American Medical Association*. 2013; 309(20):2111–2120. [PubMed: 23644995]
19. Tarnow-Mordi WO, Darlow B, Doyle L. Target Ranges of Oxygen Saturation in Extremely Preterm Infants. *New England Journal of Medicine*. 2010; 363(13):1285. [PubMed: 20860514]

20. Natanson, Charles. [accessed October 23, 2014] testimony at HHS open meeting. 2013 Aug. <http://www.hhs.gov/ohrp/newsroom/rfc/Public%20Meeting%20August%2028,%202013/supportmeetingtranscriptfinal.html>
21. Janvier A, Lorenz JM, Lantos JD. Antenatal Counselling for Parents Facing an Extremely Preterm Birth: Limitations of the Medical Evidence. *Acta Paediatrica*. 2012; 101(8):800–804. [PubMed: 22497312] Haward MF, Murphy RO, Lorenz JM. Default Options and Neonatal Resuscitation Decisions. *Journal of Medical Ethics*. 2012; 38(12):713–718. [PubMed: 23180252]
22. Kodish E, et al. Communication of Randomization in Childhood Leukemia Trials. *Journal of the American Medical Association*. 2004; 291(5):470–475. [PubMed: 14747504] Greenley RN, et al. Stability of Parental Understanding of Random Assignment in Childhood Leukemia Trials: An Empirical Examination of Informed Consent. *Journal of Clinical Oncology*. 2006; 24(6):891–897. [PubMed: 16484698]