JAMA | Original Investigation

# Association Between Oxygen Saturation Targeting and Death or Disability in Extremely Preterm Infants in the Neonatal Oxygenation Prospective Meta-analysis Collaboration

Lisa M. Askie, PhD; Brian A. Darlow, MD; Neil Finer, MD; Barbara Schmidt, MD; Ben Stenson, MD; William Tarnow-Mordi, MBChB; Peter G. Davis, MD; Waldemar A. Carlo, MD; Peter Brocklehurst, MBChB; Lucy C. Davies, MSc; Abhik Das, PhD; Wade Rich, BSHS; Marie G. Gantz, PhD; Robin S. Roberts, MSc; Robin K. Whyte, MB; Lorrie Costantini, BA; Christian Poets, MD; Elizabeth Asztalos, MD; Malcolm Battin, MD; Henry L. Halliday, MD; Neil Marlow, DM; Win Tin, MBBS; Andrew King, BA; Edmund Juszczak, MSc; Colin J. Morley, MD; Lex W. Doyle, MD; Val Gebski, MSc; Kylie E. Hunter, MPH; Robert J. Simes, MD; for the Neonatal Oxygenation Prospective Meta-analysis (NeOProM) Collaboration

**IMPORTANCE** There are potential benefits and harms of hyperoxemia and hypoxemia for extremely preterm infants receiving more vs less supplemental oxygen.

**OBJECTIVE** To compare the effects of different target ranges for oxygen saturation as measured by pulse oximetry ($Spo_2$) on death or major morbidity.

**DESIGN, SETTING, AND PARTICIPANTS** Prospectively planned meta-analysis of individual participant data from 5 randomized clinical trials (conducted from 2005-2014) enrolling infants born before 28 weeks' gestation.

**EXPOSURES** $Spo_2$ target range that was lower (85%-89%) vs higher (91%-95%).

**MAIN OUTCOMES AND MEASURES** The primary outcome was a composite of death or major disability (bilateral blindness, deafness, cerebral palsy diagnosed as ≥2 level on the Gross Motor Function Classification System, or Bayley-III cognitive or language score <85) at a corrected age of 18 to 24 months. There were 16 secondary outcomes including the components of the primary outcome and other major morbidities.

**RESULTS** A total of 4965 infants were randomized (2480 to the lower $Spo_2$ target range and 2485 to the higher $Spo_2$ range) and had a median gestational age of 26 weeks (interquartile range, 25-27 weeks) and a mean birth weight of 832 g (SD, 190 g). The primary outcome occurred in 1191 of 2228 infants (53.5%) in the lower $Spo_2$ target group and 1150 of 2229 infants (51.6%) in the higher $Spo_2$ target group (risk difference, 1.7% [95% CI, −1.3% to 4.6%]; relative risk [RR], 1.04 [95% CI, 0.98 to 1.09], $P = .21$). Of the 16 secondary outcomes, 11 were null, 2 significantly favored the lower $Spo_2$ target group, and 3 significantly favored the higher $Spo_2$ target group. Death occurred in 484 of 2433 infants (19.9%) in the lower $Spo_2$ target group and 418 of 2440 infants (17.1%) in the higher $Spo_2$ target group (risk difference, 2.8% [95% CI, 0.6% to 5.0%]; RR, 1.17 [95% CI, 1.04 to 1.31], $P = .01$). Treatment for retinopathy of prematurity was administered to 220 of 2020 infants (10.9%) in the lower $Spo_2$ target group and 308 of 2065 infants (14.9%) in the higher $Spo_2$ target group (risk difference, −4.0% [95% CI, −6.1% to −2.0%]; RR, 0.74 [95% CI, 0.63 to 0.86], $P < .001$). Severe necrotizing enterocolitis occurred in 227 of 2464 infants (9.2%) in the lower $Spo_2$ target group and 170 of 2465 infants (6.9%) in the higher $Spo_2$ target group (risk difference, 2.3% [95% CI, 0.8% to 3.8%]; RR, 1.33 [95% CI, 1.10 to 1.61], $P = .003$).

**CONCLUSIONS AND RELEVANCE** In this prospectively planned meta-analysis of individual participant data from extremely preterm infants, there was no significant difference between a lower $Spo_2$ target range compared with a higher $Spo_2$ target range on the primary composite outcome of death or major disability at a corrected age of 18 to 24 months. The lower $Spo_2$ target range was associated with a higher risk of death and necrotizing enterocolitis, but a lower risk of retinopathy of prematurity treatment.

*JAMA*. 2018;319(21):2190-2201. doi:10.1001/jama.2018.5725

Editorial page 2173

Supplemental content

**Author Affiliations:** Author affiliations are listed at the end of this article.

**Group Information:** Information about the members of the Neonatal Oxygenation Prospective Meta-analysis (NeOProM) Collaboration appears at the end of the article.

**Corresponding Author:** Lisa M. Askie, PhD, National Health and Medical Research Council Clinical Trials Centre, University of Sydney, Medical Foundation Building, Level 6, 92-94 Parramatta Rd, Camperdown, NSW 2050, Australia (lisa.askie@ctc.usyd.edu.au).

Oxygen has been used in nurseries for more than 70 years. In the 1950s, it was shown that administering unrestricted oxygen to preterm infants significantly increased their risk of severe retinopathy of prematurity (ROP).[1] Oxygen saturation as measured by pulse oximetry ($Spo_2$), which is a noninvasive measure, is now almost universal in neonatal intensive care units. Lower oxygen levels ($Spo_2$ target ≤90%) may reduce ROP,[2] but no studies predating these investigations[1,3] demonstrated impaired neurodevelopment or an increased risk of death. Higher oxygen levels ($Spo_2$ target >90%) may increase adverse pulmonary sequelae at $Spo_2$ levels higher than 95% in infants who remain dependent on oxygen for many weeks after birth.[4,5]

A total sample size of approximately 5000 infants was required to detect the small but clinically important hypothesized difference of 4% in the primary outcome of death or major disability between lower and higher $Spo_2$ target ranges. To achieve this, the Neonatal Oxygenation Prospective Meta-analysis (NeOProM) Collaboration[6] was formed in 2003 with the investigators from 5 separate randomized clinical trials prospectively planning to undertake their individual trials using similar study designs, participants, interventions, comparators and outcomes, and agreeing to provide individual participant data at trial completion for inclusion in a meta-analysis. A previous Cochrane review[7] reported the findings of an analysis of these 5 studies using aggregate data available from the published trial reports. This article reports the results from the prospectively planned meta-analysis of the individual participant data from these trials.

### Key Points

**Question** For extremely preterm infants, is targeting a lower oxygen saturation (85%-89%) compared with a higher saturation (91%-95%) associated with a difference in death or major disability by a corrected age of 24 months?

**Findings** In a prospectively designed meta-analysis of individual participant data from 4965 infants in 5 randomized clinical trials, there was no significant difference in the primary composite outcome of death or major disability between those treated with lower vs higher oxygen saturations (53.5% vs 51.6%, respectively). Lower oxygen targets were associated with increased death and necrotizing enterocolitis but reduced retinopathy of prematurity treatment.

**Meaning** Among extremely preterm infants, there was no significant difference between lower and higher oxygen saturation targets on a composite of death or major disability; secondary end points may need to be considered in decision making.

## Methods

### Data Sources and Search Strategy

The NeOProM Collaboration was a prospectively planned meta-analysis of individual participant data for the following 5 trials: the Surfactant, Positive Pressure, and Pulse Oximetry Randomized Trial (SUPPORT),[8] which was conducted from 2005 through 2011 in the United States; the Canadian Oxygen Trial,[9] which was conducted from 2006 through 2012; the Benefits Of Oxygen Saturation Targeting (BOOST) in New Zealand,[10] which was conducted from 2006 through 2012; BOOST II in the United Kingdom,[11] which was conducted from 2007 through 2014; and BOOST II in Australia,[12] which was conducted from 2006 through 2013. These studies were considered eligible for inclusion in the meta-analysis prior to the results of any of the trials being known.[13] The study protocol was published[14] (Supplement 1) in January 2011 and registered on ClinicalTrials.gov. The statistical analysis plan was finalized in September 2015 and appears in Supplement 2. The conduct of each trial was approved by institutional review boards or ethics committees and written informed consent was obtained from participating parents.

### Study Selection and Eligibility Criteria

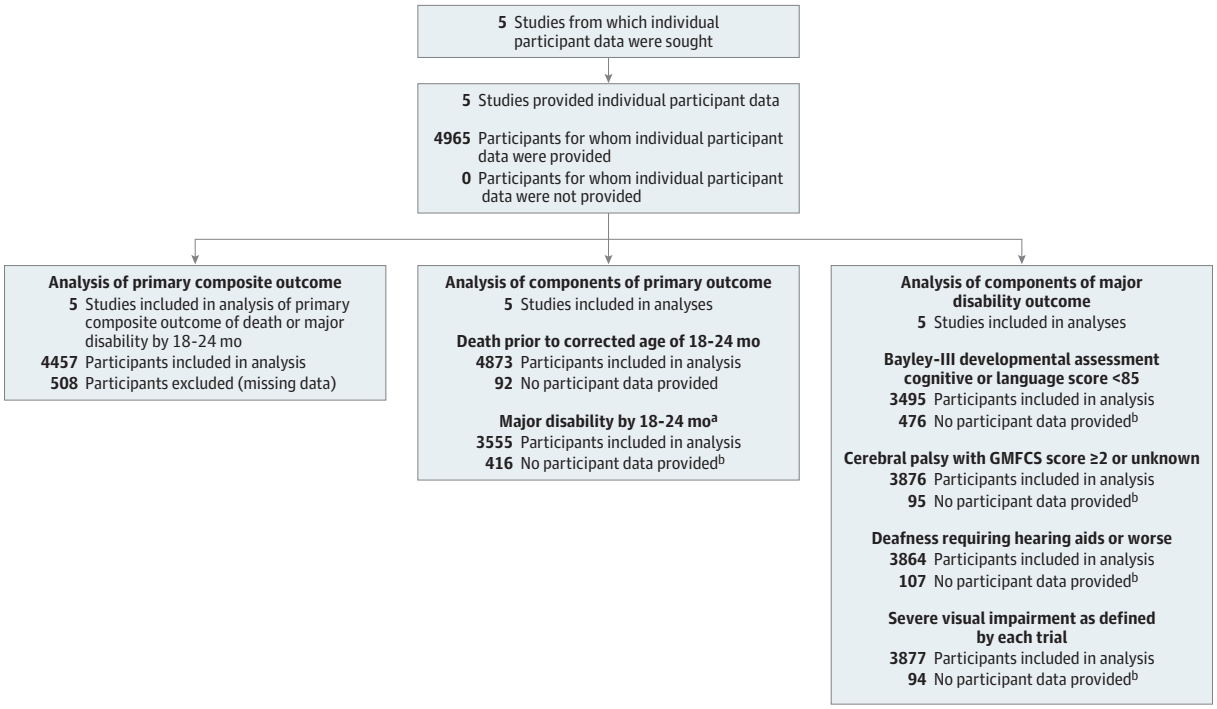
All 5 studies[15-20] were randomized, double-blind, multicenter trials with infants eligible if they were born before 28 weeks' gestation and enrolled within 24 hours of birth. Infants were randomized within each trial to target either a lower (85%-89%) or higher (91%-95%) $Spo_2$ range. To ensure that parents, caregivers, and outcome assessors remained masked to treatment allocation, each trial used Masimo pulse oximeters that had been modified to display and store oxygen saturations between 88% and 92% that were either 3% above or 3% below the actual values. True values were displayed if the actual $Spo_2$ decreased below 84% or increased above 96%. Caregivers were instructed to adjust the concentration of inspired oxygen to maintain the displayed $Spo_2$ between 88% and 92%, thus producing 2 treatment groups with actual target saturations of either 85% to 89% or 91% to 95% (eFigure 1 in Supplement 3).

During the trials, an artifact was identified in the calibration software of the oximeters that had the potential to influence the achieved oxygen saturation patterns.[21] Three of the trials (BOOST II in the United Kingdom, BOOST II in Australia, and the Canadian Oxygen Trial) changed their oximeters to incorporate the revised oximeter software. Based on advice from their data and safety monitoring committees, 2 trials (BOOST II in the United Kingdom and BOOST II in Australia) were terminated by their respective trial steering committees after a pooled interim analysis of mortality data was undertaken[22] in subgroups by oximeter software type when 81% and 95%, respectively, of their planned trial recruitment sample sizes had been met.

### Data Extraction

A list of requested variables was sent to each trial group based on the statistical analysis plan prior to the sharing of any individual participant data for use in the combined meta-analysis. These variables included randomization and baseline characteristics (including subgroup variables) while infants were in the hospital as well as 18- to 24-month follow-up information from individual participants (a full list of prespecified variables appear in Supplement 3). Deidentified data were provided by the trial groups between March

Figure 1. Participant Flow Diagram




[a] Major disability was prespecified (published in the Neonatal Oxygenation Prospective Meta-analysis protocol; Supplement 1) and includes any of the following: Bayley-III developmental assessment cognitive score of less than 85, language score of less than 85, or both; severe visual impairment; cerebral palsy with Gross Motor Function Classification System (GMFCS)[23] level 2 or higher, at age 18 to 24 months corrected for prematurity; or deafness requiring hearing aids.

[b] The maximum number of infants available for major disability assessment at 18 to 24 months was 3971 because 902 infants were known to have died prior to the age of 18 to 24 months. There were an additional 92 infants with unknown death status at this time point who could not be assessed for major disability outcomes.

and April 2016. Data were checked for accuracy with published reports, trial protocols, and data collection sheets. Inconsistencies were discussed with individual investigators and discrepancies were resolved by consensus. Each trial verified its own finalized data set prior to inclusion in the study database. Data from the 5 included trials were collected and synthesized centrally after publication of the main results from each trial.

## Key Outcome Definitions

The primary outcome was a composite of death or major disability at a corrected age of 18 to 24 months. Major disability comprised any of the following: Bayley Scales of Infant and Toddler Development version 3 (Bayley-III)[23] cognitive or language score of less than 85; severe visual loss (cannot fixate or is legally blind with visual acuity <6/60 in both eyes); cerebral palsy with the Gross Motor Function Classification System level 2 or higher[24]; or deafness requiring hearing aids. When a Bayley-III assessment was unavailable, some trials used alternative sources of information for classifying cognitive delay such as a Bayley-II Mental Developmental Index score of less than 70 or another validated assessment tool (eg, Griffiths test), a pediatric assessment, or a parent-reported measure of neurodevelopmental impairment (eg, able to speak <5-10 words). To assess the statistical effects of inclusion of these alternate measures of disability, a prespecified supportive analysis of the primary outcome also was undertaken (**Figure 1** and Supplement 3).

Secondary outcomes were the components of the primary outcome (death prior to corrected age of 24 months and major disability); death prior to postmenstrual age of 36 weeks; death prior to hospital discharge; the individual components of the major disability outcome (developmental delay, severe visual impairment, deafness, cerebral palsy); ROP treated by laser photocoagulation, cryotherapy, or anti-vascular endothelial growth factor injection in 1 or both eyes; severe necrotizing enterocolitis leading to abdominal surgery or death; oxygen treatment at postmenstrual age of 36 weeks; postmenstrual age when each of the following respiratory support measures ceased: endotracheal intubation, continuous positive airway pressure, oxygen treatment, or home oxygen (if received); patent ductus arteriosus (PDA) diagnosed by ultrasound and receiving any treatment; PDA receiving surgical treatment; *z* scores for infant body weight at postmenstrual age of 36 weeks, at hospital discharge, and at corrected age of 18 to 24 months; 1 or more readmissions to the hospital by corrected age of 18 to 24 months; and time to death.

### Assessing the Risk of Bias

The 5 trials were assessed for risk of bias using the Cochrane Collaboration domains[25] and consensus was reached via discussion with the full study group.

### Statistical Analysis

The preplanned total sample size was 5230 infants. Because 2 of the trials were stopped early, a meta-analysis of individual participant data was undertaken of the 4965 infants recruited overall, which provided approximately 80% power (with a 2-sided *P* value of .05) to detect a minimum absolute risk difference of 4% in the primary composite outcome of death or major disability by a corrected age of 18 to 24 months, corresponding to a minimally important number needed to treat of 25 infants to prevent 1 major adverse outcome.[14] This minimal difference was derived via discussion with clinical experts, and no formal assessments were undertaken.

The analysis was performed on an intention-to-treat basis using all data from each trial included in a single model. The $I^2$ statistic[26] was used to assess heterogeneity for all primary and secondary outcomes. No statistical methods were used to deal with the small proportion of missing data, but sensitivity analyses were undertaken for the primary outcome by using alternative measures of disability when Bayley-III outcomes were missing. Binary end points were analyzed using log binomial regression in a generalized estimating equations model with an exchangeable correlation structure to account for multiple births. Models were adjusted for trial as a fixed effect because the methods used for the prospective meta-analysis meant all 5 trials were very similar with respect to their included participants, interventions, and outcome definitions. Sensitivity analyses using random-effects models also were undertaken.

The results are presented as risk differences and relative risks (RRs) with 95% CIs and 2-sided *P* values. If these models failed to converge, Poisson models with a robust variance estimator were used. Continuous outcomes were analyzed using linear regression in models for generalized estimating equations and presented as mean differences. Time to death was assessed between treatment groups using proportional hazard models and displayed using Kaplan-Meier survival curves.[27]

Relative risks and hazard ratios were computed such that values greater than 1 favored the higher target group. Subgroup analyses for gestational age (<26 weeks vs ≥26 weeks), inborn (indicates infant was born in the treating center) or outborn, use of any antenatal corticosteroids, sex, small for gestational age (SGA; <10th percentile using either the prespecified charts from Kramer et al[28] or the post hoc curves from Alexander et al[29] as in the SUPPORT trial[30]), multiple birth, type of delivery (vaginal or cesarean), time of intervention commencement (<6 hours vs ≥6 hours after birth), and type of oximeter software (original vs revised) were prespecified and performed for primary and secondary outcomes by including a treatment × subgroup interaction term in the model.

Two-sided *P* values of less than .05 were considered to indicate statistical significance, with no adjustment for multiple comparisons. Therefore, because of the potential for type I error, the prespecified secondary outcomes and the subgroup analyses should be considered exploratory. The statistical analyses were performed using SAS version 9.3 (SAS Institute Inc).

## Results

### Study Identification and Selection

Characteristics of the 5 studies appear in Supplement 1 and in eTable 1 in Supplement 3. Individual participant data from 4965 infants (2480 randomized to the lower and 2485 to the higher $Spo_2$ target range), with a median gestational age of 26 weeks (interquartile range, 25-27 weeks) and a mean birthweight of 832 g (SD, 190 g) were included in the meta-analysis. Baseline characteristics of each of the included trials and the combined data appear in the **Table**. Data were available for 90% of infants for the protocol-defined primary outcome and for 96% of infants for the prespecified supportive analysis of the primary outcome, which used alternate measures of cognitive disability (Figure 1).

### Primary Outcomes

There was no significant difference between a lower $Spo_2$ target range (85%-89%) compared with a higher $Spo_2$ target range (91%-95%) on the primary composite outcome of death or major disability at a corrected age of 18 to 24 months (53.5% with a lower $Spo_2$ target vs 51.6% with a higher $Spo_2$ target; risk difference, 1.7% [95% CI, −1.3% to 4.6%]; RR, 1.04 [95% CI, 0.98 to 1.09]; *P* = .21, $I^2$ = 14%; **Figure 2**). A supportive analysis of the primary outcome, which included alternate measures of disability, also showed no significant between-group difference in the rate of death or major disability (51.2% with a lower $Spo_2$ target vs 49.3% with a higher $Spo_2$ target; risk difference, 1.7% [95% CI, −1.2% to 4.5%]; RR, 1.04 [95% CI, 0.98 to 1.09]; *P* = .20, $I^2$ = 27%; Figure 2).

### Secondary Outcomes

Of the 16 secondary outcomes, 11 were null, 2 significantly favored a lower $Spo_2$ target, and 3 significantly favored a higher $Spo_2$ target. An analysis of each component of the primary outcome (Figure 2) showed that the lower $Spo_2$ target range was associated with a significantly increased incidence of death at a corrected age of 18 to 24 months (19.9% with a lower $Spo_2$ target vs 17.1% with a higher $Spo_2$ target; risk difference, 2.8% [95% CI, 0.6% to 5.0%]; RR, 1.17 [95% CI, 1.04 to 1.31]; *P* = .01, $I^2$ = 0%), but not major disability or the components of major disability. The survival analysis also showed a significant increase in risk of death by a corrected age of 18 to 24 months for the lower target group (hazard ratio, 1.17 [95% CI, 1.03 to 1.34]; *P* = .02; eTable 2 and eFigure 2 in Supplement 3).

Other secondary outcome results appear in **Figure 3**. These results show infants in the lower target group had an increase in death at other time points (postmenstrual age of 36 weeks and at hospital discharge), severe necrotizing enterocolitis (9.2% with a lower $Spo_2$ target vs 6.9% with a higher $Spo_2$ target; risk difference, 2.3% [95% CI, 0.8% to 3.8%]; RR, 1.33 [95% CI, 1.10 to 1.61]; *P* = .003), and PDA treated with surgical ligation, but a lower rate of ROP treatment (10.9% with a lower $Spo_2$

Table. Baseline Characteristics[a]

| | SUPPORT[15,16] (n = 1316) | COT[17] (n = 1201) | BOOST NZ[18] (n = 340) | BOOST II UK[19,20] (n = 973) | BOOST II AUS[19,20] (n = 1135) | $Spo_2$ Target | |
|---|---|---|---|---|---|---|---|
| | | | | | | Lower (n = 2480) | Higher (n = 2485) |
| **Mothers at Birth** | | | | | | | |
| Use of antenatal corticosteroids, No. (%) | | | | | | | |
| None | 50 (3.8) | 131 (10.9) | 38 (11.2) | 88 (9.0) | 106 (9.3) | 215 (8.7) | 198 (8.0) |
| Partial course[b] | 326 (24.8) | 259 (21.6) | 89 (26.2) | 272 (28.0) | 293 (25.8) | 609 (24.6) | 630 (25.4) |
| Full course | 939 (71.4) | 807 (67.4) | 213 (62.6) | 607 (62.4) | 727 (64.1) | 1648 (66.5) | 1645 (66.3) |
| Type of delivery, No. (%) | | | | | | | |
| Normal vaginal | 433 (32.9) | 462 (38.6) | 149 (43.8) | 593 (61.1) | 511 (45.0) | 1064 (43.0) | 1084 (43.7) |
| Instrumental vaginal | 0 | 3 (0.3) | 5 (1.5) | 0 | 18 (1.6) | 10 (0.4) | 16 (0.6) |
| Cesarean | 883 (67.1) | 732 (61.2) | 186 (54.7) | 378 (38.9) | 600 (52.9) | 1400 (56.5) | 1379 (55.5) |
| **Infants at Birth** | | | | | | | |
| Birth weight, mean (SD), g | 830 (193) | 837 (193) | 879 (194) | 821 (185) | 825 (184) | 829 (187) | 836 (192) |
| Girls, No. (%) | 604 (45.9) | 546 (45.5) | 160 (47.1) | 456 (46.9) | 546 (48.1) | 1169 (47.1) | 1143 (46.0) |
| Gestational age, wk | | | | | | | |
| Median (IQR) | 26.3 (25.3-27.1) | 26.0 (25.0-27.0) | 26.2 (25.2-27.0) | 26.1 (25.0-27.1) | 26.1 (25.1-27.0) | 26.0 (25.0-27.0) | 26.0 (25.0-27.0) |
| <26, No. (%) | 565 (42.9) | 512 (42.6) | 144 (42.4) | 431 (44.3) | 481 (42.4) | 1063 (42.9) | 1070 (43.1) |
| ≥26, No. (%) | 751 (57.1) | 689 (57.4) | 196 (57.6) | 542 (55.7) | 654 (57.6) | 1417 (57.1) | 1415 (56.9) |
| Small for gestational age, No. (%) | | | | | | | |
| Defined by trial investigators[c] | 96 (7.3) | 105 (8.7) | 30 (8.8) | 147 (15.2) | 158 (13.9) | 267 (10.8) | 269 (10.8) |
| Defined by NeOProM[d] | 210 (16.0) | 105 (8.7) | 30 (8.8) | 113 (11.6) | 158 (13.9) | 302 (12.2) | 314 (12.6) |
| Apgar score at 5 min, median (IQR)[e] | 7 (6-8) | 7 (6-8) | 8 (6-9) | | 7 (6-8) | 7 (6-8) | 7 (6-8) |
| Admission temperature, mean (SD), °C | 36.2 (0.9) | 36.4 (0.9) | 36.4 (1.0) | 36.6 (0.9) | 36.0 (1.0) | 36.3 (1.0) | 36.3 (0.9) |
| Inborn, No. (%)[f] | 1316 (100) | 1105 (92.0) | 316 (92.9) | 854 (88.0) | 1049 (92.4) | 2327 (93.8) | 2313 (93.1) |
| Inspired oxygen concentration immediately prior to randomization, median (IQR), %[e,g] | | 21 (21-25) | 21 (21-25) | | 21 (21-24) | 21 (21-25) | 21 (21-25) |
| **Infants at Randomization** | | | | | | | |
| Oximeter calibration software, No. (%) | | | | | | | |
| Original | 1316 (100) | 564 (47.0) | 340 (100) | 228 (23.4) | 692 (61.0) | 1569 (63.3) | 1571 (63.2) |
| Revised | 0 | 563 (46.9) | 0 | 745 (76.6) | 443 (39.0) | 879 (35.4) | 872 (35.1) |
| Mixed | 0 | 74 (6.2) | 0 | 0 | 0 | 32 (1.3) | 42 (1.7) |
| Time intervention started <6 h, No. (%)[e] | 1283 (99.2) | 53 (4.4) | 56 (16.5) | | 119 (10.5) | 752 (38.0) | 759 (38.3) |
| Positive airway pressure, No. (%)[e] | | | | | | | |
| With endotracheal tube[h] | 835 (63.9) | 925 (77.0) | 230 (67.6) | | 714 (63.0) | 1337 (67.3) | 1367 (68.5) |
| Without endotracheal tube[i] | 449 (34.4) | 242 (20.1) | 109 (32.1) | | 410 (36.2) | 621 (31.3) | 589 (29.5) |
| Oxygen treatment without positive airway pressure, No. (%)[e] | 11 (0.8) | 3 (0.2) | 0 | | 1 (0.1) | 9 (0.5) | 6 (0.3) |
| No respiratory support, No. (%)[e] | 12 (0.9) | 31 (2.6) | 1 (0.3) | | 9 (0.8) | 20 (1.0) | 33 (1.7) |

Abbreviations: AUS, Australia; BOOST, Benefits Of Oxygen Saturation Targeting; COT, Canadian Oxygen Trial; IQR, interquartile range; NZ, New Zealand; $Spo_2$, oxygen saturation as measured by pulse oximetry; SUPPORT, Surfactant, Positive Pressure and Pulse Oximetry Randomized Trial; UK, United Kingdom.

[a] Denominators include the total number of infants with a known outcome.

[b] Mother did not receive the full 2 doses within 48 hours before birth.

[c] Defined using trial-specific definitions.

[d] Defined as less than the 10th percentile using charts from Kramer et al.[28]

[e] Data were not available from BOOST II UK for this variable.

[f] Indicates infant was born in the treating center.

[g] Data were not available from SUPPORT for this variable.

[h] Includes all forms of positive pressure ventilation.

[i] Includes all other forms of respiratory support including continuous positive airway pressure and nasal cannula oxygen (high or low flow).

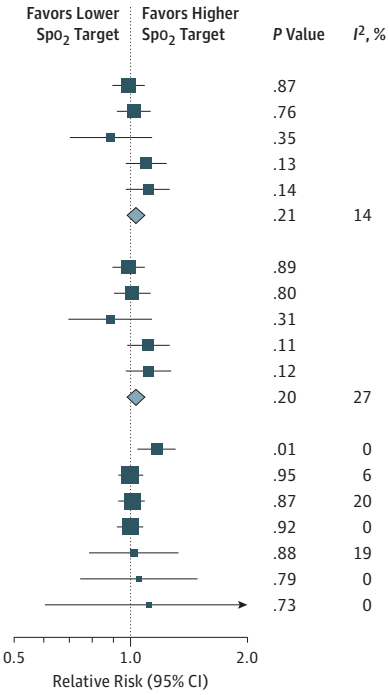
target vs 14.9% with a higher $Spo_2$ target; risk difference, −4.0% [95% CI, −6.1% to 2.0%]; RR, 0.74 [95% CI, 0.63 to 0.86], $P < .003$) and oxygen treatment at a postmenstrual age of 36 weeks. There were no significant between-group differences for other secondary outcomes (Figure 2).

## Subgroup Analyses

There were no between-group differences for the primary outcome (death or major disability) for any of the prespecified subgroup analysis factors (gestational age, outborn, use of any antenatal corticosteroids, sex, SGA, multiple pregnancy,

Figure 2. Effect of Oxygen Saturation as Measured by Pulse Oximetry ($Spo_2$) Target Levels on Composite Primary Outcome of Death or Major Disability

| Trial | No. of Infants With Event/Total No. (%) | | Risk Difference (95% CI), % | Relative Risk (95% CI) | P Value | $I^2$, % |
|---|---|---|---|---|---|---|
| | Lower $Spo_2$ Target | Higher $Spo_2$ Target | | | | |
| **Protocol-defined primary outcome**[a] | | | | | | |
| SUPPORT,[16] 2012 | 363/613 (59) | 374/624 (60) | -0.5 (-6.1 to 5.2) | 0.99 (0.90 to 1.09) | .87 | |
| COT,[17] 2013 | 298/577 (52) | 282/568 (50) | 0.8 (-4.9 to 6.6) | 1.02 (0.92 to 1.13) | .76 | |
| BOOST II in New Zealand,[18] 2014 | 62/143 (43) | 71/144 (49) | -5.4 (-17.0 to 6.2) | 0.89 (0.70 to 1.14) | .35 | |
| BOOST II in United Kingdom,[20] 2016 | 231/388 (60) | 211/385 (55) | 5.4 (-1.7 to 12.4) | 1.10 (0.97 to 1.24) | .13 | |
| BOOST II in Australia,[20] 2016 | 237/507 (47) | 212/508 (42) | 4.4 (-1.7 to 10.5) | 1.11 (0.97 to 1.26) | .14 | |
| Overall | 1191/2228 (54) | 1150/2229 (52) | 1.7 (-1.3 to 4.6) | 1.04 (0.98 to 1.09) | .21 | 14 |
| **Supportive analysis of primary outcome**[b] | | | | | | |
| SUPPORT,[16] 2012 | 364/614 (59) | 374/624 (60) | -0.4 (-6.0 to 5.2) | 0.99 (0.90 to 1.09) | .89 | |
| COT,[17] 2013 | 298/578 (52) | 283/569 (50) | 0.7 (-5.0 to 6.4) | 1.01 (0.91 to 1.13) | .80 | |
| BOOST II in New Zealand,[18] 2014 | 65/167 (39) | 75/168 (45) | -4.9 (-15.5 to 5.7) | 0.88 (0.69 to 1.13) | .31 | |
| BOOST II in United Kingdom,[20] 2016 | 245/473 (52) | 220/468 (47) | 5.1 (-1.3 to 11.6) | 1.11 (0.98 to 1.26) | .11 | |
| BOOST II in Australia,[20] 2016 | 246/545 (45) | 217/540 (40) | 4.5 (-1.3 to 10.4) | 1.11 (0.97 to 1.27) | .12 | |
| Overall | 1218/2377 (51) | 1169/2369 (49) | 1.7 (-1.2 to 4.5) | 1.04 (0.98 to 1.09) | .20 | 27 |
| **Components of primary outcome** | | | | | | |
| Death by corrected age of 18-24 mo | 484/2433 (20) | 418/2440 (17) | 2.8 (0.6 to 5.0) | 1.17 (1.04 to 1.31) | .01 | 0 |
| Primary major disability[c] | 707/1744 (41) | 732/1811 (40) | 0.03 (-3.2 to 3.3) | 1.00 (0.93 to 1.08) | .95 | 6 |
| Supportive major disability[d] | 734/1893 (39) | 751/1951 (39) | 0.2 (-2.9 to 3.2) | 1.01 (0.93 to 1.09) | .87 | 20 |
| Bayley-III score <85[e] | 647/1713 (38) | 672/1782 (38) | -0.2 (-3.4 to 3.0) | 1.00 (0.92 to 1.08) | .92 | 0 |
| Cerebral palsy[f] | 106/1910 (6) | 107/1966 (5) | 0.1 (-1.4 to 1.5) | 1.02 (0.78 to 1.33) | .88 | 19 |
| Deafness[g] | 60/1905 (3) | 60/1959 (3) | 0.2 (-0.9 to 1.3) | 1.05 (0.74 to 1.49) | .79 | 0 |
| Severe visual impairment[h] | 25/1910 (1) | 23/1967 (1) | 0.1 (-0.6 to 0.8) | 1.12 (0.60 to 2.08) | .73 | 0 |




Box sizes correspond to precision; therefore, the more precise the larger the box. Precision was ascertained by calculating the inverse of the variance for each estimate.

[a] Defined as a composite outcome of death or major disability by the age of 18 to 24 months, which was corrected for prematurity and prespecified in the published Neonatal Oxygenation Prospective Meta-analysis protocol (Supplement 1).

[b] Included using alternative sources of information for classifying major disability as used within individual trials. This may have included a Bayley-II major disability score of less than 70, another validated assessment tool (eg, the Griffiths test), a pediatrician assessment, or parent-reported measure of neurodevelopmental impairment (eg, able to speak <5-10 words), or other measures.

[c] Defined per protocol.

[d] Defined using supplementary data as noted in the "b" footnote.

[e] Developmental assessment for cognition or language.

[f] Defined by Gross Motor Function Classification System[23] level 2 or greater (higher levels = functioning more impaired) or cerebral palsy diagnosed but score unknown.

[g] Requiring hearing aids or worse.

[h] Defined by the trial investigators.

type of delivery, time intervention started, or oximeter software type; **Figure 4**). The number of prespecified subgroup analyses of secondary outcomes performed was large (n = 319; of which 17 [5%] were nominally significant), and the interaction *P* values were not formally adjusted for multiple subgroup comparisons and are thus considered exploratory.[31]

Subgroup analyses by oximeter software type (**Figure 5**) showed a significant difference in death by corrected age of 18 to 24 months for the original software (RR, 1.06 [95% CI, 0.91 to 1.23]; *P* = .47) vs the revised software (RR, 1.38 [95% CI, 1.14 to 1.68]; *P* = .001; *P* = .03 for interaction subgroup difference). A similar result was seen for death both before hospital discharge and before a postmenstrual age of 36 weeks.

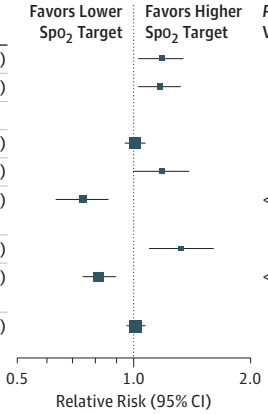
Other subgroup analyses of secondary outcomes appear in eTables 3-32 in Supplement 3. Even though there were differences in the subgroups for some of the outcomes using bivariable analyses, there was no overall pattern indicating that any particular subgroup of infants benefited more or less from the lower vs the higher $Spo_2$ target.

There was no significant difference in the association with the lower $Spo_2$ target for death at a corrected age of 18 to 24 months by known risk factors such as early gestational age, SGA, male sex, or infants born outside a tertiary center (eTables 15 and 33 in Supplement 3). The association with the lower oxygen target for severe necrotizing enterocolitis was greater for inborn infants and singletons (eTable 26 in Supplement 3).

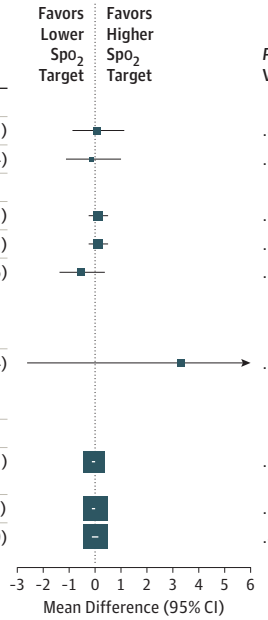
For the outcome of ROP treatment, the association with the lower $Spo_2$ target was larger among infants starting the intervention at an age of less than 6 hours (largely driven by SUPPORT results) and for those born via cesarean section (eTable 27 in Supplement 3). There was no difference in the association with the lower $Spo_2$ target for PDA among infants treated surgically for any of the prespecified subgroup vari-

Figure 3. Effect of Oxygen Saturation as Measured by Pulse Oximetry ($Spo_2$) Target Levels on Secondary Outcomes

| Dichotomous Outcomes | No. of Infants With Event/Total No. (%)[a] Lower $Spo_2$ Target | Higher $Spo_2$ Target | Risk Difference (95% CI), % | Relative Risk (95% CI) | P Value | $I^2$, % |
|---|---|---|---|---|---|---|
| Death before postmenstrual age of 36 wk | 415/2478 (17) | 354/2481 (14) | 2.5 (0.5 to 4.5) | 1.18 (1.03 to 1.34) | .01 | 0 |
| Death before discharge from hospital | 460/2478 (19) | 397/2481 (16) | 2.6 (0.5 to 4.7) | 1.17 (1.03 to 1.32) | .01 | 0 |
| Patent ductus arteriosus[b] | | | | | | |
| Treated medically or surgically | 1139/2456 (46) | 1127/2463 (46) | 0.5 (-2.3 to 3.3) | 1.01 (0.95 to 1.07) | .71 | 0 |
| Treated surgically | 281/2462 (11) | 240/2464 (10) | 1.7 (0 to 3.4) | 1.18 (1.00 to 1.39) | .046 | 13 |
| Treated retinopathy of prematurity before corrected age of 18-24 mo | 220/2020 (11) | 308/2065 (15) | -4.0 (-6.1 to -2.0) | 0.74 (0.63 to 0.86) | <.001 | 80 |
| Severe necrotizing enterocolitis[c] | 227/2464 (9) | 170/2465 (7) | 2.3 (0.8 to 3.8) | 1.33 (1.10 to 1.61) | .003 | 0 |
| Supplemental oxygen at postmenstrual age of 36 wk | 459/1846 (25) | 578/1910 (30) | -5.6 (-8.5 to -2.7) | 0.81 (0.74 to 0.90) | <.001 | 0 |
| ≥1 Readmission to hospital | 942/1754 (54) | 967/1819 (53) | 0.6 (-2.6 to 3.9) | 1.01 (0.96 to 1.07) | .64 | 0 |




| Continuous Outcomes | Lower $Spo_2$ Target No. of Infants | Mean (95% CI) | Higher $Spo_2$ Target No. of Infants | Mean (95% CI) | Mean Difference (95% CI) | P Value | $I^2$,% |
|---|---|---|---|---|---|---|---|
| Bayley-III score at corrected age of 18-24 mo | | | | | | | |
| Cognition | 1675 | 94.70 (90.16 to 99.24) | 1743 | 94.47 (90.03 to 98.91) | 0.09 (-0.90 to 1.07) | .86 | 13 |
| Language | 1633 | 90.33 (85.95 to 94.71) | 1699 | 90.20 (85.91 to 94.49) | -0.12 (-1.17 to 0.94) | .83 | 52 |
| Postmenstrual age when ceased use of positive airway pressure, wk | | | | | | | |
| With endotracheal tube | 1251 | 30.48 (28.79 to 32.17) | 1265 | 30.38 (28.71 to 32.05) | 0.08 (-0.31 to 0.47) | .68 | 0 |
| Without endotracheal tube | 1284 | 34.19 (32.32 to 36.06) | 1287 | 34.06 (32.20 to 35.92) | 0.11 (-0.31 to 0.53) | .61 | 0 |
| Postmenstrual age when ceased use of supplemental oxygen without positive airway pressure, wk | 1591 | 37.72 (35.87 to 39.57) | 1604 | 38.33 (36.45 to 40.21) | -0.54 (-1.43 to 0.36) | .24 | 0 |
| Postmenstrual age when ceased use of home oxygen, wk[d] | 262 | 71.92 (63.21 to 80.63) | 275 | 72.32 (63.77 to 80.87) | 3.33 (-2.67 to 9.34) | .28 | 0 |
| *z* score for infant body weight | | | | | | | |
| At postmenstrual age of 36 wk | 1847 | -12.75 (-12.83 to -12.67) | 1904 | -12.68 (-12.76 to -12.60) | -0.06 (-0.16 to 0.05) | .31 | 0 |
| At discharge from hospital | 1429 | -0.94 (-1.00 to -0.88) | 1475 | -0.88 (-0.94 to -0.82) | -0.06 (-0.15 to 0.02) | .13 | 0 |
| At corrected age of 18-24 mo | 1306 | -0.21 (-0.27 to -0.15) | 1383 | -0.23 (-0.30 to -0.17) | 0.01 (-0.08 to 0.09) | .88 | 0 |




Box sizes correspond to precision; therefore, the more precise the larger the box. Precision was ascertained by calculating the inverse of the variance for each estimate.

[a] Denominators include the total number of infants with a known outcome.

[b] Diagnosed by ultrasound during initial hospitalization.

[c] Treated with surgery or leading to death during initial hospitalization.

[d] Data on postmenstrual age when ceased use of home oxygen can only be calculated using the 537 infants who received home oxygen and for whom the postmenstrual age when ceased use is known.

ables (eTable 25 in Supplement 3). The association with a lower $Spo_2$ target at a postmenstrual age of 36 weeks was greater among SGA infants (eTable 30 in Supplement 3).

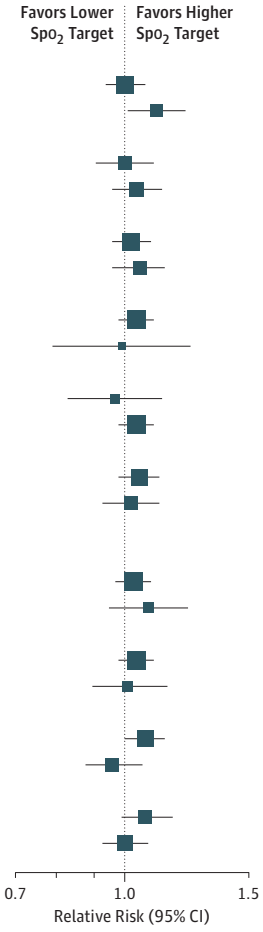
## Sensitivity Analyses and Assessments of Bias and Heterogeneity

Sensitivity analyses exploring variations in the definition of the primary outcome (Figure 2) including a Bayley-III cognitive or language score of less than 70 or by other definition variations used by the individual trials did not change the primary outcome findings. Using a random-effects model (rather than a fixed-effect model) gave the same conclusions for all outcomes with the exception of PDA treated with surgical ligation, which became nonsignificant (eTable 34 in Supplement 3).

Overall, the 5 trials were assessed as being at low risk of bias for all domains[7] (selection, performance or detection, attrition, and reporting biases) and had low levels of statistical heterogeneity for most outcomes. The outcome of ROP treatment had a high level of heterogeneity ($I^2$ = 80%), which resulted from the substantially larger treatment effect of the lower $Spo_2$ target on this outcome in the SUPPORT trial.

Figure 4. Subgroup Analyses of Primary Outcome Composite of Death or Major Disability

| Subgroup | No. of Infants With Event/Total No. (%)[a] Lower $Spo_2$ Target | Higher $Spo_2$ Target | Relative Risk (95% CI) | *P* Value | *P* Value for Interaction |
|---|---|---|---|---|---|
| Type of oximeter software[b] | | | | | |
| Original | 757/1423 (53) | 768/1443 (53) | 1.00 (0.94-1.07) | .95 | .09 |
| Revised | 419/774 (54) | 361/747 (48) | 1.11 (1.01-1.22) | .03 | |
| Age at intervention start, h | | | | | |
| <6 | 404/699 (58) | 411/711 (58) | 1.00 (0.91-1.10) | .99 | .48 |
| ≥6 | 548/1129 (49) | 516/1119 (46) | 1.04 (0.96-1.13) | .34 | |
| Gestational age at birth, wk | | | | | |
| <26 | 623/976 (64) | 610/977 (62) | 1.02 (0.96-1.09) | .51 | .63 |
| ≥26 | 568/1252 (45) | 540/1252 (43) | 1.05 (0.96-1.14) | .29 | |
| Birth location[c] | | | | | |
| Inborn | 1124/2099 (54) | 1074/2087 (52) | 1.04 (0.98-1.10) | .19 | .72 |
| Outborn | 67/129 (52) | 76/142 (54) | 0.99 (0.79-1.24) | .91 | |
| Antenatal corticosteroids | | | | | |
| No | 113/191 (59) | 111/175 (63) | 0.97 (0.83-1.13) | .71 | .35 |
| Yes | 1073/2031 (53) | 1034/2044 (51) | 1.04 (0.98-1.10) | .18 | |
| Sex | | | | | |
| Female | 729/1188 (61) | 697/1196 (58) | 1.05 (0.98-1.12) | .15 | .54 |
| Male | 462/1040 (44) | 453/1033 (44) | 1.02 (0.93-1.12) | .72 | |
| Small for gestational age | | | | | |
| Defined by trial | | | | | |
| No | 1031/1982 (52) | 999/1982 (50) | 1.03 (0.97-1.09) | .31 | .56 |
| Yes | 159/244 (65) | 148/244 (61) | 1.08 (0.95-1.23) | .22 | |
| Defined by NeOProM[d] | | | | | |
| No | 1010/1949 (52) | 966/1939 (50) | 1.04 (0.98-1.10) | .23 | .87 |
| Yes | 181/279 (65) | 184/290 (63) | 1.01 (0.90-1.15) | .84 | |
| Multiple birth | | | | | |
| Yes | 857/1599 (54) | 807/1614 (50) | 1.07 (1.00-1.14) | .05 | .08 |
| No | 334/629 (53) | 343/615 (56) | 0.96 (0.88-1.06) | .42 | |
| Type of delivery | | | | | |
| Vaginal | 521/956 (55) | 495/971 (51) | 1.07 (0.99-1.17) | .09 | .26 |
| Cesarean | 667/1269 (53) | 652/1254 (52) | 1.00 (0.93-1.08) | .93 | |




Box sizes correspond to precision; therefore, the more precise the larger the box. Precision was ascertained by calculating the inverse of the variance for each estimate. $Spo_2$ indicates oxygen saturation as measured by pulse oximetry.

[a] Denominators include the total number of infants with a known outcome.

[b] Excluded 74 infants in the Canadian Oxygen Trial who were exposed to both the original and revised software.

[c] Inborn defined as born inside the treating center; outborn, born outside the treating center (eg, transferred from another hospital).

[d] Less than 10th percentile using charts from Kramer et al.[28]

## Discussion

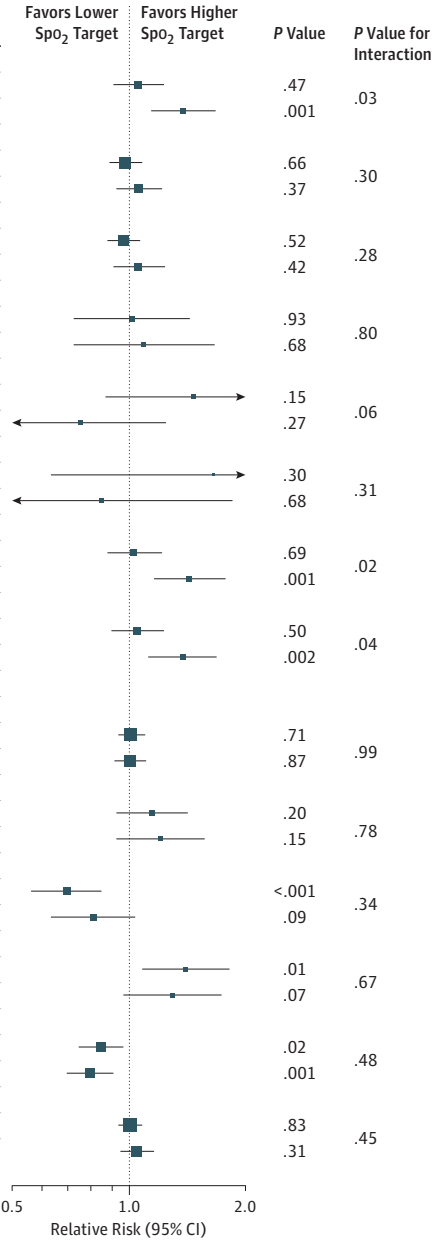
In this prospectively planned meta-analysis of individual participant data involving clinical trials of extremely preterm infants, there was no significant difference between a lower $Spo_2$ target range (85%-89%) and a higher $Spo_2$ target range (91%-95%) from soon after birth on the primary composite outcome of death or major disability at a corrected age of 18 to 24 months. However, the lower target range was associated with more deaths and cases of severe necrotizing enterocolitis and less treated ROP, but was not associated with blindness.

When evaluating outcomes within a clinical trial sample or synthesizing results from several trials in a meta-analysis, the effects associated with treatment represent averages, and the true benefits and harms may differ from those in these analyses. Furthermore, tests of associations between treatment and secondary, albeit prespecified and important, outcomes (including the individual components of the composite primary outcome), should be considered exploratory and the results interpreted with caution. In particular, the statistically significant increased risk of death would not remain significant if adjusted for multiple testing. However, death was a major component of the composite primary outcome, and a clear difference in death, in either direction, was used to assess the need for early stopping in 2 trials.[22] The current pooled estimated risk and 95% CIs for mortality from these trials thus provide the best currently available evidence to guide future clinical practice.

Prespecified subgroup analyses showed consistent results across trials for most outcomes, except for a larger association on treated ROP within the SUPPORT trial. The reasons for this result in the SUPPORT trial need to be explored more fully. One possible explanation for the heterogeneity is that

Figure 5. Subgroup Analysis by Oximeter Software Type

| Outcome | No. of Infants With Event/Total No. (%)[a] Lower $Spo_2$ Target | Higher $Spo_2$ Target | Relative Risk (95% CI) | *P* Value | *P* Value for Interaction |
|---|---|---|---|---|---|
| Death prior to corrected age of 18-24 mo | | | | | |
| Original software | 292/1542 (19) | 279/1545 (18) | 1.06 (0.91-1.23) | .47 | .03 |
| Revised software | 190/860 (22) | 137/856 (16) | 1.38 (1.14-1.68) | .001 | |
| Major disability at corrected age of 18-24 mo | | | | | |
| Original software | 465/1131 (41) | 489/1164 (42) | 0.98 (0.89-1.08) | .66 | .30 |
| Revised software | 229/584 (39) | 224/610 (37) | 1.06 (0.93-1.22) | .37 | |
| Bayley-III score <85 at corrected age of 18-24 mo[b] | | | | | |
| Original software | 435/1115 (39) | 464/1154 (40) | 0.97 (0.88-1.07) | .52 | .28 |
| Revised software | 199/569 (35) | 191/591 (32) | 1.06 (0.91-1.24) | .42 | |
| Cerebral palsy at corrected age of 18-24 mo[c] | | | | | |
| Original software | 63/1220 (5) | 62/1237 (5) | 1.02 (0.72-1.44) | .93 | .80 |
| Revised software | 42/661 (6) | 41/692 (6) | 1.09 (0.72-1.67) | .68 | |
| Deafness at corrected age of 18-24 mo[d] | | | | | |
| Original software | 36/1218 (3) | 24/1228 (2) | 1.47 (0.87-2.48) | .15 | .06 |
| Revised software | 24/658 (4) | 35/694 (5) | 0.75 (0.45-1.25) | .27 | |
| Severe visual impairment at corrected age of 18-24 mo[e] | | | | | |
| Original software | 14/1221 (1) | 9/1234 (1) | 1.65 (0.63-4.32) | .30 | .31 |
| Revised software | 11/660 (2) | 14/696 (2) | 0.85 (0.39-1.85) | .68 | |
| Death prior to postmenstrual age of 36 wk | | | | | |
| Original software | 244/1569 (16) | 237/1570 (15) | 1.03 (0.88-1.22) | .69 | .02 |
| Revised software | 169/877 (19) | 117/869 (14) | 1.43 (1.16-1.78) | .001 | |
| Death prior to discharge | | | | | |
| Original software | 276/1569 (18) | 264/1570 (17) | 1.05 (0.90-1.23) | .50 | .04 |
| Revised software | 182/877 (21) | 131/869 (15) | 1.38 (1.12-1.69) | .002 | |
| Patent ductus arteriosus during initial hospitalization[f] | | | | | |
| Treated medically or surgically | | | | | |
| Original software | 698/1550 (45) | 693/1552 (45) | 1.01 (0.94-1.10) | .71 | .99 |
| Revised software | 424/874 (49) | 408/869 (47) | 1.01 (0.92-1.11) | .87 | |
| Treated surgically | | | | | |
| Original software | 165/1556 (11) | 143/1554 (9) | 1.15 (0.93-1.42) | .20 | .78 |
| Revised software | 108/874 (12) | 89/868 (10) | 1.21 (0.93-1.57) | .15 | |
| Retinopathy of prematurity treated before corrected age of 18-24 mo | | | | | |
| Original software | 129/1275 (10) | 188/1279 (15) | 0.69 (0.56-0.85) | <.001 | .34 |
| Revised software | 86/715 (12) | 112/744 (15) | 0.81 (0.63-1.04) | .09 | |
| Necrotizing enterocolitis during initial hospitalization[g] | | | | | |
| Original software | 128/1556 (8) | 91/1557 (6) | 1.40 (1.08-1.82) | .01 | .67 |
| Revised software | 96/876 (11) | 73/866 (8) | 1.30 (0.97-1.74) | .07 | |
| Supplemental oxygen at postmenstrual age of 36 wk | | | | | |
| Original software | 254/1182 (22) | 300/1197 (25) | 0.85 (0.74-0.97) | .02 | .48 |
| Revised software | 195/640 (31) | 256/683 (38) | 0.79 (0.69-0.91) | .001 | |
| ≥1 Readmission to hospital before corrected age of 18-24 mo | | | | | |
| Original software | 618/1156 (54) | 637/1191 (54) | 1.01 (0.94-1.08) | .83 | .45 |
| Revised software | 318/569 (56) | 309/591 (52) | 1.05 (0.95-1.16) | .31 | |




Box sizes correspond to precision; therefore, the more precise the larger the box. Precision was ascertained by calculating the inverse of the variance for each estimate. $Spo_2$ indicates oxygen saturation as measured by pulse oximetry. This subgroup analysis excludes 74 infants in the Canadian Oxygen Trial who were exposed to both the original and revised software.

[a] Denominators include the total number of infants with a known outcome.

[b] Developmental assessment cognitive or language score of less than 85.

[c] Defined by Gross Motor Function Classification System[23] level 2 or greater (higher levels = functioning more impaired) or cerebral palsy diagnosed but score unknown.

[d] Requiring hearing aids or worse.

[e] Defined by the trial investigators.

[f] Diagnosed by ultrasound.

[g] Treated with surgery or leading to death during initial hospitalization.

most infants in the SUPPORT trial were randomized before birth; however, this hypothesis cannot be explored reliably in the other trials because they included too few infants recruited early.[32]

Mortality was increased in the lower $Spo_2$ target group overall, in the first reported trial that used the original software exclusively,[15] and in the prespecified subgroup analysis of original vs revised oximeter software. There has

been considerable debate among the study investigators whether the change in oximeter software was responsible for this result.[22,33-36]

A subgroup analysis undertaken by the SUPPORT trial investigators found that, in their trial, mortality in the lower $Spo_2$ target group was greater for SGA infants.[30] A prespecified subgroup analysis using a common definition of SGA[28] across the combined data set, and a post hoc analysis on the full data set using the same definition of SGA as used in the SUPPORT trial (curves by Alexander et al)[29,30] did not confirm this relationship.

The main strength of this meta-analysis is that the 5 trials were planned prospectively to be similar in design and their investigators agreed to undertake a combined pooled meta-analysis of individual participant data based on a protocol developed in advance of any trial results.[37,38] The statistical analysis plan was finalized after the trial results were known, but before any central receipt or synthesis of data. As would be expected with this study design, heterogeneity across the trials for most outcomes was low.

A previous Cochrane review[7] had synthesized the aggregate data available from the published reports of the 5 trials. In contrast, these results were derived using raw individual participant data sourced directly from the trial investigators and combined centrally, making this the most comprehensive and rigorous analyses available of these data. The methods of the analyses used for the individual participant data also permitted adjustment for the correlation of multiple births; standardization of important outcomes across trials, including the definition of major disability; and enabled testing of the effect of differences in outcome definitions via sensitivity analyses. Even though the main findings are similar to some of the Cochrane Review results, the current meta-analysis of individual participant data has provided new insights into the consistency of results across multiple subgroups that indicate the findings should not be restricted to certain groups of infants such as those born SGA or at very early gestational ages. The 2016 guidelines from the American Academy of Pediatrics noted that their recommendations at that time were made "pending additional data, including the individual patient meta-analysis (NeOProM)."[39] Thus these new findings should help inform these ongoing debates.

Implications for future research may include investigations of the effects of differences in alarm limits and targeting compliance[40] and in the level of exposure to the intervention on outcomes; measures of $Spo_2$ achieved, the proportion of time spent at various $Spo_2$ levels on outcomes (eg, via prediction models adjusted for potential confounders), or both; the oximeter software change on mortality (eg, further explanation of why a larger association was seen in this subgroup); and, using automated methods to match the relatively narrow target ranges required.

### Limitations

This study has several limitations. First, all 5 trials reported less separation in oxygen exposure between treatment groups than anticipated, largely because the lower $Spo_2$ target groups had higher than intended saturation levels.[17] Second, 2 trials (BOOST II in United Kingdom and Australia) were stopped early, which may have resulted in some overestimation of the effect on mortality in these trials.[41] However, excluding truncated studies from meta-analyses can lead to substantial bias due to underestimation of overall treatment effects.[42] Therefore, the best estimate of the association with treatment remains the overall combined results from the 5 trials.

Third, the lack of an association of $Spo_2$ target range on blindness, but with a clear difference on ROP by treatment group, may change with longer follow-up, when less severe visual impairments may become apparent. Fourth, the potential for false-positive results based on multiple comparisons from 16 secondary outcomes and hundreds of subgroup analyses means that individual comparisons, although nominally significant, should be considered exploratory and interpreted cautiously. Fifth, even though the results are generalizable across the 5 trials, caution should be exercised not to extend these findings to other settings that do not have early screening for ROP, appropriate ROP treatment, or skilled nursing care regarding alarm limits. The trials studied $Spo_2$ target ranges, not oximeter alarm limits, and these 2 concepts are not interchangeable.

## Conclusions

In this prospectively planned meta-analysis of individual participant data from extremely preterm infants, there was no significant difference between a lower $Spo_2$ target range compared with a higher $Spo_2$ target range on the primary composite outcome of death or major disability at a corrected age of 18 to 24 months. The lower $Spo_2$ target range was associated with a higher risk of death and necrotizing enterocolitis, but a lower risk of retinopathy of prematurity treatment.

ARTICLE INFORMATION

**Accepted for Publication:** April 30, 2018.

**Author Affiliations:** National Health and Medical Research Council Clinical Trials Centre, University of Sydney, Sydney, Australia (Askie, Tarnow-Mordi, Davies, Gebski, Hunter, Simes); Department of Paediatrics, University of Otago, Christchurch, New Zealand (Darlow); Department of Pediatrics, University of California, San Diego (Finer, Rich); Division of Neonatology, University of Pennsylvania, Philadelphia (Schmidt); Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada (Schmidt, Roberts, Costantini); Department of Neonatology, Royal Infirmary of Edinburgh, Edinburgh, Scotland (Stenson); Newborn Research, Royal Women's Hospital, Departments of Obstetrics and Gynaecology, and Paediatrics, University of Melbourne, Melbourne, Australia (Davis, Doyle); Clinical Sciences, Murdoch Children's Research Institute, Melbourne, Australia (Davis, Doyle); Department of Pediatrics, University of Alabama, Birmingham (Carlo); Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, England (Brocklehurst); National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, England (Brocklehurst, King, Juszczak); Statistics and Epidemiology Unit, RTI International, Rockville, Maryland (Das); Statistics and Epidemiology Unit, RTI International, Research Triangle Park, North Carolina (Gantz); Department of Pediatrics, Dalhousie University, Halifax, Nova Scotia, Canada (Whyte); Department of Neonatology, Tuebingen University Hospital, Tuebingen, Germany (Poets); Department of Paediatrics, University of Toronto, Toronto, Ontario, Canada (Asztalos); Newborn Services, Auckland City Hospital, Auckland,

New Zealand (Battin); Royal Maternity Hospital, Belfast, Ireland (Halliday); Department of Child Health, Queen's University, Belfast, Ireland (Halliday); EGA Institute for Women's Health, University College London, London, England (Marlow); Department of Neonatal Medicine, James Cook University, Middlesbrough, England (Tin); University of Cambridge, Department of Obstetrics and Gynaecology, Cambridge, England (Morley).

**Author Contributions:** Dr Askie and Ms Davies had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
*Concept and design:* All authors.
*Acquisition, analysis, or interpretation of data:* All authors.
*Drafting of the manuscript:* Askie, Darlow, Schmidt, Stenson, Davis, Carlo, Davies, Simes.
*Critical revision of the manuscript for important intellectual content:* All authors.
*Statistical analysis:* Askie, Davies, Gebski.
*Obtained funding:* Askie, Darlow, Schmidt, Tarnow-Mordi, Carlo, Brocklehurst, Gebski, Simes.
*Administrative, technical, or material support:* All authors.
*Supervision:* Askie, Gebski, Simes.

**Conflict of Interest Disclosures:** The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Askie reported receiving honoraria from the Nemours Foundation (Hot Topics in Neonatology). Dr Schmidt reported receiving honoraria from several US academic institutions, the Nemours Foundation (Hot Topics in Neonatology), and the Vermont Oxford Network for lectures on the topic of oxygen saturation targeting in extremely preterm infants. Dr Tarnow-Mordi reported receiving honoraria from the Nemours Foundation (Hot Topics in Neonatology) and the Vermont Oxford Network for speaking on topics related to the care of premature infants. Dr Davis reported receiving a fellowship from the Australian National Health and Medical Research Council. Dr Brocklehurst reported receiving personal fees from the UK Medical Research Council. Mr Rich reported receiving personal fees from Windtree Therapeutics Inc. Dr Poets reported receiving honoraria from Chiesi Farmaceutici. Dr Halliday reported being an advisor to Chiesi Farmaceutici and the joint editor-in-chief of *Neonatology*. Dr Marlow reported serving as a consultant to Shire. No other disclosures were reported.

**Funding/Support:** The data analysis was supported by grant R03HD 079867 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Department of Health and Human Services. Support for staff of the National Health and Medical Research Council (NHMRC) Clinical Trials Centre (University of Sydney, Sydney, Australia) was partly funded by NHMRC program grant 1037786. Dr Marlow receives funding from the UK Department of Health's National Institute for Health Research Biomedical Research Centre's funding scheme at University College London Hospital and the University College London. The SUPPORT trial was supported by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, and the National Heart, Lung, and Blood Institute, and from the National Institutes of Health (U10 HD21364, U10 HD21373, U10 HD21385, U10 HD21397, U10 HD27851, U10 HD27853, U10 HD27856, U10 HD27880, U10 HD27871, U10 HD27904, U10 HD34216, U10 HD36790, U10 HD40461, U10 HD40492, U10 HD40498, U10 HD40521, U10 HD40689, U10 HD53089, U10 HD53109, U10 HD53119, and U10 HD53124); and by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, cofunding from the National Heart, Lung, and Blood Institute, and grants from the National Institutes of Health (M01 RR30, M01 RR32, M01 RR39, M01 RR44, M01 RR54, M01 RR59, M01 RR64, M01 RR70, M01 RR80, M01 RR125, M01 RR633, M01 RR750, M01 RR997, M01 RR6022, M01 RR7122, M01 RR8084, M01 RR16587, UL1 RR25008, UL1 RR24139, UL1 RR24979, and UL1 RR25744). The Canadian Oxygen Trial was funded by the Canadian Institutes of Health Research (MCT-79217). The BOOST New Zealand trial was funded by th New Zealand Health Research Council (05/145) and the Child Health Research Foundation. The BOOST II UK trial was funded by the UK Medical Research Council and managed by the National Institute for Health Research (NIHR) on behalf of the MRC-NIHR partnership. The BOOST II Australia trial was supported by Australian NHMRC project grant 352386.

**Role of the Funder/Sponsor:** None of the listed funders or the Masimo Corporation had a role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Group Information:** All authors are members of the Neonatal Oxygenation Prospective Meta-analysis (NeOProM) Collaboration.

**Additional Contributions:** We thank the many people who have contributed to this project (none of whom received financial compensation for their role in the study) including: William Silverman, MD (Columbia-Presbyterian Medical Center, New York, NY), Jack Sinclair, MD (McMaster University, Hamilton, Ontario, Canada), Edmund Hey, MD, DPhil (Princess Mary Maternity Hospital, Newcastle, England), and David Henderson-Smart, PhD (University of Sydney, Sydney, Australia), who are deceased. We also thank Cynthia Cole, MD, MPH (Boston University School of Medicine, Boston, Massachusetts), who was instrumental in forming the initial collaboration; Dale Phelps, MD (University of Rochester School of Medicine and Dentistry, Rochester, New York), for advice regarding the oximeter masking; and Marion Fournier, MSc, Adrienne Kirby, MSc, Mark Donoghoe, PhD, Luke Buizen BSc(Hons), Rebecca Asher, MSc, and Anna Lene Seidler MSc (all employees of the NHMRC Clinical Trials Centre, University of Sydney, Australia), for statistical advice and preparation of the manuscript figures and tables. We also acknowledge and thank the dedicated staff who worked on the included trials and the families of the enrolled infants for giving of their time.

## REFERENCES

1. Askie LM, Henderson-Smart DJ. Restricted versus liberal oxygen exposure for preventing morbidity and mortality in preterm or low birth weight infants. *Cochrane Database Syst Rev*. 2001;4(4):CD001077.

2. Hellström A, Smith LE, Dammann O. Retinopathy of prematurity. *Lancet*. 2013;382(9902):1445-1457.

3. Tin W, Milligan DW, Pennefather P, Hey E. Pulse oximetry, severe retinopathy, and outcome at one year in babies of less than 28 weeks gestation. *Arch Dis Child Fetal Neonatal Ed*. 2001;84(2):F106-F110.

4. STOP-ROP Investigators. Supplemental therapeutic oxygen for prethreshold retinopathy of prematurity (STOP-ROP), a randomized, controlled trial, I: primary outcomes. *Pediatrics*. 2000;105(2):295-310.

5. 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.

6. Cole CH, Wright KW, Tarnow-Mordi W, Phelps DL; Pulse Oximetry Saturation Trial for Prevention of Retinopathy of Prematurity Planning Study Group. Resolving our uncertainty about oxygen therapy. *Pediatrics*. 2003;112(6 pt 1):1415-1419.

7. Askie LM, Darlow BA, Davis PG, et al. Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants. *Cochrane Database Syst Rev*. 2017;4:CD011190.

8. Clinical Trials website. Surfactant Positive Airway Pressure and Pulse Oximetry Trial (SUPPORT). https://clinicaltrials.gov/ct2/show/NCT00233324. Accessed March 12, 2018.

9. Current Controlled Trials website. Efficacy and safety of targeting lower arterial oxygen saturations to reduce oxygen toxicity and oxidative stress in very preterm infants: the Canadian Oxygen Trial. http://www.isrctn.com/ISRCTN62491227. Accessed March 12, 2018.

10. Australian New Zealand Clinical Trials Registry website. A randomised phase III study to evaluate whether a lower versus a higher oxygen saturation target in infants of <28 weeks gestation is associated with a reduction in death or disability at 2 years of age. http://www.anzctr.org.au/ACTRN12605000253606.aspx. Accessed March 12, 2018.

11. Current Controlled Trials website. Which oxygen saturation level should we use for very premature infants? a randomised controlled trial. http://www.isrctn.com/ISRCTN00842661. Accessed March 12, 2018.

12. Australian New Zealand Clinical Trials Registry website. Which oxygen saturation level should we use for very premature infants? a randomised controlled trial to investigate the effect of two slightly different oxygen levels on the health of very premature infants. http://www.anzctr.org.au/ACTRN12605000055606.aspx. Accessed March 12, 2018.

13. Ghersi D, Berlin J, Askie L. Prospective meta-analysis. In: Cochrane Handbook for Systematic Reviews of Interventions: vol 5.1.0; 2011. http://www.handbook.cochrane.org. Accessed May 10, 2018.

14. 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(6):6.

**15**. 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.

**16**. 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.

**17**. 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.

**18**. Darlow BA, Marschner SL, Donoghoe M, et al; Benefits Of Oxygen Saturation Targeting-New Zealand (BOOST-NZ) Collaborative Group. Randomized controlled trial of oxygen saturation targets in very preterm infants: two year outcomes. *J Pediatr*. 2014;165(1):30-35.e2.

**19**. 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.

**20**. 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.

**21**. 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.

**22**. Stenson B, Brocklehurst P, Tarnow-Mordi W; UK BOOST II trial; Australian BOOST II trial; New Zealand BOOST II trial. Increased 36-week survival with high oxygen saturation target in extremely preterm infants. *N Engl J Med*. 2011;364(17):1680-1682.

**23**. Bayley N. *Bayley Scales of Infant and Toddler Development*. 3rd ed. San Antonio, TX: Harcourt Assessment Inc; 2006.

**24**. Palisano RJ, Hanna SE, Rosenbaum PL, et al. Validation of a model of gross motor function for children with cerebral palsy. *Phys Ther*. 2000;80(10):974-985.

**25**. Higgins JPT, Altman DG, Sterne JAC; Cochrane Statistical Methods Group; Cochrane Bias Methods Group. Assessing risk of bias in included studies. In: Cochrane Handbook for Systematic Reviews of Interventions: vol 5.1.0; 2011. http://www.handbook.cochrane.org. Accessed May 10, 2018.

**26**. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-560.

**27**. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-481.

**28**. Kramer MS, Platt RW, Wen SW, et al; Fetal/Infant Health Study Group of the Canadian Perinatal Surveillance System. A new and improved population-based Canadian reference for birth weight for gestational age. *Pediatrics*. 2001;108(2):E35.

**29**. Alexander GR, Himes JH, Kaufman RB, Mor J, Kogan M. A United States national reference for fetal growth. *Obstet Gynecol*. 1996;87(2):163-168.

**30**. 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.

**31**. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine—reporting of subgroup analyses in clinical trials. *N Engl J Med*. 2007;357(21):2189-2194.

**32**. Rich W, Finer NN, Gantz MG, et al; SUPPORT and Generic Database Subcommittees of the Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Enrollment of extremely low birth weight infants in a clinical research study may not be representative. *Pediatrics*. 2012;129(3):480-484.

**33**. Stenson BJ. Oxygen targets for preterm infants. *Neonatology*. 2013;103(4):341-345.

**34**. Schmidt B, Roberts RS, Whyte RK, et al; Canadian Oxygen Trial Group. Impact of study oximeter masking algorithm on titration of oxygen therapy in the Canadian oxygen trial. *J Pediatr*. 2014;165(4):666-71.e2.

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

**36**. Whyte RK, Nelson H, Roberts RS, Schmidt B. Benefits of oxygen saturation targeting trials: oximeter calibration software revision and infant saturations. *J Pediatr*. 2017;182:382-384.

**37**. Simes RJ; PPP and CTT Investigators. Prospective meta-analysis of cholesterol-lowering studies: the Prospective Pravastatin Pooling (PPP) Project and the Cholesterol Treatment Trialists (CTT) Collaboration. *Am J Cardiol*. 1995;76(9):122C-126C.

**38**. Ioannidis J. Next-generation systematic reviews: prospective meta-analysis, individual-level data, networks and umbrella reviews. *Br J Sports Med*. 2017;51(20):1456-1458.

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

**40**. Schmidt B, Whyte RK, Roberts RS. Trade-off between lower or higher oxygen saturations for extremely preterm infants: the first benefits of oxygen saturation targeting (BOOST) II trial reports its primary outcome. *J Pediatr*. 2014;165(1):6-8.

**41**. Bassler D, Briel M, Montori VM, et al; STOPIT-2 Study Group. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA*. 2010;303(12):1180-1187.

**42**. Schou IM, Marschner IC. Meta-analysis of clinical trials with early stopping: an investigation of potential bias. *Stat Med*. 2013;32(28):4859-4874.