UC Davis
UC Davis Previously Published Works

Title

Oxygen Saturation Target Range for Extremely Preterm Infants: A Systematic Review and Meta-analysis

Permalink

https://escholarship.org/uc/item/7pp553t6

Journal

JAMA Pediatrics, 169(4)

ISSN

2168-6203

Authors

Manja, Veena
Lakshminrusimha, Satyan
Cook, Deborah J

Publication Date

2015-04-01

DOI

10.1001/jamapediatrics.2014.3307

Peer reviewed

HHS Public Access

Author manuscript

Oxygen Saturation Target Range for Extremely Preterm Infants:

A Systematic Review and Meta-analysis

Veena Manja, MD, **Satyan Lakshminrusimha, MD**, and **Deborah J. Cook, MSc, MD**

Division of Cardiology, Department of Medicine, Veterans Affairs Medical Center, Buffalo, New York (Manja); Department of Internal Medicine, University at Buffalo, the State University of New York, Buffalo (Manja); Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada (Manja, Cook); Division of Neonatal-Perinatal Medicine, Department of Pediatrics, Women and Children's Hospital of Buffalo and University at Buffalo, the State University of New York, Buffalo (Lakshminrusimha); Division of Critical Care Medicine, Department of Medicine, McMaster University, Hamilton, Ontario, Canada (Cook)

Abstract

IMPORTANCE—The optimal oxygen saturation (SpO_2) target for extremely preterm infants is unknown.

OBJECTIVE—To systematically review evidence evaluating the effect of restricted vs liberal oxygen exposure on morbidity and mortality in extremely preterm infants.

DATA SOURCES—MEDLINE, PubMed, CENTRAL, and CINAHL databases from their inception to March 31, 2014, and abstracts submitted to Pediatric Academic Societies from 2000 to 2014.

STUDY SELECTION—All published randomized trials evaluating the effect of restricted (SpO_2, 85%–89%) vs liberal (SpO_2, 91%–95%) oxygen exposure in preterm infants (<28 weeks' gestation at birth).

DATA EXTRACTION AND SYNTHESIS—All meta-analyses were performed using Review Manager 5.2. The Cochrane risk-of-bias tool was used to assess study quality. The summary of the findings and the level of confidence in the estimate of effect were assessed using GRADEpro. Treatment effect was analyzed using a random-effects model.

MAIN OUTCOMES AND MEASURES—Death before hospital discharge, death or severe disability before 24 months, death before 24months, neurodevelopmental outcomes, hearing loss, bronchopulmonary dysplasia, necrotizing enterocolitis, and severe retinopathy of prematurity.

RESULTS—Five trials were included in the final synthesis. These studies had a similar design with a prespecified composite outcome of death/disability at 18 to 24 months corrected for prematurity; however, this outcome has not been reported for 2 of the 5 trials. There was no difference in the outcome of death/disability before 24 months (risk ratio [RR], 1.02 [95% CI, 0.92–1.14]). Mortality before 24 months was not different (RR, 1.13 [95% CI, 0.97–1.33]); however, a significant increase in mortality before hospital discharge was found in the restricted oxygen group (RR, 1.18 [95% CI, 1.03–1.36]). The rates of bronchopulmonary dysplasia, neurodevelopmental outcomes, hearing loss, and retinopathy of prematurity were similar between the 2 groups. Necrotizing enterocolitis occurred more frequently in infants on restricted oxygen (RR, 1.24 [95% CI, 1.05–1.47]). Using the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) criteria, we found that the quality of evidence for these outcomes was moderate to low.

CONCLUSIONS AND RELEVANCE—Although infants cared for with a liberal oxygen target had significantly lower mortality before hospital discharge than infants cared for with a restricted oxygen target, the quality of evidence for this estimate of effect is low. Necrotizing enterocolitis occurred less frequently in the liberal oxygen group. We found no significant differences in death or disability at 24 months, bronchopulmonary dysplasia, retinopathy of prematurity, neurodevelopmental outcomes, or hearing loss at 24 months.

Although oxygen is crucial for survival, too much oxygen can be harmful. Monitoring and maintaining optimal oxygen saturation (SpO_2) by pulse oximetry is a critical component of treating respiratory diseases. Published literature regarding the harm from excess oxygen in neonates is especially compelling.[1] A well-known complication of oxygen therapy, particularly for infants born at less than 28 weeks' postmenstrual age (PMA), is retinopathy of prematurity (ROP).[2] Oxygen was used liberally in the 1940s and early 1950s, resulting in an "epidemic" of ROP.[3] This led to extreme caution, the restriction of oxygen use, and clinical tolerance of hypoxia in premature infants, which then led to excess mortality.[4] Furthermore, in extremely preterm babies, the level of oxygenation may influence the development of other morbidities, including adverse neurodevelopmental outcomes, hearing loss, bronchopulmonary dysplasia,[5] and necrotizing enterocolitis.[6]

In the past several decades, improvements in technology have led to the increased survival of extremely premature infants. The methods for measuring oxygenation have also improved, enabling precise titration of oxygen delivery. 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.[7] In 2007, the American Academy of Pediatrics stated that SpO_2 values between 85% and 95% and

PaO_2 values between 50 and 80 mm Hg are examples of ranges pragmatically determined by some clinicians to guide oxygen therapy for preterm infants.[8] In 2007, European guidelines recommended an SpO_2 of less than 95% with no lower limit specified.[9] Updated guidelines in 2010 recommended a range from 85% to 93%, with an upper limit of 95%.[10] The SpO_2 target varied in the policies of different neonatal intensive care units, and optimal SpO_2 targets were not known.[11] This topic has been the focus of several recent randomized clinical controlled trials (RCTs). Although the complications of hyperoxia were reduced in the low range of SpO_2, some of these studies demonstrated an increased mortality and have generated controversy.[12–17]

To add to the controversy, an audit of the pulse oximeter used in these trials[15] revealed an artifact in the algorithm. This led to an artificial elevation of SpO_2 readings maximal at a displayed value of 90%, which thus led to less frequent SpO_2 readings of 87% to 90%. This artifact would be expected to reduce the number of SpO_2 values in the target range for the lower saturation target group[18] and potentially reduce the separation between the groups.[19] A new software algorithm was installed midway through 3 trials. This artifact may affect the results and will be explored in a subgroup analysis in this review.

A recently published meta-analysis of optimal oxygenation of extremely low birth weight infants included these trials.[17] However, the risk of bias in the studies included and the confidence in the estimates of effect for each outcome as recommended by the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) working group[20] were not appraised.

The objectives of this systematic review were to assess whether targeting a lower SpO_2 range (85%–89% restricted oxygen) has an effect on mortality and severe disability, mortality alone, bronchopulmonary dysplasia, neurodevelopmental outcome, hearing loss, necrotizing enterocolitis, or severe ROP compared with targeting a higher SpO_2 range (91%–95% liberal oxygen) after accounting for the risk of bias of each included study, as well as the quality of evidence for each outcome.

Methods

Criteria for Selecting Studies

All published RCTs and submitted abstracts with sufficient information, irrespective of language of publication, publication year, publication type, and publication status, were eligible for inclusion in our review. The participants in the RCTs had to be extremely preterm infants receiving supplemental oxygen (low [85%–89%] vs high [91%–95%] SpO_2 target) at any time prior to hospital discharge and for any duration. The outcome measures included any of the following: death before hospital discharge, death or severe disability before 24 months, death before 24 months, neurodevelopmental outcomes, hearing loss, bronchopulmonary dysplasia, necrotizing enterocolitis, and severe ROP. Studies other than RCTs and studies including infants 28 weeks' PMA or older at birth were excluded from our review.

Search Methods for Identification of Studies

A comprehensive literature search was undertaken using the MEDLINE, PubMed, CENTRAL, and CINAHL databases from their inception to March 31, 2014. The Cochrane sensitivity maximizing RCT filter was applied to MEDLINE, and adaptations of it were applied to the other databases when applicable. The search terms for MEDLINE involved various combinations of the following keywords (using the search field of keyword, abstracts, MeSH headings, exploded subject headings, publication type, text word, and title): "oxygen," "oxygen saturation," "hypoxia," "anoxia," "preterm," "premature," "neonate," "newborn," "infant," "randomized controlled trials," "controlled clinical trial," "randomly," "placebo," and "randomized." No language restrictions or date limits were applied. Previous reviews, including cross references and bibliographic citations of relevant publications, were reviewed. Ongoing trials were searched on ClinicalTrial.gov (https://www.clinicaltrials.gov/). Abstracts submitted to Pediatric Academic Societies from 2000 to 2014 were searched at http://www.pas-meeting.org.

Data Collection and Analysis

Study Selection—The titles and abstracts of potentially relevant publications retrieved by the search were reviewed independently by 2 reviewers (V.M. and S.L.) in order to select all citations that might contain comparisons of interest. If a review of an abstract led to disagreement about eligibility of the study for inclusion, it was included for the full text review. Full text articles of the selected citations were reviewed by the 2 reviewers independently, and eligible studies were included in the systematic review. For excluded articles, the reason for exclusion was recorded. Any discordance was identified; disagreement was resolved by discussion and consensus without a third reviewer. Kappa (κ) was calculated to estimate agreement on selection between the reviewers (κ 0.65 was chosen a priori to indicate adequate agreement).

Data Collection Process—Data were abstracted independently by the 2 reviewers on a data abstraction form prepared specifically for this review. Any differences and disagreements in the abstracted data were discussed and resolved by consensus. Details of methodological quality, study design, analysis, and results were abstracted. For each outcome, the numeric results, the statistic used, and the *P* value were abstracted.

Software and Summary of Findings—All meta-analyses were performed using Review Manager 5.2 software.[21] The summary of the findings and the level of confidence in the estimate of effect were assessed using GRADEpro version 3.2 for Windows (http://www.who.int/hiv/topics/mtct/grade_handbook.pdf?ua=1).

Assessment of Risk of Bias of Included Studies—The Cochrane risk-of-bias tool was used to assess study quality. Critical appraisal and assignment of a quality of evidence for the included studies were conducted independently by the 2 reviewers. All discrepancies were resolved by discussion and consensus. The studies were rated to be at high risk of bias, low risk of bias, or unclear risk of bias based on sequence generation, concealment of allocation, blinding of participants/parents and personnel, blinding of outcome assessment, incomplete outcome data, and selective outcome reporting.

Assessment of Quality of Evidence and Confidence in Estimates of Effect for Each Outcome—In addition to assessing the risk of bias for each study, we assessed the quality of the evidence to support the estimate of effect for each outcome using GRADEpro. With the use of this method, an overall assessment of the quality of evidence is made for each outcome across all included studies. Randomized clinical controlled trials start out as high-level evidence; however, downgrading of the level of evidence can occur because of deficiencies in 1 or more of the following domains: risk of bias,[22] inconsistency,[23] indirectness,[24] imprecision,[25] and publication bias.[26] The 2 reviewers independently followed GRADE quality assessment criteria[27] to rate the level of confidence in estimates of effect for each outcome, classified as being high, moderate, low, or very low. A summary of findings with quality of evidence for estimates with reasons for quality assignment is presented for each outcome (Table).

Measure of the Treatment Effect—Dichotomous data are expressed as risk ratios (RRs) with 95% CIs. A random-effects model was used, and a 2-tailed $P < .05$ was considered to be statistically significant.

Dealing With Missing Data—For most outcomes, data were missing for less than 5% of the published trials. We are still awaiting the final reports of 2 major trials. We were unable to obtain further information regarding the missing data from the existing literature. We did not contact the primary investigators for the missing data. No imputations were performed. We used complete available-case analysis for all outcomes.

Assessment of Heterogeneity—Clinical heterogeneity was explored by comparing the patient populations, the variations, if any, in study methods, and the oxygenation targets. Statistical heterogeneity was evaluated both by visual inspection of the Forest plot and by using a standard χ^2 value with a significant level of $P = .10$. Heterogeneity was also assessed by the I^2 statistic for each meta-analysis. An I^2 estimate of 50% or higher with a statistically significant value for χ^2 was interpreted as substantial heterogeneity.

Planned Subgroup Analysis—The effect of the oxygenation target may vary depending on the patient or the intervention characteristics; to elucidate these differences, subgroup analysis based on the infant's sex, the mode of delivery, and the intervention characteristics of the oximetry adjustment algorithm (original or revised) was planned. The results using a fixed-effects model were explored in a sensitivity analysis.

Results

Search Strategy

The results of the search are summarized in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (eFigure in the Supplement). After title and abstract review, 54 articles were included for full text review. The weighted κ for overall agreement between the 2 reviewers for the title and abstract screening was 0.92. There was no disagreement in the selection of the final articles for the systematic review.

Included Studies

Five trials were included in this review. The eTable in the Supplement provides a brief description of the trials and the primary, secondary, and post hoc outcomes. The Surfactant, Positive Pressure, and Oxygenation Randomized Trial (SUPPORT) was conducted in the United States and was published in 2010[13]; the Canadian Oxygen Trial (COT) was a multinational trial published in 2013,[12] and the Benefits of Oxygen Saturation Targeting II (BOOST II) trials included 3 trials conducted in the United Kingdom, Australia, and New Zealand, with preliminary results published in 2013 (the combined data are presented in the Table).[15] The 18- to 22-month results for the outcome of death and neurodevelopmental impairment in SUPPORT were published in 2012.[28] The 2-year outcome for the composite of death and major disability was published for the BOOST II–New Zealand trial[29] in 2014.

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. In SUPPORT, infants were randomized within 2 hours of birth. The gestational age at randomization was between 24 weeks, 0 days and 27 weeks, 6 days. In COT, infants with a PMA of 23 weeks, 0 days through 27 weeks, 6 days were eligible for enrollment during the first 24 hours after birth. Finally, in the BOOST II trial, infants with a PMA of less than 28 weeks at birth were randomized within 24 hours of birth.

Study Quality—Using the Cochrane risk-of-bias assessment, we found that, overall, 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.

Outcomes—Our Table is a summary of the findings. The reason for downgrading the quality of evidence for each outcome is explained in the table footnotes. One justification for decreasing the quality of evidence across all outcomes to moderate is the fact that, although a distinct separation of SpO_2 was planned in the study protocols, there was significant overlap in the SpO_2 values achieved in the 2 groups (Figure 1),[19] resulting in significant crossover of infants in the intervention and comparator groups. The level of confidence in the estimate for death before hospital discharge was downgraded further to low because this was not a prespecified primary outcome in any of the trials.[27] In SUPPORT, the original protocol specified the primary outcome of death before 36 weeks' PMA; this outcome was changed to death before hospital discharge; in the BOOST trials, death before hospital discharge was only significant for the revised pulse oximeter algorithm subset and not for the original algorithm or the overall pooled result. Initially, the 2 reviewers disagreed. The quality of evidence was rated moderate by S.L. and low by V.M.; the most conservative low rating was the final decision. The quality of evidence for ROP was downgraded to low owing to the inconsistency of results between the studies.[30] Quality of evidence for all other outcomes was assessed as moderate with no disagreement between reviewers.

Forest plots for individual outcomes are shown in Figure 2, Figure 3, and Figure 4. These demonstrate unadjusted RRs. The adjusted RRs (corrected for factors such as study center,

gestational age groups, and multiple births) cited in the original studies are different from the RRs shown in these plots.

Death before hospital discharge was not a prespecified outcome in these trials. This has been reported for SUPPORT and the BOOST II trial (Figure 2A); COT did not report the outcome of death before hospital discharge and reported only the 18-month outcome. Death before hospital discharge occurred more frequently for infants randomized to the SpO$_2$ target range of 85% to 89%. The RR for this outcome is 1.18 (95% CI, 1.03–1.36) in favor of the SpO$_2$ target range of 91% to 95%.

With regard to death or disability before 2 years of age, data from SUPPORT, COT, and the BOOST II–New Zealand trial are available, and the 2 groups were not significantly different from each other with regard to this outcome (RR, 1.02 [95% CI, 0.92–1.14]), as shown in Figure 2B. The numbers of infants who died before 24 months (ie, the outcome of death before 24 months) were not significantly different between the 2 groups (RR, 1.13 [95% CI, 0.97–1.33]) (Figure 2C). The numbers of infants who developed bronchopulmonary dysplasia at 36 weeks were not significantly different between the 2 groups (RR, 0.95 [95% CI, 0.87–1.04]) (Figure 3A).

Necrotizing enterocolitis occurred significantly more frequently in the restricted oxygen group than in the liberal oxygen group (RR, 1.24 [95% CI, 1.05–1.47]) (Figure 3B). The numbers of adverse neurodevelopmental outcomes at 24 months, determined by use of the modified Gross Motor Function Classification System (RR, 1.03 [95% CI, 0.73–1.45]) (Figure 3C), were similar for both groups. The numbers of infants with hearing loss at approximately 24 months (RR, 1.32 [95% CI, 0.78–2.21]) were not different between the 2 groups (Figure 4A).

The numbers of infants who developed severe ROP were not significantly different between the 2 groups for the pooled result using the random-effects model (RR, 0.72 [95% CI, 0.50–1.04]) (Figure 4B). There was significant heterogeneity among trials for this outcome. The Q statistic had a P value of .01, and I^2 was 78%; this was also noted by visual inspection of the forest plots. Owing to unexplained heterogeneity, the level of confidence in this estimate was downgraded by 1 for inconsistency for this outcome.

Subgroup Analysis—The data were insufficient to perform subgroup analysis based on sex, mode of delivery, and oximeter algorithm. Data on sex and mode of delivery were not reported separately in these studies. Outcomes based on oximeter-algorithm assignment were not available for similar time points. The COT provided the breakdown based on the oximeter algorithm (original vs modified) for outcomes at 18 months; the BOOST II trial, on the other hand, provided these data during the index hospitalization only.

Sensitivity Analysis—As part of the sensitivity analysis, all outcomes were evaluated using the fixed-effect model. With this analysis, significantly fewer infants in the restricted oxygen group had ROP (RR, 0.75 [95% CI, 0.63–0.88]) (Figure 4C). The random- and fixed-effects models assign study weights differently, which results in varying the influence of individual studies on the final result. The fixed-effects model assigns weight based on the

size of the study and largely ignores the information in smaller studies.[31] The random-effects model assigns a disproportionately smaller weight to larger studies. In this instance, using the random-effects model, we found that the BOOST II trial contributed 36.6% to the final result; using the fixed-effects model, we found that the BOOST II trial contributed 50.3% to the final result. Because ROP was less frequent in the restricted oxygen group than in the liberal oxygen group in the BOOST II trial, the change in weighting moved the overall estimate of effect away from the center, resulting in a statistically significant difference between the 2 groups.

We believe that the random-effects model (Figure 4B) is more appropriate for assessing the association between target SpO_2 and ROP. Our goal was to estimate the mean effect in all 3 studies and not let the overall estimate be overly influenced by 1 study (the BOOST II trial). The BOOST II trial, SUPPORT, and COT were performed by researchers using similar outlines but operating independently; hence, it is unlikely that all these studies are functionally equivalent.[31] The patients and/or protocols in these studies have differed in ways that would have affected the results, and therefore we did not assume a common effect size. In addition, our goal was to generalize the results to a range of clinical scenarios, and the random-effects model is more suitable for this purpose. Results did not differ significantly using the fixed-effects model for any of the other outcomes.

Discussion

This systematic review summarizes the results of recent RCTs that compare 2 different SpO_2 target ranges and explicitly ranks the quality of the evidence supporting each outcome using the internationally accepted GRADE criteria. Using the summary estimates of the meta-analysis conducted as part of this review, we found that, although there were higher frequencies of necrotizing enterocolitis and death prior to hospital discharge in the restricted oxygen group, there were no differences in the other outcomes between the 2 groups. These quantitative estimates have to be interpreted along with an assessment of the quality of the evidence for each outcome, which corresponds to the degree of confidence that we can have in the quantitative estimates. As elucidated in our Table, using GRADE criteria, we found that the levels of confidence for the outcomes of death before hospital discharge and ROP were low. In other words, even though there was a significant difference in the pooled analysis for the outcome of death before hospital discharge between the 2 groups, owing to the reasons noted in our Table, there is significant uncertainty regarding this estimate. Similarly, although we found that there was no difference in the outcome of ROP between the 2 groups using the random-effects model, this result should be interpreted with caution given the level of evidence.

The reasons for decreasing the level of confidence in the estimate of death before hospital discharge are 2-fold. First, this was not a prespecified outcome in any of the trial protocols.[32] The SUPPORT had a primary outcome of a composite of severe ROP and death before discharge. Given the high expected rate of death among premature infants, death was included as an outcome because it competed with ROP as a risk, not because a difference in mortality was expected between the 2 groups.[14] None of the other trials had a prespecified outcome of hospital mortality. Second, although the protocols specified a distinct separation

of the SpO$_2$ target in the 2 groups, there was significant overlap (Figure 1) in the SpO$_2$ achieved, resulting in poor separation between the intervention and comparator groups.[19] The degrees of separation in SpO$_2$ achieved in the comparator and intervention arms were different in these trials (Figure 1), and they are partly explained by differences in the protocol for monitoring and ensuring compliance with this target. Therefore, these results are not a true reflection of differences in outcome due to intended SpO$_2$ targets. We acknowledge that, based on the foregoing factors, the level-of-confidence assignment was subjective.

Some differences in outcomes across trials may be partly due to differences in trial design and the population enrolled. The primary and secondary outcomes were different as were aspects of statistical analysis. As noted in the original publication by Carlo et al,[13] SUPPORT had 46 planned analyses of secondary outcomes according to treatment group with no adjustment for multiple comparisons; some of the significant differences between the 2 groups may be purely due to chance. The overall longer-term mortality rates in COT and SUPPORT were different (15.9% in COT vs 20.1% in SUPPORT).[12,28] The SUPPORT enrolled infants at a much younger age (birth to 2 hours) than did other trials and did not exclude unstable infants and infants with pulmonary hypertension. In addition, as noted in Figure 1, mortality does not correlate with SpO$_2$ targets achieved for the whole group; other factors (such as time spent <85% SpO$_2$ at an individual patient level) need to be considered to explain these results.[19]

Other confounding factors that may influence these outcomes need be considered.[19] These include the contribution, if any, of the modification of the software algorithm midway through the trial, the frequency of intermittent hypoxia (<85%) in the 2 groups in the trial,[33] the site of probe placement (preductal vs postductal), and the frequency of blood transfusion (hemoglobin A) in the 2 groups to name a few.[19] Results may also vary based on the statistical method used, as illustrated by the sensitivity analysis, with the fixed-effect model producing a significant difference in the outcome of ROP compared with that obtained by the random-effects model.

Some of the articles used in this review allude to a forthcoming individual-patient meta-analysis of the approximately 5000 infants enrolled in the 5 trials included in this review; those results may enlighten clinicians about the optimal oxygenation target, especially if the analysis follows the intention-to-treat principle. Evaluating outcomes based on the actual pulse oximetry targets achieved[34] may provide some insights that could generate hypotheses for future research.

Conclusions

Based on this systematic review, a moderate level of evidence exists that suggests no significant difference in most outcomes between SpO$_2$ targets of 85% to 89% and SpO$_2$ targets of 91% to 95%. Although no difference was observed in the outcome of ROP, the level of confidence in this estimate is low; similarly, the level of confidence for the outcome of death before hospital discharge is low. The incidence of necrotizing enterocolitis was significantly lower in the liberal oxygen group than in the restricted oxygen group. Our

review conclusions therefore differ from the conclusions of the individual trials,[13,15] as well as from the conclusion of the other published review[17] that did not assess the quality of evidence for each outcome and therefore reached different conclusions. This review highlights the importance of evaluating the risk of bias and the level of confidence in the estimates of effect for individual outcomes. Thus, our review reveals that there is still significant uncertainty about the optimal target range for SpO_2 in extremely preterm infants.

Because of the overwhelming media coverage, the debate by expert commentators,[35,36] and the attention to these results by specialty groups and the Office of Human Research Protections (http://www.hhs.gov/ohrp/detrm_letrs/YR13/jun13a.pdf), the low level of confidence in the estimate of death before hospital discharge needs special emphasis. Our review and the finding of no difference in long-term mortality suggest that concerns about the increased mortality might have been exaggerated.

Based on all the evidence to date concerning RCTs, the optimal pulse oximetry saturation target that is effective and yet safe for the extremely preterm infant remains uncertain. Information from the prospective individual patient meta-analysis[37] and practical considerations (including the difficulty in maintaining SpO_2 in a narrow range) should be included when making practice recommendations.

Acknowledgments

References

1. Tiddens H. A case of retrolental fibroplasia, treated and cured with administration of oxygen in alternating concentrations. Ophthalmologica. 1960; 139:475–480. [PubMed: 13838325]

2. Silverman WA. Oxygen therapy and retrolental fibroplasia. Am J Public Health Nations Health. 1968; 58(11):2009–2011. [PubMed: 5754943]

3. Robertson AF. Reflections on errors in neonatology: I, the "hands-off" years, 1920 to 1950. J Perinatol. 2003; 23(1):48–55. [PubMed: 12556927]

4. Bolton DP, Cross KW. Further observations on cost of preventing retrolental fibroplasia. Lancet. 1974; 1(7855):445–448. [PubMed: 4131442]

5. Jobe AH, Bancalari E. Bronchopulmonary dysplasia. Am J Respir Crit Care Med. 2001; 163(7): 1723–1729. [PubMed: 11401896]

6. Neu J, Walker WA. Necrotizing enterocolitis. N Engl J Med. 2011; 364(3):255–264. [PubMed: 21247316]

7. Greenspan JS, Goldsmith JP. Oxygen therapy in preterm infants: hitting the target. Pediatrics. 2006; 118(4):1740–1741. [PubMed: 17015567]

8. American Academy of Pediatrics; American College of Obstetricians and Gynecologists. Guidelines for Perinatal Care. Elk Grove Village, IL: American Academy of Pediatrics; 2007.

9. Sweet D, Bevilacqua G, Carnielli V, et al. Working Group on Prematurity of the World Association of Perinatal Medicine; European Association of Perinatal Medicine. European consensus guidelines on the management of neonatal respiratory distress syndrome. J Perinat Med. 2007; 35(3):175–186. [PubMed: 17480144]

10. Sweet DG, Carnielli V, Greisen G, et al. European Association of Perinatal Medicine. European consensus guidelines on the management of neonatal respiratory distress syndrome in preterm infants—2010 update. Neonatology. 2010; 97(4):402–417. [PubMed: 20551710]

11. Wright CJ, Saugstad OD. OHRP and SUPPORT: lessons in balancing safety and improving the way we care for patients. J Pediatr. 2013; 163(5):1495–1497. [PubMed: 24050739]

12. 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. [PubMed: 23644995]

13. 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. [PubMed: 20472937]

14. Polin RA, Bateman D. Oxygen-saturation targets in preterm infants. N Engl J Med. 2013; 368(22): 2141–2142. [PubMed: 23642082]

15. 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. [PubMed: 23642047]

16. Carlo WA, Bell EF, Walsh MC, SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network. Oxygen-saturation targets in extremely preterm infants. N Engl J Med. 2013; 368(20):1949–1950. [PubMed: 23593943]

17. 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. [PubMed: 24247112]

18. 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. [PubMed: 21378398]

19. Lakshminrusimha S, Manja V, Mathew B, Suresh GK. Oxygen targeting in preterm infants: a physiological interpretation. J Perinatol. 2015; 35(1):8–15. [PubMed: 25357098]

20. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the *Journal of Clinical Epidemiology*. J Clin Epidemiol. 2011; 64(4):380–382. [PubMed: 21185693]

21. RevMan [computer program]. Version 5.2. Copenhagen, Denmark: The Nordic Cochrane Centre, The Cochrane Collaboration; 2012.

22. 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. [PubMed: 21247734]

23. 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. [PubMed: 21803546]

24. 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. [PubMed: 21802903]

25. 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. [PubMed: 21839614]

26. 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. [PubMed: 21802904]

27. Gallin, JI.; Ognibene, FP., et al. Principles and Practice of Clinical Research. 3. Waltham, MA: Academic Press; 2012.

28. 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. [PubMed: 23268664]

29. 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. [PubMed: 24560181]

30. Guyatt, G.; Rennie, D.; Meade, M.; Cook, D. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. 2. New York, NY: McGraw-Hill; 2008.

31. Borenstein, M.; Hedges, LV.; Higgins, JPT.; Rothstein, HR. Introduction to Meta-analysis. Chichester, England: Wiley; 2011.

32. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Trials. 2010; 11:32. [PubMed: 20334632]

33. Di Fiore JM, Walsh M, Wrage L, et al. SUPPORT Study Group of Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Low oxygen saturation target range is associated with increased incidence of intermittent hypoxemia. J Pediatr. 2012; 161(6):1047–1052. [PubMed: 22738947]

34. Kaufman DA, Zanelli SA, Gurka MJ, Davis M, Richards CP, Walsh BK. Time outside targeted oxygen saturation range and retinopathy of prematurity. Early Hum Dev. 2014; 90(suppl 2):S35–S40. [PubMed: 25220125]

35. Lantos JD. SUPPORTing premature infants. Pediatrics. 2013; 132(6):e1661–e1663. [PubMed: 24218459]

36. Lantos JD. Learning the right lessons from the SUPPORT study controversy. Arch Dis Child Fetal Neonatal Ed. 2014; 99(1):F4–F5. [PubMed: 23970799]

37. Askie LM. Optimal oxygen saturations in preterm infants: a moving target. Curr Opin Pediatr. 2013; 25(2):188–192. [PubMed: 23399572]
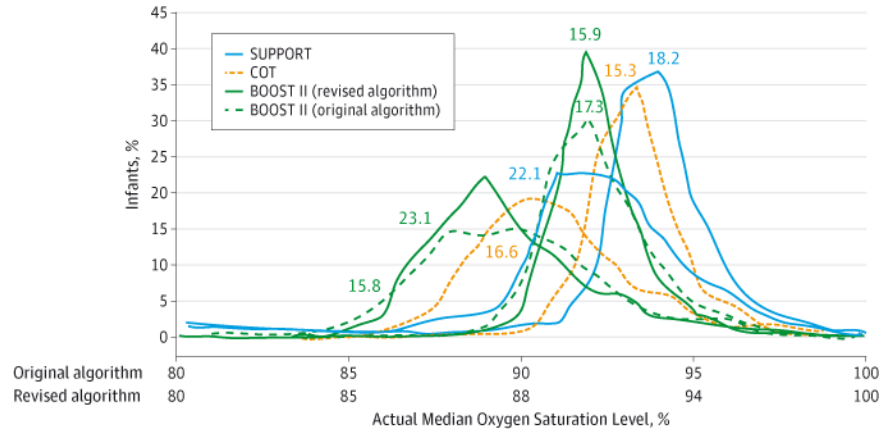
Figure 1. Distribution of Actual Median Oxygen Saturation

Distribution in the low oxygen saturation (SpO$_2$) (85%–89%) and high SpO$_2$ (91%–95%) arms in the Surfactant, Positive Pressure, and Oxygenation Randomized Trial (SUPPORT), the Canadian Oxygen Trial (COT), and the Benefits of Oxygen Saturation Targeting II (BOOST II) trials. The mortality numbers currently available are shown as percentages (note that the 18- to 22-month mortality numbers are currently not available for the BOOST II UK and Australia trials). Because the SUPPORT and the original algorithm BOOST II data are reported using the original algorithm, corresponding SpO$_2$ numbers on the revised algorithm are shown as well. A saturation of 90% in the original algorithm corresponds to a saturation of 88% on the revised algorithm. This figure was adapted from Figure 4 in Lakshminrusimha et al.[19]

A Death to hospital discharge

Study or Subgroup	Restricted Oxygen, No. Events	Total	Liberal Oxygen, No. Events	Total	Weight, %	Risk Ratio (95% CI)
BOOST II trial,[15] 2013	235	1221	203	1220	65.0	1.16 (0.98-1.37)
SUPPORT II trial,[13] 2010	130	654	107	662	35.0	1.23 (0.98-1.55)
Total (95% CI)		1875		1882	100.0	1.18 (1.03-1.36)
Total events	365		310			

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 0.17$, $df = 1$ ($P = .68$); $I^2 = 0\%$
Test for overall effect: $z = 2.39$ ($P = .02$)

Favors Restricted Oxygen / Favors Liberal Oxygen — Risk Ratio (95% CI)

B Death or severe disability before 18-24 mo

Study or Subgroup	Restricted Oxygen, No. Events	Total	Liberal Oxygen, No. Events	Total	Weight, %	Risk Ratio (95% CI)
BOOST II-New Zealand trial,[29] 2014	65	167	76	168	15.7	0.86 (0.67-1.11)
COT,[12] 2013	298	578	283	569	54.9	1.04 (0.92-1.16)
SUPPORT,[28] 2012	185	612	171	622	29.4	1.10 (0.92-1.31)
Total (95% CI)		1357		1359	100.0	1.02 (0.92-1.14)
Total events	548		530			

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 2.49$, $df = 2$ ($P = .29$); $I^2 = 20\%$
Test for overall effect: $z = 0.44$ ($P = .66$)

Favors Restricted Oxygen / Favors Liberal Oxygen — Risk Ratio (95% CI)

C Death before 18-24 mo

Study or Subgroup	Restricted Oxygen, No. Events	Total	Liberal Oxygen, No. Events	Total	Weight, %	Risk Ratio (95% CI)
BOOST II-New Zealand trial,[29] 2014	25	170	27	170	10.2	0.93 (0.56-1.53)
COT,[12] 2013	97	585	88	577	36.5	1.09 (0.83-1.42)
SUPPORT,[28] 2012	140	633	118	648	53.3	1.21 (0.98-1.51)
Total (95% CI)		1388		1395	100.0	1.13 (0.97-1.33)
Total events	262		233			

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 1.10$, $df = 2$ ($P = .58$); $I^2 = 0\%$
Test for overall effect: $z = 1.55$ ($P = .12$)

Favors Restricted Oxygen / Favors Liberal Oxygen — Risk Ratio (95% CI)

Figure 2. Mortality Associated With Restrictive or Liberal Use of Oxygen
The phrase "favors liberal oxygen" means that the negative outcome is less common in that arm and vice versa. The numbers shown in this plot are raw, unadjusted values and differ from the adjusted risk ratios provided in the references. BOOST indicates Benefits of Oxygen Saturation Targeting; COT, Canadian Oxygen Trial; and SUPPORT, Surfactant, Positive Pressure, and Oxygenation Randomized Trial.

A Bronchopulmonary dysplasia at 36 wk

Study or Subgroup	Restricted Oxygen, No. Events	Total	Liberal Oxygen, No. Events	Total	Weight, %	Risk Ratio (95% CI)
BOOST II trial,[15] 2013	160	353	172	376	33.2	0.99 (0.85-1.16)
COT,[12] 2013	164	515	171	517	27.0	0.96 (0.81-1.15)
SUPPORT,[13] 2010	205	540	237	568	39.8	0.91 (0.79-1.05)
Total (95% CI)		1408		1461	100.0	0.95 (0.87-1.04)
Total events	529		580			

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 0.63$, $df = 2$ ($P = .73$); $I^2 = 20\%$
Test for overall effect: $z = 1.09$ ($P = .28$)

Favors Restricted Oxygen / Favors Liberal Oxygen — Risk Ratio (95% CI) [0.1 – 1 – 10]

B Necrotizing enterocolitis

Study or Subgroup	Restricted Oxygen, No. Events	Total	Liberal Oxygen, No. Events	Total	Weight, %	Risk Ratio (95% CI)
BOOST II trial,[15] 2013	127	1221	97	1217	44.1	1.30 (1.01-1.68)
COT,[12] 2013	74	602	56	599	26.0	1.31 (0.95-1.83)
SUPPORT,[13] 2010	76	641	70	649	29.9	1.10 (0.81-1.49)
Total (95% CI)		2464		2465	100.0	1.24 (1.05-1.47)
Total events	277		223			

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 0.88$, $df = 2$ ($P = .65$); $I^2 = 0\%$
Test for overall effect: $z = 2.54$ ($P = .001$)

Favors Restricted Oxygen / Favors Liberal Oxygen — Risk Ratio (95% CI) [0.1 – 1 – 10]

C Neurodevelopmental outcomes (GMFCS score ≥2)

Study or Subgroup	Restricted Oxygen, No. Events	Total	Liberal Oxygen, No. Events	Total	Weight, %	Risk Ratio (95% CI)
BOOST II-New Zealand trial,[29] 2014	5	144	7	142	9.5	0.70 (0.23-2.17)
COT,[12] 2013	30	488	31	488	50.6	0.97 (0.60-1.57)
SUPPORT,[13] 2012	26	479	23	511	39.9	1.21 (0.70-2.08)
Total (95% CI)		1111		1141	100.0	1.03 (0.73-1.45)
Total events	61		61			

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 0.82$, $df = 2$ ($P = .66$); $I^2 = 0\%$
Test for overall effect: $z = 0.14$ ($P = .89$)

Favors Restricted Oxygen / Favors Liberal Oxygen — Risk Ratio (95% CI) [0.1 – 1 – 10]

Figure 3. Morbidity Associated With Restrictive or Liberal Use of Oxygen: Bronchopulmonary Dysplasia, Necrotizing Enterocolitis, and Neurodevelopmental Outcomes
The phrase "favors liberal oxygen" means that the negative outcome is less common in that arm and vice versa. BOOST indicates Benefits of Oxygen Saturation Targeting; COT, Canadian Oxygen Trial; GMFCS, Gross Motor Function Classification System; and SUPPORT, Surfactant, Positive Pressure, and Oxygenation Randomized Trial.

A Hearing loss/impairment

Study or Subgroup	Restricted Oxygen, No. Events	Restricted Oxygen, No. Total	Liberal Oxygen, No. Events	Liberal Oxygen, No. Total	Weight, %	Risk Ratio (95% CI)
BOOST II-New Zealand trial,[29] 2014	2	135	1	136	4.7	2.01 (0.18-21.96)
COT,[12] 2013	18	487	12	489	52.1	1.51 (0.73-3.09)
SUPPORT,[13] 2012	12	479	12	511	43.2	1.07 (0.48-2.35)
Total (95% CI)		1101		1136	100.0	1.32 (0.78-2.21)
Total events	32		25			

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 0.53$, $df = 2$ ($P = .77$); $I^2 = 0\%$
Test for overall effect: $z = 1.04$ ($P = .30$)

B Retinopathy of prematurity

Study or Subgroup	Restricted Oxygen, No. Events	Restricted Oxygen, No. Total	Liberal Oxygen, No. Events	Liberal Oxygen, No. Total	Weight, %	Risk Ratio (95% CI)
BOOST II trial,[15] 2013	110	1035	141	1044	36.6	0.79 (0.62-0.99)
COT,[12] 2013	64	500	66	503	32.3	0.98 (0.71-1.34)
SUPPORT,[13] 2012	41	475	91	509	31.0	0.48 (0.34-0.68)
Total (95% CI)		2010		2056	100.0	0.72 (0.50-1.04)
Total events	215		298			

Heterogeneity: $\tau^2 = 0.08$; $\chi^2 = 8.93$, $df = 2$ ($P = .01$); $I^2 = 78\%$
Test for overall effect: $z = 1.73$ ($P = .08$)

C Retinopathy of prematurity using a fixed-effects model

Study or Subgroup	Restricted Oxygen, No. Events	Restricted Oxygen, No. Total	Liberal Oxygen, No. Events	Liberal Oxygen, No. Total	Weight, %	Risk Ratio (95% CI)
BOOST II trial,[15] 2013	110	1035	141	1044	50.3	0.79 (0.62-0.99)
COT,[12] 2013	64	500	66	503	26.8	0.98 (0.71-1.34)
SUPPORT,[13] 2012	41	475	91	509	22.9	0.48 (0.34-0.68)
Total (95% CI)		2010		2056	100.0	0.75 (0.63-0.88)
Total events	215		298			

Heterogeneity: $\chi^2 = 8.93$, $df = 2$ ($P = .01$); $I^2 = 78\%$
Test for overall effect: $z = 3.47$ ($P = .001$)

Figure 4. Morbidity Associated With Restrictive or Liberal Use of Oxygen: Hearing Loss/Impairment and Retinopathy of Prematurity

The phrase "favors liberal oxygen" means that the negative outcome is less common in that arm and vice versa. BOOST indicates Benefits of Oxygen Saturation Targeting; COT, Canadian Oxygen Trial; and SUPPORT, Surfactant, Positive Pressure, and Oxygenation Randomized Trial.
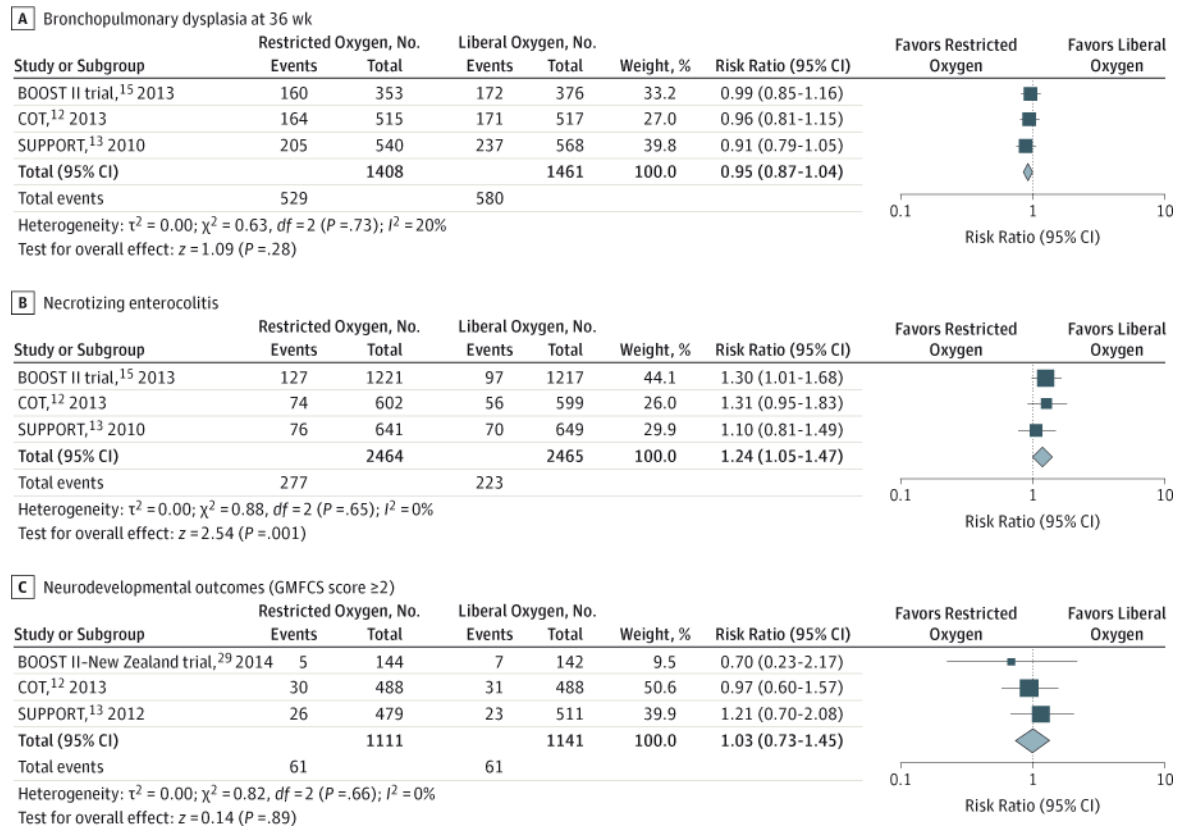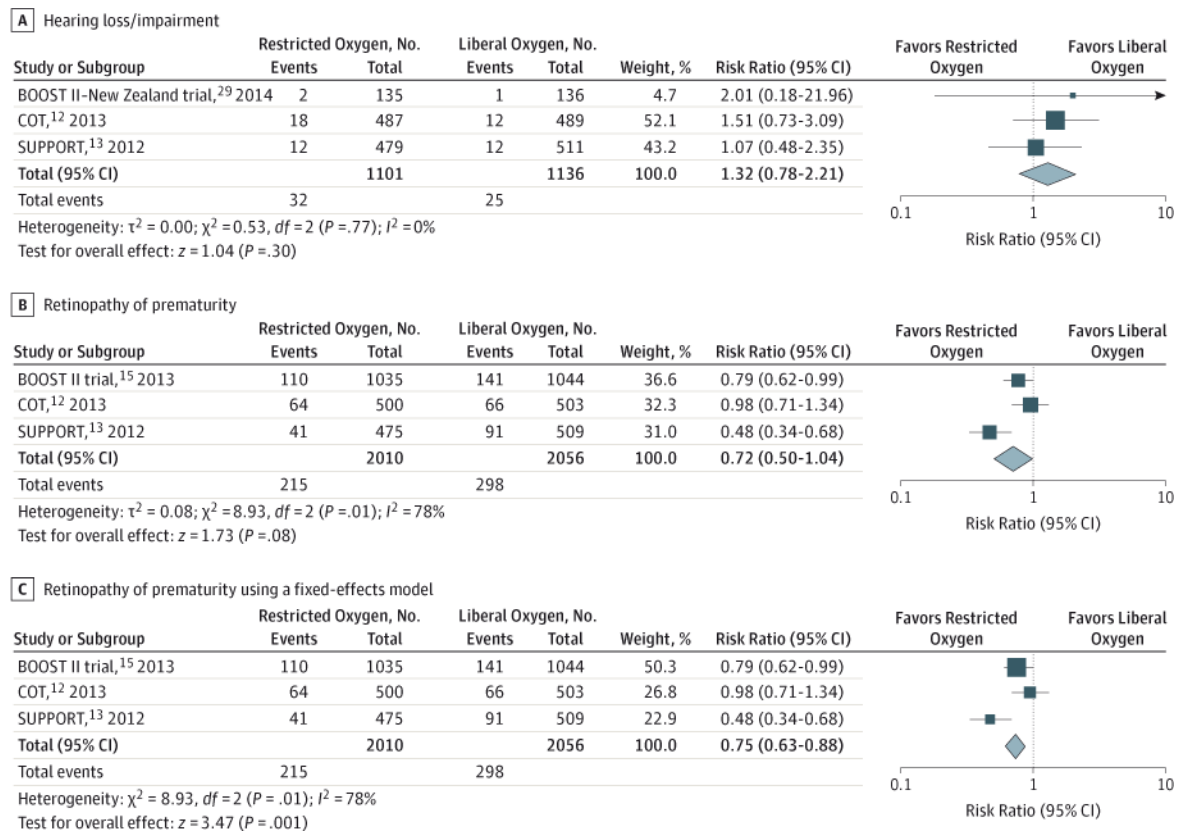
Table

GRADE Summary of Findings of Restricted *vs* Liberal Oxygen Exposure for Extremely Premature Infants

Outcomes	No. of Cases per 1000 Infants		RR (95% CI)	Participants, No.	Studies, No.	Quality of Evidence[b]
	Assumed Risk of Liberal Oxygen Exposure[a]	Corresponding Risk of Restricted Oxygen Exposure[a] (95% CI)				
Death before hospital discharge	165	194 (170–224)	1.18 (1.03–1.36)	3757	4	Low[c,d,e,f]
Death or disability at 24 mo	390	398 (367–445)	1.02 (0.94–1.14)	2716	3	Moderate[c,d]
Death before 24 mo	167	194 (164–229)	1.16 (0.98–1.37)	2783	3	Moderate[c,d]
Bronchopulmonary dysplasia	397	377 (345–413)	0.95 (0.87–1.04)	2869	5	Moderate[c,d]
Neurodevelopmental outcomes at 18–24 mo	53	55 (39–78)	1.03 (0.73–1.45)	2252	3	Moderate[c,d]
Hearing loss	22	29 (17–49)	1.32 (0.78–2.21)	2237	3	Moderate[c,d]
Necrotizing enterocolit is	90	112 (95–133)	1.24 (1.05–1.47)	4929	5	Moderate[c,d]
Retinopathy of prematurity	145	104 (72–151)	0.72 (0.5–1.04)	4066	5	Low[c,d,g]

Abbreviations: GRADE, Grades of Recommendation, Assessment, Development, and Evaluation; RR, risk ratio; SpO2, oxygen saturation.

[a] The basis for the assumed risk (eg, the median control group risk across studies) is provided. The corresponding risk (and its 95% CI) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

[b] The GRADE Working Group grades of evidence are as follows: high quality (further research is very unlikely to change our confidence in the estimate of effect), moderate quality (further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate), low quality (further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate), and very low quality (we are very uncertain about the estimate).

[c] The pulse oximeter algorithm was modified midway through the study owing to a calibration correction, and this caused a deviation from SpO2 values.

[d] The separation of SpO2 values obtained was not as planned in the study design/protocol. The median SpO2 value in the restricted arm (planned SpO2 of 85%–89%) was higher than 90% in some studies (Figure 1).

[e] This was not a prespecified outcome in the Benefits of Oxygen Saturation Targeting II trial, which was prematurely stopped because of this outcome.

[f] Only 4 of the 5 eligible trials reported on the outcome of death before hospital discharge (the Canadian Oxygen Trial group did not).

[g] There was significant unexplained heterogeneity in this outcome, which led to downgrading the quality of evidence by 1.