Achieved Oxygenation Saturations and Outcome in Extremely Preterm Infants

Benjamin J. Stenson, MD

KEYWORDS
- Oxygen • Oxygenation • Oximetry • Infant mortality • Necrotizing enterocolitis
- Retinopathy of prematurity • Infant • Extremely premature

KEY POINTS
- In the Neonatal Oxygenation Prospective Meta-analysis (NeOProM) trials, differences in outcomes between groups are likely to be related to differences in achieved peripheral capillary oxygen saturation (SpO₂).
- Mortality and necrotizing enterocolitidis were increased in infants randomized to the low SpO₂ target range groups (85%–89%), but low-target group infants who experienced these outcomes had SpO₂ distributions centered around 90% to 92%.
- Achieved SpO₂ patterns were not different between infants randomized to higher SpO₂ targets (91%–95%) who did or did not require treatment of retinopathy of prematurity (ROP).
- Trials of slightly higher SpO₂ targets than 91% to 95% may help determine whether there is further survival advantage obtainable. These will be associated with increased risk of the need for ROP treatment.

INTRODUCTION
The Neonatal Oxygenation Prospective Meta-analysis (NeOProM) trials have shown that small differences in target ranges for peripheral capillary oxygen saturation (SpO₂) in extremely preterm infants influence the risk of mortality and morbidity.¹ In older patient age groups too, the clinical community is waking up to the reality that oxygen must be administered carefully based on evidence rather than according to habit.²

The intervention in the NeOProM trials was to target SpO₂ to the ranges 85% to 89% or 91% to 95%. This was delivered by nursing staff manually adjusting fraction of inspired oxygen (FiO₂) with the aim of maintaining SpO₂ within a desired range. The NeOProM trials were masked by the use of specially adjusted oximeters so that infants enrolled in both randomization groups were targeted to the same displayed SpO₂ range.

Disclosure Statement: The author has nothing to disclose.
Neonatal Unit, Royal Infirmary of Edinburgh, 46 Little France Crescent, Edinburgh EH16 4SA, UK
E-mail address: Ben.stenson@luht.scot.nhs.uk

https://doi.org/10.1016/j.clp.2019.05.011
perinatology.theclinics.com
0095-5108/19/© 2019 Elsevier Inc. All rights reserved.
and caregivers were unaware of their group allocation. This design was pioneered in the first Benefits of Oxygen Saturation Targeting (BOOST) trial. This method should have minimized the risk that caregiver biases could influence the treatment of the infants in other ways that were unbalanced between randomization groups and makes it reasonable to conclude that differences in outcomes between groups were determined by the trial intervention and its effect on achieved SpO2. This is further supported by the little heterogeneity between the NeOProM trials for the key outcomes despite the trials being conducted independently in different settings and parts of the world.

Although there was not a statistically significant difference between groups in the NeOProM trials in the primary composite outcome of death or severe neurodevelopmental disability, infants targeted to the lower SpO2 range had a significantly increased risk of death and of severe necrotizing enterocolitis (NEC), defined as requiring surgery or causing death. There was no difference between groups in the risk of severe neurodevelopmental disability. Infants randomized to the higher target groups had an increased risk of requiring treatment of retinopathy of prematurity (ROP), but there was not a significant difference between groups in the number of infants with severe visual impairment. In fact, this was numerically higher in the infants randomized to the lower SpO2 target.

It is vital to understand the achieved SpO2 patterns that were obtained in association with these outcomes in the trials because, unlike in retrospective observational studies of achieved or targeted SpO2 in which multiple confounders may explain differences in outcomes, in these trials there can be much greater confidence that outcomes were related at least in part to differences in achieved SpO2.

SpO2 varies widely in unwell preterm infants and there is great difficulty in keeping it within a target range. Whatever target range is used, infants spend varying proportions of time inside or outside that intended range. Success in targeting SpO2 is influenced by the target range selected. There are probably multiple contributors to this observation. It is suggested that caregivers may prefer higher SpO2. It is also the case that SpO2 readings below 90% are on the steep part of the hemoglobin-oxygen dissociation curve, where small changes in alveolar oxygen tension will result in much larger fluctuations in SpO2 than they would at higher SpO2 readings. In the BOOST-II UK and BOOST-II Australia trials, with a masked intervention and both groups targeted to the same displayed SpO2, infants allocated to the lower target range spent less time in their intended range than infants allocated to the higher target range. The advantages of automated oxygen adjustment in terms of time spent within intended target range seem smaller in comparison with manual adjustment when higher SpO2 targets are used. These observations are important because the lower target range group infants in the NeOProM trials had higher than intended achieved SpO2. The differences in outcomes between groups that were observed may underestimate the differences that would have been observed if the lower target groups had been targeted more effectively and spent more time in their intended target range.

It would be a mistake to consider that all the infants in each randomization group achieved SpO2 distributions similar to those of the group as a whole. As well as considering the overall achieved SpO2 patterns of the randomization groups, there is value in considering the patterns of achieved SpO2 in the infants within the randomization groups who developed the adverse outcomes of interest.

The individual trials in the NeOProM collaboration have reported their achieved SpO2 in different ways and in varying detail. Post hoc analyses of the associations between achieved SpO2 patterns from the trials and clinical outcomes should be considered hypothesis-generating rather than high-level evidence. If observations are replicable between trials that may increase their information value.
MEDIAN PERIPHERAL CAPILLARY OXYGEN SATURATION

So far, the only information published from the trials relating achieved SpO2 to outcome has come from the SUPPORT trial\(^9,10\) and relates the risk of mortality to achieved median SpO2 and intermittent hypoxia. The subject of intermittent hypoxia is excluded from this discussion. (See Juliann M. Di Fiore and colleagues’ article, “Intermittent Hypoxemia in Preterm Infants,” in this issue.) In the SUPPORT trial,\(^9\) SpO2 data were continuously sampled every 10 seconds (Masimo, Radical, Yorba, CA, USA). Each infant’s median SpO2 was determined and histograms were presented in the trial report for the distribution of these medians. Values were not provided for the median of these values in each group in the trial report but it was stated that the actual median levels of oxygen saturation were slightly higher than targeted levels in both groups. This implies that the median SpO2 in the low-target group over the whole duration of the intervention was greater than 89% and the median SpO2 in the high-target group was greater than 95%. In other words, the low-target group in this trial had a higher risk of mortality despite achieving a median SpO2 of at least 90%.

In a post hoc analysis of SpO2 data from the first 3 days of life from the SUPPORT trial,\(^10\) the median SpO2 of infants in the low SpO2 target group (target range 85%–89%) with birthweight appropriate for gestational age was 93%. Low-target group infants who were small for gestational age in the SUPPORT trial had a median SpO2 in the first 3 days of 90%.

Median oxygen saturation for all infants was divided into quartiles. The lowest quartile for oxygen saturation was less than or equal to 92%. Appropriately grown infants with a median oxygen saturation during the first 3 days after birth greater than 92% had the highest 90-day survival rate at 90.8%. In comparison with these infants, there was a lower 90-day survival in both infants who were appropriately grown (83.2%, \(P = .0013\)) and those who were small for gestational age (64.3%, \(P < .0001\)), with median SpO2 in the first 3 days less than or equal to 92% (Fig. 1).

Fig. 1. Kaplan Meier survival analysis in appropriately grown and small for gestational age infants in the SUPPORT trial according to median achieved SpO2 during the first 3 days of life. (From Di Fiore JM, Martin RJ, Li H, et al. Patterns of oxygenation, mortality and growth status in the surfactant, positive pressure and oxygen trial cohort. J Pediatr 2017;186:49–56.e1; with permission.)
OVERALL PERIPHERAL CAPILLARY OXYGEN SATURATION DISTRIBUTION

In the 3 BOOST-II trials, SpO₂ readings were also recorded every 10 seconds throughout the duration of monitoring. For each trial, SpO₂ readings were pooled from all infants and the percentage of time that the infants in each randomization group spent at each SpO₂ was calculated. This was plotted by randomization group in the form of frequency histograms indicating the percentage of time spent at each SpO₂ (Fig. 2) and was also tabulated in the supplementary data. Separate results were given for each trial. During the trials, there was a change in the study oximeters to correct an artifact that had been identified in their calibration, and separate plots and tables were provided for the different oximeter types. Differences between randomization groups in achieved SpO₂ and in clinical outcomes were clearer after the revision of the trial oximeters.

As in the SUPPORT trial, infants in the low-target groups in the BOOST-II trials had higher than intended SpO₂. Infants in the high-target groups had SpO₂ distributions that were centered in their intended SpO₂ target range. Increased mortality was observed in infants targeted to lower SpO₂ (85%–89%) in the BOOST-II UK and BOOST-II Australia trials even though the median SpO₂ of infants in the low-target

![Fig. 2. Average frequency distributions of the time spent by each infant at each oxygen saturation level from 80% to 100% while receiving supplemental oxygen. For trials in the United Kingdom and Australia, separate distributions are provided for infants managed using the original oximeter-calibration algorithm and those managed using the revised algorithm, according to whether they were assigned to receive a higher target of oxygen saturation (91%–95%) or a lower target (85%–89%). Revised oximeters were not used in the New Zealand trial. Vertical dotted lines indicate the intended target SpO₂ ranges of the randomization groups. (From The BOOST II United Kingdom, Australia, and New Zealand Collaborative Groups, Oxygen Saturation and Outcomes in Preterm Infants. Volume 368 N Engl J Med. 2013 May 30;368(22):2094–104. https://doi.org/10.1056/NEJMoa1302298. Epub 2013 May 5. Copyright © (2013) Massachusetts Medical Society. Reprinted with permission.)](CLP1111_proof ■ 8 June 2019 ■ 1:44 pm)
groups in these trials (89%–90%) was at the upper limit of the intended low-target range of 85% to 89%.

To explore the achieved SpO2 distributions in relation to clinical outcome in the BOOST-II UK trial, there has been a further post hoc analysis to produce plots of achieved SpO2 distributions, separating the infants by randomization groups and also according to whether they died or developed severe NEC or ROP requiring treatment (Figs. 3–5). Most infants in the UK trial were treated with oximeters that had revised software and the plots are restricted to data from these infants.

For the outcomes mortality and NEC that were increased in the infants randomized to lower SpO2 targets in the NeOProM trials, all time on the oximeter was analyzed. This is because lower SpO2 readings were the focus and these were allowed by protocol in lower target group infants whether or not they were breathing supplemental oxygen. Low SpO2 readings would be modifiable whether the infant was breathing air or supplemental oxygen. For ROP, the analysis was for time that the infants were breathing supplemental oxygen because high SpO2 readings were the focus. High SpO2 readings in infants breathing air were not considered to be modifiable and to be unlikely to be associated with underlying hyperoxia, whereas high readings in infants on supplemental oxygen may indicate hyperoxia and were amenable to modification.

Fig. 3 shows data for mortality. In both randomization groups, the distribution of achieved SpO2 was lower in infants who died than in infants who survived. In both randomization groups, these differences were apparent at higher but not at lower SpO2 in which the curves overlaid each another. Consistent with the significantly increased risk of mortality observed in infants randomized to lower SpO2 in the NeO-ProM trials overall, the difference in SpO2 distribution between infants who lived and died in the BOOST-II UK trial was greater in low-target group infants than in high-target group infants. Low-target group infants who survived had a distribution of SpO2 similar to that of high-target group infants. In comparison, the SpO2 distribution of low-target group infants who died was shifted markedly to the left but was still higher than their intended target range, with a peak at 90%. The difference in achieved SpO2 between low-target group of infants who lived and died was greatest at SpO2 readings within and above the intended low-target SpO2 range and not at lower SpO2 readings.

Fig. 4 shows data for severe NEC. Severe NEC was increased in low-target group infants in the NeOProM trials. Infants in the high-target group of the BOOST-II UK trial who got severe NEC had an SpO2 distribution that was not different from unaffected infants in the same group. In contrast, infants in the low-target group who developed severe NEC differed quite clearly from unaffected infants in their SpO2 distribution. As was observed with mortality, low-target group infants who did not develop severe NEC had SpO2 distribution similar to that of high-target group infants. Low-target group infants who developed severe NEC had a SpO2 distribution that was shifted to the left. The peak of this distribution, at 92% to 93%, was 3 to 4 points higher than the intended target range for the low-target group. There was not a difference in exposure to very low SpO2 between infants who did or did not develop severe NEC in the low-target group.

Fig. 5 shows data for ROP requiring treatment. In both randomization groups, the achieved SpO2 distribution of infants who developed ROP requiring treatment was indistinguishable from that of infants in the same randomization group who did not.
Fig. 3. Achieved SpO₂ distributions from the BOOST-II UK trial, with randomization groups plotted separately according to whether the infants survived or died. Data are for all time on the oximeter. Vertical dotted lines indicate the intended target SpO₂ ranges of the randomization groups.

Fig. 4. Achieved SpO₂ distributions from the BOOST-II UK trial, with randomization groups plotted separately according to whether or not the infant developed severe NEC. Data are for all time on the oximeter. Vertical dotted lines indicate the intended target SpO₂ ranges of the randomization groups.
DISCUSSION

These analyses were not prespecified at the time that the trials were planned and consequently must be considered hypothesis-generating rather than fully explanatory. However, they are surprising and thought-provoking, and there would be value in seeing whether they are replicated in the other trials.

The data show that the risks of mortality and severe NEC, which were increased in infants randomized to the low SpO2 target range, were not associated with greater exposure than other infants to very low SpO2 readings. The difference in achieved SpO2 observed with these adverse outcomes was mainly increased time spent in the intended low-target SpO2 range and even slightly above it. The data suggest that there may be value in studying the risks and benefits of a slightly higher SpO2 target range than 91% to 95% to see if further reductions in risk of mortality and NEC are achievable.

In the NeOProM trials there was more ROP requiring treatment in infants targeted to 91% to 95%. In the BOOST-II UK trial, the achieved SpO2 patterns of the infants in this high-target group who got ROP were not different from those who did not, suggesting that their ROP was not simply a reflection of less effective targeting in some infants than others. The implication may be that it will be hard to avoid further increases in ROP requiring treatment if higher ranges are studied, even with meticulous targeting.

The intended target ranges in the NeOProM trials of 85% to 89% and 91% to 95% were the same in all 5 NeOProM trials. Consequently, although the trial data suggest that, of these 2 ranges, the higher range should be preferred, the optimal target range remains unknown because other ranges have not been studied. Some may consider that an intermediate range between the 2 is worthy of study. It has been speculated that the optimal SpO2 target range may be dynamic, changing with advancing...
The use of servocontrol systems might enable more successful targeting, lowering the associated ROP risk of higher SpO2 targets. If these achieved SpO2 data are considered alongside these hypotheses, they suggest that these speculations are approaches that should not be used outside randomized controlled trials. The adverse outcomes, death, NEC, and ROP, move in opposite directions with different SpO2 target ranges and their causation overlaps in timing, so it is likely to be challenging to minimize all of them simultaneously. Careful measurement of these competing risks is required.

If an intermediate range of 88% to 92% was targeted, then the infants targeted most successfully would have a distribution of achieved SpO2 centered around 90%, similar to that described in the low-target group infants in BOOST-II UK trial who died. They would have median SpO2 values less than 92%, which was associated with a greater risk of mortality in the SUPPORT trial.\(^\text{10}\) If, like the infants in the NeOProM trials, infants targeted to this intermediate range had a distribution of SpO2 that was a few points higher than intended, then it may still be similar to that described in infants who developed severe NEC in the BOOST-II UK trial.

Most cases of NEC and some cases of mortality occur in the early weeks after birth. The increase in these outcomes observed in the low-target group infants in the NeOProM trials should caution against uncontrolled use of lower SpO2 targets in the early weeks after birth. The achieved SpO2 distributions from the SUPPORT and BOOST-II UK trials add further to this caution. The BOOST-II UK trial achieved SpO2 histograms that do not support a hypothesis that the risk of death was associated with time spent with deep hypoxia that might be avoided by more careful targeting of the lower SpO2 range than was achieved in the trial. Both the SUPPORT trial and the BOOST-II UK trial achieved SpO2 data that suggest that mortality was increased in low-target group infants despite their having achieved SpO2 higher than the intended low-target range. Targeting this lower range has already been tested and shown to worsen outcome.

A variety of automated oxygen adjustment systems has been evaluated and the technology is becoming available for clinical use both for ventilated infants and infants receiving noninvasive support.\(^\text{15-18}\) These devices are effective and it is clear that achieved SpO2 distributions in infants on automated adjustment systems differ from those of infants treated using manual adjustment. Therefore, it is likely that introduction of automated oxygen adjustment systems into clinical practice will influence the risks of ROP, NEC, and death. Clinicians considering using these devices have little to guide them in terms of device settings. The best evidence presently available is that mortality risk will be lower in infants with achieved SpO2 distributions, such as those achieved in the high-target group survivors in the NeOProM trials. Setting automated control devices to achieve a value centered around the middle of the range 91% to 95% would probably achieve a lower SpO2 distribution than was seen in the trials and could conceivably increase the risk of mortality or NEC. Trials are required to determine the optimum target for automated adjustment systems.

**SUMMARY**

Data relating achieved SpO2 to clinical outcome from the SUPPORT and BOOST-II UK trials facilitate their interpretation and add caution regarding variation in practice in this area outside the context of further clinical trials. The best current evidence for clinical safety is to target the SpO2 range 91% to 95% by manual oxygen adjustment. This should be the baseline comparator for further studies of different
SpO2 target ranges and targeting technologies. Increased risks of mortality and NEC were observed in association with SpO2 distributions centered around 90% to 93%, suggesting value in testing whether or not a higher SpO2 target range, such as 92% to 97%, could be associated with further survival advantage. It is unlikely that a higher range could be studied without an increased risk of the need for ROP treatment; however, with advances in the treatment of this condition, the need for ROP treatment is not as important an adverse outcome as death or severe NEC, and may be a price worth paying.

ACKNOWLEDGMENTS

The author is grateful to Mr Andy King, BOOST-II UK Trial Statistician, for providing the analyses and histograms of achieved SpO2 readings from the BOOST-II UK trial.

REFERENCES


