The Problem of Randomization within a Standard of Care Range: A Case Study

James Cummings*

Albany Medical Center, Albany, New York, USA

***Corresponding author:** James Cummings, Albany Medical Center, Albany, New York, USA, Tel: 5182625421; E-mail: cumminj1@mail.amc.edu

Abstract

Background: Recent controversy in the press and medical literature surrounding a clinical trial in preterm infants raised several questions about the informed consent process in comparative effectiveness trials. An important consideration that was missing from these discussions were the implications of randomizing subjects within a standard of care, when that standard of care is not defined discretely, but by a range of physiologic measurement.

Summary: This paper discusses the risk/benefit implications when subjects are randomized within a standard of care that is defined by a range of physiologic measurement, rather than discrete therapeutic interventions. The recent controversy surrounding the informed consent process for a large, multicenter, clinical trial in preterm infants is used as the backdrop for this discussion, and a hypothetical study design built around a common clinical problem is used to further demonstrate the significant alteration in risk/benefit that might occur when subjects are randomized to narrower ranges of response within a larger standard of care range. While it may be possible to mitigate the negative effects of this type of randomization by alterations in study design, specifically closer monitoring and intervention, this potentially increased risk must clearly be addressed in the informed consent process.

Background

Comparative effectiveness studies, in which subjects are randomized to one of various alternatives within an accepted standard of care, are relatively common in clinical research. Recent controversy over one such study in preterm infants has raised questions about how risk may be altered when patients are randomized to different standards of care. These concerns may be even more relevant when standard of care is defined not by specific treatments, but by a range of physiologic responses. After a brief description of the study and the concerns raised, we examine the issue at the heart of this dialogue, focusing on the ethical considerations and the implications for future clinical research.

The Surfactant, Positive Pressure, and Pulse Oximetry Randomized Trial (SUPPORT) was conducted at 23 academic centers and enrolled 1316 preterm infants during a 5-year period [1]. One question SUPPORT asked was whether targeting higher or lower oxygen saturations altered the development of retinopathy of prematurity (ROP) a serious vascular dysgenesis of the retina that is common in preterm infants and the leading cause of blindness in developed countries. It is well known that high oxygen saturation is a major determinant of ROP. However, oxygen can also be life saving, and the optimum balance between too little and too much is unknown. The American Academy of Pediatrics recommends a target range of 85-95% [2]; however, studies suggest that keeping saturations on the higher side of this range may result in more ROP [3]. SUPPORT was designed to test the hypothesis that a lower target range (85-89%) would result in less ROP than a higher target range (91-95%), without any increase in adverse outcomes.

Results from SUPPORT, reported in a prominent clinical journal, noted that while the lower oxygen target range resulted in less ROP, it also increased the risk of mortality [1].

Allegations of Non-Compliance

After SUPPORT was published, the Office for Human Research Protection (OHRP) began an investigation of allegations of noncompliance with Department of Health and Human Services (DHHS) regulations (45 CFR part 46) [4]. In March 2013, OHRP determined that SUPPORT "failed to adequately inform parents of the reasonably foreseeable risks and discomforts of research participation" [5]. One month later, the advocacy group Public Citizen drew attention to the complaint, and The New York Times reported that OHRP had determined that major academic institutions failed to disclose the risks of the study to parents, thereby "depriving them of information needed to decide whether to participate" [6].

OHRP never specifically questioned the ethical conduct of the study. However, when the OHRP letters were released to the public, SUPPORT was criticized for ethical violations. Public Citizen sent a letter to the DHHS Secretary, expressing concern over this "highly unethical" study that had "egregious informed-consent omissions", and criticized OHRP's response as "grossly inadequate" [7]. An editorial in the Times, entitled "An Ethical Breakdown", stated that this failure to inform was "startling and deplorable" [8].

Responses to Above

At the same time, a flurry of commentary occurred in the press and medical literature [9-18]. The majority was supportive of the investigators and critical of OHRP's determination. In response, OHRP reconsidered their initial position, and subsequently suspended all compliance actions in regards to SUPPORT [19]. In addition, in recognition of the "complex" nature of study design in trials like SUPPORT, in which subjects are randomized to treatments that fall within standard of care (SOC), OHRP announced that they would plan no further action until they could provide better guidance on the rules regarding risk disclosure in such studies [19].

Randomization within the Standard of Care

At the center of the SUPPORT controversy is the concept of randomization within SOC. For OHRP, this raised the concern of differential risks between the groups that were not addressed in the informed consent. For groups like Public Citizen, this raised graver concerns, that such randomization is unethical. Although both these points have been well argued, what is missing from the discussion is the fact that there is a misperception about how SOC can be applied in a clinical research trial.

When SOC allows for a variety of treatments, there is no moral obligation to provide a specific treatment, only a treatment that falls within that SOC. In the simplest case, SOC may involve two possible therapies. Both are acceptable and care providers are justified in picking either. However, the fact that one treatment is not preferable over the other does not mean that the treatments are the same; in fact, they may carry quite different risks and benefits. That is one reason why comparative trials are done.

The ethical principle of justice, as it applies to clinical research trials, is generally interpreted as ensuring that selection of subjects is fair and equitable, so that all share the potential benefits (and burdens) of research [20]. In a randomized clinical trial, the principle of justice is further satisfied in the allocation of treatments, in that each subject has an equal chance of being allocated to a specific treatment group. However, the principle of justice must also be considered within the context of treatment choices in a clinical trial. It is implied, although not a given, that each treatment group within a clinical trial has a comparable risk/benefit profile; that the risks and benefits, real or potential, within each group are balanced and do not 'a priori' make one group preferable to another. The point of the randomized clinical trial is to see if one treatment is actually preferable to another.

That brings us to what is the fundamental concern within SUPPORT - the selection of treatment groups. Although the allocation to each group was random, and therefore just, the treatment groups might not be balanced in regards to risk/benefit. Explicit in the informed consent is that there is no perceived risk to the low oxygen group, but that it might benefit from less ROP [21]; this implies that the high oxygen group might have a relatively higher risk of ROP. Moreover, there is no benefit mentioned to the high oxygen group. Instead, the study investigators argued that since each treatment fell within the SOC, both were justifiable.

It is this SOC argument that is flawed. It fails to appreciate a critical difference when SOC involves a range of acceptable responses - in this case a range of oxygenation levels – as opposed to a discrete set of treatments. As noted above, when SOC consists of a discrete set of treatments, each treatment is equally acceptable and no treatment is preferable to another, based on current knowledge. If the treatments are to be compared in a randomized clinical trial, then subjects can be allocated to any of the treatments, since they each represent SOC.

When SOC consists of a range of acceptable measurements, the situation is different. First, we are not giving a treatment, but targeting a range of clinical responses; this range is generally based on the knowledge that responses outside this range (too high or too low) increase the risk of harm, when compared to responses within this range. Second, while we can ensure that a patient receives a prescribed treatment, we cannot ensure that a clinical response will always fall within the SOC range; we are therefore constantly titrating interventions, making corrections when the patient's response level approaches or exceeds the range limits. Third, response levels are dynamic and vary over time, both within and between individuals; this variation can be described by a frequency distribution, typically a "bell-shaped" curve in which responses in the central portion of the curve will occur most often, and values at the upper and lower extremes (i.e., near the range limits) much less often.

These characteristics of a SOC range have important implications when it comes to randomized clinical trials. When SOC refers to discrete treatments, then allocation to any particular treatment still represents that SOC. However, when the SOC refers to a range of acceptable responses, allocation to any particular subset of responses may no longer be comparable to the original SOC because it is not only the level of response, but also the distribution of responses that define that SOC.
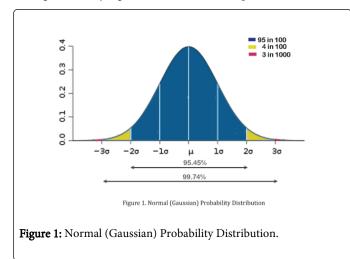
To illustrate, let's use the analogy of driving on a wide, single lane road. Your goal is not to hit either the left or right guardrail, as this would cause harm. To avoid harm, we set the SOC range as the white (right) and yellow (left) lines of the highway. If you stray over either line, a rumble strip alerts you to that fact and you correct to the center of the road. In this scenario, you will spend most of the time in or near the center of the road; because you have a relatively wide safety margin (distance between you and the edges of the road), you will rarely wander over the lines.

Now, let's divide that wide road into two lanes, each half the width of the original road. Your goal is still not to hit the guardrails, but now your SOC limits are narrower; they include a center dashed line and either the yellow (left lane), or the white (right lane) line. In this scenario, you still spend most of your time in the center of your lane, but because your lane is now only half the size as before, you have a narrower safety margin and are much more likely to stray over the solid line. Depending on how close the guardrails are to those solid lines, your chances of harm may be significantly increased.
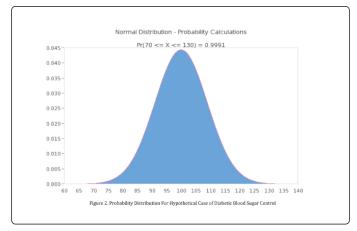
When SOC is defined by a Range of Responses – A Theoretical Clinical Trial

To examine this further, let us use a well-known medical condition as an example. Normally, the blood sugar level in our bodies is tightly regulated, in order for it not to go too low or too high, because extremely low or high blood sugar can be harmful. Low blood sugar can cause collapse, loss of consciousness and even death, while high blood sugar causes both acute effects (e.g., dehydration, weakness, susceptibility to infection) and chronic damage to a variety of organs, including the kidneys, eyes, and heart. In diabetes, the body's normal mechanisms for keeping one's blood sugar from going too high do not work, so the patient must take active steps to prevent high blood sugar; steps like carefully controlling their sugar intake or taking medications, like insulin, that will lower their blood sugar.

These types of treatment, diet control and medications, comprise the basic toolkit for diabetes SOC. However, since the treatment options for diabetics all work by lowering blood sugar, we must also be careful not to lower it too much, since very low blood sugar is harmful as well. That is why the SOC for diabetic care is to closely monitor blood sugar levels and maintain those levels within recommended limits. A range, rather than a specific level, is an acceptable SOC, since a diabetic's blood sugar level will tend to vary a bit throughout the day, based on many things including the time of the last meal, the nature of that meal, physical activity, and timing and dosing of medications. The American Diabetes Association recommends a pre-prandial (just before meal) blood sugar level between 70 and 130 mg/dl [22].

For the sake of simplicity, let us assume that we can model a diabetic's blood sugar level as a normally distributed continuous variable, which is defined by its average value (mean, or μ) and variability (standard deviation, or σ). Then, over the course of time, the patient's blood sugars will assume a normal (Gaussian) distribution, as shown in Figure 1. In Figure 1, the mean is centered on the x-axis, and the values on the y-axis give the relative frequency of each possible value for x. The average, or mean, is the most frequently occurring value, and values away from the average (measured in terms of σ) occur less and less often the further they are from the mean. In such a distribution, all values (even negative ones) are possible, although extremely high or low values are infrequent.

Figure 1. Normal (Gaussian) Probability Distribution

Figure 1: Normal (Gaussian) Probability Distribution.

Assuming that the diabetic's blood sugar values will occasionally vary outside the recommended range, let us define good diabetic control by how often one stays within the recommended range of 70-130. In Figure 2, we show how this would look if we define good control as being within the acceptable range 99.9% of the time. In Figure 2, the mean is 100 (mid-point between 70 and 130) and the standard deviation is 9, giving us a probability of being "in-range" 99.9% of the time (shaded area) [23]. For the sake of our illustration this will define the SOC for good diabetic control.

Figure 2. Probability Distribution For Hypothetical Case of Diabetic Blood Sugar Control
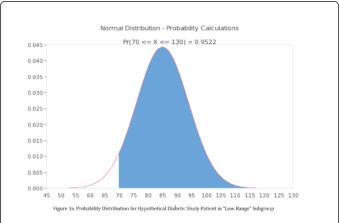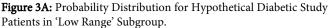
Figure 2: Probability Distribution for Hypothetical Case of Diabetic Blood Sugar Control.
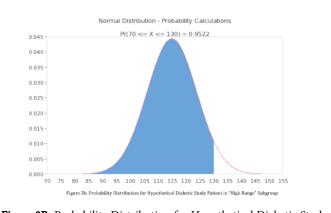
Now, let's imagine a study that is interested in determining whether avoiding the higher end of these blood sugar levels will result in less long-term organ damage. We decide to do this by simply dividing the

SOC range of blood sugar levels into two equal ranges, 70-100 and 100-130, assigning half the study subjects to each range and looking at long-term outcomes. Since the two ranges are still within the current SOC range for diabetic control, can we assume that there is no increased risk to being in such a study? Let's see what happens to the distribution curves for blood sugar levels for an individual in the study, assuming they continue to have similar variability in blood sugar levels as determined earlier (a standard deviation of 9).

Figures 3a and 3b show what happens within each of the two study subgroups. Both curves appear similar in shape to each other and to the original SOC curve, but they are shifted along the x-axis. The average, or mean, values are different; in the original SOC distribution the average value was 100, in the lower range study group (Figure 3a) it is 85, and in the higher range study group (Figure 3b) it is 115. Of course, each of these three means are within the SOC range of 70-130, but look what happens to the frequency of "outliers", or values that fall outside the SOC range. In the lower range study group, nearly 5% of values now fall below 70, while in the higher range study group an equal portion now exceed 130 (non-shaded areas).

Figure 3a. Probability Distribution for Hypothetical Diabetic Study Patient in "Low Range" Subgroup

Figure 3A: Probability Distribution for Hypothetical Diabetic Study Patients in 'Low Range' Subgroup.

Figure 3b. Probability Distribution for Hypothetical Diabetic Study Patient in "High Range" Subgroup

Figure 3B: Probability Distribution for Hypothetical Diabetic Study Patients in 'High Range' Subgroup.

By splitting the SOC range into two halves what we have effectively done is eliminate the chances of an outlier in one direction (i.e., either too high or too low) but disproportionately increase the chances of an

outlier in the other direction. In fact, in the original SOC distribution the chances of a blood sugar level falling out of range was approximately 0.1% while in each of the two study subgroups, the chances are now approximately 5%, or 50 times greater. Thus, although the study ranges were both within the original SOC range, subdividing that SOC range actually increases the risk of an out-of-range value to subjects in either group. In addition, the specific risks between the two subgroups are quite different; the low range group has a significant risk of extremely low blood sugar levels, while the high range group has a significant risk of extremely high blood sugars. That is not to say that there are not ways in which the study can reduce those risks to its subjects, perhaps by closer monitoring, more aggressive intervention, or other measures to reduce variability in blood sugar levels, but those safeguards would have to be built into the study design; regardless, it must be acknowledged that dividing the SOC range into smaller ranges increases the risk to the subjects in the study, compared to SOC.

How Does This Relate to Clinical Trials Like SUPPORT?

Similarly in SUPPORT, dividing the SOC range up into two smaller ranges actually increased the chances that the infant's oxygen level might stray above (high oxygen group) or below (low oxygen group) the SOC limits. If there is a risk associated with too high or too low an oxygen level, then the allocations provided in SUPPORT increased one's risk over SOC; thus, neither allocation group is comparable to SOC. That is not to say that allocation was unjustified, assuming that allocation also increased the chances for benefit. However, according to the SUPPORT consent, potential benefit was confined to one group.

Rebalancing Risk/Benefit in Comparative Effectiveness Trials

How does the theoretical risk/benefit profile of the allocations determine whether or not those allocations were just? Do they suggest an alternative allocation strategy for SUPPORT? Let us examine three scenarios.

The risks and benefits of either allocation group are unknown: This is the most common situation in comparative effectiveness trials. There is the potential for one group to perform "better" but since which group is not known at the onset of the study, randomization ensures just allocation. This scenario does not fit SUPPORT, since there is at least one potential benefit.

There is potential benefit to one group: This is how SUPPORT was presented. As argued earlier, both groups have increased risk compared to SOC; for the low oxygen group that risk is offset by a potential benefit (less ROP). The high oxygen group is not justified because it has increased risk but no potential benefit. This could be remedied by using a SOC group instead of a high oxygen group but that might require more infants be studied, for statistical purposes. The consideration then becomes whether subjecting more infants to potential but unknown risks (of low oxygen) would be justified by the knowledge gained.

There are potential risks and benefits to both allocated groups: Although not acknowledged in SUPPORT, this was actually the position taken by other investigators who recently conducted very similar studies [24-26]. In the informed consents for those trials, there was explicit mention that lower oxygen had the potential benefit of less

ROP and the potential risks of lower survival or greater neurodevelopmental impairment. This represents a balanced risk/benefit profile between groups, although both different than SOC. This approach would justify the allocations made in SUPPORT.

Implications for Future Research

As a result of the SUPPORT controversy, DHHS retracted its original determination against the investigators. They recognize the need to develop better guidelines concerning disclosure of risks in SOC comparative effectiveness trials and have begun soliciting input from both the public and scientific sectors [27]. In developing those guidelines, we think it is important to consider the increase in risk that occurs when allocation is made within SOC that targets a range of clinical responses rather than discrete treatments.

References

1. SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network, Carlo WA, Finer NN, Walsh MC, Rich W, et al. (2010) Target ranges of oxygen saturation in extremely preterm infants. N Engl J Med 362: 1959-1969.

2. (2007) American Academy of Pediatrics, American College of Obstetrics and Gynecologists. Guidelines for perinatal care (6th Edn) Elk Grove Village, IL: American of Pediatrics: 260-262.

3. Tin W, Milligan DW, Pennefather P, Hey E (2001) Pulse oximetry, severe retinopathy, and outcome at one year in babies of less than 28 weeks gestation. Arch Dis Child Fetal Neonatal Ed 84: F106-110.

4. (2005) Code of Federal Regulations. Title 45-Public Welfare. Part 46-Protection of Human Subjects. Subpart D: Additional Protections for Children Involved as Subjects in Research. Rockville, MD: Department of Health and Human Services. Accessed July 2014.

5. Buchanan LR (2014) (Compliance Oversight Coordinator, Department of Health & Human Services, Office for Human Research Protections). Letter to Richard B. Marchese, PhD (V.P. for Research and Economic Development, University of Alabama at Birmingham). 2013 Mar 7. Accessed July 2014.

6. Tavernise S (2013) Study of babies did not disclose risks, U.S. finds. The New York Times. Accessed July 2014.

7. Carome MA, Wolfe S (Public Citizen) (2013) Letter to the Honorable Kathleen Sibelius (Secretary, Department of Health and Human Services). Accessed July 2014.

8. (2013) An ethical breakdown. The New York Times. Accessed July 2014.

9. Carlo WA, Bell EF (2013) Walsh MC for the SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network. Oxygen saturation targets in extremely preterm infants. N Engl J Med 368: 1949-1950.

10. Wilfond BS, Magnus D, Antommaria AH, Appelbaum P, Aschner J, et al. (2013) The OHRP and SUPPORT. N Engl J Med 368: e36.

11. Magnus D, Caplan AL (2013) Risk, consent, and SUPPORT. N Engl J Med 368: 1864-1865.

12. Drazen JM, Solomon CG, Greene MF (2013) Informed consent and SUPPORT. N Engl J Med 368: 1929-1931.

13. Hudson KL, Guttmacher AE, Collins FS (2013) In support of SUPPORT--a view from the NIH. N Engl J Med 368: 2349-2351.

14. Macklin R, Shepherd L, Dreger A, Asch A, Baylis F, et al. (2013) The OHRP and SUPPORT--another view. N Engl J Med 369: e3.

15. Wright CJ Saugstad OD (2013) OHRP and SUPPORT: lessons in balancing safety and improving the way we care for patients. J Pediatr 163: 1495-1497.

16. Lantos JD (2014) Learning the right lessons from the SUPPORT study controversy. Arch Dis Child Fetal Neonatal Ed 99: F4-5.

17. Rettner R (2013) Premature baby study: families speak out, criticize consent forms. LiveScience. Accessed July 2014.

18. Benham K (2013) The parental consent dilemma: saving extremely premature babies by signing forms. Tampa Bay Times. Accessed July 2014.

19. Buchanan LR (2013) (Compliance Oversight Coordinator, Department of Health & Human Services, Office for Human Research Protections). Letter to Richard B. Marchese, PhD (V.P. for Research and Economic Development, University of Alabama at Birmingham). Accessed July 2104.

20. (1978) National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Washington, DC: US Government Printing Office 1978.

21. Carlo W, Ambalavanan N (2014) Informed consent form for the SUPPORT trial. Accessed July 2014.

22. (2014) Tight Diabetes Control. Accessed July 2014.

23. (2014) Normal Distribution Probability Calculator Tool. Accessed July 2014.

24. Stenson B, Brocklehurst P, Tarnow-Mordi W; UK. BOOST II trial; Australian BOOST II trial; New Zealand BOOST II trial. (2011) Increased 36-week survival with high oxygen saturation target in extremely preterm infants. N Engl J Med 364: 1680-1682.

25. BOOST II United Kingdom Collaborative Group; BOOST II Australia Collaborative Group; BOOST II New Zealand Collaborative Group, Stenson BJ, Tarnow-Mordi WO, Darlow BA, Simes J, et al. (2013) Oxygen saturation and outcomes in preterm infants. N Engl J Med 368: 2094-2104.

26. Schmidt B, Whyte RK, Asztalos EV, Moddemann D, Poets C, et al. (2013) Effects of targeting higher vs. lower arterial oxygen saturations on death or disability in extremely preterm infants: a randomized clinical trial. JAMA 309: 211-20.

27. (2014) Public Meeting on August 28th, 2013. Accessed July 2014.