Oxygen Saturation Targets in Preterm Infants and Outcomes at 18–24 Months: A Systematic Review

Veena Manja, MD, MS,[a,b] Ola D. Saugstad, MD, PhD,[c] Satyan Lakshminrusimha, MD, FAAP[d]

abstract

CONTEXT: The optimal oxygen saturation target for extremely preterm infants remains unclear.

OBJECTIVE: To systematically review evidence evaluating the effect of lower (85%–89%) versus higher (91%–95%) pulse oxygen saturation (SpO_2) target on mortality and neurodevelopmental impairment (NDI) at 18 to 24 months.

DATA SOURCES: Electronic databases and all published randomized trials evaluating lower versus higher SpO_2 target in preterm infants.

STUDY SELECTION: A total of 2896 relevant citations were identified; 5 trials were included in the final analysis.

DATA EXTRACTION: Data from 5 trials were analyzed for quality of evidence and risk of bias.

LIMITATIONS: Limitations include heterogeneity in age at enrollment and comorbidities between trials and change in oximeter algorithm midway through 3 trials.

RESULTS: There was no difference in the incidence of primary outcome (death/NDI at 18–24 months) in the 2 groups; risk ratio,1.05, 95% confidence interval 0.98–1.12, $P = .18$. Mortality before 18 to 24 months was higher in the lower-target group (risk ratio,1.16, 95% confidence interval 1.03–1.31, $P = .02$). Rates of NDI and severe visual loss did not differ between the 2 groups. Proportion of time infants spent outside the target range while on supplemental oxygen ranged from 8.2% to 27.4% <85% and 8.1% to 22.4% >95% with significant overlap between the 2 groups.

CONCLUSIONS: There was no difference in primary outcome between the 2 SpO_2 target groups. The collective data suggest that risks associated with restricting the upper SpO_2 target limit to 89% outweigh the benefits. The quality of evidence was moderate. We speculate that a wider target range (lower alarm limit, 89% and upper, 96%) may increase time spent within range, but the safety profile of this approach remains to be determined.

[a]*Division of Cardiology, Department of Internal Medicine, Veterans Affairs Medical Center, Buffalo, New York;* [b]*Department of Clinical Epidemiology and Biostatistics, McMaster's University, Hamilton, Ontario, Canada;* [c]*Department of Pediatric Research, Oslo University Hospital, University of Oslo, Oslo, Norway; and* [d]*Division of Neonatal-Perinatal Medicine, Department of Pediatrics, Woman and Children's Hospital of Buffalo, University at Buffalo, Buffalo, New York*

Dr Manja conceptualized and designed the study, reviewed the literature, carried out the initial analyses, and reviewed and revised the manuscript; Dr Saugstad critically reviewed and extensively revised the manuscript; Dr Lakshminrusimha was the second reviewer of the literature and drafted the initial manuscript; and all authors approved the final manuscript as submitted and agree to be accountable for all aspects of work.

DOI: 10.1542/peds.2016-1609

Accepted for publication Oct 10, 2016

To cite: Manja V, Saugstad OD, Lakshminrusimha S. Oxygen Saturation Targets in Preterm Infants and Outcomes at 18–24 Months: A Systematic Review. *Pediatrics.* 2017;139(1):e20161609

Oxygen therapy for preterm infants was introduced in the 1940s and is the most commonly used "drug" in neonatal intensive care.[1] Liberal use of oxygen in the 1940s and 1950s resulted in an increase in retinopathy of prematurity (ROP), a well-known complication of extreme prematurity.[2,3] Restriction of oxygen use in the 1960s and clinical tolerance of hypoxia in premature infants resulted in increased mortality.[4] More recently, improvements in technology have allowed precise measurement of pulse oxygen saturation (SpO_2), enabling titration of oxygen delivery. In 2007, the American Academy of Pediatrics (AAP) stated that SpO_2 between 85% and 95% and partial pressure of oxygen, arterial (PaO_2) between 50 and 80 mm Hg are examples of ranges pragmatically determined by some clinicians to guide oxygen therapy in preterm infants.[5] However, the optimal SpO_2 target in extremely premature infants has been debated for many years with varying results in previous randomized and observational studies leading to significant uncertainty.[6,7]

Between 2005 and 2007, 5 randomized controlled trials (RCTs) were initiated to resolve the uncertainty of target SpO_2 range in extremely preterm infants (<28 weeks' postmenstrual age [PMA] at birth).[8] These studies are part of the Neonatal Oxygenation Prospective Meta-analyses (NeOProM), a collaborative effort[8] examining the effect of lower-target (85% to 89%) and higher-target (91% to 95%) SpO_2 levels. The trials have recruited 4911 extremely preterm newborns and include SUPPORT (Surfactant, Positive Pressure and Pulse Oximetry Randomized Trial),[9] the 3 BOOST-II (Benefits of Oxygen Saturation Targeting-II) studies,[10] and the COT (Canadian Oxygen Trial).[11]

An audit of the pulse oximeter used in these trials revealed an artifact in the algorithm causing an artificial elevation of SpO_2 that was maximal at a displayed value of 90%, leading to less frequent readings of 87% to 90%.[12] A new revised software algorithm was installed midway through 3 trials (BOOST-II UK, BOOST-II Australia, and COT). This artifact may affect the results and will be explored in a subgroup analysis.

A meta-analysis of these studies showed an increased risk ratio (RR) for mortality and necrotizing enterocolitis (NEC), while RR for severe ROP was decreased, in lower compared with higher SpO_2 target.[13] There was no difference in the combined outcome of death and neurodevelopmental impairment (NDI) at 18 to 24 months, bronchopulmonary dysplasia (BPD), ROP, NDI, or hearing loss at 18 to 24 months. Based on the Grades of Recommendation, Assessment, Development, and Evaluation criteria, the quality of evidence for the outcomes in this analysis was moderate to low.[14] These meta-analyses were conducted before the publication of 2-year outcomes from the BOOST-II Australia/UK trials.[15] With this publication, 18–24 month outcomes are available for all the studies conducted by the NeOProM collaboration.

The objectives of this systematic review were to assess whether targeting a lower SpO_2 range (85%–89%) has an effect on mortality and NDI compared with a higher SpO_2 range (91%–95%) after accounting for the risk of bias of each included study as well as the quality of evidence for each outcome.

METHODS

The written protocol for this meta-analysis was reviewed by 2 authors (VM and SL) but was not registered online.

Criteria for Selecting Studies

All published RCTs with sufficient information were eligible for inclusion in our review. Preterm infants <28 weeks' PMA at birth receiving supplemental oxygen for any duration at any time before hospital discharge were included. The intervention of lower (85%–89%) SpO_2 target was compared with higher (91%–95%) target. The outcome measures included any of the following at 18 to 24 months: death or severe NDI, death, NDI, or visual or hearing loss. We have previously reported a meta-analysis of short-term outcomes, such as BPD, NEC, and severe ROP.[13,14] Studies other than RCTs, studies including infants ≥28 weeks' PMA at birth, and SpO_2-target range other than 85% to 89% for the lower-target and 91% to 95% for the higher-target were excluded.

Data Collection and Analysis

Study Selection

The titles and abstracts retrieved by the search were reviewed independently by the authors. Any discordance was identified; disagreement was resolved by discussion. A $\kappa \geq 0.65$ was chosen a priori to indicate adequate agreement among reviewers.

Software and Summary of Findings

All meta-analyses were carried out by using Review Manager 5.3 (RevMan; The Nordic Cochrane Center, Cochrane Collaboration, Copenhagen, Denmark, 2014). The level of confidence in the estimate of effect was assessed by using GRADEpro (Evidence Prime, Inc, Ontario, Canada). The Cochrane risk-of-bias tool was used to assess study quality.

Impact of "Tails"

The proportion of time spent outside the overall target range of 85% to 95% (<85% = lower tail; >95% = higher tail) was collected for all studies for original and revised algorithms. The association of time spent in the lower and higher tails with negative outcomes was explored.

Assessment of Quality of Evidence and Confidence in Estimates of Effect for Each Outcome

We assessed the quality of the evidence to support the estimate of effect for each outcome by using GRADEpro. By using this method, the level of evidence is assessed for the following domains: risk of bias,[16] inconsistency,[17] indirectness,[18] imprecision,[19] and publication bias.[20]

Measure of Treatment Effect

Dichotomous data are expressed as RRs with 95% confidence intervals. A random-effects model was used and a 2-tailed $P < .05$ was considered statistically significant. A fixed-effect model assumes that the true effect size is the same in all studies and the summary effect is an estimate of this effect size and assigns weight based on the size of the study and largely ignores information in smaller studies.[21] A random-effects model assigns a disproportionately smaller weight to larger studies. Our goal was to estimate the mean effect of all 5 studies and not let the overall estimate be overly influenced by 1 study. Although these 5 trials were performed by researchers using similar outlines, the investigators operated independently, and the patients and protocols differed in ways that may have affected the results. Therefore, we did not assume a common effect size and preferred the random-effects model.

Sensitivity Analysis

Results using the fixed-effects model were explored in a sensitivity analysis.

Dealing With Missing Data

The SUPPORT reported disability as Bayley Scale of Infant Development III (BSID-III) with a cutoff cognitive composite score <70. The other trials used a BSID-III composite language/cognitive cutoff score <85. Data on proportion of time spent <85% and >95% ("tails") were available for COT and BOOST-II but not for SUPPORT. The steering committee of the Neonatal Research Network provided the data from SUPPORT for disability by using a cutoff score of <85 and proportion of time spent <85% and >95%.

Assessment of Heterogeneity

Statistical heterogeneity was evaluated both by visual inspection of the Forest plot and by using a standard χ^2 test. Heterogeneity also was assessed by using the I^2 statistic for each outcome. An I^2 estimate $\geq 50\%$ with a $P < .10$ for χ^2 was interpreted as substantial.

Planned Subgroup Analysis

The effect of the SpO_2 target may vary depending on oximeter algorithm; to elucidate these differences, subgroup analysis based on oximetry calibration algorithm were planned for the outcome of death and death/NDI by 18 to 24 months of age. Analyses for visual/hearing impairment were performed by using pooled data because of small numbers.

RESULTS

Search Strategy

The results of the search are summarized in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram (Fig 1). The weighted κ for overall agreement between reviewers for the title/abstract screening was 0.88. There was no disagreement in the selection of the final articles for the systematic review.

Included Studies

Five trials were included in this review. Table 1 provides a brief description of the trials and differences that might have influenced outcomes. The SUPPORT was conducted in the United States, and published in 2010[9]; the COT was a multinational trial published in 2013[11]; and the BOOST-II included 3 trials conducted in the United Kingdom, Australia, and New Zealand, and was published in 2013.[10] The 18- to 24-month results for the outcome of death/NDI in SUPPORT were published in 2012.[22] The 2-year outcome for the composite of death/disability was published for BOOST-II New Zealand in 2014[23] and for the United Kingdom/Australia in 2016.[15]

Patient Characteristics

All 5 trials enrolled extremely premature infants <28 weeks' PMA at birth. The exact postnatal age at inclusion and the lower limit of gestation differed slightly (Table 1). The percentage of outborn infants differed between the studies; the SUPPORT trial included only inborn infants. There was a higher percentage of white infants enrolled in the BOOST-II (UK) compared with SUPPORT.

Primary Outcome

The primary outcome of the follow-up component of these 5 trials was death or NDI at 18 to 24 months' corrected age. There was no difference between lower-target and higher-target groups (46.5% and 44.4%, respectively, $P = .18$, Fig 2A).

Outcomes by Subgroup: Pulse Oximeter Algorithm Assignment

The SUPPORT and BOOST-II New Zealand trials were conducted by using the original oximeter algorithm. BOOST-II UK/Australia and COT trials revised the oximeter algorithm midway through the studies. The primary outcome of death/NDI was not different with pooled data (Fig 2A) and with original algorithm (Fig 2B). However, data from the revised algorithm demonstrated increased incidence of death/NDI with lower target (Fig 2C). Death by 18 to 24 months was significantly higher in the lower-oxygen target group with pooled data and revised algorithm but not different with the original algorithm (Fig 3). Incidence of NDI or severe visual/hearing

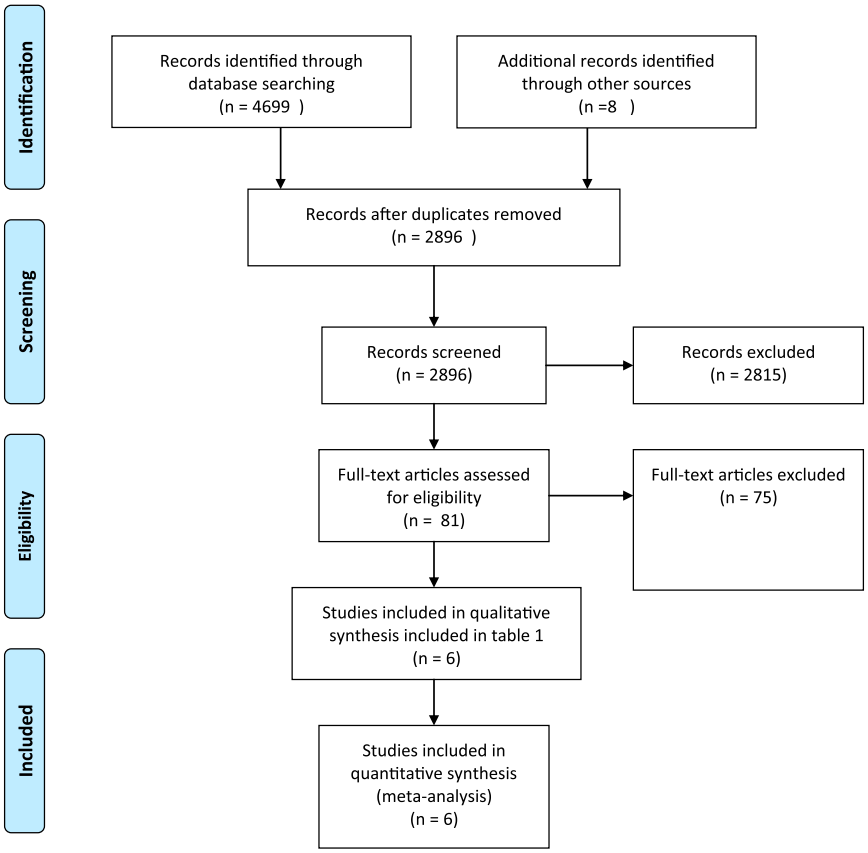


FIGURE 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram. (From Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. *PLoS Med.* 2009;6(6):e1000097. For more information, visit www.prisma-statement.org.)

impairment did not differ between the 2 groups (Figs 4 and 5).

Risk of Bias

By using the Cochrane risk-of-bias assessment, these studies were all at low risk of bias for sequence generation, concealment of allocation, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, and selective outcome reporting. In the domain of other sources of bias, assessment of adequacy of achieved SpO_2 in the 2 groups was a prespecified criterion in our protocol. Although a distinct, maximal 6% separation of SpO_2 was planned in the study protocols, there was significant overlap in the SpO_2 values achieved in the intervention and comparator groups.

Quality of Pooled Data for Each Outcome

The quality of pooled data for each outcome was high in the domains of inconsistency,[17] indirectness,[18] imprecision,[19] or publication bias.[20] The quality of evidence across all outcomes was assigned as moderate because of the overlap between the intervention and control groups.

Tails and Revision of Algorithm

It has been suggested that proportion of time spent <85% (lower tail) may be associated with adverse outcomes such as mortality. The proportion of time spent <85% and >95% with original and revised pulse oximeter algorithms from SUPPORT, COT, and BOOST-II trials while infants were on supplemental oxygen is shown in Table 2. In the COT, after the first 3 days, data from infants who received >12 hours of supplemental oxygen only were included. The proportion of time spent <85% was significantly higher in the lower-target group compared with the higher-target group. Revision of the algorithm modestly reduced the proportion of time spent <85% and increased the proportion of time spent within the target 85% to 89% range in the lower-target group but this increase did not reach statistical significance (Fig 6). We did not generate similar graphs for NEC and ROP because of the subtle differences in definitions in various trials and the impact of mortality on these outcomes.

Sensitivity Analysis

There was no difference in outcomes between random-effects and fixed-effects analysis.

Heterogeneity

Clinical and statistical heterogeneity was low for all outcomes.

Limitations

Subtle differences in inclusion criteria were observed (Table 1): no outborn infants in the SUPPORT trial; time of randomization was short in SUPPORT compared with other trials. These differences could have contributed to the heterogeneity of patients. Errors in the pulse oximeter algorithm led to revision of the algorithm midway through 3 trials and could have contributed to heterogeneous results.

DISCUSSION

The 5 trials included in this systematic review were carefully planned in a collaborative manner to answer the following question: "Is the incidence of death and/or NDI different with a target SpO_2 of 85% to 89% vs 91% to 95% in extremely premature infants?" With primary combined outcome data available in

TABLE 1 Characteristics of the 5 Trials Addressing Optimal Oxygen Saturation Targets in Extremely Preterm Infants (Percentages Shown as Lower-Target Versus Higher-Target Groups)

	BOOST-II Australia[10,15]	BOOST-II UK[10,15]	BOOST-II New Zealand[10,23]	COT[11]	SUPPORT[9,22]
Centers	15	34	5	25	16
Start date	Mar 25, 2006	Sep 29, 2007	Sep 2006	Dec 24, 2006	Feb 2005
Closure of recruitment	Dec 24, 2010	Dec 24, 2010	Dec 2009	August 25, 2010	Feb 2009
Gestational age	<28 wk	<28 wk	<28 wk	23 0/7 to 27 6/7 wk	24 0/7 to 27 6/7 wk
Postnatal age	<24 h	<24 h	<24 h	First 24 h	< 2 h
Exclusion	Major congenital anomalies; Unlikely to survive or would not be available for follow-up	Major congenital anomalies; Unlikely to survive or would not be available for follow-up	Major congenital anomalies; Unlikely to survive or would not be available for follow-up	Not considered viable, pulmonary hypertension, dysmorphic features or congenital malformations, cyanotic congenital heart disease, unlikely to follow-up	Outborn; Decision not to provide full resuscitation; major congenital anomalies;
Multiple births - % of subjects and randomization of multiples	24.3% vs 23.8% Individual randomization	28.4% vs 29.4% Individual randomization	27.1% both groups Randomized separately	33.7% vs 31.1% Individual randomization	24.6% vs 26.6% Same group
Boys	51.6% vs 52.2%	52.5% vs 53.5% (revised)	52.9% vs 52.9%	55.5% vs 54.1%	52.1% vs 56.0%
Race, white		85.7% vs 84.3%		67.3% vs 68.4%	37.0% vs 42.1%
Birth weight	817 ± 177 vs 833 ± 190	821 ± 182 vs 818 ± 189 (revised)	873 ± 202 vs 884 ± 186	827 (190) vs 844 (199)	836 (193) vs 825 (193)
GA	26.0 ± 1.2 both groups	26.0 ± 1.3 both groups	26.1 ± 1.2 both groups	25.6 (1.2) both	26 (1) both groups
Outborn %	7.7% vs 7.4%	12.6% vs 11.2%	6.5% vs 7.6%	6.9% vs 9.3%	0
No antenatal glucocorticoids	11.3% vs 7.5%	6.3% vs 8.4%	11.8% vs 10.6%	11.8% vs 10%	3.2% vs 4.4%
Born by cesarean delivery	51.9% vs 54.4%	42.6% vs 40.3%	55.9% vs 53.5%	62.6% vs 59.4%	69.3% vs 65.6% (follow-up)
SGA		15.7% vs 14.8%		9.3% vs 8.6%	6.3% vs 8.3%
Minimization procedures to balance groups	Sex, GA, center, single/ multiple, inborn/ outborn	Sex, GA, center,	Sex, GA, center, inborn/ outborn	Center, GA	Center, GA
Total subjects (lower target + higher target)	1135 (568 + 567)	973 (486 + 487)	340 (170 + 170)	1201 (602 + 599)	1316 (654 + 662)
Subjects – primary outcome determined	1094%–96.4% (549 + 545)	941%–96.7% (473 + 468)	335% –98.5% (167 + 168)	1147% –95.5% (578 + 569)	1234%–93.8% (612 + 622)
Original algorithm	674 (335 + 339)	218 (107 + 111)	335 (167 + 168)	275 + 264[a]	1234 (612 + 622)
Revised algorithm	420 (214 + 206)	723 (366 + 357)	0	272 + 266[a]	0
Upper alarm threshold[b]	94% recommended	94% recommended	93% recommended	94% mandated (unless off supplemental oxygen)	95% suggested
Lower alarm threshold[c]	86% recommended	Left to individual centers	87% recommended	86% mandated	85% suggested
Discontinuing study pulse oximeters	36 wk PMA or stable in ambient air	36 wk PMA or stable in ambient air	36 wk PMA (at least first 2 wk of postnatal age), or SpO_2 >96% for >95% of time in ambient air for 3 d	36 wk PMA even if infant was in ambient air; if on respiratory support or oxygen at 35 wk, until 40 wk PMA (or discharge home)	36 wk PMA or stable in ambient air for 72 h
Assessment of outcome	Up to a corrected age of 2 y	Up to a corrected age of 2 y	Up to a corrected age of 2 y	Corrected age 18 mo (18–21-mo window)	Corrected age 18–22 mo
BSID-III cutoff score for NDI	<85[d]	<85[d]	BSID-III <85 BSID-II <70[f]	BSID-III <85 BSID-II <70	BSID-III cognitive score <70[e]
Visual impairment	Legally blind with <6/60 in better eye	Legally blind or partially sighted	Legally blind with <6/60	Corrected acuity <20/200 in the better eye	Vision worse than 20/200
Motor deficit[g]	Severe cerebral palsy (GMFCS ≥2) or not walking unaided at 2 y	Severe cerebral palsy (GMFCS ≥2) or not walking unaided at 2 y	Severe cerebral palsy (GMFCS ≥2)	GMFCS ≥2 or child walks <10 steps independently at 18 mo	GMFCS ≥2 or cerebral palsy

TABLE 1 Continued

	BOOST-II Australia[10,15]	BOOST-II UK[10,15]	BOOST-II New Zealand[10,23]	COT[11]	SUPPORT[9,22]
Hearing impairment	Hearing loss requiring or too severe to benefit from aiding or a cochlear implant	Hearing loss requiring or too severe to benefit from aiding or a cochlear implant	Deafness requiring hearing aids	Prescription of hearing aids or cochlear implants	Inability to understand the oral directions of the examiner and to communicate with or without hearing amplification

GA, gestational age.

[a] In the COT trial, 70 infants were exposed to both algorithms.

[b] Upper alarm limit refers to the displayed SpO_2 value (95% displayed SpO_2 value corresponds to 92% in the lower-target group and 95%–96% in the higher-target group).

[c] Lower alarm limit refers to the displayed SpO_2 value (86% displayed SpO_2 value corresponds to 84%–85% in the lower-target group and 89% in the higher-target group).

[d] Alternative measures of cognition and language if BSID-III not arranged in BOOST-II UK and Australia trials.

[e] Data for BSID-III cognitive score <85 included in this analysis.

[f] BSID-II <70 (assessed in 25 infants) and BSID-III <85 (assessed in 238/289 infants) or <10-word vocabulary at 2 years (assessed in 3/33 infants).

[g] GMFCS: gross motor function classification system scores range from 0 (normal) to 5 (most impaired).

4751 infants, there is no difference between these 2 SpO_2 target groups.

These trials used the composite outcome of mortality and NDI. The rationale for using the composite outcome in these trials was to account for death as a competing outcome and not because a difference in mortality was expected a priori.[24] Composite outcomes, in which multiple end points are combined (with one of the end points being mortality), are frequently used as primary outcome measures in neonatal RCTs. In a review of major trials with composite outcomes, only 4% of the trials were significant for mortality but not for the primary composite outcome.[25] The results of the current meta-analysis show higher mortality with lower target compared with higher target without any difference in composite primary outcome. This has led to considerable controversy regarding recommendations, with the revised European guidelines recommending a higher-target range (90%–94%) with Grades of Recommendation, Assessment, Development, and Evaluation level of evidence "B."[26] However, adaptation of higher target has been linked to increased ROP[27] and the impact of these guidelines on mortality and long-term visual impairment and NDI needs to be closely observed.

Although a pooled analysis of the 3 BOOST-II trials did not show a difference in primary outcome,[28] 2 recent meta-analyses of all 5 studies[29,30] reported an increased death/NDI with the lower target (RR 1.07 with 95% confidence interval 1.00–1.14, $P = .04$). However, these analyses included NDI as reported by individual study investigators with a BSID-III cognitive cutoff score of <70 for the SUPPORT and <85 for other trials. We have used uniform criteria for BSID-III cutoff scores for NDI (<85% for all studies) and demonstrated no significant difference for this primary outcome (Fig 2A).

The quality of evidence was rated as high in the following domains: consistency, directness, precision, and lack of publication bias. The heterogeneity was low for all outcomes. The risk-of-bias category was graded as moderate due to a lack of separation in the target saturations.[31] A maximal separation of 6% was intended by using a masking algorithm that allowed gradual reduction to no separation outside of the target SpO_2 range.[31–33] We acknowledge that quality of evidence is subjective and should not be penalized, if maintaining SpO_2 within a target range is exclusively due to clinical and practical factors. However, as pointed out by the COT investigators,[33] the masking algorithm may have played a role in reducing the separation between the 2 groups. In the lower-target group, the displayed SpO_2 decreased from 88% to 84% when the true SpO_2 changed from 85% to 84%, creating a zone of instability and tendency for the bedside provider to increase fraction of inspired oxygen (FIO_2). This partly explains high median SpO_2 (89%–91%) in the lower-target arms of the various trials (Table 2). Similarly, in the higher-oxygen target group, the displayed SpO_2 increased from 92% to 96% when the true saturation changed from 95% to 96%, creating a zone of instability and a tendency for the provider to decrease FIO_2. However, the masking algorithm did not significantly affect the higher-target group (median SpO_2 92%–93%, Table 2). The net effect was reduced separation between the 2 groups possibly as a consequence of the masking algorithm.

In the previously published BOOST-I trial comparing 2 SpO_2 targets (91%–94% and 95%–98%), the masking algorithm was simple with display SpO_2 ± 2% throughout without an area of "correction" or instability.[7] The BOOST-I investigators achieved the intended 4% separation between the groups with this simple algorithm (a median of 93% in the "standard" saturation group with desired target SpO_2 range 91%–94% and a median of 97% in the high SpO_2 group with a desired target SpO_2 range 95%–98%). In contrast, most trials in the NeOProM collaboration achieved a median SpO_2 of 90% to 91% in the lower-target group (outside the intended target 85%–89%). The

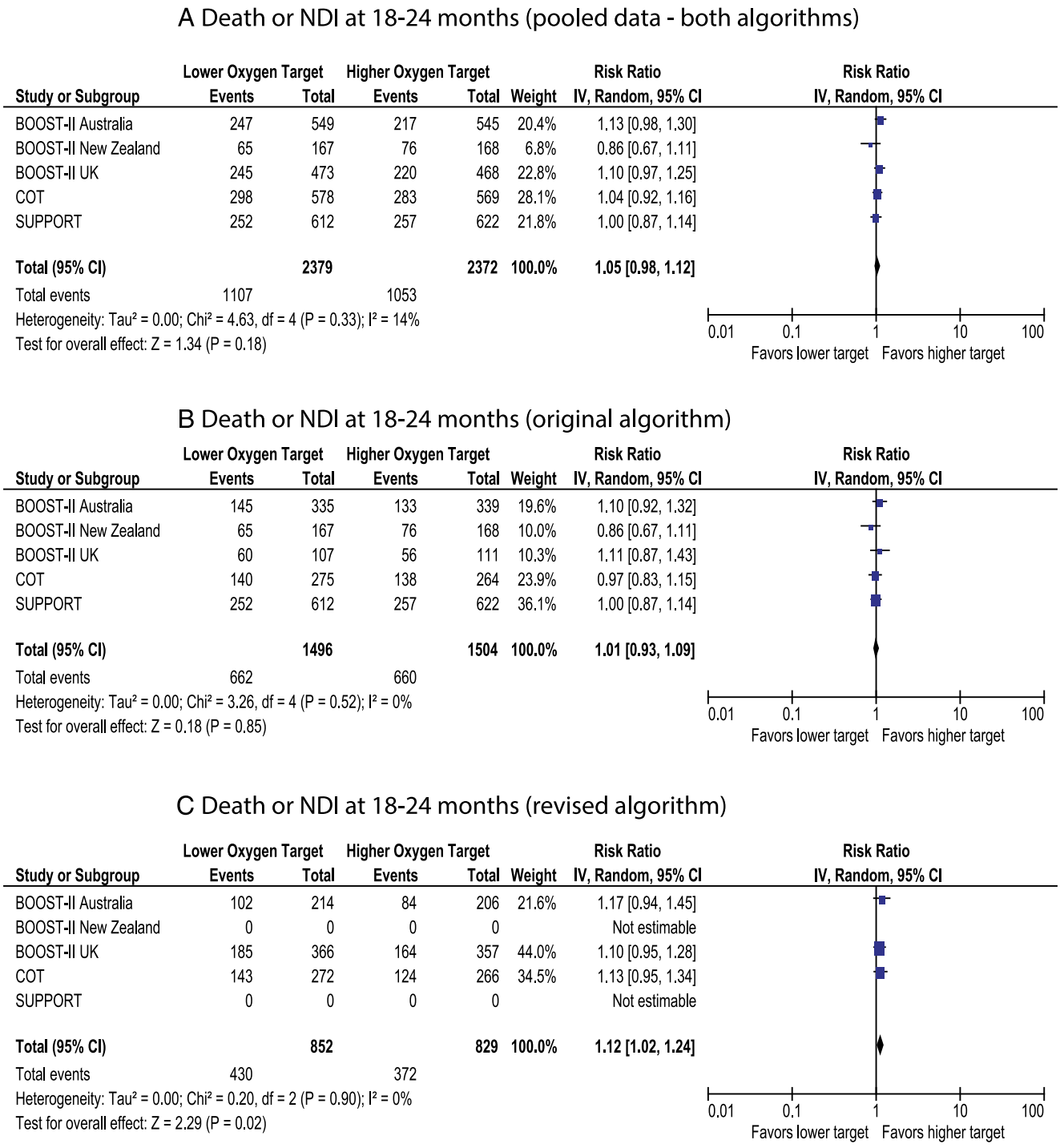
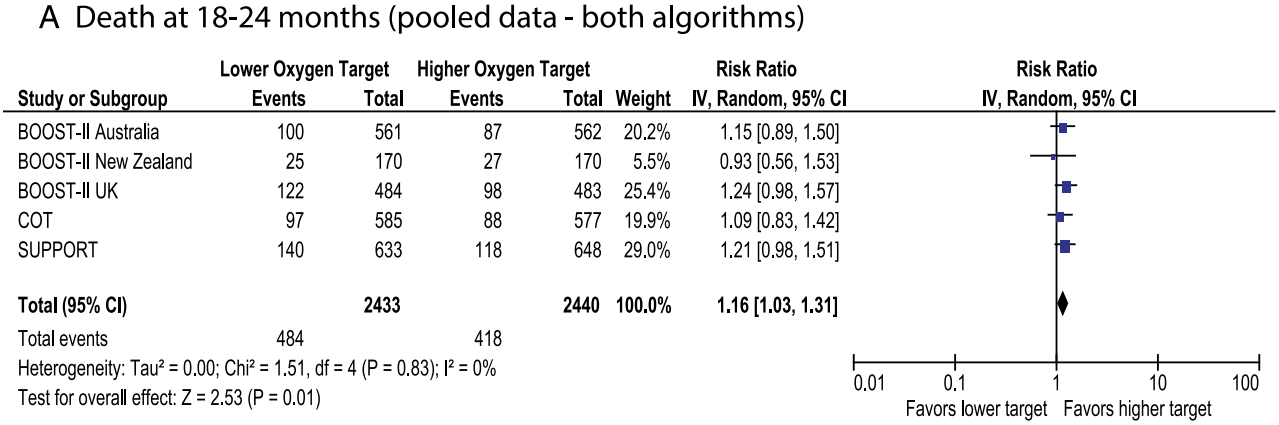


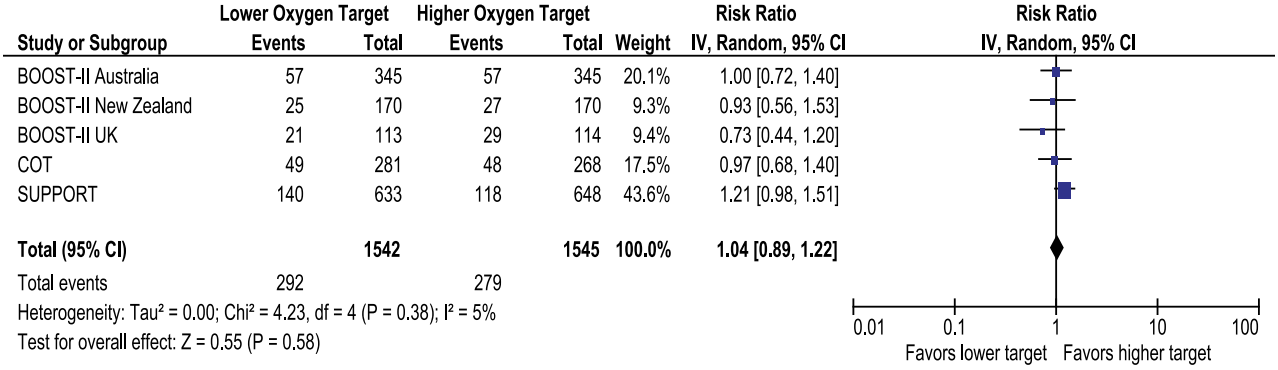
FIGURE 2

Forest plot demonstrating the incidence of death and/or NDI (BSID-III <85) at follow-up. Data for the SUPPORT trial include information provided by the steering committee of the neonatal research network (National Institute of Child Health and Human Development). A, Pooled data from both original and revised algorithms. B, Data from original algorithm only. C, Data from revised algorithm only. The pooled data with both algorithms has 70 additional babies from the COT trial who were exposed to both the original and revised software.

NeOProM trials did achieve a median SpO_2 of 92% to 93% in the high-target group (within the intended target range 91%–95%). Inability to achieve target SpO_2 within the target range appears to predominantly involve the lower-target range.

It is also possible that it is difficult to maintain SpO_2 in the lower-target range of 85% to 89% due to the inherent nature of the oxygen-hemoglobin dissociation curve[15] and not exclusively due to the masking algorithm. The higher range includes the plateau of





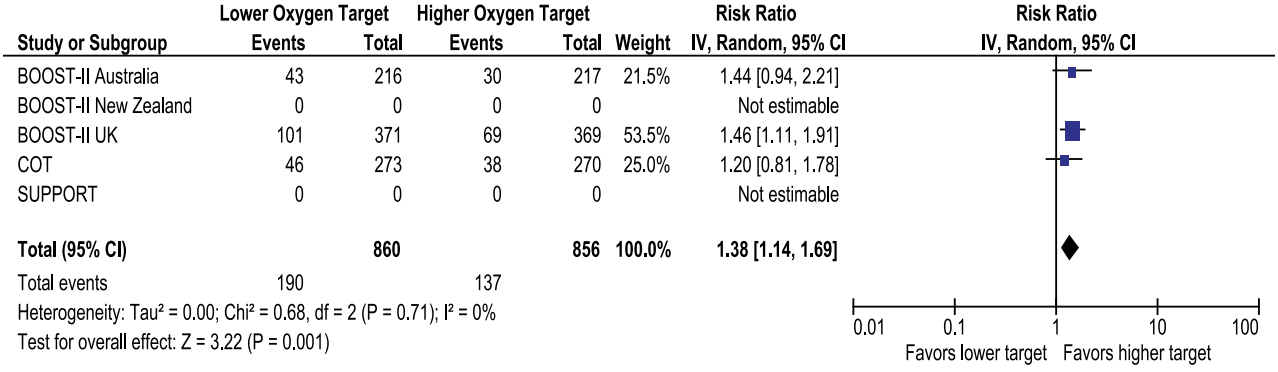


FIGURE 3
Forest plot demonstrating the incidence of death at follow-up. A, Pooled data from both original and revised algorithms. B, Data from original algorithm only. C, Data from revised algorithm only. The pooled data with both algorithms has 70 additional babies from the COT trial who were exposed to both the original and revised software.

the oxygen-hemoglobin dissociation curve, in which SpO_2 fluctuates less with changing PaO_2. In contrast, the slope of the oxygen-hemoglobin dissociation curve is steep in the 85% to 89% range, resulting in higher fluctuation in SpO_2 with small changes in PaO_2.

Finally, studies evaluating manual versus automated (closed-loop) FIO_2 control provide data on difficulties in limiting SpO_2 within a target range. Recent closed-loop FIO_2 control studies[34,35] suggest that the proportion of time spent within

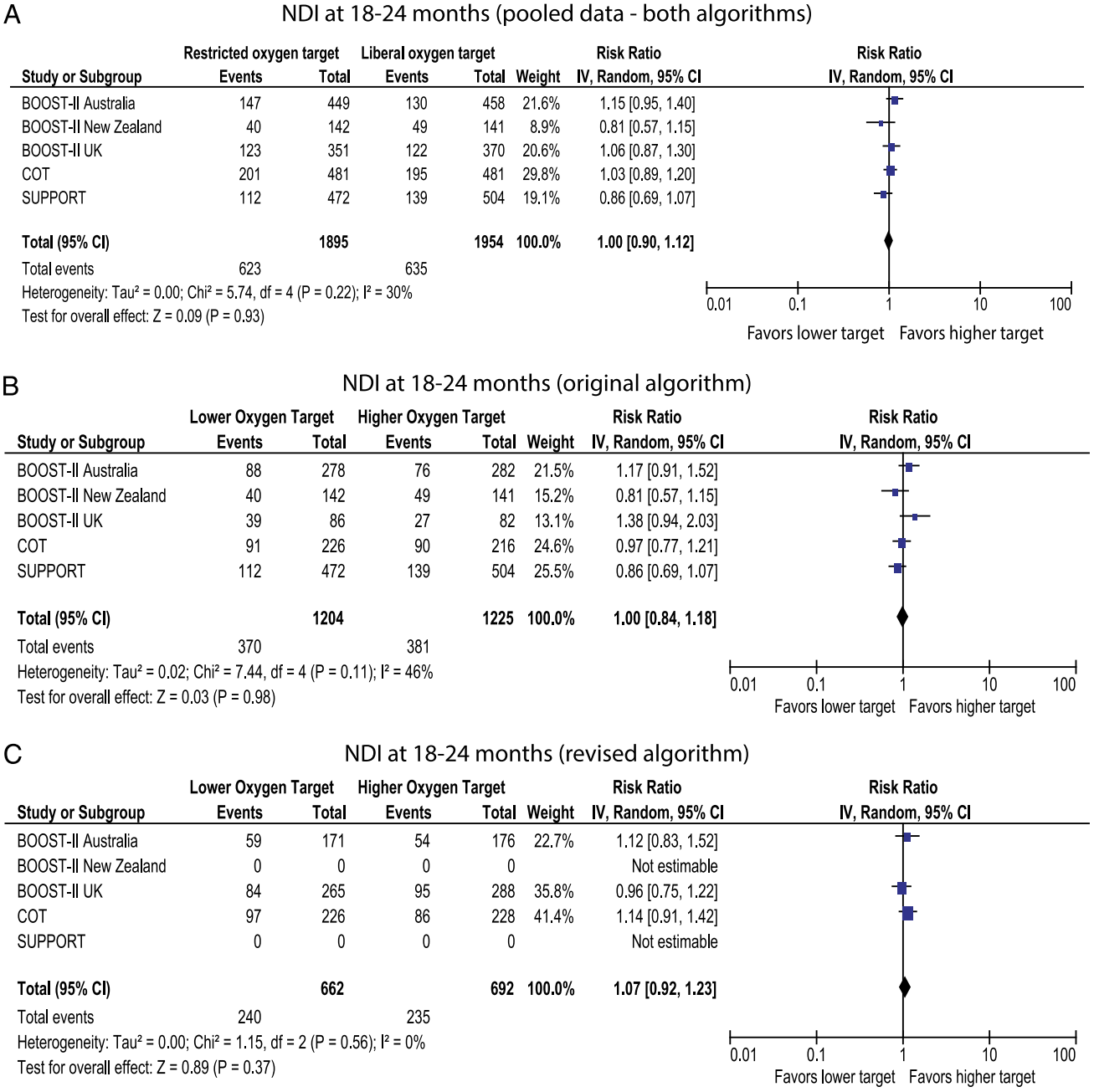


FIGURE 4

Forest plots demonstrating the incidence of NDI (BSID-III <85) at follow-up. Data for the SUPPORT trial include information provided by the steering committee of the neonatal research network (National Institute of Child Health and Human Development). A, Pooled data from both original and revised algorithms. B, Data from original algorithm only. C, Data from revised algorithm only. The pooled data with both algorithms has additional babies from the COT trial who were exposed to both the original and revised software.

range to be considerably higher with automated control (62% during automated and 57% with manual with target SpO_2 of 91%–95% and 72.8% during automated and 59.6% with manual with target SpO_2 of 90%–95%). However, the proportion of time spent within range with manual adjustment in these studies is higher than that reported within the 91% to 95% target arm in the BOOST-II UK and Australia trial (43.4% and 48.7%, respectively). This further suggests that the masking algorithm may have played a role in reducing the amount of time spent within the target range.

The lower-target groups had higher SpO_2 than intended and yet had significantly increased mortality compared with the higher-target group without significant heterogeneity for this outcome. Would the composite primary outcome or its components differ if the NeOProM studies had achieved the intended

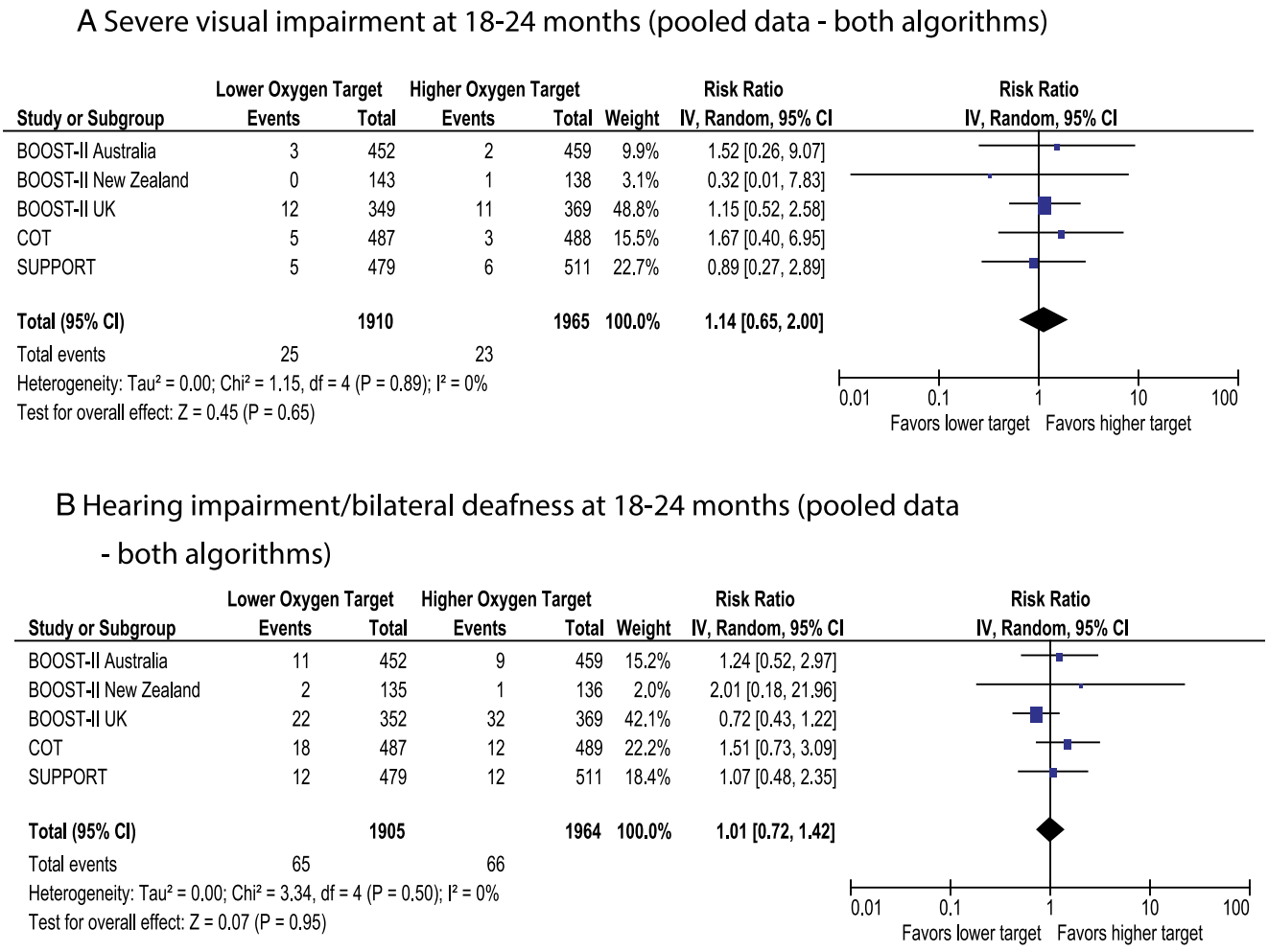


FIGURE 5
Forest plots demonstrating the incidence of severe visual impairment and hearing loss at follow-up.

separation and the lower-target group had median SpO_2 of ~87%? One can speculate that the effect size of mortality and/or the combined outcome of mortality/NDI may be higher in the lower-target group if this group had spent more time in the 85% to 89% range (as intended). However, a subgroup analysis of the COT trial showed that centers with more separation observed lower rates of death/NDI in the 85% to 89% than the 91% to 95% target range.[36] Such post hoc analysis should be interpreted with caution and the planned individual patient data meta-analysis might clarify the impact of separation on mortality.

From a physiologic perspective, revision of the algorithm had only a minor impact on SpO_2 and led to slightly better separation of the low- and high-target groups in the BOOST-II UK trial[15] (median SpO_2 in Table 2). The influence of this revision on outcome, especially mortality, varied between trials. In the SUPPORT trial, mortality at discharge was higher in the lower-target group with the original algorithm. In the BOOST-II UK trial, mortality trended lower in the lower-target group with the original algorithm and shifted to a significantly higher mortality after algorithm revision. In the BOOST-II Australia and COT trials, revision of the algorithm did not significantly change outcome. These differences in outcomes with algorithm revision are difficult to explain and we based the remainder of the discussion on pooled analysis (original + revised algorithms). We hope that the planned individual patient data meta-analysis of the NeOProM studies will enhance our understanding of the impact of revision of algorithm on outcomes.

Although the studies were planned to be identical with similar patient population and SpO_2 targets, many geographic and methodological differences may have contributed to differences in outcome (Table 1). The early time of enrollment and randomization in the SUPPORT trial potentially led to inclusion of many sicker infants. It was recently reported that the increased mortality in the lower-target arm of the SUPPORT trial was predominantly seen in infants small for gestational

age (SGA).[37] Respiratory diagnoses (respiratory distress syndrome, BPD, and pulmonary hypoplasia) accounted for 48.6% of mortality among SGA infants in the SUPPORT. The incidence of SGA infants was similar in both arms of the study in all the trials but their percentage was higher in BOOST-II UK compared with COT and SUPPORT and could have contributed to higher mortality/morbidity in that trial. Exclusion of infants with pulmonary hypertension from the COT trial may potentially have contributed to lack of difference in mortality, as low SpO_2 target may exacerbate pulmonary hypertension.[38] Outborn infants, at risk for higher mortality and morbidity, were excluded from SUPPORT. Finally, oximeter alarm settings and implementation differed between the studies (Table 1) and could have influenced outcomes.

If the median saturations were not well separated, what other factors could have influenced mortality in the lower-target group? The proportion of time spent <85% and >95% significantly differed between the lower-target and higher-target groups (as expected). Revision of the algorithm decreased the lower tail (Table 2 and Fig 6). Although profound hypoxemia is associated with adverse outcome, we could not demonstrate an association between time spent <85% SpO_2 at a study level and mortality. However, no conclusions can be derived from this observation without performing individual patient data analysis. It is possible that intermittent, profound desaturations and SpO_2 range that an individual patient spends irrespective of whether on supplemental oxygen or not may influence outcome.

Major neonatal morbidities, including ROP, have a strong genetic component.[39,40] Variations in genetic factors[41] may play a role in differences in mortality, NDI, and visual/hearing impairment observed between BOOST-II UK compared with other studies. The percentage of white infants was higher in BOOST-II UK compared with COT and SUPPORT (Table 1) and might have played a role in increasing morbidity and mortality.[42]

TABLE 2 Median Saturations and Proportion of Time Spent <85% (Lower Tail) and >95% (Higher Tail) Based on Pulse Oximeter Algorithm (Original or Revised) While on Supplemental Oxygen

Study	Target SpO_2 Arm	Median SpO_2, %		Tail, Proportion of Time Spent at This SpO_2 Level, %	Original, %	Revised, %
		Original	Revised			
COT[a]	85%–89% (low)	91	91	<85	20.2	18.5
				>95	17.5	15.2
	91%–95% (high)	93	93	<85	9.1	8.2
				>95	22.4	21.2
BOOST-II UK[b]	85%–89% (low)	91	90	<85	25.7	22.1
				>95	16.1	13.9
	91%–95% (high)	92	93	<85	15.0	12.3
				>95	18.7	20.4
BOOST-II Australia[b]	85%–89% (low)	90	89	<85	27.4	24.1
				>95	11.1	8.1
	91%–95% (high)	93	92	<85	13.5	10.8
				>95	18.6	16.4
BOOST-II New Zealand[b]	85%–89% (low)	91		<85	21.1	Only oximeters with the original algorithm were used in these studies.
				>95	15.5	
	91%–95% (high)	93		<85	10.8	
				>95	22.2	
SUPPORT[b]	85%–89% (low)	90[c]		<85	10.9	
				>95	33.0	
	91%–95% (high)	93[c]		<85	16.0	
				>95	26.7	

[a] All study days with >12 hours of supplemental O_2.
[b] Time when the infant was receiving oxygen.
[c] Estimate based on visual inspection of the graphs.

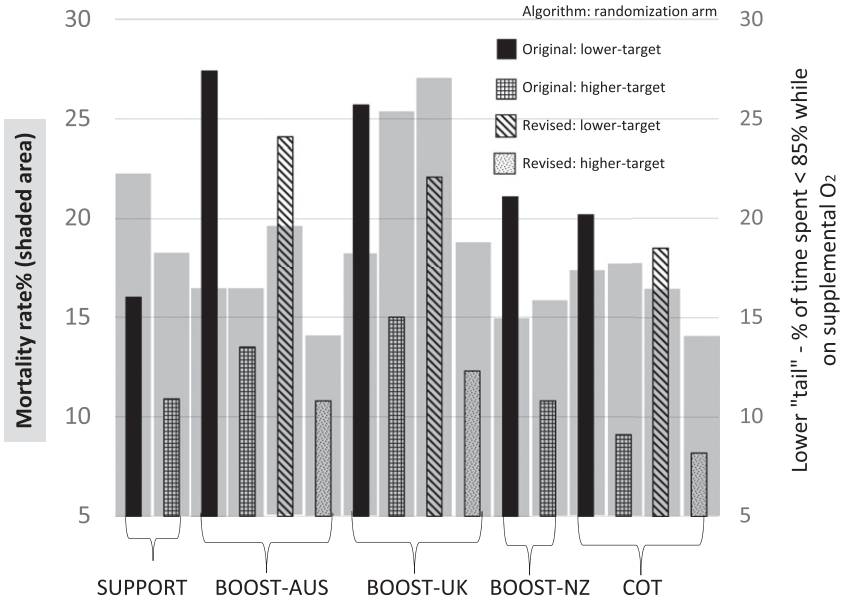


FIGURE 6
Combination chart showing mortality (shown as shaded gray area) and bar diagrams showing proportion of time spent <85% while on supplemental oxygen. Lower-oxygen target groups spent more time <85% SpO_2 compared with higher-oxygen target groups.

Primary Outcome

Based on this systematic review, a moderate level of evidence exists that suggests no significant difference in the primary outcome of death/NDI between SpO_2 targets of 85% to 89% and 91% to 95% by using pooled data. We have previously reported that the incidence of NEC is more common in the lower-target group. The lower-target group has a tendency toward reduced severe ROP.[13] Interestingly, the higher incidence of ROP did not translate to increased severe visual impairment (defined as bilateral legal blindness, Table 1) at 18 to 24 months in the higher-target group. However, severe ROP and intervention for severe ROP result in other forms of visual impairment (eg, unilateral blindness) and nonvisual morbidity,[28] affecting quality-of-life outcomes for extremely preterm infants and must be taken into account while creating practical guidelines.

Conclusions and Practical Recommendations

The SUPPORT, BOOST-II, and COT are well-conducted trials addressing a very important question in neonatology. The study oximeters (errors in the original algorithm and the effect of masking algorithm on maintaining target range) may have at least partly influenced the results, casting doubt on the validity of these findings. The higher mortality (19.9% vs 17.1%) and reduced severe ROP in the lower-target group was not accompanied by a significant change in severe visual impairment (1.3% vs 1.2%). Although the primary outcome was similar between the 2 groups, these data do not support restricting the upper limit of SpO_2 target range to 89% in preterm infants. Oxygen saturation targets between 91% and 95% appear safer, but are associated with increased incidence of ROP.[27] Practical considerations, such as difficulty in maintaining SpO_2 in a narrow 5% range, have resulted in subjects randomized to 91% to 95% spending 13.9% to 22.4% of the time with SpO_2 >95% and subjects randomized to 85% to 89% spending 20.2% to 27.4% of the time <85% while on supplemental oxygen. Although the safer range appears to be 91% to 95%, avoidance of hyperoxia is warranted through education, innovation, and identification of alternate methods of monitoring (such as transcutaneous PO_2)[43] and automated, closed-loop adjustment of FIO_2.[35,36]

Some experts have suggested reducing the lower limit of SpO_2 alarm to the high 80s.[32] Target ranges should not be equated to alarm settings. Alarm limits (lower limit of 89% and upper limit of 96%) have not been rigorously studied, but may offer a practical solution. Such settings can potentially limit the amount of time an extremely preterm infant spends in extreme hypoxemia (<85%) and hyperoxemia (>95%) but need further investigation.

The recently published AAP Clinical Report states that the ideal physiologic target range is a compromise among negative outcomes associated with either hyperoxemia or hypoxemia.[44] This report mentions that recent RCTs suggest that targeted SpO_2 range of 90% to 95% may be safer than 85% to 89% at least for some infants. The AAP report concludes that the ideal SpO_2 range for extremely low birth weight infants remains unknown.[44] The revised 2016 European guidelines recommend SpO_2 target between 90% to 94% (quality of evidence: moderate; strength of recommendation: weak) and alarm limits at 89% and 95% (quality of evidence: very low; strength of recommendation: weak) for preterm infants receiving oxygen.[26]

Individual preterm infants have different mechanisms of susceptibility to injury and resilience to hypoxia and hyperoxia.[45] Factors such as corrected gestational age,[46] growth status,[37] pulmonary hypertension,[38] and ROP[47] influence the risk of hyperoxia and hypoxia. Results from the planned meta-analysis of individual patient data may clarify optimal targets for individual patients based on clinical characteristics and comorbidities and lead to an era of "precision-medicine" in neonatology.[43] In the meantime, we can state with moderate confidence that a lower alarm limit of 89% and an upper alarm limit of 96% offer a practical solution pending further studies and analyses.

ACKNOWLEDGMENTS

We thank Drs Marie Gantz, Rosemary Higgins, and Waldemar Carlo and the Neonatal Research Network Steering Committee for providing additional data (infants with NDI based on BSID-III cognitive score <70 and proportion of time spent <85% and >95%) from the SUPPORT trial.

ABBREVIATIONS

AAP: American Academy of Pediatrics
BOOST-II: Benefits of Oxygen Saturation Targeting-II
BPD: bronchopulmonary dysplasia
BSID: Bayley Scale of Infant Development
COT: Canadian Oxygen Trial
FIO_2: fraction of inspired oxygen
NDI: neurodevelopmental impairment
NEC: necrotizing enterocolitis
NeOProM: Neonatal Oxygenation Prospective Meta-analyses
PaO_2: partial pressure of oxygen, arterial
PMA: postmenstrual age
RCT: randomized controlled trial
ROP: retinopathy of prematurity
RR: risk ratio
SGA: small for gestational age
SpO_2: pulse oxygen saturation
SUPPORT: Surfactant, Positive Pressure and Pulse Oximetry Randomized Trial

Address correspondence to Satyan Lakshminrusimha, MD, Division of Neonatal-Perinatal Medicine, Women and Children's Hospital of Buffalo, Buffalo, NY 14222. E-mail: slakshmi@buffalo.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2017 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: Funded by 5 R01 HD072929 – Optimal oxygenation in neonatal lung injury (Dr Lakshminrusimha). Funded by the National Institutes of Health (NIH).

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.

REFERENCES

1. Vento M, Saugstad OD. Oxygen as a therapeutic agent in neonatology: a comprehensive approach. *Semin Fetal Neonatal Med*. 2010;15(4):185–189
2. Silverman WA. Oxygen therapy and retrolental fibroplasia. *Am J Public Health Nations Health*. 1968;58(11):2009–2011
3. Robertson AF. Reflections on errors in neonatology: I. The "Hands-Off" years, 1920 to 1950. *J Perinatol*. 2003;23(1):48–55
4. Bolton DP, Cross KW. Further observations on cost of preventing retrolental fibroplasia. *Lancet*. 1974;1(7855):445–448
5. American Academy of Pediatrics, American College of Obstetricians and Gynecologists, March of Dimes Birth Research Foundation. *Guidelines for Perinatal Care*. Elk Grove Village, IL: American Academy of Pediatrics; 2007
6. Greenspan JS, Goldsmith JP. Oxygen therapy in preterm infants: hitting the target. *Pediatrics*. 2006;118(4):1740–1741
7. Askie LM, Henderson-Smart DJ, Irwig L, Simpson JM. Oxygen-saturation targets and outcomes in extremely preterm infants. *N Engl J Med*. 2003;349(10):959–967
8. Askie LM, Brocklehurst P, Darlow BA, Finer N, Schmidt B, Tarnow-Mordi W; NeOProM Collaborative Group. NeOProM: Neonatal Oxygenation Prospective Meta-analysis Collaboration study protocol. *BMC Pediatr*. 2011;11(1):6
9. Carlo WA, Finer NN, Walsh MC, et al; SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network. Target ranges of oxygen saturation in extremely preterm infants. *N Engl J Med*. 2010;362(21):1959–1969
10. Stenson BJ, Tarnow-Mordi WO, Darlow BA, et al; BOOST II United Kingdom Collaborative Group; BOOST II Australia Collaborative Group; BOOST II New Zealand Collaborative Group. Oxygen saturation and outcomes in preterm infants. *N Engl J Med*. 2013;368(22):2094–2104
11. Schmidt B, Whyte RK, Asztalos EV, et al; Canadian Oxygen Trial (COT) Group. Effects of targeting higher vs lower arterial oxygen saturations on death or disability in extremely preterm infants: a randomized clinical trial. *JAMA*. 2013;309(20):2111–2120
12. Johnston ED, Boyle B, Juszczak E, King A, Brocklehurst P, Stenson BJ. Oxygen targeting in preterm infants using the Masimo SET Radical pulse oximeter. *Arch Dis Child Fetal Neonatal Ed*. 2011;96(6):F429–F433
13. Saugstad OD, Aune D. Optimal oxygenation of extremely low birth weight infants: a meta-analysis and systematic review of the oxygen saturation target studies. *Neonatology*. 2014;105(1):55–63
14. Manja V, Lakshminrusimha S, Cook DJ. Oxygen saturation target range for extremely preterm infants: a systematic review and meta-analysis. *JAMA Pediatr*. 2015;169(4):332–340
15. Tarnow-Mordi W, Stenson B, Kirby A, et al; BOOST-II Australia and United Kingdom Collaborative Groups. Outcomes of two trials of oxygen-saturation targets in preterm infants. *N Engl J Med*. 2016;374(8):749–760
16. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol*. 2011;64(4):407–415
17. Guyatt GH, Oxman AD, Kunz R, et al; GRADE Working Group. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *J Clin Epidemiol*. 2011;64(12):1294–1302
18. Guyatt GH, Oxman AD, Kunz R, et al; GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol*. 2011;64(12):1303–1310
19. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision. *J Clin Epidemiol*. 2011;64(12):1283–1293
20. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol*. 2011;64(12):1277–1282
21. Borenstein M, Hedges LV, Higgins JPT, et al. *Introduction to Meta-Analysis*. Hoboken, NJ: Wiley; 2011
22. Vaucher YE, Peralta-Carcelen M, Finer NN, et al; SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network. Neurodevelopmental outcomes in the early CPAP and pulse oximetry trial. *N Engl J Med*. 2012;367(26):2495–2504
23. Darlow BA, Marschner SL, Donoghoe M, et al. Randomized controlled trial of oxygen saturation targets in very preterm infants: two year outcomes. *J Pediatr*. 2014;165(1):30–35.e2
24. Polin RA, Bateman D. Oxygen-saturation targets in preterm infants. *N Engl J Med*. 2013;368(22):2141–2142
25. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA*. 2003;289(19):2554–2559

26. Sweet DG, Carnielli V, Greisen G, et al. European consensus guidelines on the management of respiratory distress syndrome – 2016 update. *Neonatology*. 2016;111(2):107–125

27. Manley BJ, Kuschel CA, Elder JE, Doyle LW, Davis PG. Higher rates of retinopathy of prematurity after increasing oxygen saturation targets for very preterm infants: experience in a single center. *J Pediatr*. 2016;168:242–244

28. Cummings JJ, Lakshminrusimha S, Polin RA. Oxygen-saturation targets in preterm infants. *N Engl J Med*. 2016;375(2):186–187

29. Stenson BJ. Oxygen saturation targets for extremely preterm infants after the NeOProM Trials. *Neonatology*. 2016;109(4):352–358

30. Tarnow-Mordi W, Stenson B, Kirby A. Oxygen-saturation targets in preterm infants. *N Engl J Med*. 2016;375(2):187–188

31. Lakshminrusimha S, Manja V, Mathew B, Suresh GK. Oxygen targeting in preterm infants: a physiological interpretation. *J Perinatol*. 2015;35(1):8–15

32. Sola A, Golombek SG, Montes Bueno MT, et al. Safe oxygen saturation targeting and monitoring in preterm infants: can we avoid hypoxia and hyperoxia? *Acta Paediatr*. 2014;103(10):1009–1018

33. Schmidt B, Roberts RS, Whyte RK, et al. Impact of study oximeter masking algorithm on titration of oxygen therapy in the Canadian oxygen trial. *J Pediatr*. 2014;165(4):666–671.e2

34. Lal M, Tin W, Sinha S. Automated control of inspired oxygen in ventilated preterm infants: crossover physiological study. *Acta Paediatr*. 2015;104(11):1084–1089

35. van Kaam AH, Hummler HD, Wilinska M, et al. Automated versus manual oxygen control with different saturation targets and modes of respiratory support in preterm infants. *J Pediatr*. 2015;167(3):545–50.e1–2

36. Schmidt B, Whyte RK, Shah PS, et al; Canadian Oxygen Trial (COT) Group. Effects of targeting higher or lower oxygen saturations in centers with more versus less seperation between median saturations. *J Pediatr*. 2016;178:288–291.e2

37. Walsh MC, Di Fiore JM, Martin RJ, Gantz M, Carlo WA, Finer N. Association of oxygen target and growth status with increased mortality in small for gestational age infants: further analysis of the Surfactant, Positive Pressure and Pulse Oximetry Randomized Trial. *JAMA Pediatr*. 2016;170(3):292–294

38. Lakshminrusimha S, Manja V, Steinhorn RH. Interaction of target oxygen saturation, bronchopulmonary dysplasia, and pulmonary hypertension in small for gestational age preterm neonates. *JAMA Pediatr*. 2016;170(8):807–808

39. Bhandari V, Gruen JR. What is the basis for a genetic approach in neonatal disorders? *Semin Perinatol*. 2015;39(8):568–573

40. Bizzarro MJ, Hussain N, Jonsson B, et al. Genetic susceptibility to retinopathy of prematurity. *Pediatrics*. 2006;118(5):1858–1863

41. Bhandari V, Bizzarro MJ, Shetty A, et al; Neonatal Genetics Study Group. Familial and genetic susceptibility to major neonatal morbidities in preterm twins. *Pediatrics*. 2006;117(6):1901–1906

42. Ravelli AC, Schaaf JM, Mol BW, et al. Antenatal prediction of neonatal mortality in very premature infants. *Eur J Obstet Gynecol Reprod Biol*. 2014;176:126–131

43. Quine D, Stenson BJ. Does the monitoring method influence stability of oxygenation in preterm infants? A randomised crossover study of saturation versus transcutaneous monitoring. *Arch Dis Child Fetal Neonatal Ed*. 2008;93(5):F347–F350

44. Cummings JJ, Polin RA; Committee on Fetus and Newborn. Oxygen targeting in extremely low birth weight infants. *Pediatrics*. 2016;138(2):e20161576

45. Synnes A, Miller SP. Oxygen therapy for preterm neonates: the elusive optimal target. *JAMA Pediatr*. 2015;169(4):311–313

46. Hergenhan A, Steurer M, Berger TM. Gestational age-adapted oxygen saturation targeting and outcome of extremely low gestational age neonates (ELGANs). *Swiss Med Wkly*. 2015;145:w14197

47. Supplemental Therapeutic Oxygen for Prethreshold Retinopathy Of Prematurity (STOP-ROP), a randomized, controlled trial. I: primary outcomes. *Pediatrics*. 2000;105(2):295–310