Cochrane Library

Cochrane Database of Systematic Reviews

Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants (Review)

Askie LM, Darlow BA, Davis PG, Finer N, Stenson B, Vento M, Whyte R

Askie LM, Darlow BA, Davis PG, Finer N, Stenson B, Vento M, Whyte R.
Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants.
Cochrane Database of Systematic Reviews 2017, Issue 4. Art. No.: CD011190.
DOI: 10.1002/14651858.CD011190.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Lisa M Askie[1], Brian A Darlow[2], Peter G Davis[3], Neil Finer[4], Ben Stenson[5], Maximo Vento[6], Robin Whyte[7]

[1]NHMRC Clinical Trials Centre, University of Sydney, Camperdown, Australia. [2]Department of Paediatrics, University of Otago, Christchurch, New Zealand. [3]The University of Melbourne, Melbourne, Australia. [4]Department of Pediatrics, University of California San Diego, San Diego, California, USA. [5]Neonatal Unit, Simpson Centre for Reproductive Health, Royal Infirmary of Edinburgh, Edinburgh, UK. [6]Health Research Institute La Fe, Division of Neonatology, University & Polytechnic Hospital La Fe, Valencia, Spain. [7]Department of Neonatal Pediatrics, Halifax Dalhousie University, IWK Health Centre, Halifax, Canada

Contact address: Lisa M Askie, NHMRC Clinical Trials Centre, University of Sydney, Locked Bag 77, Camperdown, NSW, 2050, Australia. laskie@ctc.usyd.edu.au.

Editorial group: Cochrane Neonatal Group.
Publication status and date: New, published in Issue 4, 2017.

Citation: Askie LM, Darlow BA, Davis PG, Finer N, Stenson B, Vento M, Whyte R. Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants. *Cochrane Database of Systematic Reviews* 2017, Issue 4. Art. No.: CD011190. DOI: 10.1002/14651858.CD011190.pub2.

ABSTRACT

Background

The use of supplemental oxygen in the care of extremely preterm infants has been common practice since the 1940s. Despite this, there is little agreement regarding which oxygen saturation (SpO_2) ranges to target to maximise short- or long-term growth and development, while minimising harms. There are two opposing concerns. Lower oxygen levels (targeting SpO_2 at 90% or less) may impair neurodevelopment or result in death. Higher oxygen levels (targeting SpO_2 greater than 90%) may increase severe retinopathy of prematurity or chronic lung disease.

The use of pulse oximetry to non-invasively assess neonatal SpO_2 levels has been widespread since the 1990s. Until recently there were no randomised controlled trials (RCTs) that had assessed whether it is better to target higher or lower oxygen saturation levels in extremely preterm infants, from birth or soon thereafter. As a result, there is significant international practice variation and uncertainty remains as to the most appropriate range to target oxygen saturation levels in preterm and low birth weight infants.

Objectives

1. What are the effects of targeting lower versus higher oxygen saturation ranges on death or major neonatal and infant morbidities, or both, in extremely preterm infants?

2. Do these effects differ in different types of infants, including those born at a very early gestational age, or in those who are outborn, without antenatal corticosteroid coverage, of male sex, small for gestational age or of multiple birth, or by mode of delivery?

Search methods

We used the standard search strategy of Cochrane Neonatal to search the Cochrane Central Register of Controlled Trials (CENTRAL 2016, Issue 4), MEDLINE via PubMed (1966 to 11 April 2016), Embase (1980 to 11 April 2016) and CINAHL (1982 to 11 April 2016). We also searched clinical trials databases, conference proceedings and the reference lists of retrieved articles for randomised controlled trials.

Selection criteria

Randomised controlled trials that enrolled babies born at less than 28 weeks' gestation, at birth or soon thereafter, and targeted SpO_2 ranges of either 90% or below or above 90% via pulse oximetry, with the intention of maintaining such targets for at least the first two weeks of life.

Data collection and analysis

We used the standard methods of Cochrane Neonatal to extract data from the published reports of the included studies. We sought some additional aggregate data from the original investigators in order to align the definitions of two key outcomes. We conducted the meta-analyses with Review Manager 5 software, using the Mantel-Haenszel method for estimates of typical risk ratio (RR) and risk difference (RD) and a fixed-effect model. We assessed the included studies using the Cochrane 'Risk of bias' and GRADE criteria in order to establish the quality of the evidence. We investigated heterogeneity of effects via pre-specified subgroup and sensitivity analyses.

Main results

Five trials, which together enrolled 4965 infants, were eligible for inclusion. The investigators of these five trials had prospectively planned to combine their data as part of the NeOProM (Neonatal Oxygen Prospective Meta-analysis) Collaboration. We graded the quality of evidence as high for the key outcomes of death, major disability, the composite of death or major disability, and necrotising enterocolitis; and as moderate for blindness and retinopathy of prematurity requiring treatment.

When an aligned definition of major disability was used, there was no significant difference in the composite primary outcome of death

or major disability in extremely preterm infants when targeting a lower (SpO_2 85% to 89%) versus a higher (SpO_2 91% to 95%) oxygen saturation range (typical RR 1.04, 95% confidence interval (CI) 0.98 to 1.10; typical RD 0.02, 95% CI -0.01 to 0.05; 5 trials, 4754 infants) (high-quality evidence). Compared with a higher target range, a lower target range significantly increased the incidence of death at 18 to 24 months corrected age (typical RR 1.16, 95% CI 1.03 to 1.31; typical RD 0.03, 95% CI 0.01 to 0.05; 5 trials, 4873 infants) (high-quality evidence) and necrotising enterocolitis (typical RR 1.24, 95% 1.05 to 1.47; typical RD 0.02, 95% CI 0.01 to 0.04; 5 trials, 4929 infants; $I^2 = 0\%$) (high-quality evidence). Targeting the lower range significantly decreased the incidence of retinopathy of prematurity requiring treatment (typical RR 0.72, 95% CI 0.61 to 0.85; typical RD -0.04, 95% CI -0.06 to -0.02; 5 trials, 4089 infants; $I^2 = 69\%$) (moderate-quality evidence). There were no significant differences between the two treatment groups for major disability including blindness, severe hearing loss, cerebral palsy, or other important neonatal morbidities.

A subgroup analysis of major outcomes by type of oximeter calibration software (original versus revised) found a significant difference in the treatment effect between the two software types for death (interaction P = 0.03), with a significantly larger treatment effect seen for those infants using the revised algorithm (typical RR 1.38, 95% CI 1.13 to 1.68; typical RD 0.06, 95% CI 0.01 to 0.10; 3 trials, 1716 infants). There were no other important differences in treatment effect shown by the subgroup analyses using the currently available data.

Authors' conclusions

In extremely preterm infants, targeting lower (85% to 89%) SpO_2 compared to higher (91% to 95%) SpO_2 had no significant effect on the composite outcome of death or major disability or on major disability alone, including blindness, but increased the average risk of mortality by 28 per 1000 infants treated. The trade-offs between the benefits and harms of the different oxygen saturation target ranges may need to be assessed within local settings (e.g. alarm limit settings, staffing, baseline outcome risks) when deciding on oxygen saturation targeting policies.

PLAIN LANGUAGE SUMMARY

Targeting lower or higher oxygen levels in preterm infants

Review question: Is it better to target a lower or higher level of oxygen for babies born very early?

Background: Giving additional ('supplemental') oxygen to babies born very early ('extremely preterm infants') who have breathing difficulties has been common practice since the 1940s. Despite this, there is little agreement as to what levels of oxygen will maximise short- or long-term survival and development. Technology ('pulse oximetry') that can easily measure the level of oxygen in a baby's blood (oxygen saturation) has been in widespread use since the 1990s. Despite this, until recently there were no randomised trials

that had tested whether it is better to target lower or higher oxygen saturation levels in extremely preterm infants, from birth or soon thereafter. As a result there is a great deal of variation in the target ranges aimed for in different newborn care units around the world.

Study characteristics: The studies we included were randomised trials that enrolled babies born at less than 28 weeks' gestation, at birth or soon thereafter, and targeted oxygen saturation (SpO_2) ranges of either 85% to 89% or 91% to 95%, for at least the first two weeks of life.

Key results: We included five trials, which together enrolled 4965 infants, in this review. There were benefits and harms associated with both the target ranges tested. Neither the lower nor the higher target range had a significant effect on the rate of death or major disability (the main outcome), on major disability alone or on blindness. However, infants in whom the lower oxygen range was targeted had, on average, a 2.8% increased risk of death, compared to the infants in whom the higher oxygen range was targeted. They also had a 2.2% increase in the rate of a serious bowel condition known as necrotising enterocolitis. Conversely, the infants in whom the lower oxygen range was targeted had a 4.2% decrease in the rate of a serious eye problem, retinopathy of prematurity, requiring surgery or other treatments. The trade-offs between these benefits and harms may need to be assessed within local settings when deciding on oxygen saturation targeting policies.

Quality of evidence: We rated the quality of the evidence as high for the key outcomes of death, major disability, the composite of death or major disability, and necrotising enterocolitis. We rated the quality of evidence as moderate for the two eye-related outcomes (blindness, retinopathy of prematurity requiring treatment), giving us confidence that the overall results are reliable.

SUMMARY OF FINDINGS FOR THE MAIN COMPARISON [Explanation]

Lower compared to higher targeted oxygen saturations (no subgroups) in preterm infants

Patient or population: extremely preterm infants
Setting: neonatal intensive care units
Intervention: lower oxygen saturation targets
Comparison: higher oxygen saturations targets (no subgroups)

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with higher targeted oxygen saturations (no subgroups)	Risk with lower targeted oxygen saturations				
Death or major disability by 18 to 24 months corrected age (aligned definition)	Study population		RR 1.04 (0.98 to 1.10)	4754 (5 RCTs)	⊕⊕⊕⊕ HIGH	-
	493 per 1000	513 per 1000 (483 to 542)				
Death to 18 to 24 months corrected age	Study population		RR 1.16 (1.03 to 1.31)	4873 (5 RCTs)	⊕⊕⊕⊕ HIGH	-
	171 per 1000	199 per 1000 (176 to 224)				
Major disability by 18 to 24 months corrected age (aligned definition)	Study population		RR 1.01 (0.93 to 1.09)	3867 (5 RCTs)	⊕⊕⊕⊕ HIGH	-
	383 per 1000	387 per 1000 (356 to 417)				
Retinopathy of prematurity requiring treatment	Study population		RR 0.72 (0.61 to 0.85)	4089 (5 RCTs)	⊕⊕⊕◯ MODERATE [1]	-
	148 per 1000	106 per 1000 (90 to 125)				

Necrotising enterocolitis	Study population		RR 1.24 (1.05 to 1.47)	4929 (5 RCTs)	⊕⊕⊕⊕ HIGH	-
	90 per 1000	112 per 1000 (95 to 133)				
Blindness	Study population		RR 1.13 (0.65 to 1.97)	3875 (5 RCTs)	⊕⊕⊕◯ MODERATE [2]	-
	12 per 1000	13 per 1000 (8 to 23)				

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).

CI: confidence interval;**RCT:** randomised controlled trial; **RR:** risk ratio

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1]Downgraded to moderate for inconsistency due to moderate heterogeneity (I^2 = 72%).
[2]Downgraded to moderate for imprecision due to low event rates.

BACKGROUND

Description of the condition

The administration of supplemental oxygen has a long history in neonatal care (Wilson 1942; Tin 2007). Oxygen was used liberally for the first time in neonates in the 1930s when an oxygen unit was described for preterm infants (Raju 1999). The use of oxygen in preterm and low birth weight infants suffering respiratory insufficiency has resulted in significant healthcare benefits, such as reduced mortality and spastic diplegia (Avery 1960; McDonald 1963), but has also been associated with significant deleterious effects such as retinopathy of prematurity and lung toxicity (Duc 1992).

Improvements in technology in the past few decades have led to the increased survival of preterm and low birth weight infants. One of these advances is the ability to measure oxygen levels more accurately. Despite the exceedingly common use of supplemental oxygen in this population of infants, there is little consensus as to the optimal levels of oxygen for maximising short- or long-term growth and development, while minimising harmful effects (Poets 1998; McIntosh 2001; Silverman 2004).

Adverse consequences of liberal and restricted use of oxygen

The adverse consequences of liberal oxygen use were recognised in the early 1940s. Terry 1942 described a type of blindness in preterm infants characterised by a thick fibrotic membrane in the retrolental space. In 1951, the role of supplemental oxygen in the aetiology of 'retrolental fibroplasia' was first suggested (Campbell 1951). By 1954, retrolental fibroplasia had blinded about 10,000 infants (Silverman 1980; Silverman 2004). From 1954 to 1956, three randomised trials (Lanman 1954; Patz 1954; Kinsey 1956), enrolling 341 infants, proved that breathing unrestricted concentrations of inspired oxygen was a major cause of retrolental fibroplasia (Askie 2009). Throughout this period, oxygen administration was guided by the clinical observations of skin colour, as well as the rate, regularity and work of breathing. It was not until the 1960s and 1970s that sampling of blood gases, transcutaneous oxygen monitoring and later pulse oximetry became available for more precise monitoring of oxygen levels (Walsh 2009).

An early prospective cohort study, reported in 1977, was unable to establish a causal relationship between arterial oxygen tension and (what is now known as) retinopathy of prematurity (ROP), but did reveal that the most relevant factors for developing ROP were birth weight less than 1200 grams and length of exposure to supplemental oxygen (Kinsey 1977).

As a consequence of the retrolental fibroplasia blindness epidemic in the 1960s, the use of oxygen was drastically limited, usually to less than 40% inspired oxygen, even for preterm infants with respiratory distress, allowing them to become severely hypoxaemic and leading to a substantial increase in the incidence of cerebral palsy (Usher 1961). In the next 20 years over 150,000 premature babies died of hypoxic respiratory failure (Avery 1960; McDonald 1964; Cross 1973; Bolton 1974). It is estimated that for every infant whose sight was saved, 16 died (Avery 1960; Silverman 2004), and many others developed spastic diplegia (McDonald 1964).

Description of the intervention

Multiple attempts have been made to establish the optimal oxygen levels in preterm infants, using a variety of technologies, in order to circumvent the adverse consequences of either restricted or liberal use of supplemental oxygen.

However, what constitutes an 'appropriate' level of oxygen for infants born preterm, who would otherwise be in-utero, remains unknown. The fetus is relatively hypoxic with haemoglobin (Hb) oxygen saturations of 65%, 55%, and 45% in the aorta, pulmonary artery and pulmonary vein, respectively. However, it should be noted that fetal blood contains almost only fetal haemoglobin (HbF), which has an extraordinary affinity for oxygen and is therefore capable of capturing sufficient oxygen from the intervillous space to support fetal growth and metabolism (Gao 2010; Vento 2013).

In the 1980s and early 1990s the use of transcutaneous oxygen monitoring became available. A study of transcutaneous oxygen monitoring (TcO_2) in preterm infants confirmed that ROP occurred more often when longer periods of time were spent with a TcO_2 above 80 mm Hg, but did not determine if another limit was safer (Flynn 1992). A partial pressure of arterial oxygen (PaO_2) range of 50 to 80 mm Hg became widely accepted as an appropriate level to target (AAP 1988; McIntosh 2001; AAP 2002), but this was based on professional consensus rather than on evidence.

In the 1990s the use of pulse oximetry became a standard of care and continuous monitoring has allowed more frequent titration of the oxygen concentration administered. Pulse oximetry (SpO_2) refers to the estimation of the oxygen saturation of arterial blood using a device that measures the pulsatile changes in light transmission across a tissue bed. Pulse oximeters work on the principle that oxygenated and deoxygenated haemoglobin absorbs light of different wavelengths (red and infrared). The oximeter emits light of these two wavelengths and measures absorption in the pulsatile element of the blood flow, thus producing a measure of the oxygen saturation of arterial blood separate from the non-pulsatile venous blood (Williams 1998). Pulse oximeters lack the heat-related side effects of transcutaneous oxygen monitors.

Despite the ease of use of pulse oximeters, translation of SpO_2

values into PaO_2 can be difficult to establish. The relationship between SpO_2 and PaO_2 is dependent on various physiologic circumstances such as affinity of Hb for oxygen, which is significantly greater in fetal Hb. Thus, the higher the fetal Hb concentration the higher the SpO_2 would be for any given PaO_2 value. Castillo 2008 reported that in preterm infants, for oxygen saturation values between 85% and 93% the mean measured PaO_2 was 56 ± 14.7 mm Hg. Within this SpO_2 range, 87% of the samples had PaO_2 values of 40 to 80 mm Hg, 8.6% had values of less than 40 mm Hg, and 4.6% had values greater than 80 mm Hg. When the SpO_2 was greater than 93% the mean PaO_2 was 107.3 ± 59.3 mm Hg with 60% of values greater than 80 mm Hg. The Supplemental Therapeutic Oxygen for Prethreshold Retinopathy of Prematurity trial used pulse oximetry to target a lower (89% to 94%) or higher (96% to 99%) oxygen saturation range in 649 preterm infants with prethreshold ROP who were 35 weeks postmenstrual age at randomisation (STOP ROP 2000). The higher range caused more adverse respiratory events including pneumonia, chronic lung disease requiring oxygen, and diuretic therapy. There was no statistically significant difference in the rate of progression to threshold ROP. The results of this trial are included in a separate Cochrane Review entitled: *Supplemental oxygen for the treatment of pre-threshold retinopathy of prematurity* (Lloyd 2003). In the Benefits of Oxygen Saturation Targeting (BOOST) trial (Askie 2003), 358 infants born at less than 30 weeks' gestation were randomly assigned, from three weeks or more after birth (at 32 weeks' postmenstrual age) until they breathed air, to target an SpO_2 range of either 91% to 94% or 95% to 98% using masked offset oximeters. This trial found no evidence that higher SpO_2 targeting improved growth or development, but it did increase days of oxygen therapy and use of healthcare resources. The authors concluded that further large randomised trials were needed to determine how targeting different SpO_2 levels from the day of birth affects ROP, chronic lung disease, growth, disability, and mortality (Askie 2003; Silverman 2004).

In transposing oxygen tensions of 50 to 80 mm Hg into equivalent arterial oxygen saturation, most clinicians have targeted functional SpO_2 at 90% to 95% with a minimum acceptable SpO_2 of 85% (Anderson 2004). Hence the dichotomising of SpO_2 into 'higher' or 'lower' target ranges above or below a cut point of 90% appears reasonable. In the early 2000s, there were several observational studies that found lower SpO_2 was associated with less severe ROP; improved short-term respiration, growth and development outcomes; and either no apparent effect or a decrease in mortality (Tin 2001; Chow 2003; Anderson 2004).

It should be recognised that an intention to *target* a certain SpO_2 range does not guarantee that an infant's *actual* SpO_2 will always be maintained within that range. Most studies report that preterm infants receiving supplemental oxygen in a specified target range only remain in that range for about 30% to 50% of the time (Hagadorn 2006; Lim 2014).

To address the continuing uncertainty regarding the appropriate levels of oxygen saturation targeting for preterm infants with sufficient confidence, the Neonatal Oxygen Prospective Meta-analysis (NeOProM) Collaboration was formed in 2003. NeOProM is a prospective meta-analysis collaboration (Askie 2011), which includes five randomised trials (ACTRN12605000055606; ACTRN12605000253606; NCT00233324; ISRCTN00842661; ISRCTN62491227). The investigators of these five trials prospectively agreed to conduct their trials using very similar protocols, and made a commitment to combine their individual participant data once their own trial's results were published. Representatives from each of these five trials and the NeOProM Collaboration are authors on this review.

Summary

There are two opposing concerns. Less inspired oxygen (targeting SpO_2 at 90% or less) may increase the risk of death from chronic hypoxaemia or impaired neurodevelopment (Newburger 1984; Skinner 1999; Subhedar 2000). More inspired oxygen (targeting SpO_2 greater than 90%) may increase severe ROP (Hellstrom 2013), or chronic lung disease (Warner 1998; Tin 2001; Sun 2002; Chow 2003; Anderson 2004). However, uncertainty remains as to the most appropriate range to target for blood oxygen levels in preterm and low birth weight infants.

Two other related Cochrane Reviews have summarised the findings on gradual versus abrupt (Askie 2001a), and early versus late discontinuation of oxygen therapy (Askie 2001b), in preterm or low birth weight infants. Meta-analyses of the available aggregate data from the five NeOProM trials were published by Saugstad 2014, Manja 2015, Stenson 2016, and Manja 2017 and a sub-set of data relating to retinopathy of prematurity outcomes by Fang 2016.

How the intervention might work

Oxygen is the most common therapy used in the care of very preterm infants. It has been associated with significant improvements in neonatal survival and reduced disability (Avery 1960). However, preterm infants are highly sensitive to the harmful biochemical and physiological effects of supplemental oxygen. Toxic oxygen radicals are increased in hyperoxia (Maltepe 2009), and in re-oxygenation after hypoxaemia. Preterm infants are vulnerable to oxidative stress because they lack antioxidant protection (Saugstad

2001) from plasma radical scavengers, such as beta-carotene, and antioxidant enzymes, such as glutathione peroxidase, and their red blood cells and cells of other organs (e.g. lungs) lack superoxide dismutase.

Targeting a higher oxygen level contributes to bronchopulmonary dysplasia (Warner 1998; Jobe 2001; Vento 2009; Kapadia 2013). Relatively recent epidemiological/observational studies (Tin 2001; Sun 2002; Chow 2003; Anderson 2004), and small randomised trials from the 1950s (Askie 2009), have suggested that targeting lower oxygen saturation levels may reduce severe ROP. The effects on death or neurodisability of targeting either lower or higher oxygen saturation levels from birth have not yet been fully assessed.

Why it is important to do this review

Extreme prematurity of less than 28 weeks' gestation affects approximately 1% of births (Centre for Epi 2012). Although approximately 80% of these infants are discharged home alive (Chow 2013), they often sustain severe morbidity (Doyle 2010), including chronic lung disease, poor growth, respiratory illness, hospital re-admissions, visual deficits, cerebral palsy, neurodevelopmental disability and cognitive, educational, and behavioural impairment (Anderson 2003). It is essential to determine whether the range of targeted SpO_2 levels affects the occurrence of such outcomes and, if possible, to determine the optimal range for management of the very vulnerable preterm infant. Very preterm infants account for a high proportion of the costs and disability from neonatal intensive care (Sutton 1999). Reducing these morbidities would enhance quality of life for these infants and benefit their families and communities (Saigal 2000).

OBJECTIVES

1. What are the effects of targeting lower versus higher oxygen saturation ranges on death or major neonatal and infant morbidities, or both, in extremely preterm infants?

2. Do these effects differ in different types of infants, including those born at a very early gestational age, or in those who are outborn, without antenatal corticosteroid coverage, of male sex, small for gestational age or of multiple birth, or by mode of delivery?

METHODS

Criteria for considering studies for this review

Types of studies

We included randomised controlled trials. We excluded quasi-randomised trials and cluster-randomised trials.

Types of participants

Participants in the eligible trials were infants either inborn or outborn before 28 weeks' gestation.

Types of interventions

The intervention was the used of pulse oximetry to target either a lower (SpO_2 less than or equal to 90%) or higher (SpO_2 greater than 90%) oxygen saturation range by 24 hours after birth, maintaining these ranges for at least the first two weeks of life. There was no minimum required level of compliance for keeping within the target ranges. Oxygen targeting could be achieved by either manual or machine-assisted methods.

Types of outcome measures

We assessed longer-term outcomes in infancy from 18 months corrected for gestational age onwards, depending on the measurement time point used by individual trials.

Primary outcomes

- Composite outcome of death or major disability by 18 to 24 months corrected for gestational age (aligned definition, trialist defined)

Secondary outcomes

- Death (to discharge, to 18 to 24 months corrected for gestational age follow-up)
- Major disability by 18 to 24 months corrected for gestational age (aligned definition, trialist defined)
- Retinopathy of prematurity (ROP) treatment by laser photocoagulation, cryotherapy or bevacizumab treatment
- Measures of respiratory support, defined as (a) supplemental oxygen requirement at 36 weeks postmenstrual age (trialist defined), (b) days of endotracheal intubation, (c) days of continuous positive airway pressure (CPAP), (d) days of supplemental oxygen, (e) days on home oxygen
- Patent ductus arteriosus requiring medical treatment (defined as using cyclo-oxygenase inhibitors) or surgical treatment
- Necrotising enterocolitis
- Weight at 36 weeks postmenstrual age, discharge home and 18 or 24 months corrected for gestational age
- Proportion of infants re-admitted to hospital up to 18 to 24 months corrected for gestational age

- Cerebral palsy with Gross Motor Functioning Classification System (GMFCS) level 2 or higher, or Manual Ability Classification System (MACS) level 2 or higher at 18 to 24 months corrected for gestational age
- Blindness
- Severe hearing loss
- Quantitative Bayley III scores (Composite Cognitive Score (CCS) and Composite Language Score (CLS) scores)

Search methods for identification of studies

We used the standard search methods of Cochrane Neonatal.

Electronic searches

We searched the Cochrane Central Register of Controlled Trials (CENTRAL 2016, Issue 4), MEDLINE via PubMed (1966 to 11 April 2016), Embase (1980 to 11 April 2016) and CINAHL (1982 to 11 April 2016). We used the following search strategy:

Search strategy for search dates 1940 to January 1975

Search terms: (oxygen OR oxygen saturation OR hypoxia OR retinopathy of prematurity OR retrolental fibroplasia OR hyperoxia) AND ((Clinical Trial[ptyp] OR Controlled Clinical Trial[ptyp] OR Randomized Controlled Trial[ptyp]) AND ("1940/01/01"[PDat] : "1975/01/01"[PDat]) AND Humans[Mesh] AND (infant, newborn[MeSH] OR infant[MeSH]))

Search strategy for search dates February 1975 to 11 April 2016

Search terms: (oxygen OR oxygen saturation OR hypoxia OR retinopathy of prematurity OR retrolental fibroplasia OR hyperoxia)
We also used the following database-specific terms:
CENTRAL: (infant or newborn or neonate or neonatal or premature or very low birth weight or low birth weight or VLBW or LBW)
PubMed: ((infant, newborn[MeSH] OR newborn OR neonate OR neonatal OR premature OR low birth weight OR VLBW OR LBW or infan* or neonat*) AND (randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized [tiab] OR placebo [tiab] OR drug therapy [sh] OR randomly [tiab] OR trial [tiab] OR groups [tiab]) NOT (animals [mh] NOT humans [mh]))
Embase: (infant, newborn or newborn or neonate or neonatal or premature or very low birth weight or low birth weight or VLBW or LBW or Newborn or infan* or neonat*) AND (human not animal) AND (randomized controlled trial or controlled clinical trial or randomized or placebo or clinical trials as topic or randomly or trial or clinical trial)
CINAHL: (infant, newborn OR newborn OR neonate OR neonatal OR premature OR low birth weight OR VLBW OR LBW or Newborn or infan* or neonat*) AND (randomized controlled trial OR controlled clinical trial OR randomized OR placebo OR clinical trials as topic OR randomly OR trial OR PT clinical trial)
We searched for any ongoing or recently completed and unpublished trials, using the World Health Organization portal (www.who.int/ictrp).
We did not apply any language, date or publication status restrictions.

Searching other resources

We searched previous reviews and cross-references, abstracts, and conference and symposia proceedings. We contacted expert informants and carried out journal handsearching. We searched the abstracts of the relevant perinatal meetings (including Society for Pediatric Research, Neonatal Register) for the years 1985 to the present, using the following keywords: 'oxygen saturation'. For abstract books that did not include keywords, we limited the search to the relevant sections, such as pulmonology and neonatology.

Data collection and analysis

We used the methods of Cochrane Neonatal for data collection and analysis.

Selection of studies

LA, BD, and PD independently reviewed the results of the search and selected studies for inclusion.

Data extraction and management

We used a data extraction form specifically designed for this review. We collected information on the following outcome variables:
- composite outcome of death or major disability by 18 to 24 months corrected for gestational age;
- death (to discharge, to 18 to 24 months follow-up);
- major disability by 18 to 24 months corrected for gestational age;
- ROP treatment by laser photocoagulation, cryotherapy or bevacizumab treatment (performed if threshold ROP occurs);
- measures of respiratory support, defined as (a) supplemental oxygen requirement at 36 weeks postmenstrual age, (b) days of endotracheal intubation, (c) days of CPAP, (d) days of supplemental oxygen, (e) days on home oxygen; patent ductus arteriosus requiring medical treatment (defined as using cyclo-oxygenase inhibitors) or surgical treatment;
- necrotising enterocolitis (trialist defined);
- weight at 36 weeks postmenstrual age, discharge home and 18 and 24 months corrected for gestational age;

- re-admissions to hospital up to 18 to 24 months corrected for gestational age;
- cerebral palsy with GMFCS level 2 or higher or MACS level 2 or higher at 18 to 24 months corrected for gestational age;
- blindness (< 6/60 vision, 1.3 logMAR in both eyes);
- severe hearing loss;
- quantitative Bayley III scores.

We resolved differences in assessment by discussion or by involving the remaining review authors.

For each study, one review author (LA) extracted, assessed, and coded all data for each included study and entered final data into Review Manager 5 (RevMan 2014). A second review author (RW) checked these data for accuracy and each of the authors from the included trials also checked the accuracy of their own trial data. We resolved discrepancies through discussion. We contacted the authors of the original reports to provide further details when information regarding any of the above was unclear.

Assessment of risk of bias in included studies

We used the standard methods of Cochrane Neonatal. LA and RW independently assessed the risk of bias for each study using the criteria outlined in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). RW prepared the initial 'Risk of bias' tables for discussion. Disagreements were resolved by discussion or by involving the remaining review authors and members of the NeOProM Collaboration.

We assessed the methodological quality of the studies using the following criteria:

- Sequence generation (checking for possible selection bias). For each included study, we categorised the method used to generate the allocation sequence as:
  - low risk (any truly random process, e.g. random number table; computer random number generator);
  - high risk (any non-random process, e.g. odd or even date of birth; hospital or clinic record number);
  - unclear risk.
- Allocation concealment (checking for possible selection bias). For each included study, we categorised the method used to conceal the allocation sequence as:
  - low risk (e.g. telephone or central randomisation; consecutively numbered, sealed, opaque envelopes);
  - high risk (open random allocation; unsealed or non-opaque envelopes; alternation; date of birth);
  - unclear risk.
- Blinding (checking for possible performance bias). For each included study, we categorised the methods used to blind study participants and personnel from knowledge of which intervention a participant received. We assessed blinding separately for different outcomes or classes of outcomes. We categorised the methods as:
  - low risk, high risk or unclear risk for participants;
  - low risk, high risk or unclear risk for personnel;
  - low risk, high risk or unclear risk for outcome assessors.
- Incomplete outcome data (checking for possible attrition bias through withdrawals, dropouts, protocol deviations). For each included study and for each outcome, we described the completeness of data including attrition and exclusions from the analysis. We noted whether attrition and exclusions were reported, the numbers included in the analysis at each stage (compared with the total randomised participants), reasons for attrition or exclusion where reported, and whether missing data were balanced across groups or were related to outcomes. We categorised the methods as:
  - low risk (less than 20% missing data);
  - high risk (20% or more missing data);
  - unclear risk.
- Selective reporting bias. For each included study, we described how we investigated the possibility of selective outcome reporting bias and what we found. We assessed the methods as:
  - low risk (where it is clear that all of the study's prespecified outcomes and all expected outcomes of interest to the review have been reported);
  - high risk (where not all the study's prespecified outcomes have been reported; one or more reported primary outcomes were not prespecified; outcomes of interest are reported incompletely and so cannot be used; study fails to include results of a key outcome that would have been expected to have been reported);
  - unclear risk.
- Other sources of bias. For each included study, we described any important concerns we had about other possible sources of bias (for example, whether there was a potential source of bias related to the specific study design or whether the trial was stopped early due to some data-dependent process or interim analyses issues, or both). We assessed whether each study was free of other problems that could put it at risk of bias as follows:
  - low risk;
  - high risk;
  - unclear risk.
- Overall risk of bias (described in Table 8.5c in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011)).

We made explicit judgements regarding whether studies were at high risk of bias, according to the criteria given in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). With reference to (1) to (6) above, we assessed the likely magnitude and direction of the bias and whether we considered it is likely to have an impact on the findings. If need be, we planned to explore the impact of the level of bias through undertaking sensitivity analyses (see Sensitivity analysis below).

Measures of treatment effect

We used the standard methods of Cochrane Neonatal to analyse data.

We performed statistical analyses using Review Manager 5 software (RevMan 2014). We analysed dichotomous data using the risk ratio (RR), risk difference (RD), and the number needed to treat for an additional beneficial outcome (NNTB) or number needed to treat for an additional harmful outcome (NNTH). We included the 95% confidence interval (CI) for all estimates. For the purposes of the analysis, we considered the lower target range group to be the treatment or experimental group and the higher target range group to be the control group.

We analysed continuous data using the mean difference (MD) or the standardised mean difference (SMD) to combine trials that measured the same outcome but used different methods.

Unit of analysis issues

The unit of analysis was each infant randomised.

Dealing with missing data

For the included studies we noted levels of attrition. We explored the impact of including studies with high levels of missing data in the overall assessment of treatment effect using sensitivity analyses. We analysed all outcomes on an intention-to-treat basis, i.e. we included all participants randomised to each group in the analyses. We did not replace missing data by imputation: the denominator for each outcome in each trial was the number randomised minus any participants whose outcomes were known to be missing.

Assessment of heterogeneity

We examined heterogeneity between trials by first assessing differences in trial methodologies and clinical heterogeneity. If we judged clinical heterogeneity to be absent, we then quantified the impact of heterogeneity using the I^2 statistic (less than 25%, no heterogeneity; 25% to 49%, low heterogeneity; 50% to 74%, moderate heterogeneity; and 75% or more, high heterogeneity) (Higgins 2003; Higgins 2011). If heterogeneity was detected, we explored the possible causes of statistical heterogeneity using prespecified subgroup analysis (for example, differences in study quality, participant or intervention characteristics).

Assessment of reporting biases

We assessed possible publication bias and other biases using symmetry/asymmetry of funnel plots, if there were sufficient trials to allow these analyses.

For included trials that were recently performed (and therefore prospectively registered), we explored possible selective reporting of study outcomes by comparing the primary and secondary outcomes in the reports with the primary and secondary outcomes proposed at trial registration, using the website www.who.int/ictrp. If we found such discrepancies, we contacted the primary investigators to interpret variances with outcomes prespecified at trial registration.

Data synthesis

We conducted the meta-analysis using Review Manager 5 software (RevMan 2014), supplied by Cochrane. We used the Mantel-Haenszel method for estimates of typical risk ratio and risk difference. We analysed any continuous measures using the inverse variance method. We used the fixed-effect model for all meta-analyses.

Quality of evidence

We used the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, as outlined in the *GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations* (Schünemann 2013), to assess the quality of evidence for the following major outcomes: death or major disability by 18 to 24 months corrected age (aligned definition); death to 18 to 24 months corrected age; major disability by 18 to 24 months corrected age (aligned definition); retinopathy of prematurity requiring treatment; necrotising enterocolitis; blindness.

Two authors (LA, RW) independently assessed the quality of the evidence for each of the outcomes above. The full author group further considered the 'Risk of bias' assessments and discussed these in detail at a NeOProM Collaborators meeting on 29 April 2016 in order to reach consensus. We considered evidence from randomised controlled trials as high quality but downgraded the evidence by one level for serious (or by two levels for very serious) limitations based upon the following: design (risk of bias), consistency across studies, directness of the evidence, precision of estimates and presence of publication bias. We used the GRADEpro GDT Guideline Development Tool to create a 'Summary of findings' table to report the quality of the evidence.

The GRADE approach results in an assessment of the quality of a body of evidence of one of four grades:

- High: We are very confident that the true effect lies close to that of the estimate of the effect.
- Moderate: We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
- Low: Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
- Very low: We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Subgroup analysis and investigation of heterogeneity

The effect of the intervention (lower versus higher oxygen saturation targeting) may be different due to certain characteristics of

either the infant or the way the intervention was delivered. If possible, we explored these effects in the following subgroup analyses.

Participant baseline characteristics

- Gestational age (less than 26 weeks/26 weeks or more)
- Inborn or outborn status
- Antenatal steroids (any: yes/no)
- Sex (male/female)
- Small for gestational age (yes/no)
- Multiples (singleton/multiple)
- Mode of delivery (vaginal/caesarean)

Intervention characteristics

- Oximeter calibration software (original or revised)

We limited subgroup analyses to the primary outcome (death or major disability), the individual components of the primary outcome (death, major disability), and two key secondary outcomes (necrotising enterocolitis, retinopathy of prematurity requiring treatment). All subgroup analyses were pre-specified before any meta-analyses of combined data were undertaken.

Sensitivity analysis

We planned sensitivity analyses for situations where this might affect the interpretation of significant results (for example, where there are risks of bias associated with the quality of some of the included trials or missing outcome data).

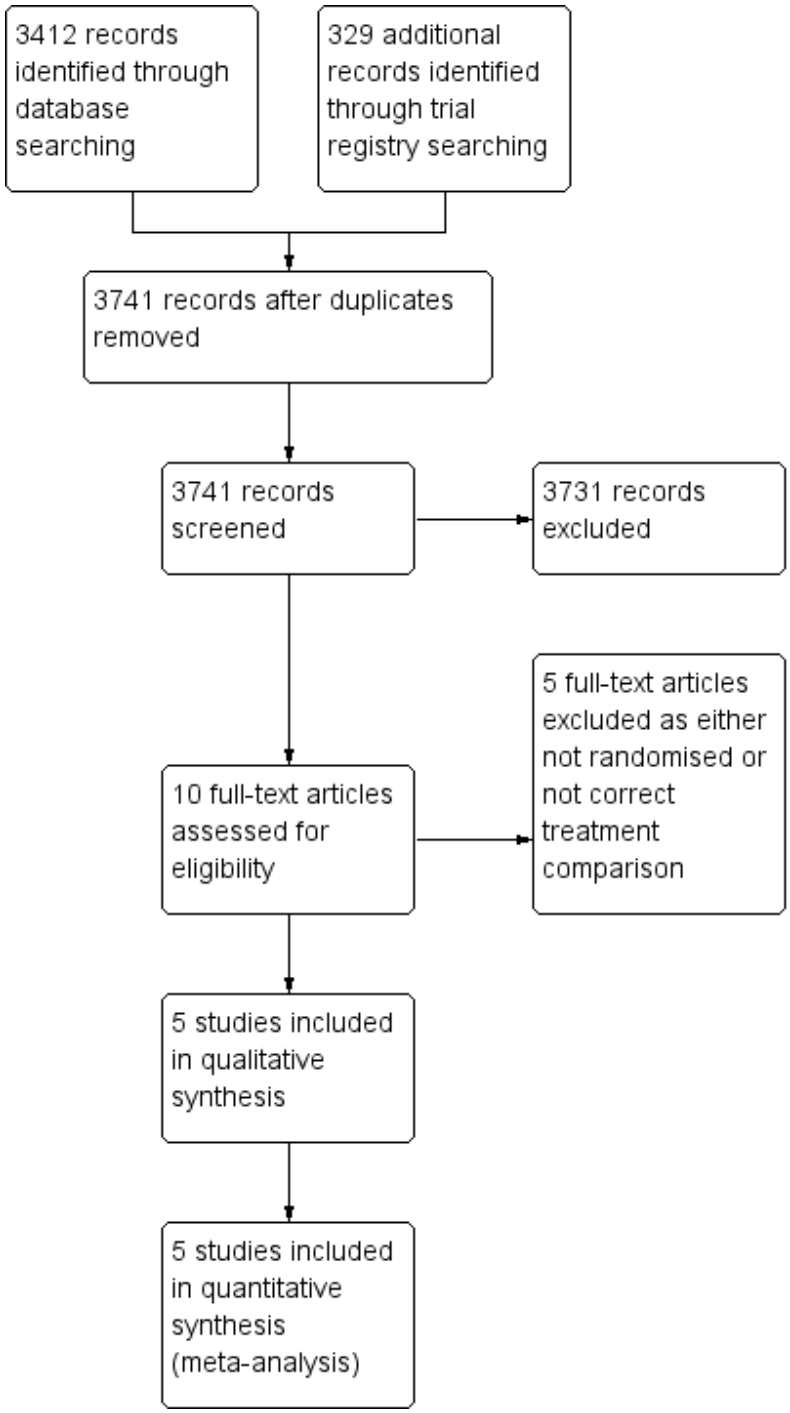
RESULTS

Description of studies

Results of the search

We searched using the search strategy outlined previously from 1940 to 11 April 2016 and identified 3412 potential published studies and 329 trial registration records. LA and BD screened the citations and abstracts of these 3741 records in duplicate for the initial search covering the period 1940 to 14 May 2014, and LA and PD screened these for the updated search covering the period from 1 January 2014 to 11 April 2016 (see Figure 1). There were no disagreements regarding study eligibility that required resolution.

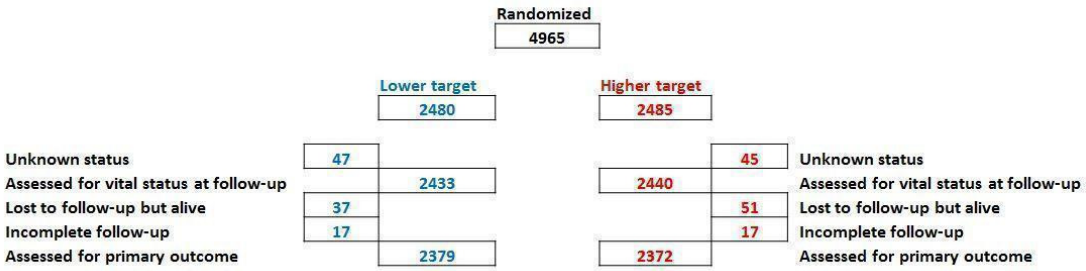
Figure 1. Study flow diagram



Included studies

We identified five trials meeting the inclusion criteria of the review (n = 4965 infants) (Vaucher 2012; Schmidt 2013; BOOST NZ 2014; BOOST-II UK 2016; BOOST-II Australia 2016). Details are included in the Characteristics of included studies table and Figure 2.

Figure 2. Participant flow chart for the combined five NeOProM trials



The inclusion criteria were similar between trials. All trials enrolled preterm infants of less than 28 weeks' gestation, with one specifying a minimum gestation of 24 weeks' (Vaucher 2012) and another 23 weeks' (Schmidt 2013). Three trials also required infants to be less than 24 hours old (Schmidt 2013; BOOST NZ 2014; BOOST-II Australia 2016), one required the infants to be less than 12 hours old (or less than 24 hours if outborn) (BOOST-II UK 2016), and one trial required infants to be enrolled by two hours of age (Vaucher 2012). In all five trials the intervention and comparator were the same, i.e. lower oxygen saturation targeting

(SpO2 85% to 89%) versus higher oxygen saturation targeting

(SpO2 91% to 95%), although recommendations for alarm limit settings differed between trials (see Characteristics of included studies table).

Excluded studies

We assessed four studies (one published study, one conference proceeding, and two trial registration records) in full but excluded them as they were either not randomised trials or did not fulfil the other eligibility criteria for the interventions being compared, but this had not been clear from the title or abstract of the identified record (see the Characteristics of excluded studies table for the reasons for these exclusions). There were no other identified ongoing studies or studies awaiting clarification that were potentially eligible for inclusion.

Risk of bias in included studies

We assessed methodological quality using the criteria of Cochrane Neonatal and the findings are summarised in Figure 3.

Figure 3. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

Allocation

Random sequence generation: all five of the included trials met this criterion.

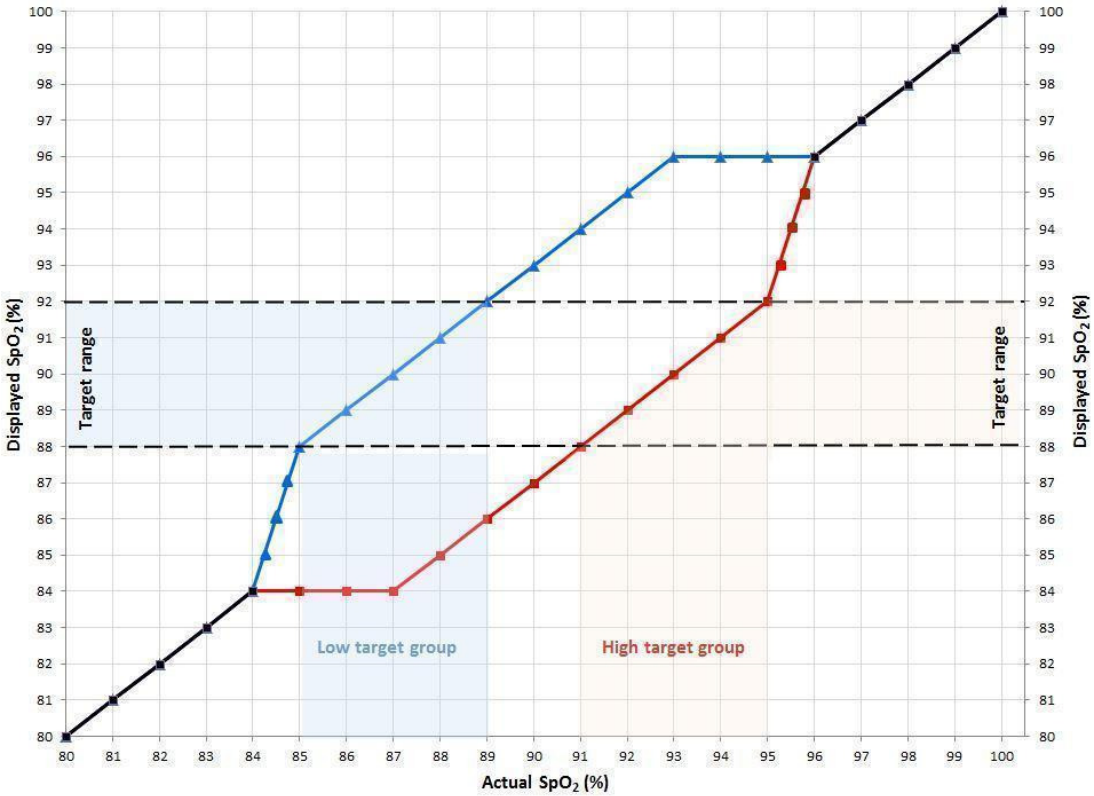
Allocation concealment: this was adequate in all five of the trials.

All five trials were judged as having a low risk of bias for this criterion.

Blinding

Blinding of participants and personnel: this was adequate in all five of the trials with all using the same masked oximeters (see Figure 4 for a diagramatic representation of how masking was achieved within the trials).

Figure 4. Oximeter offset to achieve masking as used in the five NeOProM trials



Blinding of outcome assessments: this was adequate in all five of the trials with outcome assessors remaining masked to treatment allocation.

All five trials were judged as having a low risk of bias for this criterion.

Incomplete outcome data

We assessed incomplete outcome data (attrition bias) as low risk of bias in all five trials. The primary outcome was a composite outcome of death or major disability by 18 to 24 months corrected for gestational age. Survival status was known for 97% to 100% of all infants in each trial. Major disability was defined as severe visual loss, deafness requiring hearing aids, cerebral palsy by various measures, or as a Bayley-III Developmental Assessment (BSID-III) cognitive score < 85 and/or language score < 85. There was variation between trials in the completeness of primary outcome data gathered according to the BSID-III assessments. By this measure between 79% and 96% had adequate data for the analysis of the composite primary outcome within the included trials (see Characteristics of included studies table). In two trials (Vaucher 2012 and Schmidt 2013, with 94% and 96% recovered data respectively), there was no imputation of missing data. In the three other trials various alternative measures of developmental status were substituted (BOOST NZ 2014; BOOST-II UK 2016; BOOST-II Australia 2016). With this qualification, all trials reported greater than 94% recovery of data for the primary analysis. These data, with protocol defined and alternative measures, were used in the definitions of the primary outcome and of major disabilities.

Selective reporting

We assessed selective reporting (reporting bias) as low risk in all five trials, with all trials reporting their pre-specified primary outcome and main secondary outcomes, and the trials identified any post hoc analyses in their trial reports.

Other potential sources of bias

Investigator concerns resulting from the significantly increased mortality risk with the lower SpO_2 target range in the SUPPORT Trial publication led to an unscheduled safety analysis when 1135 of the planned 1200 (95%) BOOST-II Australia and 973 of the planned 1200 (81%) BOOST-II UK infants had been recruited (BOOST-II Australia 2016; BOOST-II UK 2016). A decision was made to terminate recruitment in both the BOOST-II UK and BOOST-II Australia trials based on a pre-specified rule. There was an 8.5% excess in 36-week mortality in the low target group monitored with oximeters incorporating the revised calibration software (data pooled from both studies, P < 0.001 with a significant treatment by software subgroup interaction, P = 0.006) (Stenson 2011). The early stopping of these two trials (with 95% and 81% of their final planned sample sizes at that point) raises the question of whether this overestimates treatment effect, and thus the risk of bias was categorised as 'unclear' for the BOOST-II UK and BOOST-II Australia trials. The other three trials were assessed as low risk of bias for this criterion.

Effects of interventions

See: **Summary of findings for the main comparison** Oxygen saturation targeting in preterm infants

After considerable discussion between the review authors, we reached consensus regarding the quality of evidence assessments and data extraction from the five included trials.

Lower versus higher targeted oxygen saturations (no subgroups)

For the purpose of these analyses, we considered the lower target range group the intervention group and we considered the higher target range group the control group.

The SUPPORT trial used a different Bayley-III cut point (< 70) and components (composite cognitive score only) as part of their definition of major disability compared with the other four trials (which used Bayley-III < 85 on either the composite cognitive or language scores). The SUPPORT trial provided unpublished outcome data using the same Bayley-III cut point and components for inclusion in this review in order to better align the definitions of major disability across all five trials. For the outcomes where these data were used, we used the term 'aligned definition'. However, the definition of major disability was not 'fully' aligned as each trial used slightly different methods for assigning an outcome of major disability. The term 'trialist defined' used throughout this review indicates analyses that used data as published in the trial reports. For the two outcomes where these data are relevant ('death or major disability' and 'major disability'), both the 'aligned definition' and 'trialist defined' versions are presented. However, our primary analyses are based on the meta-analyses of updated data, using the aligned definition of major disability. The SUPPORT trial also provided unpublished data for inclusion in this review to align the definition of 'retinopathy of prematurity requiring treatment' across all trials.

We graded the quality of evidence for outcomes as high (low risk of bias, low inconsistency, no indirectness, low imprecision, and low risk of reporting bias) unless otherwise stated below.

Primary outcome

Death or major disability by 18 to 24 months corrected age (aligned definition) (outcome 1.1)

Using the aligned definition of this outcome, there was no significant difference between the groups in the incidence of death or major disability by 18 to 24 months corrected age (typical risk ratio (RR) 1.04, 95% confidence interval (CI) 0.98 to 1.10; typical risk difference (RD) 0.02, 95% CI -0.01 to 0.05; 5 trials, 4754 infants; I^2 = 27%) (high-quality evidence) (Analysis 1.1).

Death or major disability by 18 to 24 months corrected age (trialist defined) (outcome 1.2)

The meta-analysis showed a statistically significant increased risk of death or major disability at 18 to 24 months corrected age for the lower target group when using the trialists' own definitions of major disability (typical RR 1.07, 95% CI 1.00 to 1.14; typical RD 0.03, 95% CI 0.00 to 0.06; 5 trials, 4751 infants; I^2 = 1%) (Analysis 1.2).

Secondary outcomes

Death to 18 to 24 months corrected age (outcome 1.3)

The meta-analysis showed a statistically significant increased risk of death at 18 to 24 months corrected age for the lower target group (typical RR 1.16, 95% CI 1.03 to 1.31; typical RD 0.03, 95% CI 0.01 to 0.05; 5 trials, 4873 infants; I^2 = 0%) (high-quality evidence) (Analysis 1.3). The number needed to treat for an additional harmful outcome (NNTH) with lower targeting to result in one additional death by 18 to 24 months corrected age is 31 (95% CI 16 to 168).

Major disability by 18 to 24 months corrected age (aligned definition) (outcome 1.4)

There was no significant difference between the groups in the incidence of major disability by 18 to 24 months corrected age when using the aligned definition of major disability (see Outcome 1.2 above) across trials (typical RR 1.01, 95% CI 0.93 to 1.09; 5 trials, 3867 infants; I^2 = 22%) (high-quality evidence) (Analysis 1.4).

Major disability by 18 to 24 months corrected age (trialist defined) (outcome 1.5)

There was no significant difference between the groups in the incidence of major disability by 18 to 24 months corrected age when using the trialists' own definitions of major disability (typical RR 1.04, 95% CI 0.94 to 1.14; 5 trials, 3864 infants; I^2 = 0%) (Analysis 1.5).

Death to discharge (outcome 1.6)

The meta-analysis showed a statistically significant increased risk of death before discharge for the lower target group (typical RR 1.16, 95% CI 1.03 to 1.31; typical RD 0.03, 95% CI 0.00 to 0.05; 5 trials, 4958 infants; I^2 = 0%) (Analysis 1.6). The NNTH with lower targeting to result in one additional death to discharge is 34 (95% CI 17 to 180).

Severe retinopathy of prematurity (ROP) or retinal therapy (trialists defined) (outcome 1.7)

The meta-analysis showed a statistically significant reduced risk of severe ROP or retinal therapy for the lower target group (typical RR 0.72, 95% CI 0.61 to 0.85; typical RD -0.04, 95% CI -0.06 to -0.02; 5 trials, 4089 infants; I^2 = 69%) (moderate-quality evidence) (Analysis 1.7). We rated the quality of evidence for this outcome as moderate due to inconsistency (moderate heterogeneity). The number needed to treat for an additional beneficial outcome (NNTB) with lower oxygen targeting to prevent one additional case of severe ROP or retinal therapy is 34 (95% CI 21 to 63).

Patent ductus arteriosus requiring medical or surgical treatment (outcome 1.8)

There was no significant difference between the groups in the incidence of patent ductus arteriosus requiring medical or surgical treatment (typical RR 1.00, 95% CI 0.95 to 1.06; 5 trials, 4928 infants; I^2 = 0%) (Analysis 1.8).

Necrotising enterocolitis (outcome 1.9)

Three trials (BOOST Australia, NZ, and UK) used a definition of 'necrotising enterocolitis requiring surgery or leading to death' for this outcome. The other two trials used either a modified Bell's staging (SUPPORT) or surgical or X-ray diagnoses (COT) to define this outcome. The meta-analysis showed a statistically significant increased risk of necrotising enterocolitis for the lower target group (typical RR 1.24, 95% CI 1.05 to 1.47; typical RD 0.02, 95% CI 0.01 to 0.04; 5 trials, 4929 infants; I^2 = 0%) (high-quality evidence) (Analysis 1.9). The NNTH with lower targeting to result in one additional case of necrotising enterocolitis is 37 (95% CI 19 to 178).

Cerebral palsy with Gross Motor Functioning Classification System (GMFCS) level 2 or higher at 18 to 24 months corrected age (outcome 1.10)

There was no significant difference between the groups in the incidence of cerebral palsy with GMFCS level 2 or higher at 18 to 24 months corrected age (typical RR 1.02, 95% CI 0.79 to 1.32; 5 trials, 3877 infants; I^2 = 20%) (Analysis 1.10).

Blindness (outcome 1.11)

There was no significant difference between the groups in the incidence of blindness (typical RR 1.13, 95% CI 0.65 to 1.97; 5 trials, 3875 infants; I^2 = 0%) (Analysis 1.11). We rated the quality of evidence for this outcome as moderate for imprecision due to low event rates.

Severe hearing loss (outcome 1.12)

There was no significant difference between the groups in the incidence of severe hearing loss (typical RR 1.02, 95% CI 0.73 to 1.43; 5 trials, 3869 infants; I^2 = 0%) (Analysis 1.12). We rated the quality of evidence for this outcome as moderate for imprecision due to low event rates.

Proportion of infants re-admitted to hospital up to 18 to 24 months corrected age (outcome 1.13)

There was no significant difference between the groups in the proportion of infants re-admitted to hospital up to 18 to 24 months corrected age (RR 1.08, 95% CI 0.93 to 1.26; 1 trial, 295 infants; I^2 not applicable) (Analysis 1.13). We rated the quality of evidence for this outcome as moderate for imprecision as data were available from only one trial.

Weight (grams) at discharge home (outcome 1.14)

There was no significant difference between the groups in the weight (grams) of infants at discharge home (mean difference (MD) -52.00, 95% CI -214.25 to 110.25; P = 0.53; 1 trial, 295 infants; I^2 not applicable) (Analysis 1.14). We rated the quality of evidence for this outcome as moderate for imprecision as data were available from only one trial.

Weight (kilograms) at 18 or 24 months corrected age (outcome 1.15)

There was no significant difference between the groups in the weight (grams) of infants at discharge home (MD 0.80, 95% CI -0.24 to 1.84; 1 trial, 280 infants; I^2 not applicable) (Analysis 1.15). We rated the quality of evidence for this outcome as moderate for imprecision as data were available from only one trial.

Days of endotracheal intubation (outcome 1.16)

There was no significant difference between the groups in days of endotracheal intubation (MD 0.28, 95% CI -1.16 to 1.72; 2 trials, 1386 infants; I^2 = 0%) (Analysis 1.16).

Days of continuous positive airway pressure (CPAP) (outcome 1.17)

There was no significant difference between the groups in days of CPAP (MD -0.04, 95% CI -1.38 to 1.30; 3 studies, 2526 infants; I^2 = 0%) (Analysis 1.17).

Days of supplemental oxygen (outcome 1.18)

The meta-analysis showed a statistically significant reduction in days of supplemental oxygen for the lower target group (MD -8.78, 95% CI -12.02 to -5.54; P < 0.00001; 3 trials, 2507 infants; I^2 = 4%) (Analysis 1.18).

Supplemental oxygen requirement at 36 weeks postmenstrual age (outcome 1.19)

This outcome was determined using a physiologic test in the Vaucher 2012 and BOOST-II UK 2016 trials while the other three trials determined the need for supplemental oxygen at 36 weeks postmenstrual age using the infant's assigned study oximeter.
The meta-analysis showed a statistically significant reduced risk of supplemental oxygen requirement at 36 weeks postmenstrual age for the lower target group (typical RR 0.87, 95% CI 0.81 to 0.94; typical RD -0.06, 95% CI -0.09 to -0.03; 5 trials, 4175 infants; I^2 = 44%) (Analysis 1.19). We rated the quality of evidence for this outcome as moderate due to inconsistency (moderate heterogeneity). The NNTB with lower oxygen targeting to prevent one additional infant receiving supplemental oxygen at 36 weeks postmenstrual age is 20 (95% CI 14 to 44).

Days on home oxygen (outcome 1.20)

For the infants who went home on supplemental oxygen (n = 237), there was no significant difference between the groups in the days of home oxygen (MD -24.17, 95% CI -57.99 to 9.66; P = 0.16; 2 trials, 237 infants; I^2 = 61%) (Analysis 1.20). We rated the quality of evidence for this outcome as low due to inconsistency (moderate heterogeneity) and imprecision.

Quantitative Bayley III scores (Composite Cognitive Score (CCS)) (outcome 1.21)

There was no significant difference between the groups in quantitative Bayley III scores (Composite Cognitive Score (CCS)) (MD 0.55, 95% CI -0.91 to 2.00; P = 0.46; 2 trials, 1892 infants; I^2 = 0%) (Analysis 1.21).

Quantitative Bayley III scores (Composite Language Score (CLS)) (outcome 1.22)

There was no significant difference between the groups in quantitative Bayley III scores (Composite Language Score (CLS)) (MD 0.20, 95% CI -2.03 to 2.43; 1 trial, 903 infants; P = 0.86; I^2 not applicable) (Analysis 1.22). We rated the quality of evidence for this outcome as moderate for imprecision as data were available from only one trial.

Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by gestational age)

Death or major disability by 18 to 24 months corrected age (trialist defined) (outcome 2.1)

There was no statistically significant difference in the incidence of death or major disability by 18 to 24 months corrected age (trialist defined) in the subgroup of infants born at < 26 weeks' gestation (RR 1.09, 95% CI 0.89 to 1.32; 1 trial, 537 infants; I^2 not applicable) compared with infants born at ≥ 26 weeks' gestation (RR 1.17, 95% CI 0.86 to 1.60; 1 trial, 697 infants; I^2 not applicable): test for subgroup difference P = 0.69. We rated the quality of evidence for this outcome as moderate for imprecision as data were available from only one trial (Analysis 2.1).

Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by sex)

Death or major disability by 18 to 24 months corrected age (trialist defined) (outcome 3.1)

There was no statistically significant difference in the incidence of death or major disability by 18 to 24 months corrected age (trialist defined) in the subgroup of male infants (RR 1.13, 95% CI 0.96 to 1.33; 1 trial, 503 infants; I^2 not applicable) compared with female infants (RR 1.07, 95% CI 0.87 to 1.31; 1 trial, 438 infants; I^2 not applicable): test for subgroup difference P = 0.66. We rated the quality of evidence for this outcome as moderate for imprecision as data were available from only one trial (Analysis 3.1).

Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by multiples)

Death or major disability by 18 to 24 months corrected age (trialist defined) (outcome 4.1)

There was no statistically significant difference in the incidence of death or major disability by 18 to 24 months corrected age (trialist defined) in the subgroup of singleton infants (RR 1.10, 95% CI 0.94 to 1.29; 1 trial, 670 infants; I^2 not applicable) compared with infants from multiple births (RR 1.11, 95% CI 0.88 to 1.39; 1 trial, 271 infants; I^2 not applicable): test for subgroup difference P = 0.94. We rated the quality of evidence for this outcome as moderate for imprecision as data were available from only one trial (Analysis 4.1).

Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by oximeter calibration software)

Death or major disability by 18 to 24 months corrected age (aligned definition) (outcome 5.1)

There was no statistically significant difference in the incidence of death or major disability by 18 to 24 months corrected age (aligned definition) in the subgroup of infants who used the original oximeter calibration algorithm (typical RR 1.00, 95% CI 0.94 to 1.07; 5 trials, 3003 infants; I^2 = 0%) compared with the infants who used the revised algorithm (typical RR 1.13, 95% CI 1.02 to 1.24; 3 trials, 1681 infants; I^2 = 0%): test for subgroup difference P = 0.06 (Analysis 5.1).

Death or major disability by 18 to 24 months corrected age (trialist defined) (outcome 5.2)

There was no statistically significant difference in the incidence of death or major disability by 18 to 24 months corrected age (trialist defined) in the subgroup of infants who used the original oximeter calibration algorithm (typical RR 1.04, 95% CI 0.95 to 1.13; 5 trials, 3000 infants; I^2 = 0%) compared with infants who used the revised oximeter calibration algorithm (typical RR 1.13, 95% CI 1.02 to 1.24; 3 trials, 1681 infants; I^2 = 0%): test for subgroup difference P = 0.24 (Analysis 5.2).

Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by oximeter calibration software)

Death by 18 to 24 months corrected age (outcome 6.1)

There was a statistically significant difference in the incidence of death by 18 to 24 months corrected age in the subgroup of infants who used the original oximeter calibration algorithm (typical RR 1.05, 95% CI 0.91 to 1.22; 5 trials, 3087 infants; I^2 = 5%) compared with infants who used the revised oximeter calibration algorithm (typical RR 1.38, 95% CI 1.13 to 1.68; 3 trials, 1716 infants; I^2 = 0%): test for subgroup difference P = 0.03 (Analysis 6.1).

Major disability by 18 to 24 months corrected age (aligned definition) (outcome 6.2)

There was no statistically significant difference in the incidence of major disability by 18 to 24 months corrected age (aligned definition) in the subgroup of infants who used the original oximeter calibration algorithm (typical RR 0.99, 95% CI 0.90 to 1.09; 5 trials, 2529 infants; I^2 = 40%) compared with infants who used

the revised oximeter calibration algorithm (typical RR 1.05, 95% CI 0.91 to 1.22; 3 trials, 1438 infants; I^2 = 0%): test for subgroup difference P = 0.49 (Analysis 6.2).

Major disability by 18 to 24 months corrected age (trialist defined) (outcome 6.3)

There was no statistically significant difference in the incidence of major disability by 18 to 24 months corrected age (trialist defined) in the subgroup of infants who used the original oximeter calibration algorithm (typical RR 1.02, 95% CI 0.89 to 1.17; 5 trials, 2526 infants; I^2 = 30%) compared with infants who used the revised oximeter calibration algorithm (typical RR 1.05, 95% CI 0.91 to 1.22; 3 trials, 1438 infants; I^2 = 0%): test for subgroup difference P = 0.78 (Analysis 6.3).

Death to discharge (outcome 6.4)

There was a statistically significant difference in the incidence of death to discharge in the subgroup of infants who used the original oximeter calibration algorithm (typical RR 1.06, 95% CI 0.90 to 1.26; 4 trials, 2575 infants; I^2 = 37%) compared with infants who used the revised oximeter calibration algorithm (typical RR 1.45, 95% CI 1.15 to 1.84; 2 trials, 1182 infants; I^2 = 0%): test for subgroup difference P = 0.04 (Analysis 6.4).

Severe retinopathy of prematurity or retinal therapy (trialist defined) (outcome 6.5)

There was no statistically significant difference in the incidence of severe retinopathy of prematurity or retinal therapy in the subgroup of infants who used the original oximeter calibration algorithm (typical RR 0.67, 95% CI 0.53 to 0.84; 4 trials, 2085 infants; I^2 = 67%) compared with infants who used the revised oximeter calibration algorithm (typical RR 0.77, 95% CI 0.56 to 1.05; 2 trials, 988 infants; I^2 = 55%): test for subgroup difference P = 0.47 (Analysis 6.5).

Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by gestational age)

Death by 18 to 24 months corrected age (outcome 7.1)

There was no statistically significant difference in the incidence of death by 18 to 24 months corrected age in the subgroup of infants born at < 26 weeks' gestation (RR 1.22, 95% CI 0.95 to 1.57; 1 trial, 550 infants; I^2 = not applicable) compared with infants born at ≥ 26 weeks' gestation (RR 1.25, 95% CI 0.84 to 1.86; 1 trial, 731 infants; I^2 = not applicable): test for subgroup difference P = 0.91. We rated the quality of evidence for this outcome as moderate for imprecision as data were available from only one trial (Analysis 7.1).

Major disability by 18 to 24 months corrected age (trialist defined) (outcome 7.2)

There was no statistically significant difference in the incidence of major disability by 18 to 24 months corrected age in the subgroup of infants born at < 26 weeks' gestation (RR 0.84, 95% CI 0.52 to 1.37; 1 trial, 367 infants; I^2 = not applicable) compared with infants born at ≥ 26 weeks' gestation (RR 1.07, 95% CI 0.59 to 1.93; 1 trial, 609 infants; I^2 = not applicable): test for subgroup difference P = 0.54. We rated the quality of evidence for this outcome as moderate for imprecision as data were available from only one trial (Analysis 7.2).

DISCUSSION

Summary of main results

When compared to targeting a higher oxygen saturation range (SpO_2 91% to 95%) from soon after birth in extremely preterm infants, targeting a lower oxygen saturation range (SpO_2 85% to 89%) had no significant effect on the primary composite outcome of death or major disability (using an aligned definition) (P = 0.18), but significantly increased the incidence of death at 18 to 24 months corrected age (P = 0.01), death before hospital discharge (P = 0.02), and necrotising enterocolitis (P = 0.01). Targeting the lower range significantly decreased the rate of severe or treated retinopathy of prematurity (ROP) (P = 0.004) and supplemental oxygen use at 36 weeks postmenstrual age (P = 0.0002), but had no significant effect on blindness (P = 0.65), patent ductus arteriosus requiring treatment (P = 0.87), or major disability (when an aligned definition across trials was used) at 18 to 24 months corrected age (P = 0.80).

Subgroup analyses of major outcomes by oximeter calibration software (original versus revised) showed a statistically significant difference in death by 18 to 24 months corrected age: original algorithm (P = 0.52) versus revised algorithm (P = 0.001) (test for subgroup difference P = 0.03). A similar result was seen for death before hospital discharge (test for subgroup difference P = 0.04). There were no other participant or intervention characteristics within the pre-specified subgroup analyses that showed significant differences in the treatment effect, although data for many are not yet currently available from the published trials.

The five trials included in this review were prospectively planned to be similar with regards to the enrolled participants, the interventions compared, and the outcomes measured and they were designed to be included in a meta-analysis once the individual trials were completed. As would thus be expected, the heterogeneity seen across the trials for most outcomes was low. Similarly there was an overall low risk of bias across all five trials. In the meta-

analysis the combined sample size of 4965 provides high-quality, robust evidence for important outcomes with regards to oxygen saturation targeting in extremely preterm infants.

Overall completeness and applicability of evidence

Data were available and complete from all five included trials for the primary outcome (death or major disability at 18 to 24 months corrected age), and the components of the primary outcome: death, major disability, cerebral palsy, blindness, and severe hearing loss, as well as other important neonatal outcomes including death before hospital discharge, severe ROP, patent ductus arteriosus requiring treatment, and necrotising enterocolitis. Upon request, trialists provided unpublished data (SUPPORT for retinopathy requiring treatment and major disability with definition aligned with the other trials, and COT for death before discharge) to improve the completeness of the included data for the major outcomes.

Other secondary outcomes (1.13 to 1.23) and subgroup analyses (small for gestational age, multiples, antenatal steroids, inborn status) data are either incomplete or not available from the current trial publications. The included trials are all members of the NeOProM Collaboration, which is currently collating and analysing the individual participant data (IPD) from all five trials, including data for the outcomes pre-specified, but not yet available in this review. It is anticipated that once the NeOProM analyses have been published, this review will be updated with the additional information.

Supplemental oxygen delivery and monitoring in extremely preterm infants in the first days/weeks of life can be particularly challenging. How well a target oxygen saturation range is maintained is dependent on many factors including infant stability, alarm limit settings and compliance with these settings, pulse oximeter properties (e.g. averaging times, alarm trigger delay times), and nursing staff ratios and experience levels, all of which can contribute to alarm burden/fatigue. It should be noted that the trials included in this review did not specifically assess the effects of *actual* target ranges achieved, they assessed the *intention* to target one of two oxygen saturation ranges. It should also be re-iterated that the included trials assessed the effect of different targeting policies, not the effects of different alarm settings. Future technologies, such as automated oxygen titration, and more complex alarm strategies, such as variations in alarm escalation regimes and cell phone notifications, may all contribute to how well a desired target range is maintained for an individual infant. However, until we have better monitoring systems that accurately provide information regarding tissue oxygenation, we will need to continue to rely on thorough review of oxygen saturation histograms to better target oxygen saturation in extremely preterm infants.

The reasons for the difference in mortality seen when oximeters with the original calibration software, compared to the revised software, were used cannot be explained simply with the currently available data. Further exploration of oximeter saturation data from the five NeOProM trials will be needed to assess whether this difference was due to the change in oximeter software resulting in more effective targeting of SpO_2, whether trial centres became more effective at targeting SpO_2 over time, or whether there are other explanations for the differences seen. Until further analyses are undertaken, the best estimate of the effects of different target ranges should be focused on the combined results from all trials.

Quality of the evidence

Overall, we assessed all five included trials as being at low risk of bias (see Characteristics of included studies) as all employed random sequence generation, had adequate allocation concealment, were blinded to participants, personnel and outcome assessors, had low levels of missing data, and have (or will have, within the NeOProM analyses) reported all pre-specified outcomes (low risk of reporting bias). Although data are currently unavailable for some of the secondary and subgroup analyses, we have not downgraded these outcomes for 'missingness' as they are known to be available and will be included in updated versions of this review once the NeOProM analyses are available.

As noted previously, this prospectively planned meta-analysis has low levels of statistical heterogeneity (inconsistency) for most outcomes. The 'severe retinopathy of prematurity or retinal therapy' outcome had a moderate level of heterogeneity (I^2 = 69%), which resulted from the substantially larger treatment effect of lower targeting on this outcome in the SUPPORT trial. Reasons for the larger treatment effect for this outcome seen in the SUPPORT trial need to be further explored, possibly by examining 'achieved' (rather than 'intended') saturation patterns across trials. The other outcome that had a moderate level of heterogeneity (I^2 = 44%) was 'supplemental oxygen at 36 weeks postmenstrual age'. There were differences between the trials in how the need for supplemental oxygen at 36 weeks was assessed, with two trials (SUPPORT and BOOST-II UK) using a physiologic test to determine this need, while the other three trials did not use such a test. The outcomes of blindness (1.11) and severe hearing loss (1.12) had low event rates (43 and 131 events respectively from 3870 infants) so while there was no heterogeneity seen for these outcomes (I^2 = 0% for both), we downgraded both for imprecision due to the resulting wide 95% confidence intervals.

Despite being well planned and conducted, all five trials each only achieved a 2% to 3% difference in median oxygen saturations between the two treatment groups. There was also variability (clinical heterogeneity) between trials in the oximeter alarm settings (see Characteristics of included studies table) and in achieved SpO_2

distributions.

All five trials directly assessed the same comparisons (SpO_2 85% to 89% versus 91% to 95%) and the combined sample size of over 3800 for all the main outcomes resulted in the quality of evidence remaining graded as high with regards to the indirectness and imprecision criteria.

Given the above quality assessments, for the main outcomes (primary outcomes 1.1. and 1.2 and key secondary outcomes 1.3 to 1.12) more evidence is unlikely to change our confidence in the estimates of the effects, and these outcomes are thus graded as high-quality overall.

Other secondary outcomes (1.13 to 1.23) and subgroup analyses (small for gestational age, multiples, antenatal steroids, inborn status) data are either incomplete or not available from the current trial publications resulting in grading the evidence as moderate for these outcomes at present. Further planned evidence from the NeOProM Collaboration will likely change this assessment of moderate quality for these outcomes in the near future.

Potential biases in the review process

The inclusion of representatives from the five included trials, the NeOProM Collaboration, and an independent author in the review team meant that there was considerable debate, discussion, and difference of opinion regarding many issues covered within this review. Where consensus could not be reached within the author team regarding the interpretation of the review's findings using the methods outlined, we have included the differing views.

Agreements and disagreements with other studies or reviews

The data included in this Cochrane Review allow the most complete meta-analysis of the five NeOProM trials published to date, including unpublished information sourced directly from the trialists for the 'aligned' definition of major disability-related outcomes (Analysis 1.1; Analysis 1.4) and treated ROP (Analysis 1.7), thereby superceding the previous meta-analyses by Manja 2017, Manja 2015, Saugstad 2014, Fang 2016 and Stenson 2016. This review was undertaken with direct input from the investigators whose trials were included, and involved extensive discussion and consultation. An author independent of the trialists and the NeOProM Collaboration (MV) provided additional input and was available to resolve disagreements when necessary. The resulting Characteristics of included studies table, including the 'Risk of bias' assessments, are thus based on detailed information sourced directly from the trialists, which was not available to the authors of the other published reviews. These assessments of the quality of evidence (high to moderate) and strength of recommendations (GRADE criteria) thus differ from those published by Manja 2015 (moderate to low).

AUTHORS' CONCLUSIONS

Implications for practice

When using an aligned definition for major disability, this meta-analysis of five randomised trials found no significant difference in the primary composite outcome of death or major disability in extremely preterm infants when targeting a lower (SpO_2 85% to 89%) versus higher (SpO_2 91% to 95%) oxygen saturation range. Compared with a higher target range, a lower target range significantly increased the incidence of death (at both discharge and 18 to 24 months corrected age) and necrotising enterocolitis. Conversely, targeting the lower range significantly decreased the incidence of severe or treated retinopathy of prematurity. There were no significant differences between the two treatment groups for major disability including blindness, severe hearing loss, cerebral palsy or other important neonatal morbidities. A pre-specified subgroup analysis of major outcomes by type of oximeter calibration software (original versus revised) found a significant difference in the treatment effect between the two software types for death, with a significantly larger treatment effect seen for those infants using the revised calibration software. There were no other significant differences in treatment effect in any of the pre-specified subgroup analyses, although data for many of these analyses are not available within the published trial reports.

The size of the observed treatment effects found in this meta-analysis give estimates that for every 31 to 40 extremely preterm infants targeted at a lower oxygen range, on average there would be one additional death and one additional case of necrotising enterocolitis, but there would be one less infant with severe/treated ROP, and eight less days of oxygen exposure per patient. The treatment effect estimates found in this review show the *relative* effects of the different target ranges on outcomes. The number needed to treat for an additional beneficial/harmful outcome (NNTB/NNTH) estimates included in this review are based on the *absolute* risk reductions found within the included trials. Whilst the baseline risks for each of these outcomes may differ in different settings, and thus trade-offs between the benefits and harms of the different target ranges can be assessed at a local level (Schmidt 2014), the combined data included in this review constitute the largest dataset as yet compiled to address this clinical question. As such, these results provide important information for clinicians, researchers, parents, and other stakeholders to make judgements regarding the choice of oxygen saturation targets. It should, however, be noted that the findings of this meta-analysis should not be extrapolated to oxygen saturation ranges, or practice settings, outside those tested within the included trials. Similarly, data from this review do not provide evidence of an ability to personalise oxygen saturation targets for individual infants.

Recent guidance from the American Academy of Pediatrics Committee on Fetus and Newborn advises that "recent RCTs suggest

that a targeted oxygen saturation range of 90% to 95% may be safer than 85% to 89%" (Cummings 2016). The results from this review concur with this conclusion.

Implications for research

Further meta-analyses using individual participants from the NeOProM Collaboration will provide data for the remaining secondary outcomes and planned subgroup analyses.

More detailed examination of the included trials' pulse oximetry data (either individually or collectively) will be important to ascertain why such a small change in the target range (of approximately 2% to 3% saturation points) produced the significant differences in important outcomes, such as death and necrotising enterocolitis, which were seen. This could include investigation of potential effect modifying factors such as the proportion of time spent at lower SpO_2 levels, or other combinations of participant- and intervention-level factors. The implications for nursing practice when implementing these findings in practice may warrant further study, including the potential for new technologies such as automated inspired oxygen adjustment, the effects of nursing staff ratios, and alarm policies.

ACKNOWLEDGEMENTS

We wish to acknowledge the following people who have contributed to the preparation of this review in many important ways over many years: Kylie Hunter, Lucy Davies, Cynthia Cole, Win Tin, the investigators/statisticians/data managers/research nurses of the included trials, and all the members of the NeOProM Collaboration. We also wish to pay tribute to several (sadly now deceased) colleagues who have challenged and inspired us to continue this work over several decades: David Henderson-Smart, Jack Sinclair, Edmund Hey, and Bill Silverman.

REFERENCES

References to studies included in this review

BOOST-II Australia 2016 *{published data only}*

Stenson BJ, Tarnow-Mordi WO, Darlow BA, Simes J, Juszczak E, Askie L, 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. *New England Journal of Medicine* 2013;**368**(22):2094–104. [DOI: 10.1056/NEJMoa1302298; PUBMED: 23642047]

* Tarnow-Mordi W, Stenson B, Kirby A, Juszczak E, Donoghoe M, Deshpande S, et al. BOOST-II Australia and United Kingdom Collaborative Groups. Outcomes of two trials of oxygen-saturation targets in preterm infants. *New England Journal of Medicine* 2016;**374**(8):749–60. [DOI: 10.1056/NEJMoa1514212; PUBMED: 26863265]

BOOST-II UK 2016 *{published data only}*

BOOST II United Kingdom Collaborative Group, BOOST II Australia Collaborative Group, BOOST II New Zealand Collaborative Group, Stenson BJ, Tarnow-Mordi WO, Darlow BA, Simes J, Juszczak E, Askie L, et al. Oxygen saturation and outcomes in preterm infants. *New England Journal of Medicine* 2013;**368**(22):2094–104.

* BOOST-II Australia and United Kingdom Collaborative Groups, Tarnow-Mordi W, Stenson B, Kirby A, Juszczak E, Donoghoe M, Deshpande S, et al. Outcomes of two trials of oxygen-saturation targets in preterm infants. *New England Journal of Medicine* 2016;**374**(8):749–60.

BOOST NZ 2014 *{published and unpublished data}*

BOOST II United Kingdom Collaborative Group, BOOST II Australia Collaborative Group, BOOST II New Zealand Collaborative Group, Stenson BJ, Tarnow-Mordi WO, Darlow BA, Simes J, Juszczak E, Askie L, et al. Oxygen saturation and outcomes in preterm infants. *New England Journal of Medicine* 2013;**368**(22):2094–104.

* Darlow BA, Marschner SL, Donoghoe M, Battin MR, Broadbent RS, Elder MJ, 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. *Journal of Pediatrics* 2014;**165**(1):30–5.

Schmidt 2013 *{published and unpublished data}*

* Schmidt B, Whyte RK, Asztalos EV, Moddemann D, Poets C, Rabi Y, 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):2111–20. [DOI: 10.1001/jama.2013.5555; NCT00637169]

Vaucher 2012 *{published and unpublished data}*

Carlo WA, Finer NN, Walsh MC, Rich W, Gantz MG, Laptook AR, et al. SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network. Target ranges of oxygen saturation in extremely preterm infants. *New England Journal of Medicine* 2010; **362**(21):1959–69. [DOI: 10.1056/NEJMoa0911781; NCT00233324]

* Vaucher YE, Peralta-Carcelen M, Finer NN, Carlo WA, Gantz MG, Walsh MC, et al. SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network. Neurodevelopmental outcomes in the early CPAP and pulse oximetry trial. *New England Journal of Medicine* 2012;**367**(26):2495–504. [DOI: 10.1056/NEJMoa1208506; NCT00233324]

References to studies excluded from this review

Arora 2013 *{published data only}*

Arora V, Cayabyab R, Durand M, Ramanathan R. Graded oxygen saturation targets for premature infants in relation to outcomes. *Journal of Investigative Medicine* 2013;**61**(1):205.

Bard 1996 *{published data only}*

Bard H, Belanger S, Fouron JC. Comparison of effects of 95% and 90% oxygen saturations in respiratory distress syndrome. *Archives of Disease in Childhood. Fetal and Neonatal Edition* 1996;**75**(2):F94–6. PUBMED: 8949690]

NCT00845624 *{unpublished data only}*

NCT00845624. Time outside target oxygen saturation range in preterm infants and long term outcomes and preterm infants [Duration of time outside, below, and above the targeted oxygen saturation range in preterm infants]. https://clinicaltrials.gov/show/NCT00845624 (first received 16 February 2009).

NCT01590316 *{unpublished data only}*

NCT01590316. SafeBoosC - a phase II trial [SafeBoosC – Safeguarding the brain of our smallest children – an investigator–initiated randomised, blinded, multinational, phase II feasibility clinical trial on near–infrared spectroscopy monitoring combined with defined treatment guidelines versus standard monitoring and treatment as usual in premature infants]. https://clinicaltrials.gov/show/NCT01590316 (first received 1 May 2012).

Additional references

AAP 1988

American Academy of Pediatrics. Committee on Fetus and Newborn, ACOG Committee on Obstetrics: Maternal and Fetal Medicine. *Guidelines for Perinatal Care*. 2nd Edition. Elk Grove Village, Ill: American Academy of Pediatrics, 1988.

AAP 2002

American Academy of Pediatrics, American College of Obstetricians and Gynecologists. *Guidelines for Perinatal Care*. 5th Edition. American Academy of Pediatrics, 2002.

ACTRN12605000055606

ACTRN12605000055606. BOOST II: Benefits Of Oxygen Saturation Targeting Study [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 (first received 14 July 2005).

ACTRN12605000253606

ACTRN12605000253606. BOOST NZ [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 (first received 24 August 2005).

Anderson 2003

Anderson P, Doyle LW, Victorian Infant Collaborative Study Group. Neurobehavioral outcomes of school-age children born extremely low birth weight or very preterm in the 1990s. *JAMA* 2003;**289**(24):3264–72. [DOI: 10.1001/jama.289.24.3264; PUBMED: 12824207]

Anderson 2004

Anderson CG, Benitz WE, Madan A. Retinopathy of prematurity and pulse oximetry: a national survey of recent practices. *Journal of Perinatology* 2004;**24**(3):164–8. [DOI: 10.1038/sj.jp.7211067; PUBMED: 14999216]

Askie 2001a

Askie LM, Henderson-Smart DJ. Gradual versus abrupt discontinuation of oxygen in preterm or low birth weight infants. *Cochrane Database of Systematic Reviews* 2001, Issue 4. [DOI: 10.1002/14651858.CD001075]

Askie 2001b

Askie LM, Henderson-Smart DJ. Early versus late discontinuation of oxygen in preterm or low birth weight infants. *Cochrane Database of Systematic Reviews* 2001, Issue 4. [DOI: 10.1002/14651858.CD001076]

Askie 2003

Askie LM, Henderson-Smart DJ, Irwig L, Simpson JM. Oxygen-saturation targets and outcomes in extremely preterm infants. *New England Journal of Medicine* 2003;**349**(10):959–67. [DOI: 10.1056/NEJMoa023080; PUBMED: 12954744]

Askie 2009

Askie LM, Henderson-Smart DJ, Ko H. Restricted versus liberal oxygen exposure for preventing morbidity and mortality in preterm or low birth weight infants. *Cochrane Database of Systematic Reviews* 2009, Issue 1. [DOI: 10.1002/14651858.CD001077.pub2; PUBMED: 19160188]

Askie 2011

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 Pediatrics* 2011;**11**:6. [DOI: 10.1186/1471-2431-11-6; PUBMED: 21235822]

Avery 1960

Avery ME. Recent increase in mortality from hyaline membrane disease. *Journal of Pediatrics* 1960;**57**:553–9. [PUBMED: 13685272]

Bolton 1974

Bolton DP, Cross KW. Further observations on cost of preventing retrolental fibroplasia. *Lancet* 1974;**1**(7855): 445–8. [PUBMED: 4131442]

Campbell 1951

Campbell K. Intensive oxygen therapy as a possible cause of retrolental fibroplasia; a clinical approach. *Medical Journal of Australia* 1951;**2**(2):48–50. [PUBMED: 14874698]

Castillo 2008
Castillo A, Sola A, Baquero H, Neira F, Alvis R, Deulofeut R, et al. Pulse oxygen saturation levels and arterial oxygen tension values in newborns receiving oxygen therapy in the neonatal intensive care unit: is 85% to 93% an acceptable range?. *Pediatrics* 2008;**121**(5):882–9. [DOI: 10.1542/peds.2007-0117; PUBMED: 18450890]

Centre for Epi 2012
Centre for Epidemiology and Evidence. New South Wales Mothers and Babies Report 2010. Sydney: NSW Ministry of Health 2012.

Chow 2003
Chow LC, Wright KW, Sola A, CSMC Oxygen Administration Study Group. Can changes in clinical practice decrease the incidence of severe retinopathy of prematurity in very low birth weight infants?. *Pediatrics* 2003;**111**(2):339–45. [PUBMED: 12563061]

Chow 2013
Chow SW. *The Report of the Australian and New Zealand Neonatal Network, 2010.* Sydney: ANZNN, 2013.

Cross 1973
Cross KW. Cost of preventing retrolental fibroplasia?. *Lancet* 1973;**2**(7835):954-6. [PUBMED: 4126572]

Cummings 2016
Cummings JJ, Polin RA, AAP Committee on Fetus and Newborn. Oxygen targeting in extremely low birth weight infants. *Pediatrics* 2016;**138**(2):e20161576. [DOI: 10.1542/peds.2016-1576]

Doyle 2010
Doyle LW, Roberts G, Anderson PJ, Victorian Infant Collaborative Study Group. Outcomes at age 2 years in infants <28 weeks' born in Victoria in 2005. *Journal of Pediatrics* 2010;**156**(1):49–53. [DOI: 10.1016/j.jpeds.2009.07.013; PUBMED: 19783004]

Duc 1992
Duc G, Sinclair JC. Oxygen administration. In: Sinclair JC, Bracken MB editor(s). *Effective Care of the Newborn*. New York: Oxford University Press, 1992:178–99.

Fang 2016
Fang JL, Sorita A, Carey WA, Colby CE, Murad MH, Alahdab F. Interventions to prevent retinopathy of prematurity: a meta-analysis. *Pediatrics* 2016;**137**(4):e20153387. [DOI: 10.1542/peds.2015-3387; PUBMED: 269622040]

Flynn 1992
Flynn JT, Bancalari E, Snyder ES, Goldberg RN, Feuer W, Cassady J, et al. A cohort study of transcutaneous oxygen tension and the incidence and severity of retinopathy of prematurity. *New England Journal of Medicine* 1992;**326**(16):1050–4. [DOI: 10.1056/NEJM199204163261603; PUBMED: 1549150]

Gao 2010
Gao Y, Raj JU. Regulation of the pulmonary circulation in the fetus and newborn. *Physiological Reviews* 2010;**90**(4):1291–335. [DOI: 10.1152/physrev.00032.2009; PUBMED: 20959617]

GRADEpro GDT [Computer program]
Grade Working Group, McMaster University. GRADEpro [www.gradepro.org]. Version Version 14 September 2014. Hamilton (ON): Grade Working Group, McMaster University, 2014.

Hagadorn 2006
Hagadorn JI, Furey AM, Nghiem TH, Schmid CH, Phelps DL, Pillers DA, et al. AVIOx Study Group. Achieved versus intended pulse oximeter saturation in infants born less than 28 weeks' gestation: the AVIOx study. *Pediatrics* 2006;**118**(4):1574–82. [DOI: 10.1542/peds.2005-0413]

Hellstrom 2013
Hellstrom A, Smith LE, Dammann O. Retinopathy of prematurity. *Lancet* 2013;**382**(9902):1445–57. [DOI: 10.1016/S0140-6736(13)60178-6; PUBMED: 23782686]

Higgins 2003
Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**(7414):557–60. [DOI: 10.1136/bmj.327.7414.557]

Higgins 2011
Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

ISRCTN00842661
ISRCTN00842661. BOOST-II UK [Which oxygen saturation level should we use for very premature infants? A randomised controlled trial]. http://isrctn.org/ISRCTN00842661 (first received 7 July 2006). [DOI 10.1186/ISRCTN00842661]

ISRCTN62491227
ISRCTN62491227. COT [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://isrctn.org/ISRCTN62491227 (first received 22 August 2006). [DOI 10.1186/ISRCTN62491227]

Jobe 2001
Jobe AH, Bancalari E. Bronchopulmonary dysplasia. American Journal of Respiratory and Critical Care Medicine 2001; Vol. 163, issue 7:1723–9. [DOI: 10.1164/ajrccm.163.7.2011060; PUBMED: 11401896]

Kapadia 2013
Kapadia VS, Chalak LF, Sparks JE, Allen JR, Savani RC, Wyckoff MH. Resuscitation of preterm neonates with limited versus high oxygen strategy. *Pediatrics* 2013;**132**(6):e1488–96. [DOI: 10.1542/peds.2013-0978]

Kinsey 1956
Kinsey VE. Retrolental fibroplasia; cooperative study of retrolental fibroplasia and the use of oxygen. *A.M.A. Archives of Ophthalmology* 1956;**56**(4):481–543. [PUBMED: 13361620]

Kinsey 1977
Kinsey VE, Arnold HJ, Kalina RE, Stern L, Stahlman M, Odell G, et al. PaO_2 levels and retrolental fibroplasia:

a report of the cooperative study. *Pediatrics* 1977;**60**(5): 655–68. [PUBMED: 578921]

Lanman 1954

Lanman JT, Guy LP, Dancis J. Retrolental fibroplasia and oxygen therapy. *Journal of the American Medical Association* 1954;**155**(3):223–6. [PUBMED: 13151906]

Lim 2014

Lim K, Wheeler KI, Gale TJ, Jackson HD, Kihlstrand JF, Sand C, et al. Oxygen saturation targeting in preterm infants receiving continuous positive airway pressure. *Journal of Pediatrics* 2014;**164**(4):730–6. [DOI: 10.1016/j.jpeds.2013.11.072; PUBMED: 24433828]

Lloyd 2003

Lloyd J, Askie L, Smith J, Tarnow-Mordi W. Supplemental oxygen for the treatment of prethreshold retinopathy of prematurity. *Cochrane Database of Systematic Reviews* 2003, Issue 2. [DOI: 10.1002/14651858.CD003482]

Maltepe 2009

Maltepe E, Saugstad OD. Oxygen in health and disease: regulation of oxygen homeostasis - clinical implications. *Pediatric Research* 2009;**65**(3):261-8. [DOI: 10.1203/PDR.0b013e31818fc83f]

Manja 2015

Manja V, Lakshminrusimha S, Cook DJ. Oxygen saturation target range for extremely preterm infants: a systematic review and meta-analysis. *JAMA Pediatrics* 2015;**169** (4):332–40. [DOI: 10.1001/jamapediatrics.2014.3307; PUBMED: 25664703]

Manja 2017

Manja V, Saugstad OD, Lakshminrusimha S. Oxygen saturation targets in preterm infants and outcomes at 18-24 months: a systematic review. *Pediatrics* 2017;**139**(1): e20161609. [DOI: 10.1542/peds.2016-1609]

McDonald 1963

McDonald AD. Cerebral palsy in children of low birth weight. *Archives of Disease in Childhood* 1963;**38**(202): 579–88. [PUBMED: 14092062]

McDonald 1964

McDonald AD. Oxygen treatment of premature babies and cerebral palsy. *Developmental Medicine and Child Neurology* 1964;**6**:313–4. [PUBMED: 14155197]

McIntosh 2001

McIntosh N. High or low oxygen saturation for the preterm baby. *Archives of Disease in Childhood. Fetal and Neonatal Edition* 2001;**84**(3):F149. [PUBMED: 11320038]

NCT00233324

NCT00233324. Surfactant Positive Airway Pressure and Pulse Oximetry Trial [Surfactant Positive Airway Pressure and Pulse Oximetry Trial (SUPPORT) in Extremely Low Birth Weight Infants]. clinicaltrials.gov/show/NCT00233324 (first received 03 October 2005).

Newburger 1984

Newburger JW, Silbert AR, Buckley LP, Fyler DC. Cognitive function and age at repair of transposition of the great arteries in children. *New England Journal of Medicine* 1984;**310**(23):1495–9. [DOI: 10.1056/NEJM198406073102303; PUBMED: 6717539]

Patz 1954

Patz A. Oxygen studies in retrolental fibroplasia. IV. Clinical and experimental observations. *American Journal of Ophthalmology* 1954;**38**(3):291–308. [PUBMED: 13188932]

Poets 1998

Poets CF. When do infants need additional inspired oxygen? A review of the current literature. *Pediatric Pulmonology* 1998;**26**(6):424–8. [PUBMED: 9888217]

Raju 1999

Raju TN. The Nobel chronicles. 1949: Walter Rudolf Hess (1881-1973); and Antônio Egas Moniz (1874-1955). *Lancet* 1999;**353**(9160):1281. [PUBMED: 10217124]

RevMan 2014 [Computer program]

The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager 5 (RevMan 5). Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Saigal 2000

Saigal S, Burrows E, Stoskopf BL, Rosenbaum PL, Streiner D. Impact of extreme prematurity on families of adolescent children. *Journal of Pediatrics* 2000;**137**(5):701–6. [DOI: 10.1067/mpd.2000.109001; PUBMED: 11060538]

Saugstad 2001

Saugstad OD. Update on oxygen radical disease in neonatology. *Current Opinion in Obstetrics & Gynecology* 2001;**13**(2):147–53. [PUBMED: 11315869]

Saugstad 2014

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]

Schmidt 2014

Schmidt B, Whyte RK, Roberts RS. Trade-off between lower or higher oxygen saturations for preterm infants: the first Benefits Of oxygen Saturation Targeting (BOOST) II trial reports its primary outcome. *Journal of Pediatrics* 2014; **165**(1):6–8.

Schünemann 2013

Schünemann H, Broż ek J, Guyatt G, Oxman A, editors. GRADE Working Group. GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations. Available from https://gdt.gradepro.org/app/handbook/handbook.html. Updated October 2013.

Silverman 1980

Silverman WA. *Retrolental Fibroplasia: a Modern Parable*. Grune & Stratton, 1980.

Silverman 2004

Silverman WA. A cautionary tale about supplemental oxygen: the albatross of neonatal medicine. *Pediatrics* 2004; **113**(2):394–6. [PUBMED: 14754955]

Skinner 1999

Skinner JR, Hunter S, Poets CF, Milligan DW, Southall D, Hey EN. Haemodynamic effects of altering arterial oxygen

saturation in preterm infants with respiratory failure. *Archives of Disease in Childhood. Fetal and Neonatal Edition* 1999;**80**(2):F81–7. [PUBMED: 10325781]

Stenson 2011
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. *New England Journal of Medicine* 2011;**364**(17):1680–2. [DOI: 10.1056/NEJMc1101319; PUBMED: 21524227]

Stenson 2016
Stenson BJ. Oxygen saturation targets for extremely preterm infants after the NeOProM trials. *Neonatology* 2016;**109**(4):352–8. [DOI: 10.1159/000444913]

STOP ROP 2000
Supplemental Therapeutic Oxygen for Prethreshold Retinopathy Of Prematurity (STOP-ROP), a randomized, controlled trial. I: primary outcomes. *Pediatrics* 2000;**105**(2):295–310. [PUBMED: 10654946]

Subhedar 2000
Subhedar NV, Shaw NJ. Changes in pulmonary arterial pressure in preterm infants with chronic lung disease. *Archives of Disease in Childhood. Fetal and Neonatal Edition* 2000;**82**(3):F243–7. [PUBMED: 10794795]

Sun 2002
Sun SC. Relation of target SpO^2 levels and clinical outcome in ELBW infants on supplemental oxygen. *Pediatric Research* 2002;**51**:350a.

Sutton 1999
Sutton L, Bajuk B. Population-based study of infants born at less than 28 weeks' gestation in New South Wales, Australia, in 1992-3. New South Wales Neonatal Intensive Care Unit Study Group. *Paediatric and Perinatal Epidemiology* 1999;**13**(3):288–301. [PUBMED: 10440049]

Terry 1942
Terry TL. Extreme prematurity and fibroblastic overgrowth of persistent vascular sheath behind each crystalline lens: 1. Preliminary report. *American Journal of Ophthalmology* 1942;**25**:203–5. [PUBMED: 16693360]

Tin 2001
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. *Archives of Disease in Childhood. Fetal and Neonatal Edition* 2001;**84**(2):F106–10. [PUBMED: 11207226]

Tin 2007
Tin W, Gupta S. Optimum oxygen therapy in preterm babies. *Archives of Disease in Childhood. Fetal and Neonatal Edition* 2007;**92**(2):F143–7. [DOI: 10.1136/adc.2005.092726; PUBMED: 17337663]

Usher 1961
Usher RH. Clinical investigation of the respiratory distress syndrome of prematurity. Interim report. *New York State Journal of Medicine* 1961;**61**:1677–96. [PUBMED: 13779372]

Vento 2009
Vento M, Moro M, Escrig R, Arruza L, Villar G, Izquierdo I, et al. Preterm resuscitation with low oxygen causes less oxidative stress, inflammation, and chronic lung disease. *Pediatrics* 2009;**124**(3):e439–49. [PUBMED: 19661049]

Vento 2013
Vento M, Teramo K. Evaluating the fetus at risk for cardio-pulmonary compromise - what are the effects of hypoxia/hyperoxia - is there a role for in-utero resuscitation. *Seminars in Fetal and Neonatal Medicine* 2013;**18**(6):324–9.

Walsh 2009
Walsh BK, Brooks TM, Grenier BM. Oxygen therapy in the neonatal care environment. *Respiratory Care* 2009;**54**(9):1193–202. [PUBMED: 19712496]

Warner 1998
Warner BB, Stuart LA, Papes RA, Wispe JR. Functional and pathological effects of prolonged hyperoxia in neonatal mice. *American Journal of Physiology* 1998;**275**(1 Pt 1):L110–7. [PUBMED: 9688942]

Williams 1998
Williams AJ. ABC of oxygen: assessing and interpreting arterial blood gases and acid-base balance. *BMJ* 1998;**317**(7167):1213–6. [PUBMED: 9794863]

Wilson 1942
Wilson JL, Long SB, Howard PJ. Respiration of premature infants: response to variations of oxygen and to increased carbon dioxide in inspired air. *American Journal of Diseases of Children* 1942;**63**(6):1080–5. [DOI: 10.1001/archpedi.1942.02010060064002]

* *Indicates the major publication for the study*

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by year of study]*

Vaucher 2012

Methods	Randomised, multicentre (USA) trial with a 2-by-2 factorial design
Participants	1316 infants born between 24 weeks 0 days and 27 weeks 6 days of gestation. Infants were enrolled prior to birth and were thus all inborn at a trial centre. Enrollment was undertaken from February 2005 until February 2009. Follow-up assessments began in November 2006 and ended in July 2011
Interventions	Infants were monitored with target ranges of oxygen saturation of 85% to 89% or 91% to 95% using oximeters with concealed saturation offsets of +3% in actual range 85% to 92% (low target) and -3% in range 88% to 95% (high target), with true readings displayed 84% and below and 96% and above (see Figure 4). Caregivers were asked to adjust the concentration of oxygen to maintain displayed saturations between 88% and 92% when the infant was receiving supplemental oxygen. Alarms were suggested to be set so that an alarm would sound at displayed saturation values of 85% and 95%, but they could be changed for individual patients. Infants were also randomly assigned to continuous positive airway pressure or intubation and surfactant. Intervention was initiated within 2 hours of birth and continued until 36 weeks of postmenstrual age or until the infant was breathing ambient air, whichever occurred first. Infants who were returned to supplemental oxygen were reassigned to the study oximeter. All infants in this trial were managed with oximeters using the original calibration software
Outcomes	Co-primary outcomes: survival at discharge from hospital without severe ROP (threshold ROP and/or the need for surgical intervention) assessed until diagnosis or resolution; and death or survival with neurodevelopmental impairment at 18 to 22 months corrected age Neurodevelopmental impairment was defined as having any of the following: * BSID-III cognitive or language score < 70 * GMFCS level 2 or higher * Moderate to severe cerebral palsy * Hearing impairment * Bilateral visual impairment Secondary outcomes: severe retinopathy of prematurity, death before discharge, death by 36 weeks postmenstrual age, BPD defined by use of supplemental oxygen at 36 weeks, BPD physiological definition at 36 weeks, intraventricular haemorrhage grade 3 or 4, periventricular leukomalacia, necrotising enterocolitis stage $\geq$ 2, pneumothorax, postnatal corticosteroids for BPD, death by 7 days, death by 14 days, late-onset sepsis, patent ductus arteriosus requiring medical treatment, patent ductus arteriosus requiring surgical treatment, any air leaks in first 14 days
Notes	Funded by: the USA Eunice Kennedy Shriver National Institute of Child Health and Human Development; the National Heart, Lung and Blood Institute; and the National Institutes of Health Trial registration ID: NCT00233324

Risk of bias		*Risk of bias*
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Permuted-block randomisation was used, with stratification according to study centre and gestational age (24 weeks 0 days to 25 weeks 6 days or 26 weeks 0 days to 27 weeks 6 days). Multiple births were randomised to the same group
Allocation concealment (selection bias)	Low risk	Sealed, sequentially numbered with central tracking opaque envelopes. Oximeter allocation was identifiable (via colour-coded dots) to designated research staff but not to clinical staff. Bedside adjustment of supplemental oxygen was performed only by clinical staff
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Blinding was maintained by oximeter design.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Parents and assessors were unaware of allocation.
Incomplete outcome data (attrition bias) All outcomes	Low risk	Of the 1316 infants enrolled, 1234 (93.8%) had adequate data for the analysis of the composite primary outcome at 18 to 22 months corrected age 35 infants were of unknown status (21 low target group, 14 high target group) and 47 had incomplete or no follow-up (21 low target group, 26 high target group). If Bayley scores were missing, children were excluded from the primary outcome analysis No participants were excluded after randomisation. All outcome analyses followed the principle of intention-to-treat. The follow-up rate and the mean corrected age at neurodevelopmental assessment were similar for all treatment groups (in the 2-by-2 factorial design)
Selective reporting (reporting bias)	Low risk	The predetermined sample size of 1310 infants was achieved. The original study protocol specified a composite primary outcome of death before 36 weeks of post-

Vaucher 2012 (*Continued*)

		menstrual age or severe ROP, but this was changed to death before discharge or severe ROP before any data analyses were performed. All other outcomes pre-specified in the registration record were reported, including assessment of the need for oxygen at 36 weeks postmenstrual age and safety outcomes
Other bias	Low risk	The baseline characteristics of the 2 treatment groups were similar

Schmidt 2013

Methods	Randomised, multicentre (Canada, USA, Argentina, Finland, Germany, and Israel) trial
Participants	1201 infants born with gestational ages of 23 weeks 0 days through 27 weeks 6 days, enrolled within 24 hours after birth, and either born in or transferred into a trial NICU. Enrollment was undertaken from December 2006 until August 2010. Follow-up assessments began in October 2008 and ended in August 2012
Interventions	Infants were monitored with target ranges of oxygen saturation of 85% to 89% or 91% to 95% using oximeters with concealed saturation offsets of +3% in actual range 85% to 92% (low target) and -3% in range 88% to 95% (high target), with true readings displayed 84% and below and 96% and above (see Figure 4). Caregivers were asked to adjust the concentration of oxygen to maintain displayed saturations between 88% and 92% when the infant was receiving supplemental oxygen. Alarms were set so that an alarm would sound at displayed saturation values of 86% and 94%. Intervention was initiated within 24 hours of birth and continued until 36 weeks of postmenstrual age irrespective of supplemental oxygen therapy, and until 40 weeks in infants receiving oxygen therapy at 35 weeks. The oximeters used in this trial were modified with a revised calibration software in early 2009. 47% of infants in this trial were managed with oximeters using the original calibration software, 47% with oximeters using the revised calibration software, and 6% were exposed to both
Outcomes	Primary outcome: death or survival with major disability at 18 to 21 months corrected age. Major disability was defined as having any of the following: * Cognitive score < 85 or language score < 85 on BSID-III * Severe visual loss * Cerebral palsy with GMFCS level 2 or higher * Deafness requiring hearing aids Secondary outcomes: retinopathy of prematurity, brain injury, patent ductus arteriosus, necrotising enterocolitis, bronchopulmonary dysplasia, duration of use of positive airway pressure and supplemental oxygen, hospital re-admissions for respiratory disease, chronic use of respiratory medications, and mean composite cognitive, language and motor scores

Notes	Funded by Canadian Institutes of Health Research Trial registration ID: ISRCTN62491227, NCT00637169

Risk of bias ***Risk of bias***

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	A computer-generated randomisation scheme at a remote co-ordinating centre assigned the infants to treatment groups in a 1:1 ratio. Randomisation was stratified by study centre and balanced within randomly sized blocks of 2 or 4 patients. Siblings within multiple births were randomised individually
Allocation concealment (selection bias)	Low risk	Study oximeters were labelled with sequential participant numbers according to the randomisation scheme. The allocation remained unknown to the members of the clinical and research teams and all staff at the co-ordinating centre
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Blinding was maintained by oximeter design. There is evidence that the algorithm used for blinding caused a difference in nursing behaviour with high versus low oximeters, which reduced separation and which could have resulted in detection or co-intervention bias
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Parents and assessors were unaware of allocation.
Incomplete outcome data (attrition bias) All outcomes	Low risk	Of the 1201 infants enrolled, 1147 (95.5%) had adequate data for the analysis of the composite primary outcome at 18 to 21 months corrected age 39 infants were of unknown status (17 low target group, 22 high target group) and 15 had incomplete or no follow-up (7 low target group, 8 high target group). If Bayley scores were missing, children were excluded from the primary outcome analysis No participants were excluded after randomisation. All outcome analyses followed the principle of intention-to-treat. The fol-

Schmidt 2013 (*Continued*)

		low-up rate and the mean corrected age at neurodevelopmental assessment were similar for both treatment groups
Selective reporting (reporting bias)	Low risk	The predetermined sample size of 1201 infants was achieved. All outcomes pre-specified in the registration record were reported
Other bias	Low risk	There were imbalances in surfactant administration and in oxygen therapy before randomisation. Otherwise the baseline characteristics were similar in both groups

BOOST NZ 2014

Methods	Randomised, multicentre (New Zealand) trial.
Participants	340 infants born at less than 28 weeks' gestation, enrolled within 24 hours after birth, and either born in or transferred into a trial NICU. Enrollment was undertaken from September 2006 until December 2009. Follow-up assessments began in March 2009 and ended in June 2012
Interventions	Infants were monitored with target ranges of oxygen saturation of 85% to 89% or 91% to 95% using oximeters with concealed saturation offsets of +3% in actual range 85% to 92% (low target) and -3% in range 88% to 95% (high target), with true readings displayed 84% and below and 96% and above (see Figure 4). Caregivers were asked to adjust the concentration of oxygen to maintain displayed saturations between 88% and 92% when the infant was receiving supplemental oxygen. Alarm limits were recommended (but not mandated) to be set so that an alarm would sound at displayed saturation values of 87% and 93%. Intervention was initiated within 24 hours of birth, continued for at least two weeks and was discontinued when infants no longer required oxygen (pre-specified definition) or otherwise at 36 weeks. All infants in this trial were managed with oximeters using the original calibration software
Outcomes	Primary outcome: death or survival with major disability at 24 months corrected age. Major disability was defined as having any of the following: * Cognitive score < 85 or language score < 85 on BSID-III, or MDI < 70 on the BSID-II assessment * Severe visual loss * Cerebral palsy defined as GMFCS level 2 or higher * Deafness requiring hearing aids In 33 infants where Bayley scores were unavailable and there were no other events defining major disability, an alternative definition of disability (use of < 10 words) was used Secondary outcomes: severe ROP (≥ stage 3, or retinal surgery), oxygen dependency or respiratory support at 36 weeks' gestational age, days of continuous positive airway pressure, days of endotracheal intubation, days of oxygenation in both hospital and days at home, a patent ductus arteriosus diagnosed by echocardiography and requiring treatment, necrotising enterocolitis requiring surgery or a cause of death, weight at 2

BOOST NZ 2014 (*Continued*)

	years' corrected age, readmissions to hospital by 2 years' corrected age, any respiratory illness (home oxygen, asthma, chest symptoms more than once a week) between discharge and 2 years, cerebral palsy, mean cognition score and mean language score on BSID-III, brain injury (intraventricular haemorrhage ≥ grade 3, periventricular leukomalacia, or porencephalic cysts or contributing to death), and death from pulmonary causes from 4 weeks of age to 2 years' corrected age
Notes	Funded by: New Zealand Heath Research Council and the Child Health Research Foundation (Cure Kids) Trial registration ID: ACTRN12605000253606

Risk of bias ***Risk of bias***

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Computer-generated randomisation lists were prepared by an independent statistician. Stratification was by NICU, sex, gestation < 26 or ≥ 26 weeks, and inborn or outborn. Siblings within multiple births were randomised individually
Allocation concealment (selection bias)	Low risk	Central telephone randomisation by independent statistician.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Blinding was maintained by oximeter design.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Parents and assessors were unaware of allocation.
Incomplete outcome data (attrition bias) All outcomes	Low risk	Of the 340 infants enrolled, 335 (98.5%) had adequate data for the analysis of the composite primary outcome at 24 months corrected age No infants were of unknown status and 5 had incomplete or no follow-up (3 low target group, 2 high target group). Where Bayley scores were missing in a child without cerebral palsy, blindness, or deafness, "major disability" was defined as < 10 words by parent report at the paediatric assessment (n = 33 children). If none of these data were available, the primary endpoint was considered missing No participants were excluded after randomisation. All outcome analyses followed

BOOST NZ 2014 (*Continued*)

		the principle of intention-to-treat. The follow-up rate and the mean corrected age at neurodevelopmental assessment were similar for both treatment groups
Selective reporting (reporting bias)	Low risk	The predetermined sample size of 320 infants was exceeded, with a final sample size of 340 being achieved. All outcomes prespecified in the registration record were reported
Other bias	Low risk	The baseline characteristics of the 2 treatment groups were similar

BOOST-II Australia 2016

Methods	Randomised, multicentre (Australia) trial.
Participants	1135 infants born at less than 28 weeks' gestation, enrolled within 24 hours after birth, and either born in or transferred into a trial NICU. Enrollment was undertaken from March 2006 until December 2010. Follow-up assessments began in August 2008 and ended in August 2013
Interventions	Infants were monitored with target ranges of oxygen saturation of 85% to 89% or 91% to 95% using oximeters with concealed saturation offsets of +3% in actual range 85% to 92% (low target) and -3% in range 88% to 95% (high target), with true readings displayed 84% and below and 96% and above (see Figure 4). Caregivers were asked to adjust the concentration of oxygen to maintain displayed saturations between 88% and 92% when the infant was receiving supplemental oxygen. Alarm limits were recommended to be set so that an alarm would sound at displayed saturation values of 86% and 94%. Intervention was initiated within 24 hours of birth and discontinued when infants no longer required oxygen (pre-specified definition) or otherwise at 36 weeks. Infants who were returned to supplemental oxygen were reassigned to the study oximeter. The oximeters used in this trial were modified with a revised calibration algorithm in early 2009. 62% of infants in this trial were managed with oximeters with the original calibration algorithm and 38% with oximeters using the revised calibration algorithm
Outcomes	Primary outcome: death or survival with major disability at 24 months corrected age. Major disability was defined as having any of the following: * Cognitive score < 85 or language score < 85 on BSID-III * Severe visual loss * Cerebral palsy with inability to walk at 2 years corrected age * Deafness requiring hearing aids In 85 infants where Bayley scores were unavailable and there were no other events defining major disability, an alternative definition of disability (use of < 10 words, delayed development < 12 months, other severe impairment) was used Secondary outcomes: death at discharge, death at 36 weeks' postmenstrual age, treated retinopathy of prematurity, necrotising enterocolitis requiring surgery or leading to death,

BOOST-II Australia 2016 (*Continued*)

	severe intraventricular haemorrhage (≥ grade 3), other brain injury, patent ductus arteriosus (requiring medical or surgical treatment), oxygen dependency at 36 weeks' postmenstrual age, bronchopulmonary dysplasia (physiological definition)
Notes	Funded by the Australian National Health and Medical Research Council Trial registration ID: ACTRN12605000055606

Risk of bias ***Risk of bias***

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Computer-generated randomisation lists were prepared by an independent statistician and were stratified according to sex, gestational age, centre, single birth or multiple births, and whether birth took place in the hospital where enrolment took place. Siblings within multiple births were randomised individually
Allocation concealment (selection bias)	Low risk	Central telephone randomisation by independent statistician.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Blinding was maintained by oximeter design.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Parents and assessors were unaware of allocation.
Incomplete outcome data (attrition bias) All outcomes	Low risk	Of the 1135 infants enrolled, 1094 (96.4%) had adequate data for the analysis of the composite primary outcome at 24 months corrected age 12 infants were of unknown status (7 low target group, 5 high target group) and 29 had incomplete or no follow-up (12 low target group, 17 high target group). When Bayley III scores were missing, alternative measures of disability were used, including Bayley II scales, paediatric health status assessment, or a Short Health Status Questionnaire collected via phone call to parents or a GP visit (n = 85 children). If none of these data were available, the primary endpoint was considered missing No participants were excluded after ran-

BOOST-II Australia 2016 (*Continued*)

		domisation. All outcome analyses followed the principle of intention-to-treat. The follow-up rate and the mean corrected age at neurodevelopmental assessment were similar for both treatment groups
Selective reporting (reporting bias)	Low risk	All outcomes pre-specified in the registration record were reported
Other bias	Unclear risk	Investigator concerns resulting from the significantly increased mortality risk with the lower SpO_2 target range in the SUPPORT Trial publication led to an unscheduled safety analysis when 1135 of the planned 1200 infants (95%) had been recruited. A decision was made to terminate recruitment in both the BOOST-II UK and BOOST-II Australia trials based on a pre-specified rule. There was an 8.5% excess in 36-week mortality in the low target group monitored with an oximeter incorporating the revised calibration software (data pooled from both studies, $P < 0.001$ with a significant treatment by software subgroup interaction, $P = 0.006$). The early stopping of the trial (with 81% of the final planned sample size at that point) raises the question of whether this overestimates treatment effect

BOOST-II UK 2016

Methods	Randomised, multicentre (UK) trial.
Participants	973 infants born at less than 28 weeks' gestation, enrolled within 24 hours after birth, and either born in or transferred into a trial NICU. Enrollment was undertaken from September 2007 until December 2010. Follow-up assessments began in December 2009 and ended in August 2014
Interventions	Infants were monitored with target ranges of oxygen saturation of 85% to 89% or 91% to 95% using oximeters with concealed saturation offsets of +3% in actual range 85% to 92% (low target) and -3% in range 88% to 95% (high target), with true readings displayed 84% and below and 96% and above (see Figure 4). Caregivers were asked to adjust the concentration of oxygen to maintain displayed saturations between 88% and 92% when the infant was receiving supplemental oxygen. Upper alarm limits were recommended to be set so that an alarm would sound at a displayed saturation value of 94%. No lower alarm limit was specified. Intervention was initiated within 24 hours of birth and discontinued when infants no longer required oxygen (pre-specified definition) or otherwise at 36 weeks. Infants who were returned to supplemental oxygen were

BOOST-II UK 2016 (*Continued*)

	reassigned to the study oximeter. The oximeters used in this trial were modified with a revised calibration in early 2009. 23% of infants in this trial were managed with oximeters using the original calibration software, and 77% with oximeters using the revised calibration software
Outcomes	Primary outcome: death or survival with major disability at 24 months corrected age. Major disability was defined as having any of the following: * Cognitive score < 85 or language score < 85 on BSID-III * Severe visual loss * Cerebral palsy with inability to walk at 2 years corrected age * Deafness requiring (or too severe to benefit from) hearing aids In 176 infants where Bayley scores were unavailable and there were no other events defining major disability, an alternative definition of disability (incomplete BSID-III score, Denver Developmental Screening Test, Griffiths Mental Development Scales, Schedule of Growing Skills, WPPS-III, PARCA-R, paediatric assessment, GP assessment, parental report review of all data) was used Secondary outcomes: death at discharge, death at 36 weeks' postmenstrual age, treated retinopathy of prematurity, necrotising enterocolitis requiring surgery or leading to death, severe intraventricular haemorrhage (≥ grade 3), other brain injury, patent ductus arteriosus (requiring medical or surgical treatment), oxygen dependency at 36 weeks' postmenstrual age, bronchopulmonary dysplasia (physiological definition)
Notes	Funded by the UK Medical Research Council and managed by the UK National Institute for Health Research Trial registration ID: ISRCTN00842661

Risk of bias ***Risk of bias***

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	A computer-generated minimisation procedure was used to balance study group assignment according to sex, gestational age, and centre. Siblings within multiple births were randomised individually
Allocation concealment (selection bias)	Low risk	Central randomisation by computer.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Blinding was maintained by oximeter design.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Parents and assessors were unaware of allocation.
Incomplete outcome data (attrition bias) All outcomes	Low risk	Of the 973 infants enrolled, 941 (96.7%) had adequate data for the analysis of the composite primary outcome at 24 months corrected age

		6 infants were of unknown status (2 low target group, 4 high target group) and 26 had incomplete or no follow-up (11 low target group, 15 high target group). When Bayley III scores were missing, alternative measures of disability were used, including Bayley II scales, paediatric health status assessment, or a Short Health Status Questionnaire collected via phone call to parents or a GP visit (n = 176 children). If none of these data were available, the primary endpoint was considered missing No participants were excluded after randomisation. All outcome analyses followed the principle of intention-to-treat. The follow-up rate and the mean corrected age at neurodevelopmental assessment were similar for both treatment groups
Selective reporting (reporting bias)	Low risk	All outcomes pre-specified in the registration record were reported
Other bias	Unclear risk	Investigator concerns resulting from the significantly increased mortality risk with the lower SpO_2 target range in the SUPPORT Trial publication led to an unscheduled safety analysis when 973 of the planned 1200 infants (81%) had been recruited. A decision was made to terminate recruitment in both the BOOST-II UK and BOOST-II Australia trials based on a pre-specified rule. There was an 8.5% excess in 36-week mortality in the low target group monitored with an oximeter incorporating the revised calibration software (data pooled from both studies, $P < 0.001$ with a significant treatment by software subgroup interaction, $P = 0.006$). The early stopping of the trial (with 81% of the final planned sample size at that point) raises the question of whether this overestimates treatment effect

BPD: Bronchopulomary dysplasia
BSID-III: Bayley Scale of Infant Development-version 3
GMFCS: Gross Motor Functioning Classification System
GP: general practitioner
MDI: Mental Developmental Index
NICU: neonatal intensive care unit
ROP: retinopathy of prematurity
PARCA-R: Parent Report of Children's Abilities-Revised questionnaire
WPPS-III: Wechsler Preschool and Primary Scale of Intelligence-version 3

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Arora 2013	Not a randomised controlled trial. This was a 'before and after' study of a change in policy
Bard 1996	Not a randomised controlled trial.
NCT00845624	Not a randomised controlled trial. This is a prospective cohort study to determine if time spent out of the targeted oxygen saturation range in preterm infants is associated with long-term outcomes such as ROP. Planned to enrol 102 preterm infants < 1500 grams or 32 weeks' gestation
NCT01590316	Intervention not targeting higher or lower oxygen. This is a randomised, blinded, multinational, phase II feasibility clinical trial involving 165 preterm infants born at gestational ages of up to 27 weeks and 6 days, assessing the effect of viewing/not viewing cerebral NIRS oximetry monitoring on brain injury (EEG, ultrasound), mortality and other biomarkers

EEG: electroencephalogram
NIRS: near-infrared spectroscopy
ROP: retinopathy of prematurity

DATA AND ANALYSES

Comparison 1. Lower versus higher targeted oxygen saturations (no subgroups)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Death or major disability by 18 to 24 months corrected age (aligned definition)	5	4754	Risk Ratio (M-H, Fixed, 95% CI)	1.04 [0.98, 1.10]
2 Death or major disability by 18 to 24 months corrected age (trialist defined)	5	4751	Risk Ratio (M-H, Fixed, 95% CI)	1.07 [1.00, 1.14]
3 Death to 18 to 24 months corrected age	5	4873	Risk Ratio (M-H, Fixed, 95% CI)	1.16 [1.03, 1.31]
4 Major disability by 18 to 24 months corrected age (aligned definition)	5	3867	Risk Ratio (M-H, Fixed, 95% CI)	1.01 [0.93, 1.09]
5 Major disability by 18 to 24 months corrected age (trialist defined)	5	3864	Risk Ratio (M-H, Fixed, 95% CI)	1.04 [0.94, 1.14]
6 Death to discharge	5	4958	Risk Ratio (M-H, Fixed, 95% CI)	1.16 [1.03, 1.31]
7 Severe retinopathy of prematurity or retinal therapy (trialist defined)	5	4089	Risk Ratio (M-H, Fixed, 95% CI)	0.72 [0.61, 0.85]
8 Patent ductus arteriosus requiring medical or surgical treatment	5	4928	Risk Ratio (M-H, Fixed, 95% CI)	1.00 [0.95, 1.06]
9 Necrotising enterocolitis	5	4929	Risk Ratio (M-H, Fixed, 95% CI)	1.24 [1.05, 1.47]
10 Cerebral palsy with GMFCS level 2 or higher at 18 to 24 months corrected age	5	3877	Risk Ratio (M-H, Fixed, 95% CI)	1.02 [0.79, 1.32]
11 Blindness	5	3875	Risk Ratio (M-H, Fixed, 95% CI)	1.13 [0.65, 1.97]
12 Severe hearing loss	5	3869	Risk Ratio (M-H, Fixed, 95% CI)	1.02 [0.73, 1.43]
13 Proportion of infants re-admitted to hospital up to 18 to 24 months corrected age	1	295	Risk Ratio (M-H, Fixed, 95% CI)	1.08 [0.93, 1.26]
14 Weight (grams) at discharge home	1	295	Mean Difference (IV, Fixed, 95% CI)	-52.0 [-214.25, 110.25]
15 Weight (kilograms) at 18 or 24 months corrected age	1	280	Mean Difference (IV, Fixed, 95% CI)	0.80 [-0.24, 1.84]
16 Days of endotracheal intubation	2	1386	Mean Difference (IV, Fixed, 95% CI)	0.28 [-1.16, 1.72]
17 Days of CPAP	3	2526	Mean Difference (IV, Fixed, 95% CI)	-0.04 [-1.38, 1.30]
18 Days of supplemental oxygen	3	2507	Mean Difference (IV, Fixed, 95% CI)	-8.78 [-12.02, -5.54]
19 Supplemental oxygen requirement at 36 weeks postmenstrual age	5	4175	Risk Ratio (M-H, Fixed, 95% CI)	0.87 [0.81, 0.94]
20 Days on home oxygen	2	237	Mean Difference (IV, Fixed, 95% CI)	-24.17 [-57.99, 9.66]

21 Quantitative Bayley III scores (Composite Cognitive Score (CCS))	2	1892	Mean Difference (IV, Fixed, 95% CI)	0.55 [-0.91, 2.00]
22 Quantitative Bayley III scores (Composite Language Score (CLS))	1	903	Mean Difference (IV, Fixed, 95% CI)	0.20 [-2.03, 2.43]

Comparison 2. Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by gestational age)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Death or major disability by 18 to 24 months corrected age (trialist defined)	1		Risk Ratio (M-H, Fixed, 95% CI)	Subtotals only
1.1 < 26 weeks	1	537	Risk Ratio (M-H, Fixed, 95% CI)	1.09 [0.89, 1.32]
1.2 ≥ 26 weeks	1	697	Risk Ratio (M-H, Fixed, 95% CI)	1.17 [0.86, 1.60]

Comparison 3. Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by sex)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Death or major disability by 18 to 24 months corrected age (trialist defined)	1	941	Risk Ratio (M-H, Fixed, 95% CI)	1.10 [0.97, 1.26]
1.1 Male	1	503	Risk Ratio (M-H, Fixed, 95% CI)	1.13 [0.96, 1.33]
1.2 Female	1	438	Risk Ratio (M-H, Fixed, 95% CI)	1.07 [0.87, 1.31]

Comparison 4. Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by multiples)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Death or major disability by 18 to 24 months corrected age (trialist defined)	1		Risk Ratio (M-H, Fixed, 95% CI)	Subtotals only
1.1 Singleton	1	670	Risk Ratio (M-H, Fixed, 95% CI)	1.10 [0.94, 1.29]
1.2 Multiple	1	271	Risk Ratio (M-H, Fixed, 95% CI)	1.11 [0.88, 1.39]

Comparison 5. Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by oximeter calibration software)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Death or major disability by 18 to 24 months corrected age (aligned definition)	5	4684	Risk Ratio (M-H, Fixed, 95% CI)	1.04 [0.98, 1.10]
1.1 Original algorithm	5	3003	Risk Ratio (M-H, Fixed, 95% CI)	1.00 [0.94, 1.07]
1.2 Revised algorithm	3	1681	Risk Ratio (M-H, Fixed, 95% CI)	1.13 [1.02, 1.24]
2 Death or major disability by 18 to 24 months corrected age (trialist defined)	5	4681	Risk Ratio (M-H, Fixed, 95% CI)	1.07 [1.00, 1.15]
2.1 Original algorithm	5	3000	Risk Ratio (M-H, Fixed, 95% CI)	1.04 [0.95, 1.13]
2.2 Revised algorithm	3	1681	Risk Ratio (M-H, Fixed, 95% CI)	1.13 [1.02, 1.24]

Comparison 6. Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by oximeter calibration software)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Death by 18 to 24 months corrected age	5	4803	Risk Ratio (M-H, Fixed, 95% CI)	1.16 [1.03, 1.30]
1.1 Original	5	3087	Risk Ratio (M-H, Fixed, 95% CI)	1.05 [0.91, 1.22]
1.2 Revised	3	1716	Risk Ratio (M-H, Fixed, 95% CI)	1.38 [1.13, 1.68]
2 Major disability by 18 to 24 months corrected age (aligned definition)	5	3967	Risk Ratio (M-H, Fixed, 95% CI)	1.01 [0.93, 1.09]
2.1 Original	5	2529	Risk Ratio (M-H, Fixed, 95% CI)	0.99 [0.90, 1.09]
2.2 Revised	3	1438	Risk Ratio (M-H, Fixed, 95% CI)	1.05 [0.91, 1.22]
3 Major disability by 18 to 24 months corrected age (trialist defined)	5	3964	Risk Ratio (M-H, Fixed, 95% CI)	1.04 [0.94, 1.14]
3.1 Original	5	2526	Risk Ratio (M-H, Fixed, 95% CI)	1.02 [0.89, 1.17]
3.2 Revised	3	1438	Risk Ratio (M-H, Fixed, 95% CI)	1.05 [0.91, 1.22]
4 Death to discharge	4	3757	Risk Ratio (M-H, Fixed, 95% CI)	1.18 [1.03, 1.36]
4.1 Original	4	2575	Risk Ratio (M-H, Fixed, 95% CI)	1.06 [0.90, 1.26]
4.2 Revised	2	1182	Risk Ratio (M-H, Fixed, 95% CI)	1.45 [1.15, 1.84]
5 Severe retinopathy of prematurity or retinal therapy	4	3073	Risk Ratio (M-H, Fixed, 95% CI)	0.70 [0.58, 0.84]
5.1 Original	4	2085	Risk Ratio (M-H, Fixed, 95% CI)	0.67 [0.53, 0.84]
5.2 Revised	2	988	Risk Ratio (M-H, Fixed, 95% CI)	0.77 [0.56, 1.05]

Comparison 7. Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by gestational age)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Death by 18 to 24 months corrected age	1		Risk Ratio (M-H, Fixed, 95% CI)	Subtotals only
1.1 < 26 weeks	1	550	Risk Ratio (M-H, Fixed, 95% CI)	1.22 [0.95, 1.57]
1.2 ≥ 26 weeks	1	731	Risk Ratio (M-H, Fixed, 95% CI)	1.25 [0.84, 1.86]
2 Major disability by 18 to 24 months corrected age (trialist defined)	1	976	Risk Ratio (M-H, Fixed, 95% CI)	0.93 [0.64, 1.35]
2.1 < 26 weeks	1	367	Risk Ratio (M-H, Fixed, 95% CI)	0.84 [0.52, 1.37]
2.2 ≥ 26 weeks	1	609	Risk Ratio (M-H, Fixed, 95% CI)	1.07 [0.59, 1.93]

Analysis 1.1. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 1 Death or major disability by 18 to 24 months corrected age (aligned definition).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 1 Death or major disability by 18 to 24 months corrected age (aligned definition)

Study or subgroup	Lower oxygen saturation n/N	Higher oxygen saturation n/N	Risk Ratio M-H,Fixed,95% CI	Weight	Risk Ratio M-H,Fixed,95% CI
Vaucher 2012	363/613	374/624		31.7 %	0.99 [0.90, 1.08]
Schmidt 2013	298/578	283/569		24.4 %	1.04 [0.92, 1.16]
BOOST NZ 2014	65/167	76/168		6.5 %	0.86 [0.67, 1.11]
BOOST-II UK 2016	245/473	220/468		18.9 %	1.10 [0.97, 1.25]
BOOST-II Australia 2016	247/549	217/545		18.6 %	1.13 [0.98, 1.30]
Total (95% CI)	**2380**	**2374**		**100.0 %**	**1.04 [0.98, 1.10]**

Total events: 1218 (Lower oxygen saturation), 1170 (Higher oxygen saturation)

Heterogeneity: $Chi^2 = 5.50$, df = 4 (P = 0.24); $I^2 = 27\%$

Test for overall effect: Z = 1.35 (P = 0.18)

Test for subgroup differences: Not applicable

Analysis 1.2. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 2 Death or major disability by 18 to 24 months corrected age (trialist defined).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 2 Death or major disability by 18 to 24 months corrected age (trialist defined)

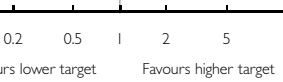
Study or subgroup	Lower oxygen saturation n/N	Higher oxygen saturation n/N	Risk Ratio M-H,Fixed,95% CI	Weight	Risk Ratio M-H,Fixed,95% CI
Vaucher 2012	185/612	171/622		17.5 %	1.10 [0.92, 1.31]
Schmidt 2013	298/578	283/569		29.4 %	1.04 [0.92, 1.16]
BOOST NZ 2014	65/167	76/168		7.8 %	0.86 [0.67, 1.11]
BOOST-II UK 2016	245/473	220/468		22.8 %	1.10 [0.97, 1.25]
BOOST-II Australia 2016	247/549	217/545		22.5 %	1.13 [0.98, 1.30]
Total (95% CI)	**2379**	**2372**		**100.0 %**	**1.07 [1.00, 1.14]**

Total events: 1040 (Lower oxygen saturation), 967 (Higher oxygen saturation)

Heterogeneity: $Chi^2 = 4.04$, df = 4 (P = 0.40); $I^2 = 1\%$

Test for overall effect: Z = 2.01 (P = 0.044)

Test for subgroup differences: Not applicable



Analysis 1.3. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 3 Death to 18 to 24 months corrected age.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 3 Death to 18 to 24 months corrected age

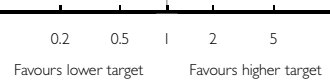
Study or subgroup	Lower oxygen saturation n/N	Higher oxygen saturation n/N	Risk Ratio M-H,Fixed,95% CI	Weight	Risk Ratio M-H,Fixed,95% CI
Vaucher 2012	140/633	118/648		27.9 %	1.21 [0.98, 1.51]
Schmidt 2013	97/585	88/577		21.2 %	1.09 [0.83, 1.42]
BOOST NZ 2014	25/170	27/170		6.5 %	0.93 [0.56, 1.53]
BOOST-II UK 2016	122/484	98/483		23.5 %	1.24 [0.98, 1.57]
BOOST-II Australia 2016	100/561	87/562		20.8 %	1.15 [0.89, 1.50]
Total (95% CI)	**2433**	**2440**		**100.0 %**	**1.16 [1.03, 1.31]**

Total events: 484 (Lower oxygen saturation), 418 (Higher oxygen saturation)

Heterogeneity: $Chi^2 = 1.51$, df = 4 (P = 0.83); $I^2 = 0.0\%$

Test for overall effect: Z = 2.50 (P = 0.012)

Test for subgroup differences: Not applicable



Analysis 1.4. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 4 Major disability by 18 to 24 months corrected age (aligned definition).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 4 Major disability by 18 to 24 months corrected age (aligned definition)

Study or subgroup	Lower oxygen saturation n/N	Higher oxygen saturation n/N	Risk Ratio M-H,Fixed,95% CI	Weight	Risk Ratio M-H,Fixed,95% CI
Vaucher 2012	223/473	256/506		33.5 %	0.93 [0.82, 1.06]
Schmidt 2013	201/488	195/489		26.4 %	1.03 [0.89, 1.20]
BOOST NZ 2014	40/142	49/141		6.7 %	0.81 [0.57, 1.15]
BOOST-II UK 2016	123/351	122/370		16.1 %	1.06 [0.87, 1.30]
BOOST-II Australia 2016	147/449	130/458		17.4 %	1.15 [0.95, 1.40]
Total (95% CI)	**1903**	**1964**		**100.0 %**	**1.01 [0.93, 1.09]**

Total events: 734 (Lower oxygen saturation), 752 (Higher oxygen saturation)

Heterogeneity: $Chi^2 = 5.13$, df = 4 (P = 0.27); $I^2 = 22\%$

Test for overall effect: Z = 0.25 (P = 0.80)

Test for subgroup differences: Not applicable

Analysis 1.5. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 5 Major disability by 18 to 24 months corrected age (trialist defined).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 5 Major disability by 18 to 24 months corrected age (trialist defined)

Study or subgroup	Lower oxygen saturation n/N	Higher oxygen saturation n/N	Risk Ratio M-H,Fixed,95% CI	Weight	Risk Ratio M-H,Fixed,95% CI
Vaucher 2012	45/472	53/504		9.4 %	0.91 [0.62, 1.32]
Schmidt 2013	201/488	195/489		35.9 %	1.03 [0.89, 1.20]
BOOST NZ 2014	40/142	49/141		9.1 %	0.81 [0.57, 1.15]
BOOST-II UK 2016	123/351	122/370		21.9 %	1.06 [0.87, 1.30]
BOOST-II Australia 2016	147/449	130/458		23.7 %	1.15 [0.95, 1.40]
Total (95% CI)	**1902**	**1962**		**100.0 %**	**1.04 [0.94, 1.14]**

Total events: 556 (Lower oxygen saturation), 549 (Higher oxygen saturation)

Heterogeneity: $Chi^2 = 3.61$, df = 4 (P = 0.46); $I^2 = 0.0\%$

Test for overall effect: Z = 0.72 (P = 0.47)

Test for subgroup differences: Not applicable

Analysis 1.6. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 6 Death to discharge.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 6 Death to discharge

Study or subgroup	Lower oxygen saturation n/N	Higher oxygen saturation n/N	Risk Ratio M-H,Fixed,95% CI	Weight	Risk Ratio M-H,Fixed,95% CI
Vaucher 2012	130/654	107/662		26.8 %	1.23 [0.98, 1.55]
Schmidt 2013	95/602	87/599		22.0 %	1.09 [0.83, 1.42]
BOOST NZ 2014	21/170	24/170		6.1 %	0.88 [0.51, 1.51]
BOOST-II UK 2016	115/484	96/483		24.2 %	1.20 [0.94, 1.52]
BOOST-II Australia 2016	99/567	83/567		20.9 %	1.19 [0.91, 1.56]
Total (95% CI)	**2477**	**2481**		**100.0 %**	**1.16 [1.03, 1.31]**

Total events: 460 (Lower oxygen saturation), 397 (Higher oxygen saturation)

Heterogeneity: $Chi^2 = 1.60$, df = 4 (P = 0.81); $I^2 = 0.0\%$

Test for overall effect: Z = 2.40 (P = 0.017)

Test for subgroup differences: Not applicable

Analysis 1.7. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 7 Severe retinopathy of prematurity or retinal therapy (trialist defined).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 7 Severe retinopathy of prematurity or retinal therapy (trialist defined)

Study or subgroup	Lower targeting n/N	HIgher targeting n/N	Risk Ratio M-H,Fixed,95% CI	Weight	Risk Ratio M-H,Fixed,95% CI
Vaucher 2012	36/482	93/514		29.9 %	0.41 [0.29, 0.59]
Schmidt 2013	63/500	66/503		21.9 %	0.96 [0.70, 1.33]
BOOST NZ 2014	11/158	12/150		4.1 %	0.87 [0.40, 1.91]
BOOST-II UK 2016	67/395	86/403		28.3 %	0.79 [0.60, 1.06]
BOOST-II Australia 2016	37/487	48/497		15.8 %	0.79 [0.52, 1.19]
Total (95% CI)	**2022**	**2067**		**100.0 %**	**0.72 [0.61, 0.85]**

Total events: 214 (Lower targeting), 305 (HIgher targeting)

Heterogeneity: Chi^2 = 12.90, df = 4 (P = 0.01); I^2 =69%

Test for overall effect: Z = 3.97 (P = 0.000072)

Test for subgroup differences: Not applicable

Analysis 1.8. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 8 Patent ductus arteriosus requiring medical or surgical treatment.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 8 Patent ductus arteriosus requiring medical or surgical treatment

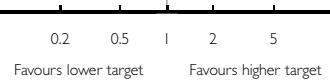
Study or subgroup	Lower oxygen saturation n/N	Higher oxygen saturation n/N	Risk Ratio M-H,Fixed,95% CI	Weight	Risk Ratio M-H,Fixed,95% CI
Vaucher 2012	307/641	324/648		26.7 %	0.96 [0.86, 1.07]
Schmidt 2013	324/602	332/599		27.5 %	0.97 [0.88, 1.08]
BOOST NZ 2014	104/170	90/170		7.4 %	1.16 [0.96, 1.39]
BOOST-II UK 2016	198/482	186/483		15.4 %	1.07 [0.91, 1.25]
BOOST-II Australia 2016	280/567	277/566		22.9 %	1.01 [0.90, 1.14]
Total (95% CI)	**2462**	**2466**		**100.0 %**	**1.00 [0.95, 1.06]**

Total events: 1213 (Lower oxygen saturation), 1209 (Higher oxygen saturation)

Heterogeneity: $Chi^2 = 3.88$, df = 4 (P = 0.42); $I^2 = 0.0\%$

Test for overall effect: Z = 0.16 (P = 0.87)

Test for subgroup differences: Not applicable



Analysis 1.9. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 9 Necrotising enterocolitis.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 9 Necrotising enterocolitis

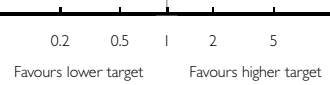
Study or subgroup	Lower oxygen saturation n/N	Higher oxygen saturation n/N	Risk Ratio M-H,Fixed,95% CI	Weight	Risk Ratio M-H,Fixed,95% CI
Vaucher 2012	76/641	70/649		31.2 %	1.10 [0.81, 1.49]
Schmidt 2013	74/602	56/599		25.2 %	1.31 [0.95, 1.83]
BOOST NZ 2014	15/170	12/170		5.4 %	1.25 [0.60, 2.59]
BOOST-II UK 2016	71/484	52/480		23.4 %	1.35 [0.97, 1.89]
BOOST-II Australia 2016	41/567	33/567		14.8 %	1.24 [0.80, 1.94]
Total (95% CI)	**2464**	**2465**		**100.0 %**	**1.24 [1.05, 1.47]**

Total events: 277 (Lower oxygen saturation), 223 (Higher oxygen saturation)

Heterogeneity: $Chi^2 = 0.98$, df = 4 (P = 0.91); $I^2 = 0.0\%$

Test for overall effect: Z = 2.55 (P = 0.011)

Test for subgroup differences: Not applicable



Analysis 1.10. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 10 Cerebral palsy with GMFCS level 2 or higher at 18 to 24 months corrected age.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 10 Cerebral palsy with GMFCS level 2 or higher at 18 to 24 months corrected age

Study or subgroup	Lower oxygen saturation n/N	Higher oxygen saturation n/N	Risk Ratio M-H,Fixed,95% CI	Weight	Risk Ratio M-H,Fixed,95% CI
Vaucher 2012	20/479	20/511		18.3 %	1.07 [0.58, 1.96]
Schmidt 2013	30/488	31/488		29.4 %	0.97 [0.60, 1.57]
BOOST NZ 2014	5/144	7/142		6.7 %	0.70 [0.23, 2.17]
BOOST-II UK 2016	35/353	24/370		22.2 %	1.53 [0.93, 2.52]
BOOST-II Australia 2016	16/446	25/456		23.4 %	0.65 [0.35, 1.21]
Total (95% CI)	**1910**	**1967**		**100.0 %**	**1.02 [0.79, 1.32]**

Total events: 106 (Lower oxygen saturation), 107 (Higher oxygen saturation)

Heterogeneity: $Chi^2 = 5.02$, df = 4 (P = 0.29); $I^2 = 20\%$

Test for overall effect: Z = 0.14 (P = 0.88)

Test for subgroup differences: Not applicable

Analysis 1.11. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 11 Blindness.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 11 Blindness

Study or subgroup	Lower oxygen saturation n/N	Higher oxygen saturation n/N	Risk Ratio M-H,Fixed,95% CI	Weight	Risk Ratio M-H,Fixed,95% CI
Vaucher 2012	5/479	6/511		25.2 %	0.89 [0.27, 2.89]
Schmidt 2013	5/487	3/488		13.0 %	1.67 [0.40, 6.95]
BOOST NZ 2014	0/143	1/138		6.6 %	0.32 [0.01, 7.83]
BOOST-II UK 2016	12/349	11/369		46.5 %	1.15 [0.52, 2.58]
BOOST-II Australia 2016	3/452	2/459		8.6 %	1.52 [0.26, 9.07]
Total (95% CI)	**1910**	**1965**		**100.0 %**	**1.13 [0.65, 1.97]**

Total events: 25 (Lower oxygen saturation), 23 (Higher oxygen saturation)

Heterogeneity: $Chi^2 = 1.15$, df = 4 (P = 0.89); $I^2 = 0.0\%$

Test for overall effect: Z = 0.43 (P = 0.66)

Test for subgroup differences: Not applicable

Analysis 1.12. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 12 Severe hearing loss.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 12 Severe hearing loss

Study or subgroup	Lower oxygen saturation n/N	Higher oxygen saturation n/N	Risk Ratio M-H,Fixed,95% CI	Weight	Risk Ratio M-H,Fixed,95% CI
Vaucher 2012	12/479	12/511		17.9 %	1.07 [0.48, 2.35]
Schmidt 2013	18/487	12/489		18.5 %	1.51 [0.73, 3.09]
BOOST NZ 2014	2/135	1/136		1.5 %	2.01 [0.18, 21.96]
BOOST-II UK 2016	22/352	32/369		48.2 %	0.72 [0.43, 1.22]
BOOST-II Australia 2016	11/452	9/459		13.8 %	1.24 [0.52, 2.97]
Total (95% CI)	**1905**	**1964**		**100.0 %**	**1.02 [0.73, 1.43]**

Total events: 65 (Lower oxygen saturation), 66 (Higher oxygen saturation)

Heterogeneity: $Chi^2 = 3.34$, df = 4 (P = 0.50); $I^2 = 0.0\%$

Test for overall effect: Z = 0.11 (P = 0.91)

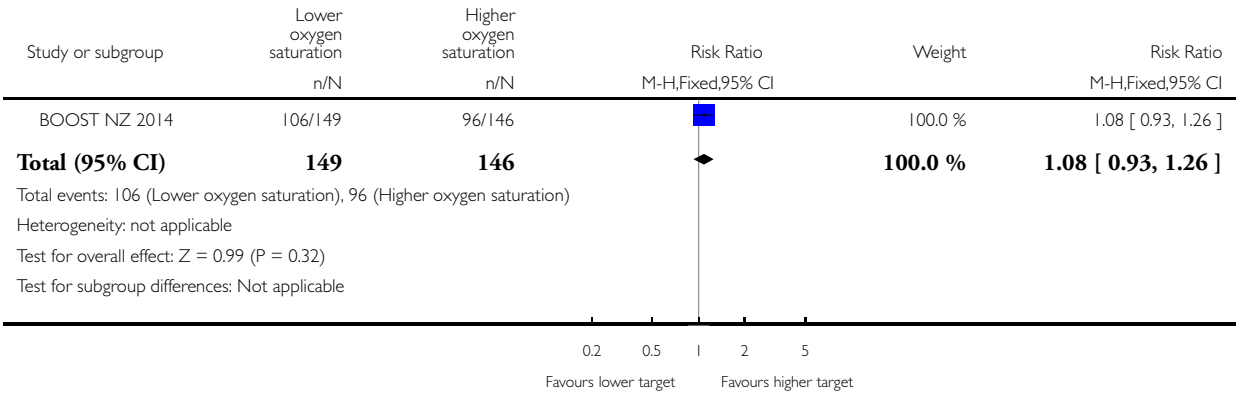
Test for subgroup differences: Not applicable

Analysis 1.13. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 13 Proportion of infants re-admitted to hospital up to 18 to 24 months corrected age.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 13 Proportion of infants re-admitted to hospital up to 18 to 24 months corrected age

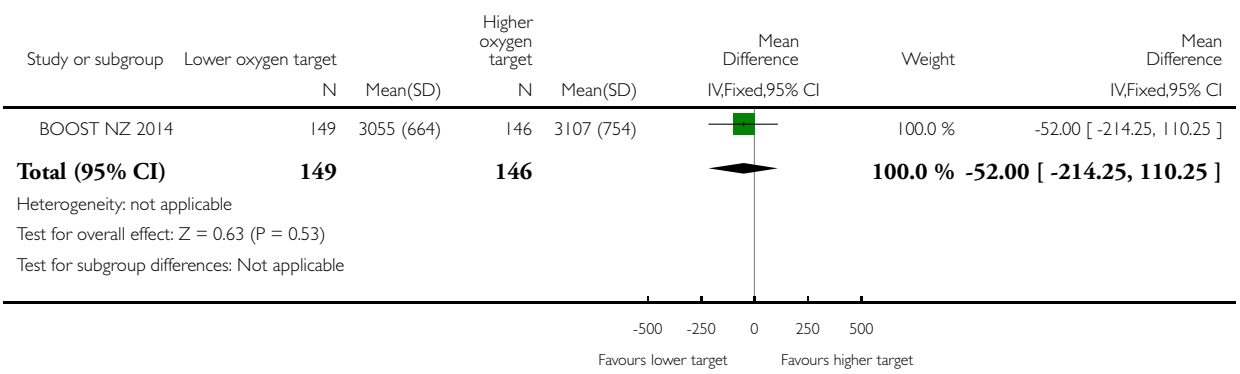


Analysis 1.14. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 14 Weight (grams) at discharge home.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 14 Weight (grams) at discharge home



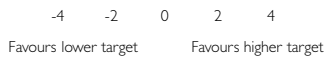
Analysis 1.15. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 15 Weight (kilograms) at 18 or 24 months corrected age.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 15 Weight (kilograms) at 18 or 24 months corrected age

Study or subgroup	Lower oxygen target		Higher oxygen target		Mean Difference	Weight	Mean Difference
	N	Mean(SD)	N	Mean(SD)	IV,Fixed,95% CI		IV,Fixed,95% CI
BOOST NZ 2014	139	11 (4.16)	141	10.2 (4.73)		100.0 %	0.80 [-0.24, 1.84]
Total (95% CI)	**139**		**141**			**100.0 %**	**0.80 [-0.24, 1.84]**
Heterogeneity: not applicable							
Test for overall effect: Z = 1.50 (P = 0.13)							
Test for subgroup differences: Not applicable							

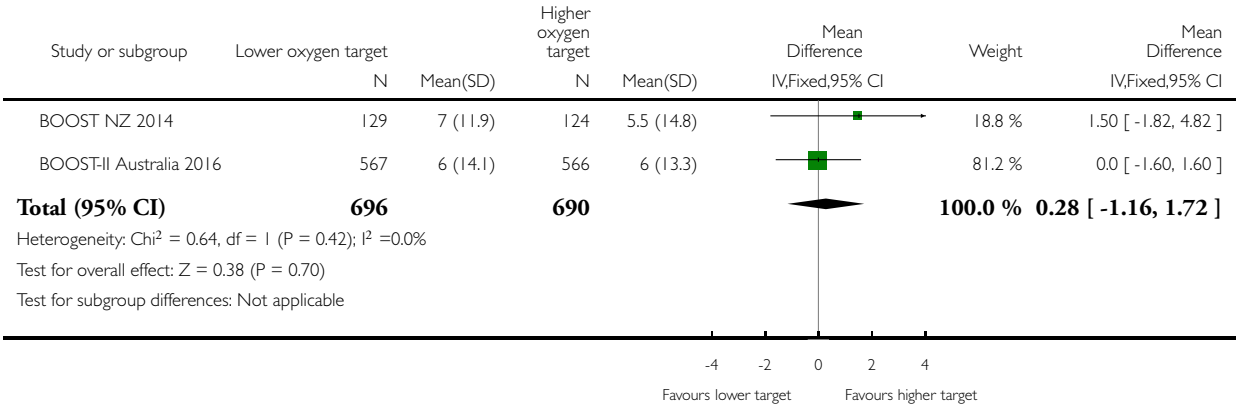


Analysis 1.16. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 16 Days of endotracheal intubation.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 16 Days of endotracheal intubation



Analysis 1.17. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 17 Days of CPAP.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 17 Days of CPAP

Study or subgroup	Lower oxygen target		Higher oxygen target		Mean Difference	Weight	Mean Difference
	N	Mean(SD)	N	Mean(SD)	IV,Fixed,95% CI		IV,Fixed,95% CI
Vaucher 2012	534	17.1 (13.9)	564	17 (14.2)		65.0 %	0.10 [-1.56, 1.76]
BOOST NZ 2014	149	48 (20.7)	146	49 (14.8)		10.7 %	-1.00 [-5.10, 3.10]
BOOST-II Australia 2016	567	37 (23.7)	566	37 (23)		24.3 %	0.0 [-2.72, 2.72]
Total (95% CI)	**1250**		**1276**			**100.0 %**	**-0.04 [-1.38, 1.30]**

Heterogeneity: $Chi^2 = 0.24$, df = 2 (P = 0.89); $I^2 = 0.0\%$

Test for overall effect: Z = 0.06 (P = 0.95)

Test for subgroup differences: Not applicable

Analysis 1.18. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 18 Days of supplemental oxygen.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

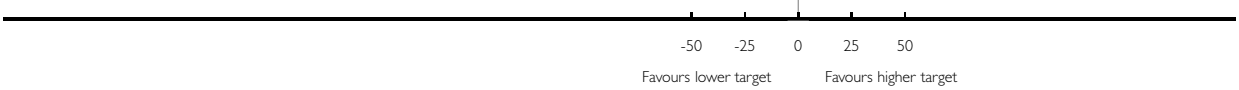
Outcome: 18 Days of supplemental oxygen

Study or subgroup	Lower oxygen target		Higher oxygen target		Mean Difference	Weight	Mean Difference
	N	Mean(SD)	N	Mean(SD)	IV,Fixed,95% CI		IV,Fixed,95% CI
Vaucher 2012	534	59.8 (37)	564	67.4 (35.6)		56.8 %	-7.60 [-11.90, -3.30]
BOOST NZ 2014	139	41 (51.1)	138	58 (51.1)		7.2 %	-17.00 [-29.04, -4.96]
BOOST-II Australia 2016	566	32 (43.7)	566	41 (48.9)		36.0 %	-9.00 [-14.40, -3.60]
Total (95% CI)	**1239**		**1268**			**100.0 %**	**-8.78 [-12.02, -5.54]**

Heterogeneity: $Chi^2 = 2.09$, df = 2 (P = 0.35); $I^2 = 4\%$

Test for overall effect: Z = 5.31 (P < 0.00001)

Test for subgroup differences: Not applicable



Analysis 1.19. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 19 Supplemental oxygen requirement at 36 weeks postmenstrual age.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 19 Supplemental oxygen requirement at 36 weeks postmenstrual age

Study or subgroup	Lower oxygen saturation n/N	Higher oxygen saturation n/N	Risk Ratio M-H,Fixed,95% CI	Weight	Risk Ratio M-H,Fixed,95% CI
Vaucher 2012	203/540	265/568		28.1 %	0.81 [0.70, 0.93]
Schmidt 2013	164/515	171/517		18.5 %	0.96 [0.81, 1.15]
BOOST NZ 2014	40/153	58/147		6.4 %	0.66 [0.47, 0.92]
BOOST-II UK 2016	193/372	212/392		22.4 %	0.96 [0.84, 1.10]
BOOST-II Australia 2016	188/477	230/494		24.5 %	0.85 [0.73, 0.98]
Total (95% CI)	**2057**	**2118**		**100.0 %**	**0.87 [0.81, 0.94]**

Total events: 788 (Lower oxygen saturation), 936 (Higher oxygen saturation)

Heterogeneity: $Chi^2 = 7.18$, df = 4 (P = 0.13); $I^2 = 44\%$

Test for overall effect: Z = 3.78 (P = 0.00015)

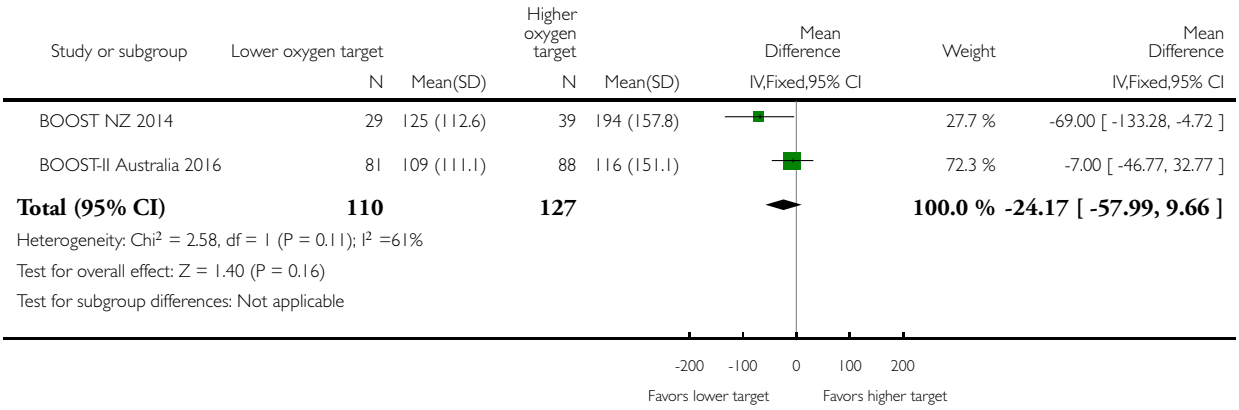
Test for subgroup differences: Not applicable

Analysis 1.20. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 20 Days on home oxygen.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 20 Days on home oxygen

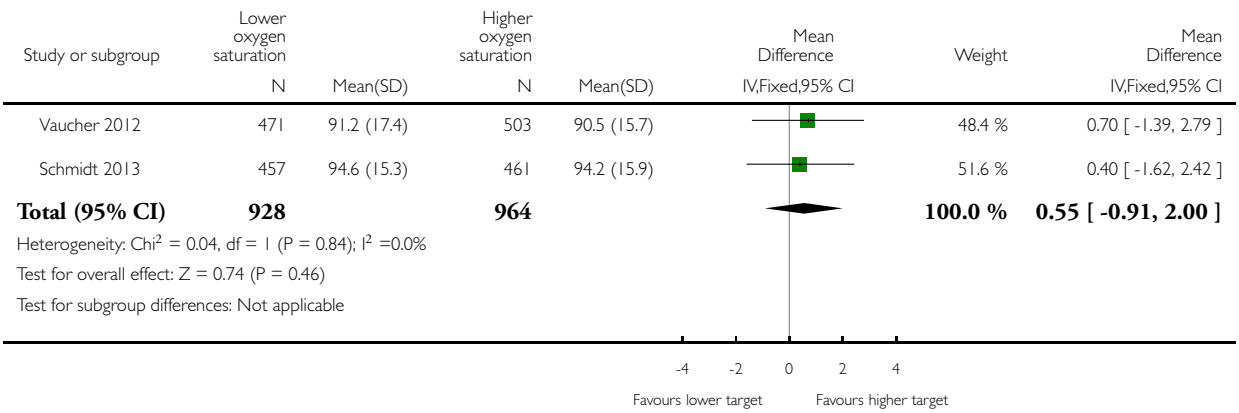


Analysis 1.21. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 21 Quantitative Bayley III scores (Composite Cognitive Score (CCS)).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 21 Quantitative Bayley III scores (Composite Cognitive Score (CCS))



Analysis 1.22. Comparison 1 Lower versus higher targeted oxygen saturations (no subgroups), Outcome 22 Quantitative Bayley III scores (Composite Language Score (CLS)).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 1 Lower versus higher targeted oxygen saturations (no subgroups)

Outcome: 22 Quantitative Bayley III scores (Composite Language Score (CLS))

Study or subgroup	Lower oxygen target		Higher oxygen target		Mean Difference	Weight	Mean Difference
	N	Mean(SD)	N	Mean(SD)	IV,Fixed,95% CI		IV,Fixed,95% CI
Schmidt 2013	449	88.6 (17.3)	454	88.4 (16.9)		100.0 %	0.20 [-2.03, 2.43]
Total (95% CI)	**449**		**454**			**100.0 %**	**0.20 [-2.03, 2.43]**

Heterogeneity: not applicable

Test for overall effect: Z = 0.18 (P = 0.86)

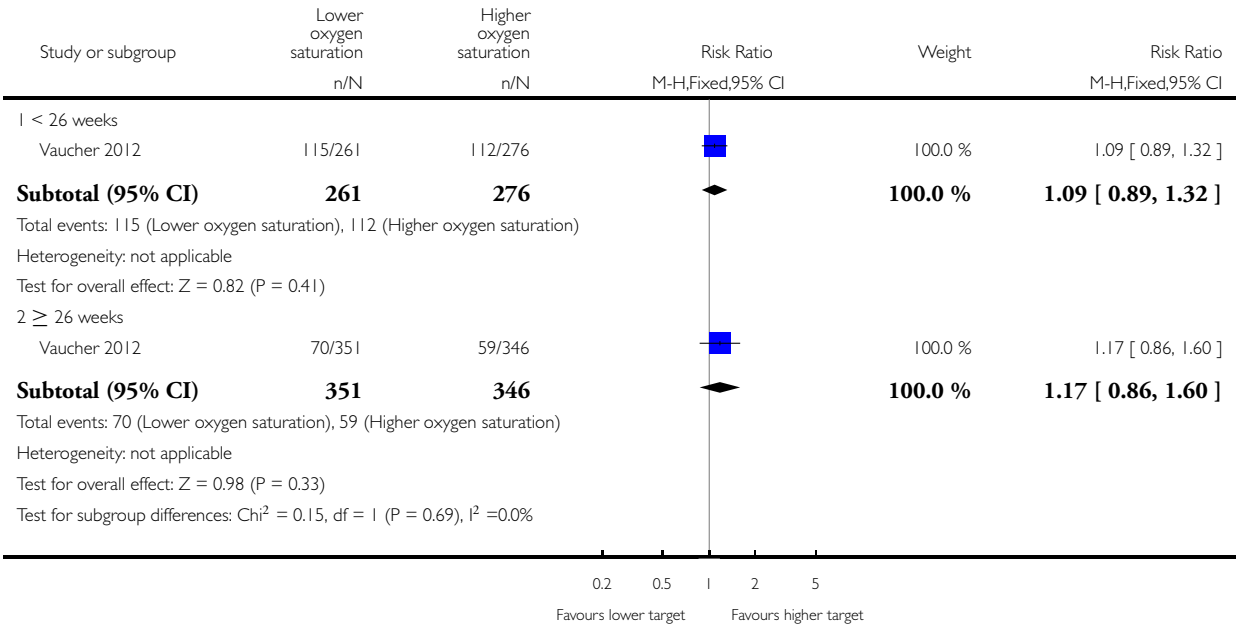
Test for subgroup differences: Not applicable

Analysis 2.1. Comparison 2 Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by gestational age), Outcome 1 Death or major disability by 18 to 24 months corrected age (trialist defined).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 2 Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by gestational age)

Outcome: 1 Death or major disability by 18 to 24 months corrected age (trialist defined)

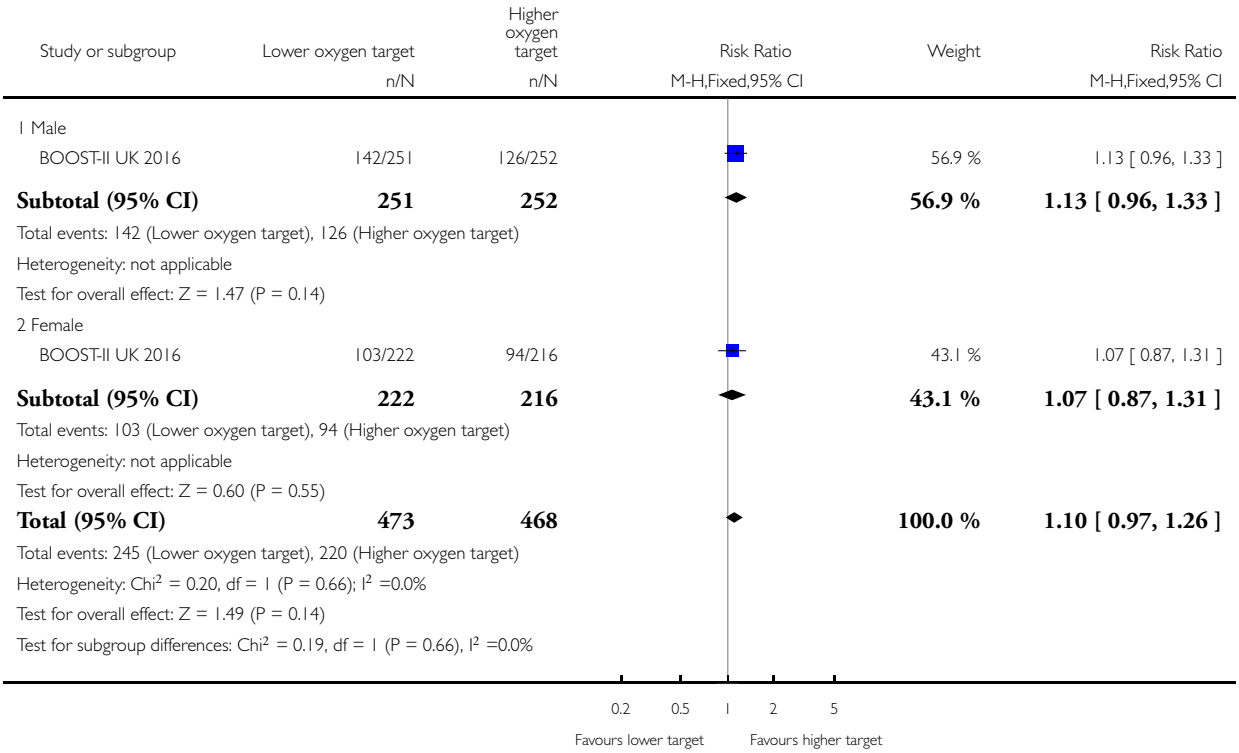


Analysis 3.1. Comparison 3 Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by sex), Outcome 1 Death or major disability by 18 to 24 months corrected age (trialist defined).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 3 Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by sex)

Outcome: 1 Death or major disability by 18 to 24 months corrected age (trialist defined)

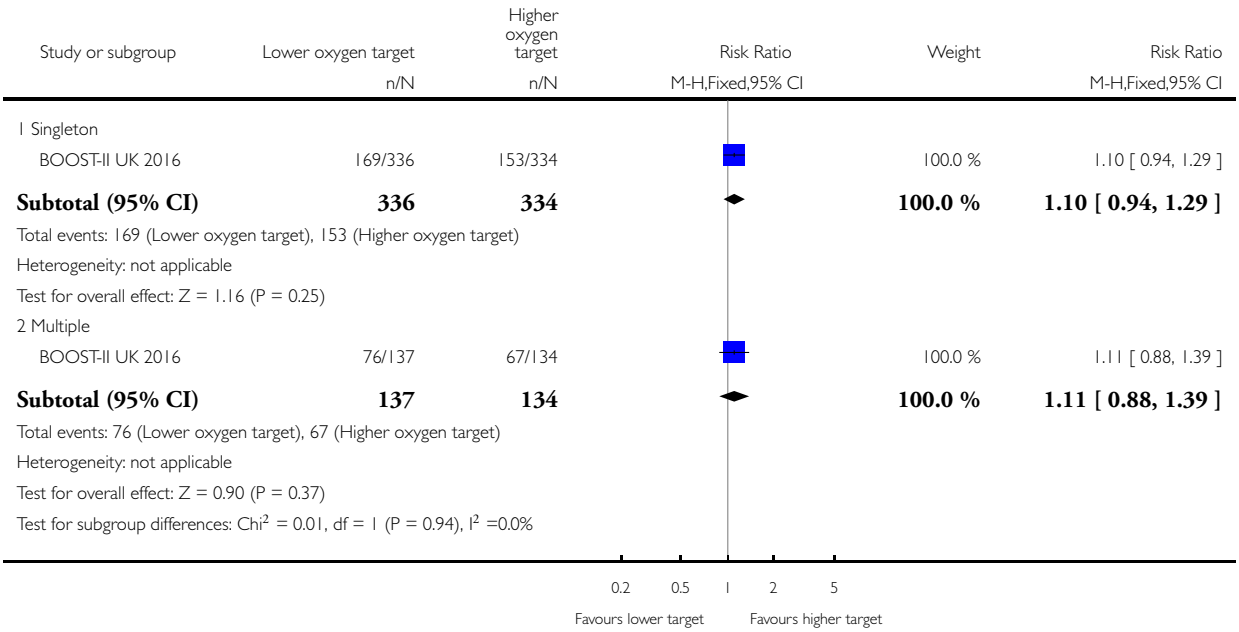


Analysis 4.1. Comparison 4 Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by multiples), Outcome 1 Death or major disability by 18 to 24 months corrected age (trialist defined).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 4 Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by multiples)

Outcome: 1 Death or major disability by 18 to 24 months corrected age (trialist defined)

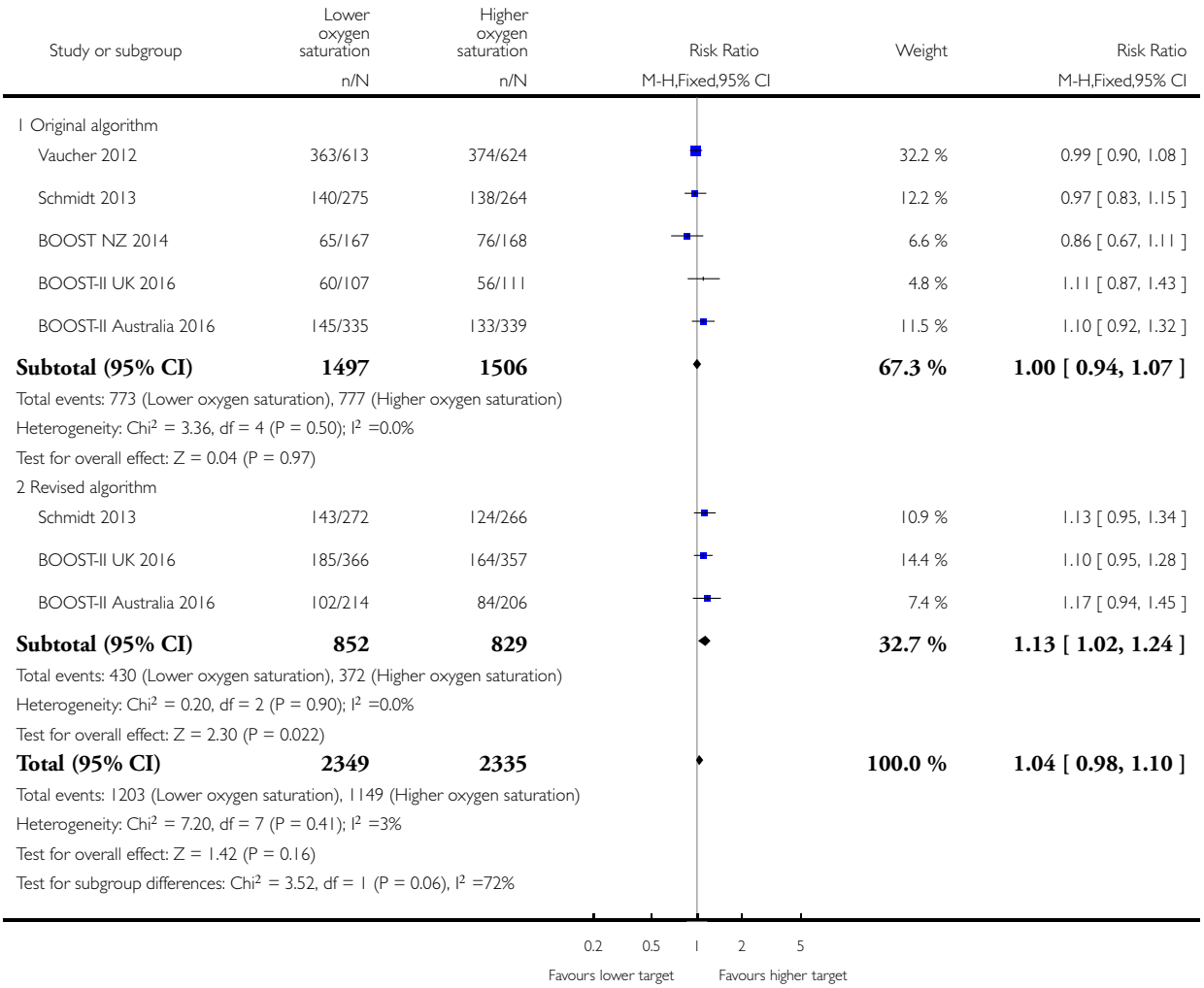


Analysis 5.1. Comparison 5 Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by oximeter calibration software), Outcome 1 Death or major disability by 18 to 24 months corrected age (aligned definition).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 5 Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by oximeter calibration software)

Outcome: 1 Death or major disability by 18 to 24 months corrected age (aligned definition)

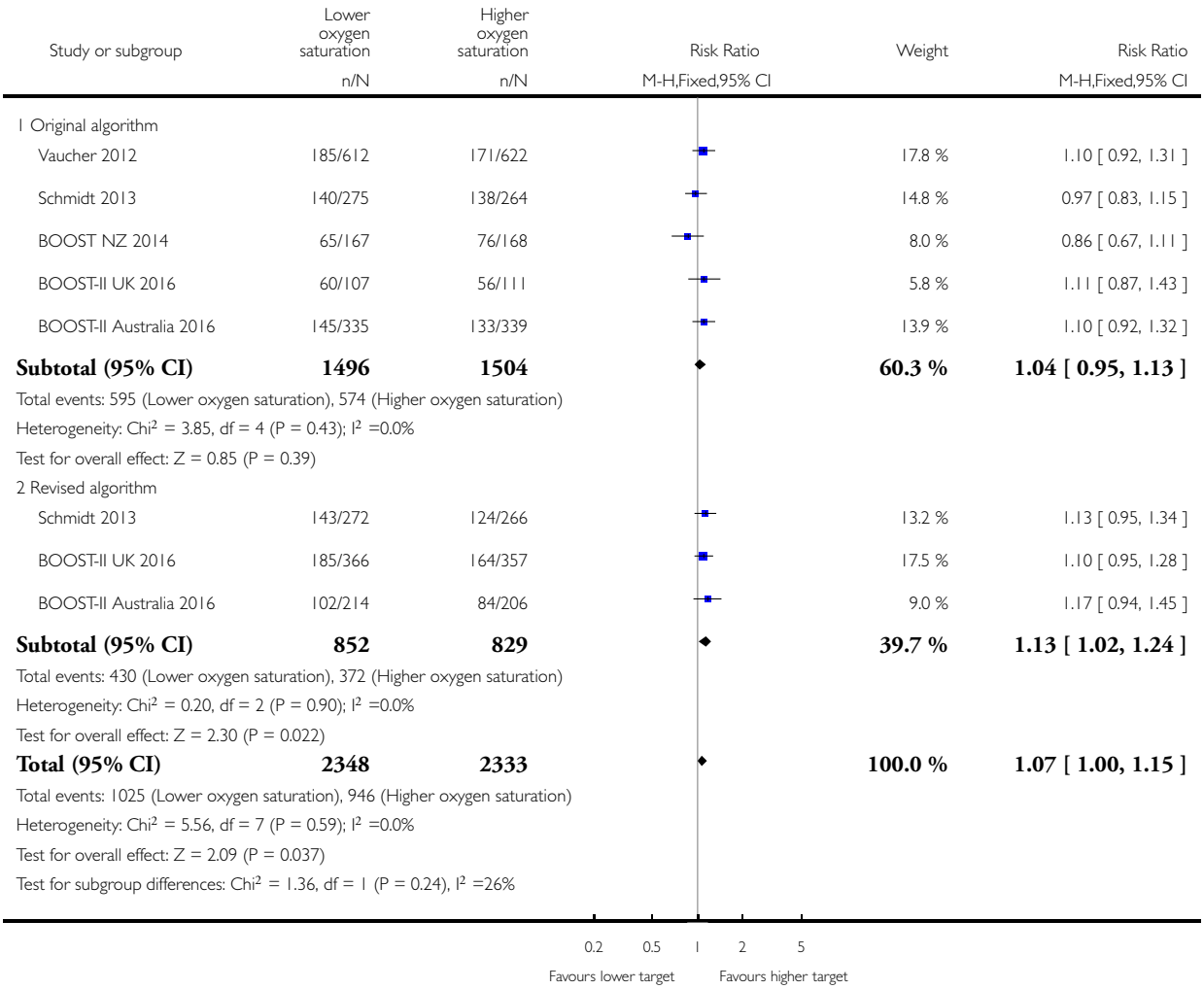


Analysis 5.2. Comparison 5 Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by oximeter calibration software), Outcome 2 Death or major disability by 18 to 24 months corrected age (trialist defined).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 5 Lower versus higher targeted oxygen saturations (primary outcome, subgrouped by oximeter calibration software)

Outcome: 2 Death or major disability by 18 to 24 months corrected age (trialist defined)

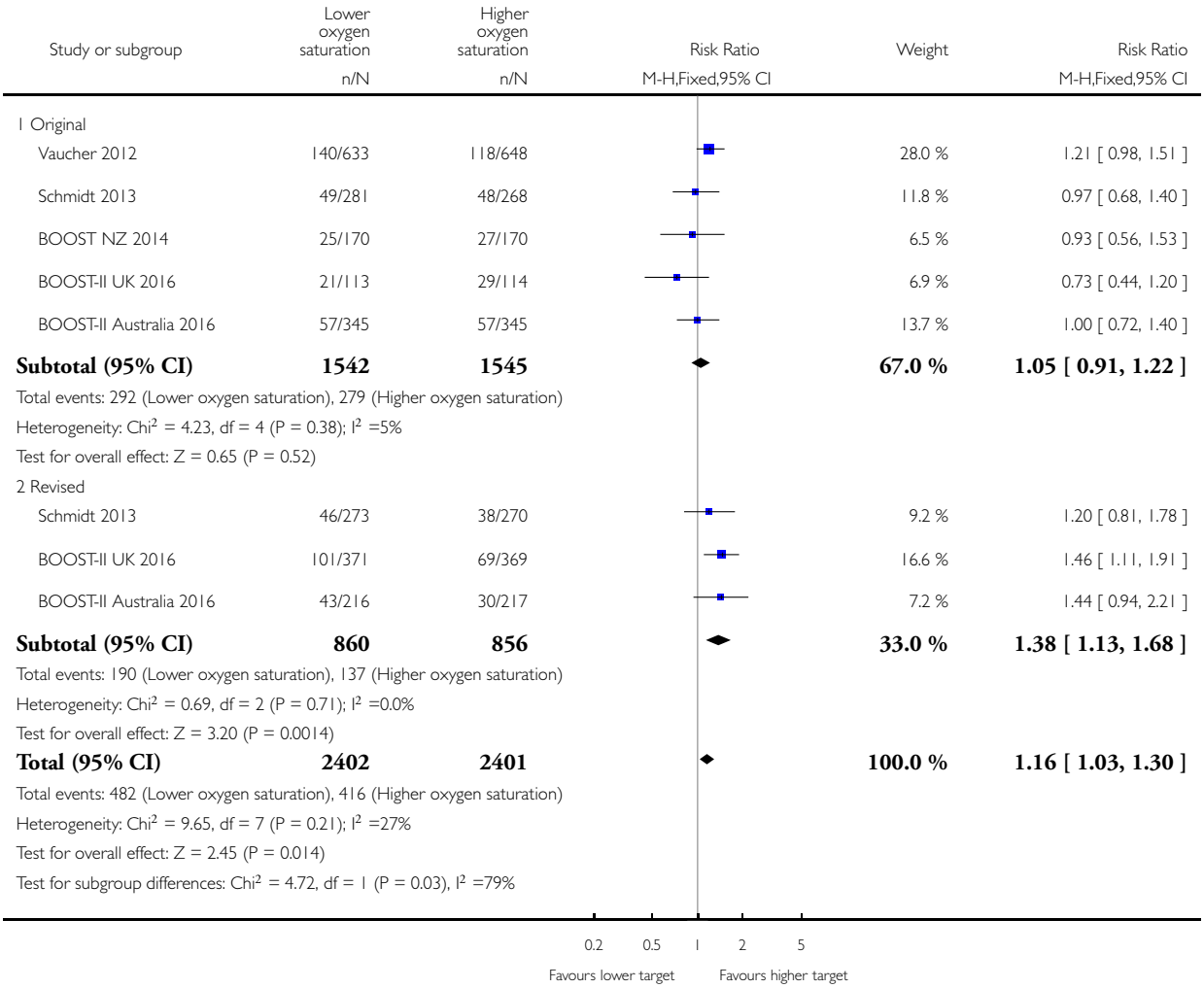


Analysis 6.1. Comparison 6 Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by oximeter calibration software), Outcome 1 Death by 18 to 24 months corrected age.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 6 Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by oximeter calibration software)

Outcome: 1 Death by 18 to 24 months corrected age

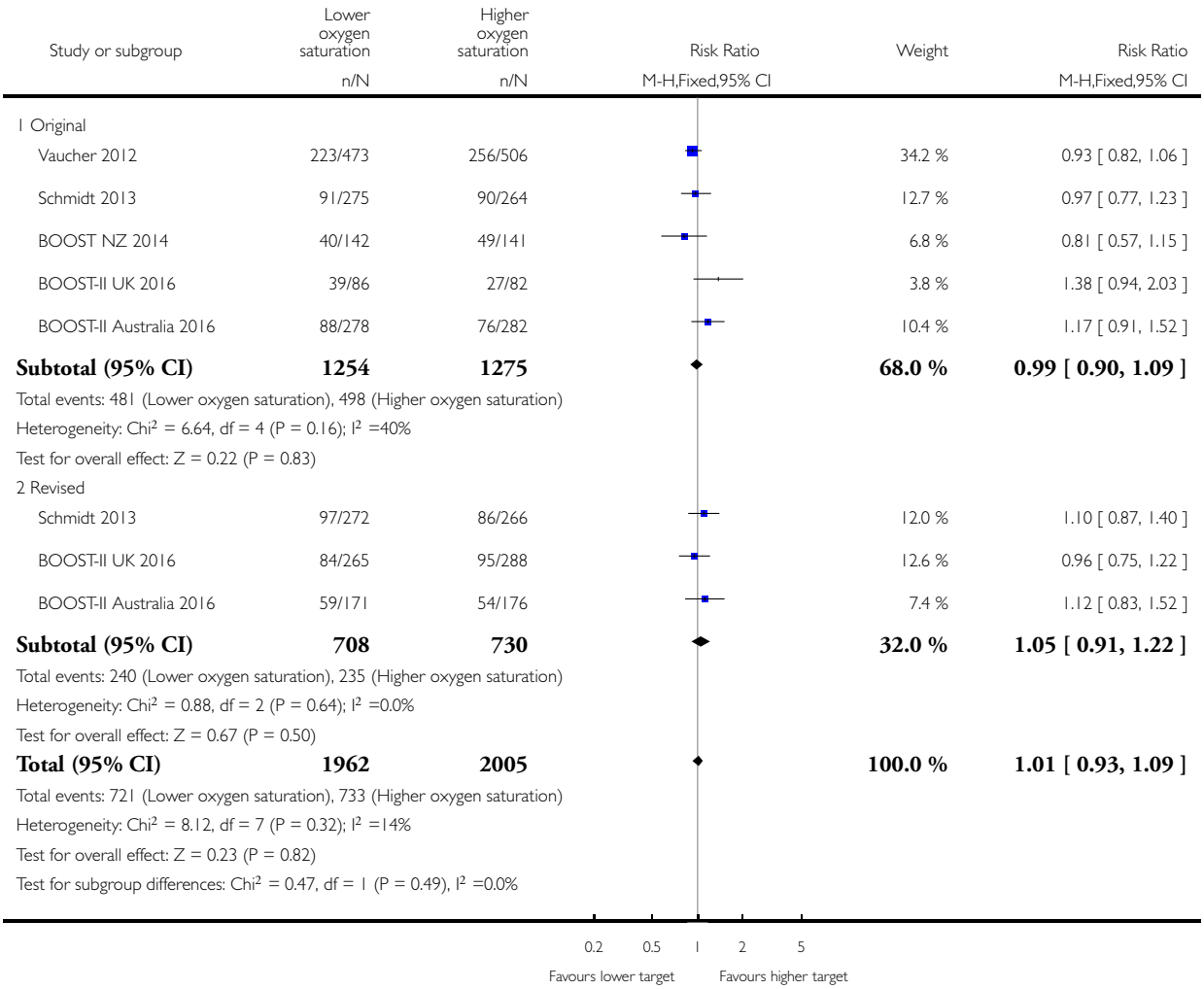


Analysis 6.2. Comparison 6 Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by oximeter calibration software), Outcome 2 Major disability by 18 to 24 months corrected age (aligned definition).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 6 Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by oximeter calibration software)

Outcome: 2 Major disability by 18 to 24 months corrected age (aligned definition)

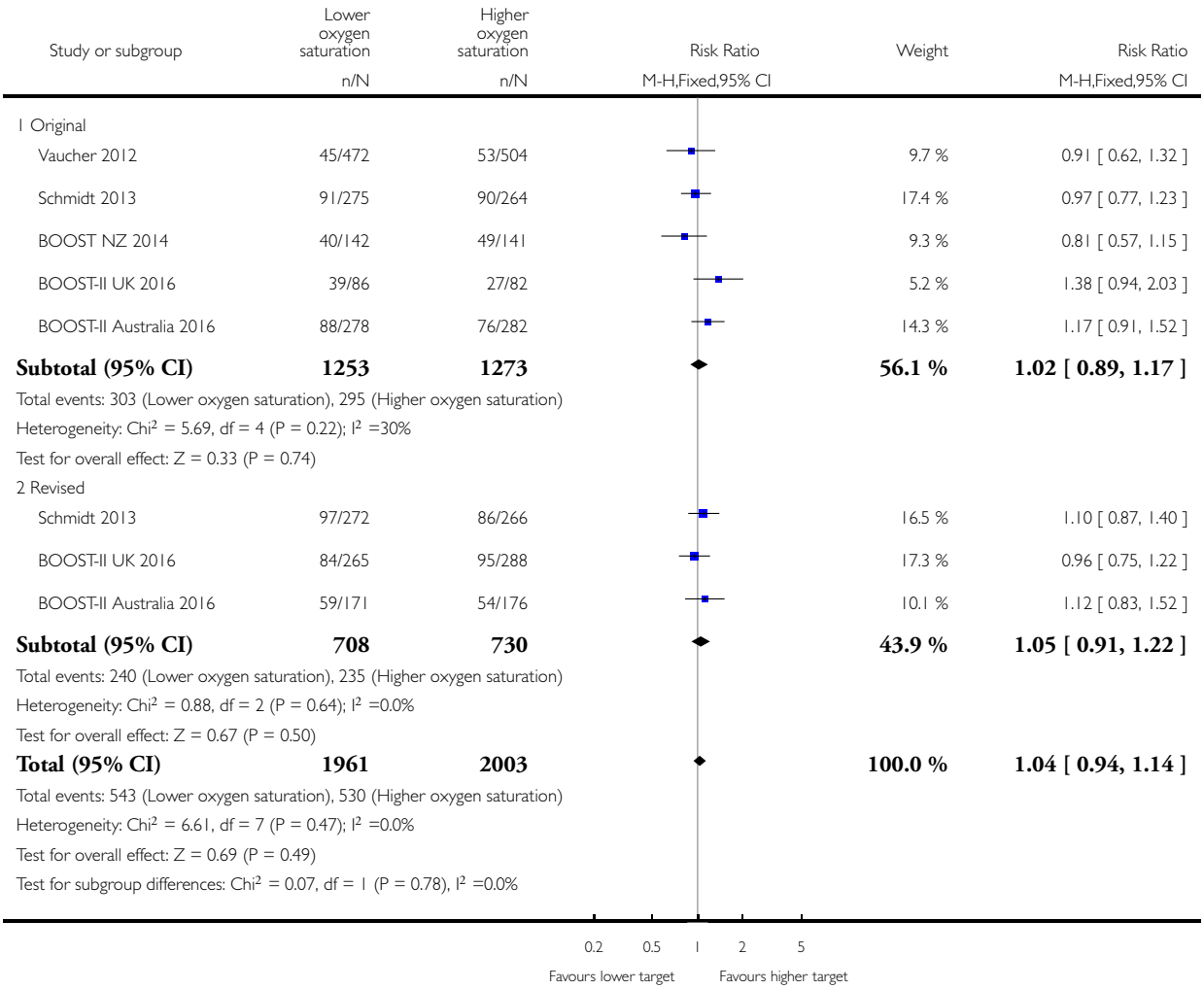


Analysis 6.3. Comparison 6 Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by oximeter calibration software), Outcome 3 Major disability by 18 to 24 months corrected age (trialist defined).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 6 Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by oximeter calibration software)

Outcome: 3 Major disability by 18 to 24 months corrected age (trialist defined)

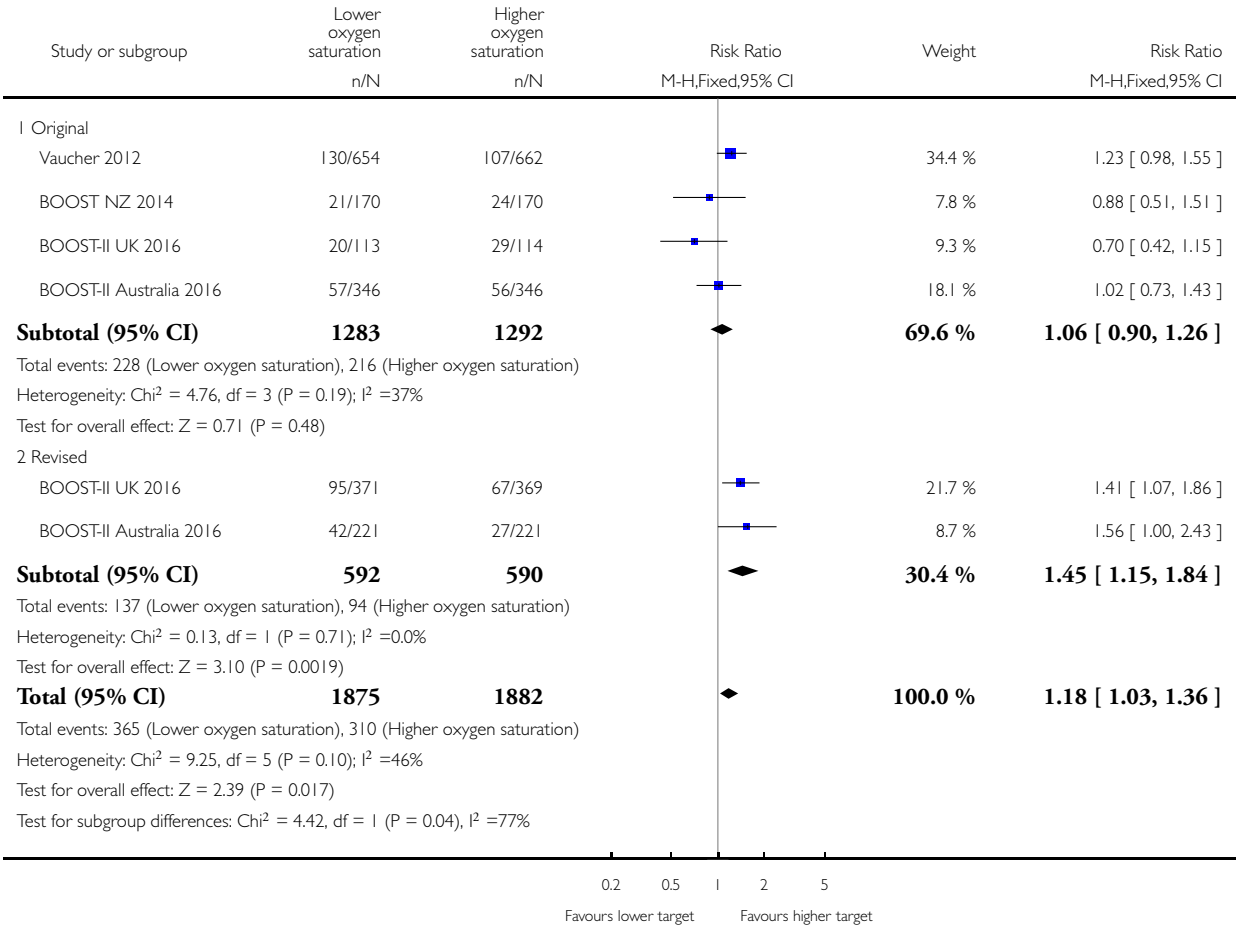


Analysis 6.4. Comparison 6 Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by oximeter calibration software), Outcome 4 Death to discharge.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 6 Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by oximeter calibration software)

Outcome: 4 Death to discharge

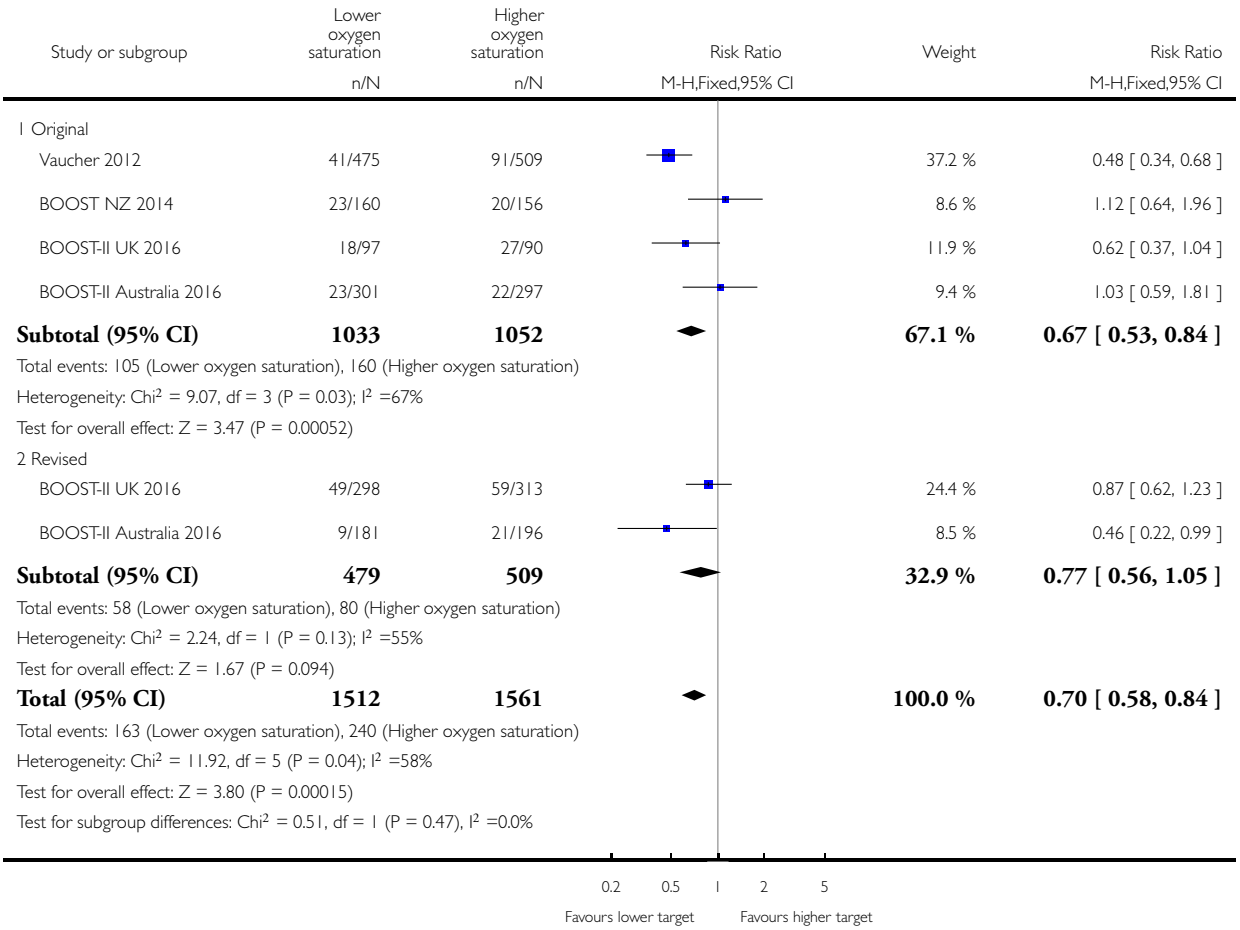


Analysis 6.5. Comparison 6 Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by oximeter calibration software), Outcome 5 Severe retinopathy of prematurity or retinal therapy.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 6 Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by oximeter calibration software)

Outcome: 5 Severe retinopathy of prematurity or retinal therapy

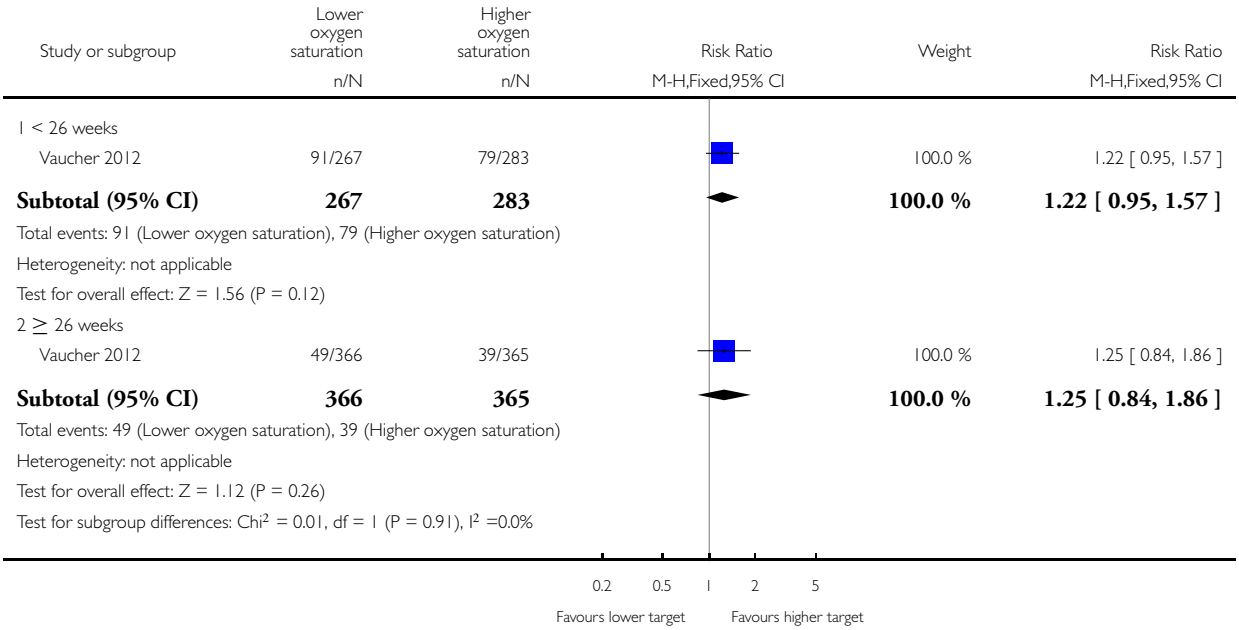


Analysis 7.1. Comparison 7 Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by gestational age), Outcome 1 Death by 18 to 24 months corrected age.

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 7 Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by gestational age)

Outcome: 1 Death by 18 to 24 months corrected age

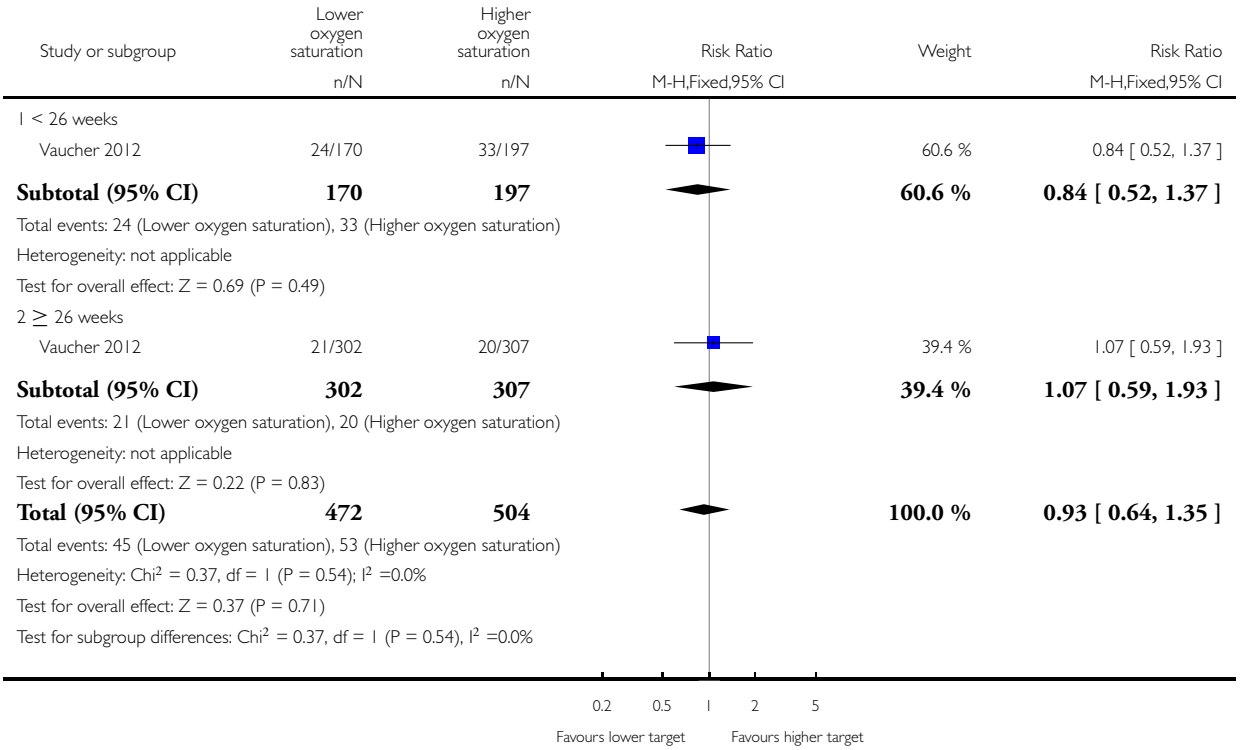


Analysis 7.2. Comparison 7 Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by gestational age), Outcome 2 Major disability by 18 to 24 months corrected age (trialist defined).

Review: Effects of targeting lower versus higher arterial oxygen saturations on death or disability in preterm infants

Comparison: 7 Lower versus higher targeted oxygen saturations (secondary outcomes, subgrouped by gestational age)

Outcome: 2 Major disability by 18 to 24 months corrected age (trialist defined)



CONTRIBUTIONS OF AUTHORS

All authors contributed to the conception and design of the study, and the interpretation of data, reviewed and commented on the intellectual content, and gave final approval of the document to be published. Additional tasks undertaken by specific authors included:

Lisa Askie: screened the search results, extracted and entered data into RevMan, checked the extracted data for accuracy, undertook the initial 'Risk of bias' assessments, contributed to the subsequent 'Risk of bias' assessments discussions, and drafted the manuscript.

Brian Darlow: screened the search results, checked the extracted data for accuracy, and contributed to the 'Risk of bias' assessments discussions.

Peter Davis: screened the search results, checked the extracted data for accuracy, and contributed to the 'Risk of bias' assessments discussions.

Neil Finer: checked the extracted data for accuracy and contributed to the 'Risk of bias' assessments discussions.

Ben Stenson: checked the extracted data for accuracy and contributed to the 'Risk of bias' assessments discussions.

Maximo Vento: checked the extracted data for accuracy and contributed to the 'Risk of bias' assessments discussions.

Robin Whyte: checked the extracted data for accuracy, undertook the initial 'Risk of bias' assessments, and contributed to the subsequent 'Risk of bias' assessments discussions.

DECLARATIONS OF INTEREST

Six members of the authorship team were investigators in the included studies and the NeOProM Collaboration. One member was included for his expertise in the field but had no affiliation with the included studies.

Lisa Askie is a member of the BOOST II Australia writing committee and the NeOProM Collaboration.

Brian Darlow is a member of the BOOST-NZ trial management committee, the BOOST II Australia trial management committee, and the NeOProM Collaboration.

Peter Davis is a member of the BOOST-II Australia trial management committee and the NeOProM Collaboration.

Neil Finer is a member of the SUPPORT trial management committee and the NeOProM Collaboration.

Ben Stenson is a member of the BOOST-II UK steering committee and the NeOProM Collaboration.

Maximo Vento has no conflicts of interest to declare.

Robin Whyte is a member of the COT trial management committee and the NeOProM Collaboration.

SOURCES OF SUPPORT

Internal sources

- NHMRC Clinical Trials Centre, University of Sydney, Australia.

A/Prof Askie's time on this Cochrane Review was partly supported by a National Health and Medical Research Council Program Grant (1037786).

- Eunice Kennedy Shriver National Institute of Child Health and Human Development National Institutes of Health, Department of Health and Human Services, USA.

A/Prof Askie's time on this Cochrane Review was partly supported by a grant (R03HD 079867) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development National Institutes of Health.

External sources

- Eunice Kennedy Shriver National Institute of Child Health and Human Development National Institutes of Health, Department of Health and Human Services, USA.

Editorial support of the Cochrane Neonatal Review Group has been funded with Federal funds from the Eunice Kennedy Shriver National Institute of Child Health and Human Development National Institutes of Health, Department of Health and Human Services, USA, under Contract No. HHSN275201100016C.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

The title of the review was changed to the "Effects of targeting lower versus higher arterial oxygen saturations..." from the original "...higher versus lower ..." to better reflect that, for the purposes of this review, lower oxygen targeting was considered the experimental treatment.