ORIGINAL ARTICLE

Outcomes of Two Trials of Oxygen-Saturation Targets in Preterm Infants

The BOOST-II Australia and United Kingdom Collaborative Groups

ABSTRACT

BACKGROUND

The safest ranges of oxygen saturation in preterm infants have been the subject of debate.

METHODS

In two trials, conducted in Australia and the United Kingdom, infants born before 28 weeks' gestation were randomly assigned to either a lower (85 to 89%) or a higher (91 to 95%) oxygen-saturation range. During enrollment, the oximeters were revised to correct a calibration-algorithm artifact. The primary outcome was death or disability at a corrected gestational age of 2 years; this outcome was evaluated among infants whose oxygen saturation was measured with any study oximeter in the Australian trial and those whose oxygen saturation was measured with a revised oximeter in the U.K. trial.

RESULTS

After 1135 infants in Australia and 973 infants in the United Kingdom had been enrolled in the trial, an interim analysis showed increased mortality at a corrected gestational age of 36 weeks, and enrollment was stopped. Death or disability in the Australian trial (with all oximeters included) occurred in 247 of 549 infants (45.0%) in the lower-target group versus 217 of 545 infants (39.8%) in the higher-target group (adjusted relative risk, 1.12; 95% confidence interval [CI], 0.98 to 1.27; P=0.10); death or disability in the U.K. trial (with only revised oximeters included) occurred in 185 of 366 infants (50.5%) in the lower-target group versus 164 of 357 infants (45.9%) in the higher-target group (adjusted relative risk, 1.10; 95% CI, 0.97 to 1.24; P=0.15). In post hoc combined, unadjusted analyses that included all oximeters, death or disability occurred in 492 of 1022 infants (48.1%) in the lower-target group versus 437 of 1013 infants (43.1%) in the higher-target group (relative risk, 1.11; 95% CI, 1.01 to 1.23; P=0.02), and death occurred in 222 of 1045 infants (21.2%) in the lower-target group versus 185 of 1045 infants (17.7%) in the higher-target group (relative risk, 1.20; 95% CI, 1.01 to 1.43; P=0.04). In the group in which revised oximeters were used, death or disability occurred in 287 of 580 infants (49.5%) in the lower-target group versus 248 of 563 infants (44.0%) in the higher-target group (relative risk, 1.12; 95% CI, 0.99 to 1.27; P=0.07), and death occurred in 144 of 587 infants (24.5%) versus 99 of 586 infants (16.9%) (relative risk, 1.45; 95% CI, 1.16 to 1.82; P=0.001).

CONCLUSIONS

Use of an oxygen-saturation target range of 85 to 89% versus 91 to 95% resulted in nonsignificantly higher rates of death or disability at 2 years in each trial but in significantly increased risks of this combined outcome and of death alone in post hoc combined analyses. (Funded by the Australian National Health and Medical Research Council and others; BOOST-II Current Controlled Trials number, ISRCTN00842661, and Australian New Zealand Clinical Trials Registry number, ACTRN12605000055606.)

The members of the writing committee (William Tarnow-Mordi, M.B., Ch.B., Ben Stenson, M.D., Adrienne Kirby, M.Sc., Edmund Juszczak, M.Sc., Mark Donoghoe, Ph.D., Sanjeev Deshpande, M.B., B.S., Colin Morley, M.D., Andrew King, B.A., Lex W. Doyle, M.D., Brian W. Fleck, M.D., Peter G. Davis, M.D., Henry L. Halliday, M.D., Wendy Hague, Ph.D., Pamela Cairns, M.D., Brian A. Darlow, M.D., Alistair R. Fielder, M.D., Val Gebski, M.Sc., Neil Marlow, D.M., Karen Simmer, Ph.D., Win Tin, M.B., B.S., Alpana Ghadge, Ph.D., Cathy Williams, M.B., B.S., Anthony Keech, M.D., Stephen P. Wardle, M.D., Zsuzsoka Kecskes, M.D., Martin Kluckow, Ph.D., Glen Gole, M.D., Nicholas Evans, M.D., Girvan Malcolm, Ph.D., Melissa Luig, M.D., Ian Wright, M.B., B.S., Jacqueline Stack, M.B., Ch.B., Kenneth Tan, M.B., B.S., Margo Pritchard, Ph.D., Peter H. Gray, M.D., Scott Morris, M.B., B.S., Ph.D., Bevan Headley, M.B., B.S., Peter Dargaville, M.D., R. John Simes, M.D., and Peter Brocklehurst, M.B., Ch.B.) assume responsibility for the overall content and integrity of this article. The affiliations of the members of the writing group are listed in the Appendix. Address reprint requests to Dr. Tarnow-Mordi at WINNER Centre, NHMRC Clinical Trials Centre, University of Sydney, Locked Bag 77, Camperdown, NSW 1450, Australia, or at williamtm@med.usyd.edu.au; or to Dr. Stenson at Neonatal Unit, Royal Infirmary of Edinburgh, EH16 4SA, United Kingdom, or at ben.stenson@nhslothian.scot.nhs.uk.

Drs. Tarnow-Mordi and Stenson and Drs. Simes and Brocklehurst contributed equally to this article.

This article was published on February 10, 2016, and updated on February 25, 2016, at NEJM.org.

N Engl J Med 2016;374:749-60.
DOI: 10.1056/NEJMoa1514212

The determination of the range of oxygen saturation that minimizes the competing risks of death, retinopathy of prematurity, and later disability in preterm infants is important.[1,2] The U.K. and Australian Benefits of Oxygen Saturation Targeting (BOOST)–II trials are two of five comparative effectiveness trials of the targeting of oxygen saturation in infants born before 28 weeks' gestation.[3-8] These trials, known collectively as the Neonatal Oxygen Prospective Meta-analysis (NeOProM) Collaboration, were designed to compare the effects of a lower oxygen-saturation target range (85 to 89%) versus a higher target range (91 to 95%) on a primary outcome of death or major disability at 18 to 24 months, with age corrected for prematurity.[3]

Observational data had suggested that targeting an oxygen saturation below 90% was associated with a lower risk of severe retinopathy, with no difference in the rate of cerebral palsy or survival,[9] and that the long-accepted "physiologic" targets of oxygen saturation may be too high.[2] The trials therefore aimed to evaluate the hypothesis that targeting an oxygen saturation of 85 to 89% versus 91 to 95% would reduce the incidence of severe retinopathy with no effects on mortality or disability.[1]

After the trials were initiated, the U.K. investigators found that Masimo Radical pulse oximeters, which were widely used and were used in all NeOProM trials, had an artifact in their calibration algorithm.[10] Approximately a third fewer oxygen-saturation values between 87 and 90% were displayed than expected, and values above 87% were shifted up by 1 to 2 percentage points. Masimo confirmed this artifact[10] and provided revised software that removed it. The revised oximeters performed similarly to other common oximeters.[10] After the Surfactant, Positive Pressure, and Oxygenation Randomized Trial (SUPPORT) showed increased mortality among infants in the lower-target group,[4] the trial management committees closed recruitment to U.K. and Australian BOOST-II trials early, because their independent data and safety monitoring committees informed them that a pooled safety analysis of the two trials showed a 65% greater relative risk of death at 36 weeks' postmenstrual age (weeks since the first day of the mother's last normal menstrual period) among infants in the lower-target group than among infants in the higher-target group when revised oximeters were used (relative risk, 1.65; 99.73% confidence interval [CI], 1.09 to 2.49; risk difference, 8.5 percentage points; number needed to harm [i.e., the number of infants needed in the lower-target group for one extra death to occur], 12; $P<0.001$) but no significant between-group difference in mortality when the original oximeters were used (test for interaction between revised and original oximeters, $P=0.006$).[11] A post hoc analysis of combined data from the two trials confirmed the higher in-hospital mortality among infants in the lower-target group in association with the use of the revised oximeters ($P=0.002$).[8] Here, we report the outcomes of the Australian and U.K. BOOST-II trials in children up to a corrected age of 2 years.

METHODS

PATIENTS

The planned sample size in each trial was 1200 infants. The Australian trial, which involved 15 centers, was started on March 25, 2006. The U.K. trial, which involved 34 centers, was started on September 29, 2007. Both trials closed recruitment on December 24, 2010. Infants were eligible for inclusion in the trial if they were born in the preceding 24 hours and before 28 weeks' gestation. Infants who had major congenital abnormalities or were unlikely to survive or to be followed up were ineligible. Infants from multiple births underwent randomization individually. Randomization was performed centrally by computer and was performed separately for each trial. Minimization procedures were used to balance group assignments according to sex, gestational age, and center and, in the Australian trial, according to whether the infant was a singleton or part of a multiple birth and whether the infant was enrolled in the hospital of birth. In each trial, the original primary analysis population was defined as all enrolled infants. In 2010, the U.K. trial steering committee, the members of which were unaware of any of the outcome data, revised the protocol on the recommendation of its data and safety monitoring committee to change the primary analysis population to 1200 infants who would be evaluated with the use of the revised oximeters. The initial and updated protocols and statistical analysis plans are available with the full text of this article at NEJM.org.

ENROLLMENT AND TREATMENT

In both trials, infants were randomly assigned to an oxygen-saturation target range of 85 to 89% (lower-target group) or 91 to 95% (higher-target group). The study oximeters were modified so that the observers were unaware of the true reading; readings in the range of 85 to 95% displayed an oxygen saturation that was up to 3 percentage points higher than the actual oxygen saturation in the lower-target group and 3 percentage points lower than actual oxygen saturation in the higher-target group. Thus, for example, a displayed reading of 90% matched an actual reading of 87% in the lower-target group and 93% in the higher-target group. The clinical staff targeted displayed readings of 88 to 92% to achieve the intended target ranges. At readings above 95% and below 85%, the oximeters reverted to the true reading. In both trials, an upper displayed oxygen-saturation alarm threshold of 94% was recommended. A lower displayed alarm threshold of 86% was recommended in the Australian trial; in the U.K. trial, this decision was left to the individual centers. Only study oximeters were used until 36 weeks' postmenstrual age. If infants were in stable condition while breathing ambient air before then, oximetry was discontinued. If oximetry was resumed before 36 weeks, a study oximeter was used. Masimo supplied the study oximeters under lease.

ASSESSMENTS

Outcomes were evaluated without knowledge of the treatment-group assignments. Initially, cognitive impairment was to be assessed as a Mental Development Index cutoff score of less than 70 on a Bayley Scales of Infant Development, Second Edition (BSID-II), assessment. Before the assessments began, the BOOST-II and NeOProM trial protocols were revised to adopt the Bayley Scales of Infant Development, Third Edition (Bayley-III), and used a cutoff score of less than 85, because this approximately matched a cognitive score of less than 70 on the BSID-II. Scores on the Bayley-III are assessed relative to a standardized mean (±SD) of 100±15, with higher scores indicating better performance.[12]

Serious neurosensory disability in the U.K. trial or major disability in the Australian trial (both hereafter called "disability") were defined as a score of less than 85 on the Bayley-III cognitive or language assessment, as being legally blind or partially sighted in the U.K. trial or legally blind with less than 6/60 vision in the better eye in the Australian trial, or as having severe cerebral palsy (Gross Motor Function Classification System [GMFCS] level ≥2[13] [on a scale of 0 to 5, with higher scores indicating greater impairment] or not walking unaided at 2 years) or deafness (hearing loss requiring or too severe to benefit from aiding or a cochlear implant). Bayley-III assessments could not always be arranged. Therefore, to minimize the risk of bias from postrandomization exclusions, alternative measures of cognition and language were prespecified in revised statistical analysis plans before the data were analyzed (Tables S1 through S3 in the Supplementary Appendix, available at NEJM.org).[7,14] If these outcomes remained unknown and the child was not blind or deaf and did not have cerebral palsy, data on the primary outcome were judged as missing.

Prespecified secondary outcomes in children up to a corrected age of 2 years included death, a Bayley-III cognitive or language score of less than 85, blindness, cerebral palsy, deafness, developmental delay of 12 months or more on a pediatric assessment, and, in the Australian trial, late-onset infection, respiratory illness, death attributed to pulmonary causes by a clinician, days of endotracheal intubation, days of treatment with continuous positive airway pressure, days of treatment with supplemental oxygen in the hospital and at home, and the number of hospital readmissions. The prespecified secondary outcomes of treatment for severe retinopathy, use of oxygen at 36 weeks' postmenstrual age, patent ductus arteriosus, necrotizing enterocolitis resulting in surgery or death, grade III or IV intraventricular hemorrhage, and brain injury were reported previously for both trials.[8] The definitions of the secondary outcomes are provided in the protocols, which include the statistical analysis plans.

STUDY OVERSIGHT

Each trial was designed and conducted separately with separate oversight committees. The joint writing committee members vouch for the accuracy and completeness of the data and analyses and for the fidelity of the reporting of each trial to its protocol. An ethics committee at each center approved the study before it began. A parent of each child provided written informed consent.

STATISTICAL ANALYSIS

Infants who were randomly assigned to the higher oxygen-saturation target range were expected to have a 30 to 40% incidence of the primary outcome.[11,12] Assuming an incidence of 35%, we calculated that each trial would need to enroll 1200 infants for the study to have 80% power to detect an 8 percentage-point absolute difference between the groups, at a two-sided 5% level of significance.

The statistical analyses were performed independently by the respective trial analysts. The primary and secondary analyses in each trial were prespecified, but pooled analyses and retrospective comparisons of the time spent in the assigned oxygen-saturation range were not. Data were analyzed according to the randomly assigned study group, regardless of deviation from protocol. In the Australian trial, the primary analysis population comprised all enrolled infants. In the U.K. trial, the primary analysis population comprised infants whose oxygen-saturation levels were evaluated with the revised oximeters; those for whom the original oximeters were used were included in secondary analyses. Demographic and clinical characteristics were summarized with counts and percentages for categorical variables or with means and standard deviations for normally distributed continuous variables. For statistical comparisons, we calculated the relative risk and 95% confidence interval for the primary outcome. For other dichotomous outcomes, 95% confidence intervals were used in the Australian trial, and 99% confidence intervals were used in the U.K. trial.

In the Australian trial, the effects of the oxygen-saturation target on the primary outcome were assessed with the use of generalized estimating equations adjusted for correlation between infants from multiple births. Binary secondary outcomes were assessed similarly, and count data (e.g., hospital readmissions) were analyzed by Poisson generalized linear models with a log link and are shown as relative rates. Before the analysis, we planned to report key outcomes for all infants and separately for those who were evaluated with the original oximeters and those who were evaluated with revised oximeters (see the protocols and statistical analysis plans).

In both trials, the denominator for events was the number of infants for whom each outcome was known. The relative risk was calculated as the incidence in the lower-target group divided by incidence in the higher-target group. The results were adjusted for clustering due to multiple births and, in the U.K. trial, for minimization factors with the use of generalized linear models with a log link. We performed sensitivity analyses that excluded data from infants for whom alternatives to the Bayley-III assessments had been used. For the statistical analyses, SAS, version 9.3 (SAS Institute), was used in the Australian trial, and Stata/SE, version 13.1 for Windows (StataCorp), was used in the U.K. trial. For the combined analyses, SAS, version 9.3, and RevMan, version 5.3 (Cochrane Collaboration), were used, with trial data unadjusted for multiple births or other descriptive variables. Tests for interaction were used to detect heterogeneity with respect to the primary outcome, with respect to disability, and with respect to death and included all infants with data, stratified according to trial and oximeter.

Figure 1 (facing page). Enrollment and Randomization in the BOOST-II Trials.

In the Australian trial, the other reasons for nonrecruitment of 465 infants were as follows: parents were not approached (248 infants), parents were under too much stress (54), infant was transferred to or from another hospital during the eligibility period (44), parents were not available (41), no interpreter was available (26), consent was obtained after the infant was 24 hours of age (14), insufficient study monitors were available (4), investigators were awaiting ethics approval for the protocol amendment (4), and miscellaneous (30).

RESULTS

PATIENT POPULATION AND PRIMARY OUTCOMES

We enrolled 1135 infants in the Australian trial and 973 infants in the U.K. trial (Fig. 1). The characteristics of the randomly assigned groups were similar at trial entry (Table 1, and Table S4 in the Supplementary Appendix). The primary outcome was determined for 1094 of 1135 infants (96.4%) in the primary analysis population of the Australian trial (all infants) and 723 of 745 infants (97.0%) in the primary analysis population of the U.K. trial (infants for whom revised oximeters only were used). In the Australian trial, the primary outcome occurred in 247 of 549 infants (45.0%) in the lower-target group

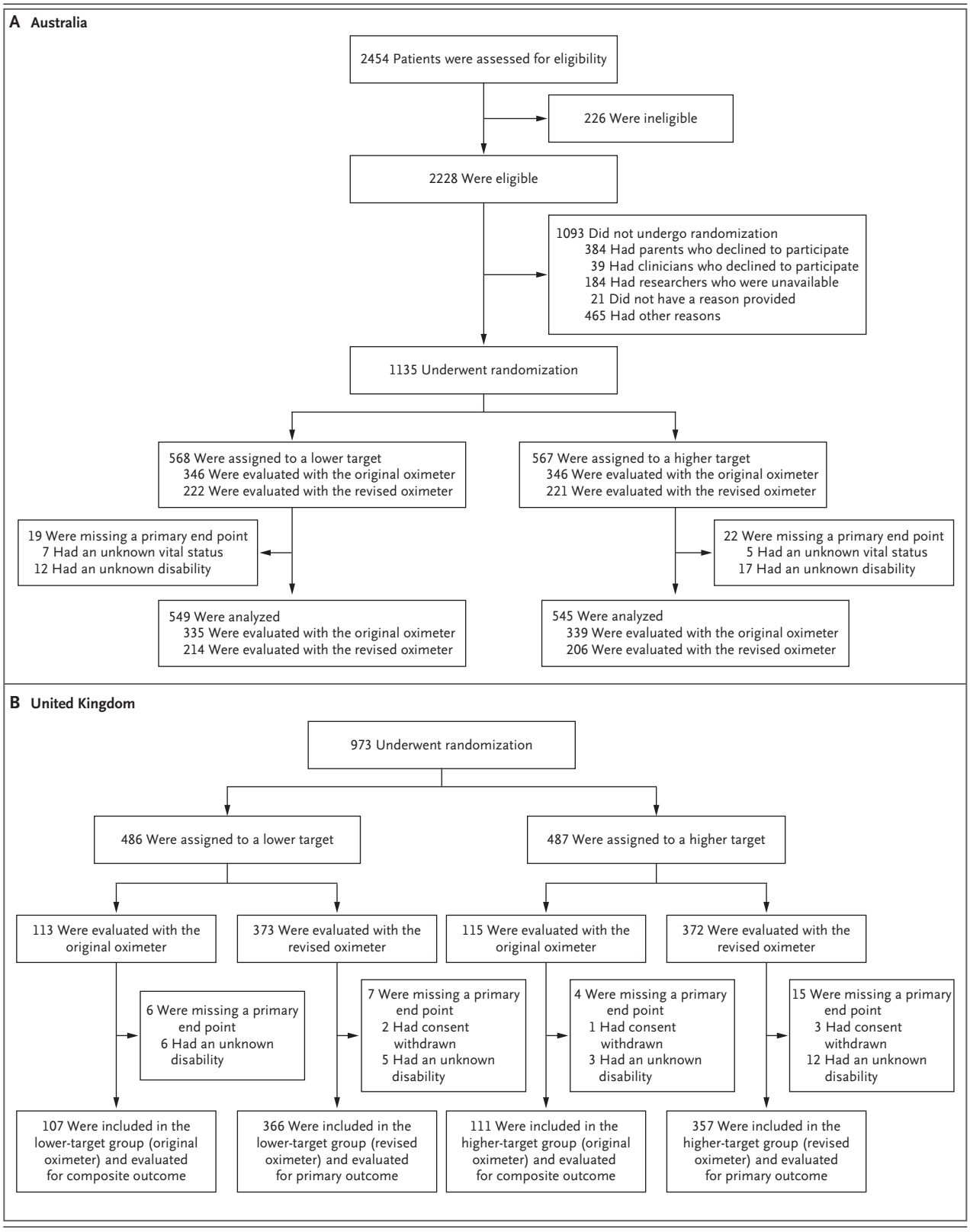


Table 1. Baseline Characteristics of the Primary Analysis Population in Each BOOST-II Trial.*

Characteristic	Australia (All Oximeters)		United Kingdom (Revised Oximeters)	
	Lower-Target Group (N=568)	Higher-Target Group (N=567)	Lower-Target Group (N=366)	Higher-Target Group (N=357)
Male sex — no. (%)	293 (51.6)	296 (52.2)	192 (52.5)	191 (53.5)
Birth weight — g	817±177	833±190	821±182	818±189
Gestational age — wk	26±1.2	26±1.2	26±1.3	26±1.3
Multiple birth — no. (%)	138 (24.3)	135 (23.8)	104 (28.4)	105 (29.4)
Born outside of a study center — no. (%)	44 (7.7)	42 (7.4)	46 (12.6)	40 (11.2)
Antenatal glucocorticoid treatment — no./total no. (%)†				
None	64/566 (11.3)	42/561 (7.5)	23/364 (6.3)	30/356 (8.4)
Incomplete course	143/566 (25.3)	150/561 (26.7)	107/364 (29.4)	104/356 (29.2)
Complete course	303/566 (53.5)	320/561 (57.0)	234/364 (64.3)	222/356 (62.4)
Birth by cesarean section — no./total no. (%)	294/566 (51.9)	306/563 (54.4)	156/366 (42.6)	144/357 (40.3)
Temperature at admission to the neonatal unit — °C	36.0±1.1	36.1±0.9	36.6±0.9	36.7±0.9

* Plus–minus values are means ±SD. The characteristics at baseline were similar in the two treatment groups in each trial, with the exception of a lack of antenatal treatment with glucocorticoids, which was more common among infants in the lower-target group than in the higher-target group in the Australian trial (P=0.03). However, the relative risks of the primary outcome in the lower-target group versus the higher-target group were unchanged after adjustment for use of antenatal glucocorticoids (Tables S12 and S13 in the Supplementary Appendix). The mean (±SD) corrected ages at the follow-up assessment were 26.0±5.1 months in the lower-target group and 25.7±4.5 months in the higher-target group in the Australian trial and 28.7±7.1 and 29.9±7.6 months, respectively, in the U.K. trial.

† An incomplete course of antenatal glucocorticoid treatment was defined as a duration of treatment of less than 24 hours; a complete course was one that was administered for 24 hours or longer. Data on infants who received a course of glucocorticoid treatment more than 7 days before birth are not included in the table.

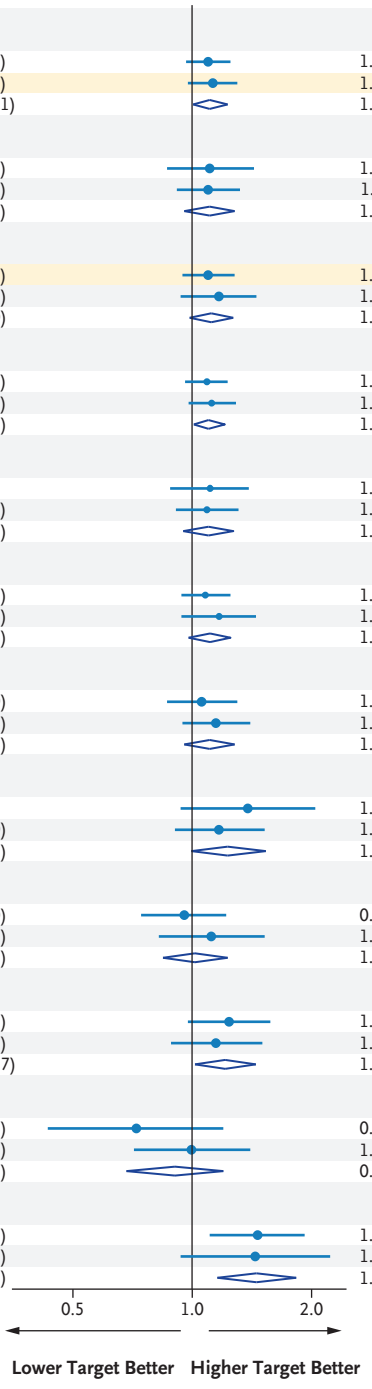
Figure 2 (facing page). Unadjusted Relative Risks of the Primary Outcome in Each Trial and of the Principal Components of the Primary Outcome.

The primary outcome in each trial was death or disability at a corrected gestational age of 2 years; this outcome was evaluated among infants whose oxygen saturation was measured with any study oximeter in the Australian trial and those whose oxygen saturation was measured with a revised oximeter in the U.K. trial. The row in which the data for the primary outcome in each trial appears is highlighted. The relative risks of the individual components of the primary outcome (death and disability) are also shown. The relative risks are unadjusted for multiple births and for variables used for minimization at randomization. $P_{het\ by\ trial}$ is the P value for heterogeneity of outcomes according to trial. There was no heterogeneity between the U.K. and Australian trials with respect to any outcome. $P_{het\ by\ algorithm}$ is the P value for heterogeneity of outcomes according to oximeter calibration algorithm. There was heterogeneity between the revised and original oximeters with respect to death (P=0.01) but no other outcomes. "Bayley alternatives" refers to cognitive or language assessments other than the Bayley Scales of Infant Development, Third Edition, and include the Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III), Griffiths Mental Development Scales (GMDS), Schedule of Growing Skills (SGS), Denver Developmental Screening Test, Parent Report of Children's Abilities–Revised (PARCA-R), assessment by a pediatrician or general practitioner, or adjudication of parent-reported information. The widths of the diamonds for pooled data indicate the 95% confidence intervals for the pooled estimates of effect.

and in 217 of 545 infants (39.8%) in the higher-target group (adjusted relative risk, 1.12; 95% CI, 0.98 to 1.27; P=0.10). In the U.K. trial, the primary outcome occurred in 185 of 366 infants (50.5%) in the lower-target group and in 164 of 357 infants (45.9%) in the higher-target group (adjusted relative risk, 1.10; 95% CI, 0.97 to 1.24; P=0.15).

Alternative, surrogate measures were used in place of Bayley-III assessments to define the primary outcome in 85 of 1135 infants (7.5%) in the Australian trial and in 129 of 745 infants (17.3%) in the U.K. trial (Tables S1 through S3 in the Supplementary Appendix). Sensitivity analyses were performed according to whether disability was determined with the use of alterna-

Subgroup	Lower Oxygen-Saturation Target	Higher Oxygen-Saturation Target	Relative Risk (95% CI)	P Value	$P_{het\ by\ trial}$	$P_{het\ by\ algorithm}$
	no. of events/total no. (%)					
Death or disability (all oximeters)						
United Kingdom	245/473 (51.8)	220/468 (47.0)	1.10 (0.97–1.25)	0.14		
Australia	247/549 (45.0)	217/545 (39.8)	1.13 (0.98–1.30)	0.08		
Pooled	492/1022 (48.1)	437/1013 (43.1)	1.11 (1.01–1.23)	0.02	0.79	
Death or disability (original oximeters)						
United Kingdom	60/107 (56.1)	56/111 (50.5)	1.11 (0.87–1.43)	0.41		
Australia	145/335 (43.3)	133/339 (39.2)	1.10 (0.92–1.32)	0.29		
Pooled	205/442 (46.4)	189/450 (42.0)	1.11 (0.96–1.28)	0.18	0.96	
Death or disability (revised oximeters)						
United Kingdom	185/366 (50.5)	164/357 (45.9)	1.10 (0.95–1.28)	0.22		
Australia	102/214 (47.7)	84/206 (40.8)	1.17 (0.94–1.45)	0.16		
Pooled	287/580 (49.5)	248/563 (44.0)	1.12 (0.99–1.27)	0.07	0.65	0.86
Death or disability (excluding Bayley alternatives, all oximeters)						
United Kingdom	231/384 (60.2)	211/381 (55.4)	1.09 (0.96–1.23)	0.18		
Australia	237/503 (47.1)	212/506 (41.9)	1.12 (0.98–1.29)	0.10		
Pooled	468/887 (52.8)	423/887 (47.7)	1.10 (1.01–1.21)	0.04	0.71	
Death or disability (excluding Bayley alternatives, original oximeters)						
United Kingdom	54/82 (65.9)	53/89 (59.6)	1.11 (0.88–1.39)	0.39		
Australia	138/304 (45.4)	132/318 (41.5)	1.09 (0.91–1.31)	0.33		
Pooled	192/386 (49.7)	185/407 (45.5)	1.10 (0.95–1.27)	0.20	0.94	
Death or disability (excluding Bayley alternatives, revised oximeters)						
United Kingdom	177/302 (58.6)	158/292 (54.1)	1.08 (0.94–1.25)	0.27		
Australia	99/199 (49.7)	80/188 (42.6)	1.17 (0.94–1.45)	0.16		
Pooled	276/501 (55.1)	238/480 (49.6)	1.11 (0.98–1.25)	0.09	0.56	0.87
Disability (all oximeters)						
United Kingdom	123/351 (35.0)	122/370 (33.0)	1.06 (0.87–1.30)	0.56		
Australia	147/449 (32.7)	130/458 (28.4)	1.15 (0.95–1.40)	0.15		
Pooled	270/800 (33.8)	252/828 (30.4)	1.11 (0.96–1.28)	0.15	0.57	
Disability (original oximeters)						
United Kingdom	39/86 (45.3)	27/82 (32.9)	1.38 (0.94–2.03)	0.10		
Australia	88/278 (31.7)	76/282 (27.0)	1.17 (0.91–1.52)	0.22		
Pooled	127/364 (34.9)	103/364 (28.3)	1.23 (1.00–1.53)	0.06	0.50	
Disability (revised oximeters)						
United Kingdom	84/265 (31.7)	95/288 (33.0)	0.96 (0.75–1.22)	0.75		
Australia	59/171 (34.5)	54/176 (30.7)	1.12 (0.83–1.52)	0.45		
Pooled	143/436 (32.8)	149/464 (32.1)	1.02 (0.85–1.23)	0.83	0.43	0.20
Death (all oximeters)						
United Kingdom	122/484 (25.2)	98/483 (20.3)	1.24 (0.98–1.57)	0.07		
Australia	100/561 (17.8)	87/562 (15.5)	1.15 (0.89–1.50)	0.29		
Pooled	222/1045 (21.2)	185/1045 (17.7)	1.20 (1.01–1.43)	0.04	0.67	
Death (original oximeters)						
United Kingdom	21/113 (18.6)	29/114 (25.4)	0.73 (0.44–1.20)	0.21		
Australia	57/345 (16.5)	57/345 (16.5)	1.00 (0.72–1.40)	1.00		
Pooled	78/458 (17.0)	86/459 (18.7)	0.91 (0.69–1.20)	0.49	0.31	
Death (revised oximeters)						
United Kingdom	101/371 (27.2)	69/369 (18.7)	1.46 (1.11–1.91)	0.006		
Australia	43/216 (19.9)	30/217 (13.8)	1.44 (0.94–2.21)	0.09		
Pooled	144/587 (24.5)	99/586 (16.9)	1.45 (1.16–1.82)	0.001	0.97	0.01



tives to the Bayley-III assessment; the results of these analyses were similar to those in the primary analyses, with no between-group differences in the rate of death or disability in either trial or in the rates of disability or its components, including blindness (Fig. 2 and Table 2, and Tables S5 and S6 in the Supplementary Appendix).

Table 2. Rates and Adjusted Relative Risks of the Primary Outcome and Its Components to 2 Years.

Outcome	Lower-Target Group	Higher-Target Group	Adjusted Relative Risk (95% or 99% CI)*
	no./total no. (%)		
Australia: all oximeters			
Death or disability†	247/549 (45.0)	217/545 (39.8)	1.12 (0.98–1.27)
Disability	147/449 (32.7)	130/458 (28.4)	1.15 (0.96–1.39)
Death before assessment at corrected age of 2 years	100/561 (17.8)	87/562 (15.5)	1.15 (0.89–1.50)
Cerebral palsy with GMFCS ≥2‡	16/446 (3.6)	25/456 (5.5)	0.68 (0.36–1.28)
Bayley-III language or cognitive score <85§	124/397 (31.2)	115/416 (27.6)	1.13 (0.91–1.40)
Deafness requiring — or too severe to benefit from — a hearing aid	11/452 (2.4)	9/459 (2.0)	1.24 (0.52–2.97)
Severe visual loss, certifiable as legal blindness or partial sight	3/452 (0.7)	2/459 (0.4)	1.52 (0.26–9.04)
United Kingdom: revised oximeters			
Death or disability	185/366 (50.5)	164/357 (45.9)	1.10 (0.97–1.24)
Disability	84/265 (31.7)	95/288 (33.0)	0.97 (0.72–1.32)
Death before assessment at corrected age of 2 years	101/371 (27.2)	69/369 (18.7)	1.38 (0.99–1.93)
Cerebral palsy with GMFCS ≥2‡	25/265 (9.4)	17/287 (5.9)	1.60 (0.73–3.51)
Bayley-III language or cognitive score <85 or equivalent§	69/261 (26.4)	78/286 (27.3)	0.98 (0.69–1.39)
Deafness requiring — or too severe to benefit from — a hearing aid	15/264 (5.7)	25/287 (8.7)	0.66 (0.29–1.50)
Severe visual loss, certifiable as legal blindness or partial sight¶	8/262 (3.1)	10/289 (3.5)	0.90 (0.26–3.16)

* The confidence intervals (CIs) in the Australian trial are 95% confidence intervals; those in the U.K. trial are 99% confidence intervals, with the exception of the primary outcome (death or disability), for which a 95% confidence interval is used.
† The between-sibling correlation for the primary outcome was 0.41.
‡ Gross motor function was assessed with the use of the modified Gross Motor Function Classification System (GMFCS), with levels ranging from 1 to 5 and higher levels indicating greater impairment.
§ Scores on the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III), are assessed relative to a standardized mean (±SD) of 100±15, with higher scores indicating better performance. A score of less than 85 on the Bayley-III (or equivalent tool) is deemed to be equivalent to more than 2 standard deviations below the mean of the Bayley Scales of Infant and Toddler Development, Second Edition (BSID-II, the original scale in use when the trial was designed). Equivalent tools included the Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III), Griffiths Mental Development Scales (GMDS), Schedule of Growing Skills (SGS), Denver Developmental Screening Test, Parent Report of Children's Abilities–Revised (PARCA-R), assessment by a pediatrician or general practitioner, or adjudication of parent-reported information.
¶ The causes of severe visual loss were retinal damage (2 patients in each study group), cortical damage (4 patients in the lower-target group and 8 patients in the higher-target group), and unknown (2 patients in the lower-target group).

In combined unadjusted analyses of data from both trials (Fig. 2), death or disability occurred more frequently in the lower-target group than in the higher-target group among all 2035 infants for whom outcomes were known (492 of 1022 [48.1%] vs. 437 of 1013 [43.1%]; relative risk, 1.11; 95% CI, 1.01 to 1.23; P=0.02). The difference in the rate of death or disability was similar when it was based on only the 1774 infants whose outcome was assigned without the use of alternative measures to the Bayley-III (468 of 887 [52.8%] vs. 423 of 887 [47.7%]; relative risk, 1.10; 95% CI, 1.01 to 1.21; P=0.04). The effect of the oxygen-saturation target range on the combined outcome of death or disability did not differ materially according to oximeter calibration algorithm (relative risk, 1.11 with the original algorithm and 1.12 with the revised

algorithm; P=0.86). However, there was evidence of heterogeneity between the original and revised oximeters with regard to mortality (P=0.01). Mortality before 2 years was higher in the lower-target group than in the higher-target group, both among the 2090 infants for whom either oximeter was used (222 of 1045 [21.2%] vs. 185 of 1045 [17.7%]; relative risk, 1.20; 95% CI, 1.01 to 1.43; P=0.04) and among the 1173 infants for whom revised oximeters were used (144 of 587 [24.5%] vs. 99 of 586 [16.9%]; relative risk, 1.45; 95% CI, 1.16 to 1.82; P=0.001). There were no significant differences between the groups in either trial in the rates of disability (Table 2). The major causes of death were sepsis, necrotizing enterocolitis, intraventricular hemorrhage, and chronic lung disease (Tables S7 and S8 in the Supplementary Appendix). There were five infant deaths in the Australian trial and nine in the U.K. trial that occurred after hospital discharge but before 2 years.

SECONDARY OUTCOMES

In the U.K. trial, the baseline characteristics of the infants in the secondary analysis, which included infants for whom the original oximeters were used, were similar in the two groups (Table S9 in the Supplementary Appendix). There were no significant differences between groups in the prespecified secondary outcomes (Table S10 in the Supplementary Appendix) or in subgroup analyses (Fig. S1 and S2 in the Supplementary Appendix). There were seven unexpected adverse events, as reported previously (Table S11 in the Supplementary Appendix).[8] In the Australian trial, there were no significant between-group differences in secondary outcomes (Table S14 in the Supplementary Appendix) and no unexpected adverse events among children up to a corrected age of 2 years. As reported previously,[8] during oxygen treatment, infants in the lower-target group spent approximately 30% more time in their assigned range after the oximeter revision than before it, both in the Australian trial (31.1% vs. 23.2%, P<0.001) and in the U.K. trial (26.3% vs. 20.0%, P<0.001).

DISCUSSION

In the primary analysis populations in the Australian and U.K. trials, there were no significant differences between the study groups in the rate of the primary outcome of death or disability in each trial individually, but the rate was significantly higher in the lower-target group than in the higher-target group in the post hoc combined, unadjusted analyses of the two trials. In combined analyses that included infants for whom any study oximeter was used, the relative risk of death or disability was 12% higher in the lower-target group than in the higher-target group; among the infants for whom revised oximeters (which are similar to currently used oximeters[10]) were used, the relative risk of death was 45% higher in the lower-target group than in the higher-target group (Fig. 2).

There was no significant difference between the groups in the rates of disability at 2 years in either trial or in the combined analyses. Although disability was measured by a combination of Bayley-III and alternative measures, sensitivity analyses excluding the alternative measures resulted in no material change in the conclusions.

In the combined analysis, more infants in the higher-target group than in the lower-target group were treated for retinopathy of prematurity,[8] but there was no significant between-group difference in the rate of blindness. Furthermore, none of the NeOProM oxygen targeting trials have shown differences in the rates of disability or blindness.[5-7]

The higher mortality among infants randomly assigned to a lower oxygen-saturation target was observed with all oximeters, and the difference was even more pronounced with the revised oximeters. Separate consideration of the mortality results obtained in association with the revised oximeters is justified for several reasons: first, the oximeter revision corrected an artifact in the calibration algorithm; second, the U.K. trial prespecified infants for whom revised oximeters were used as the primary analysis population; third, in combined unadjusted analyses, there was heterogeneity with respect to mortality between infants stratified according to the type of oximeter (P=0.01); fourth, the strength of the evidence for excess mortality in the lower-target group for whom revised oximeters were used (P=0.001) makes a false positive result unlikely; and finally, we previously reported significantly higher in-hospital mortality (representing 96% of all deaths up to 2 years of age) in the lower-target group than in the higher-target group among infants for whom revised oximeters were

used (P=0.002).[8,11] These significant differences in mortality were observed despite the lower-than-planned separation in oxygen-saturation values found between the study groups.[15,16]

After SUPPORT showed excess mortality among infants randomly assigned to lower oxygen-saturation targets,[4] the BOOST trial investigators consulted their data and safety monitoring committees because of concerns about possible harm to infants in the lower-target group.[11] The results of a pooled interim safety analysis, conducted with the use of prespecified guidelines,[11] met the criteria of both protocols for stopping the study (Fig. S3 in the Supplementary Appendix). This protocol-specified process was therefore appropriate.[17] Through chance, treatment effects may be overestimated in trials that, like the BOOST II studies, are stopped early after an interim analysis[8,18]; however, they can be underestimated in trials that, like the Canadian Oxygen Trial (COT) and SUPPORT, reach completion despite interim analyses.[4,6,19] COT[6] showed that the rate of the primary outcome of death or disability was not significantly higher in the lower-target group than in the higher-target group (relative risk, 1.08; 95% CI, 0.85 to 1.37; P=0.52). However, the point estimate of the treatment effect in COT was consistent with that observed in the combined BOOST II trials and, on the basis of the confidence intervals, its results are consistent with as much as a 15% lower or 37% higher rate of death or disability. The most accurate estimates of the magnitude of treatment effects are likely to come from syntheses of all the NeOProM trials.[3,20]

Pulse oximeters estimate arterial oxygen saturation within limits of accuracy, such that 1 standard deviation equals 3%.[10,16,21] Thus, an oxygen-saturation value of 88% could reflect an arterial oxygen saturation between 85 and 91% in 68% of infants, but it may fall outside a range of 82 to 94% in up to 5% of infants. Variations in the location of the probe and the proportion of fetal hemoglobin may influence these limits[16] but should create no bias, because randomization tends to balance such factors evenly between groups. However, because pulse oximeters underestimate hypoxemia with progressively wider limits of accuracy as true oxygen-saturation values decrease from 93% to 80%,[21] substantially more infants in the lower-target group than in the higher-target group may have been exposed to values of partial pressure of arterial oxygen below the previously recommended levels,[22] with associated adverse effects.

During oxygen treatment, fewer than a third of oxygen-saturation values among infants in the lower-target group were in the assigned range, as compared with approximately half of the values among infants in the higher-target group.[8] The higher range includes the plateau of the oxyhemoglobin dissociation curve, where oxygen saturation fluctuates less with changing partial pressure of arterial oxygen and the slope of oxygen saturation versus the fraction of inspired oxygen (F_{IO_2}) is flatter,[23] which makes targeting easier. This may help explain why, during oxygen treatment, none of the NeOProM trials achieved a median value of actual oxygen saturation below 89% among infants in the lower-target groups.[4,6-8,15] The current trials and other NeOProM trials may therefore have underestimated the effect that accurately targeting an oxygen saturation of 85 to 89% has on adverse outcomes.

After the algorithm revision, infants in the lower-target groups spent approximately 30% longer in their intended oxygen-saturation range, which resulted in greater exposure to hypoxemia.[8] This observation may explain the increased difference in mortality between the groups after the revision.[21,22] However, the magnitude of the difference in mortality may also be due to chance or to changes in confounding variables over time. Although we also report results associated with the original oximeters, the revised oximeters are more relevant to current practice.[10] The consistent trend toward higher mortality among infants in the lower-target groups across trials that used revised oximeters in three continents, with no statistical heterogeneity among the trials,[8,24,25] supports the generalizability of the current data.

The total mortality in the NeOProM trials varied from 15% in the New Zealand BOOST-II trial and COT, to 17% in the Australian BOOST-II trial, to 20% in SUPPORT, to 23% in the U.K. BOOST-II trial. These differences may be explained by variations in study populations owing to differences in admission or eligibility criteria, decisions about viability, or genetic risk.[26,27] For example, the U.K. BOOST-II trial recruited more infants who were transferred into the study centers after being born elsewhere, and SUPPORT,

which enrolled infants before birth, may have included more infants who were moribund or had severe lung disease.

Targeting an intermediate oxygen-saturation range,[16,28] such as 87 to 93%, versus a higher range is an untested practice that may increase mortality, because current oximeters permit increasingly disproportionate exposure to hypoxemia as oxygen saturation decreases to below 93%.[21] At present, the most rigorously evaluated evidence[29] for policy is that targeting an oxygen saturation of 91 to 95% is safer than targeting an oxygen saturation of 85 to 89%.

In conclusion, targeting an oxygen saturation of 85 to 89%, as compared with 91 to 95%, resulted in nonsignificantly higher rates of death or disability at 2 years in the individual trials. However, the use of the lower target significantly increased the risks of this combined outcome and of death alone in post hoc combined analyses.

Supported by the Australian National Health and Medical Research Council (NHMRC) (project grant 352386 and program grant GNT1037786) and the NHMRC Clinical Trials Centre, University of Sydney. BOOST-II UK was funded by the U.K. Medical Research Council (MRC) and managed by the National Institute for Health Research (NIHR) on behalf of the MRC–NIHR partnership.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank the many parents, nurses, neonatologists, pediatricians, ophthalmologists, and other clinical staff who participated in these studies, the members of the IMPACT Network for Improving Outcomes in Mothers and Babies for promoting recruitment in the participating Australian and New Zealand sites, and the members of the two data and safety monitoring committees for their joint collaboration.

This article is dedicated to the memory of Edmund Hey, the driving force behind the U.K. BOOST-II trial, and to the memory of David Henderson-Smart, a champion for collaborative perinatal research in Australia and New Zealand.

APPENDIX

The affiliations of the members of the writing group are as follows: the WINNER Centre, National Health and Medical Research Council (NHMRC) Clinical Trials Centre (W.T.-M.), and NHMRC Clinical Trials Centre (A. Kirby, M.D., W.H., V.G., A.G., A. Keech, R.J.S.), University of Sydney, Department of Neonatology, Royal Prince Alfred Hospital (N.E., G.M.), Department of Neonatology, Westmead Hospital (M.L.), Rainbow Clinic, Neonatal Intensive Care Unit, Liverpool Hospital (J.S.), Sydney, Department of Neonatal Research (C.M.) and Department of Obstetrics and Gynaecology (L.W.D., P.G.D.), Royal Women's Hospital, and Department of Paediatrics, Monash University (K.T.), Melbourne, VIC, Centre for Neonatal Research and Education, King Edward Memorial Hospital and University of Western Australia, Perth (K.S.), Neonatal Intensive Care Unit, Canberra Hospital, Canberra, ACT (Z.K.), Department of Neonatology, Royal North Shore Hospital, St. Leonards, NSW (M.K.), Department of Ophthalmology, University of Queensland (G.G.), Department of Women's and Newborn Services, Royal Brisbane and Women's Hospital, Australian Catholic University (M.P.), and Department of Neonatology, Mater Mothers Hospital (P.H.G.), Brisbane, QLD, Paediatrics and Child Health Research, Illawarra Health and Medical Research Institute, University of Wollongong, Wollongong, NSW (I.W.), and the School of Medicine, Flinders University (S.M.), Department of Neonatology, Adelaide Women's and Children's Hospital, Adelaide, SA (B.H.), and Department of Neonatology, Hobart Hospital, Hobart, TAS (P.D.) — all in Australia; the Neonatal Unit, Department of Reproductive Medicine, Royal Infirmary of Edinburgh (B.S.), and Department of Ophthalmology, Princess Alexandra Eye Pavilion (B.W.F.), Edinburgh, National Perinatal Epidemiology Unit Clinical Trials Unit, University of Oxford, Oxford, (E.J., A. King), Department of Paediatrics, Royal Shrewsbury Hospital, Shrewsbury (S.D.), Royal Maternity Hospital, Belfast (H.L.H.), Department of Neonatology, St. Michael's Hospital (P.C.), and Centre for Child and Adolescent Health School of Social and Community Medicine, University of Bristol (C.W.), Bristol, Division of Optometry and Visual Science, City University (A.R.F.), and Institute for Women's Health, University College London (N.M., P.B.), London, Department of Neonatal Medicine, James Cook University, Middlesbrough (W.T.), and Nottingham University Hospitals NHS Trust, Nottingham (S.P.W.) — all in the United Kingdom; and Department of Paediatrics, Christchurch School of Medicine, University of Otago, Christchurch, New Zealand (B.A.D.).

REFERENCES

1. Cole CH, Wright KW, Tarnow-Mordi W, Phelps DL. Resolving our uncertainty about oxygen therapy. Pediatrics 2003; 112:1415-9.
2. Silverman WA. A cautionary tale about supplemental oxygen: the albatross of neonatal medicine. Pediatrics 2004; 113:394-6.
3. Askie LM, Brocklehurst P, Darlow BA, Finer N, Schmidt B, Tarnow-Mordi W. NeOProM: Neonatal Oxygenation Prospective Meta-analysis Collaboration study protocol. BMC Pediatr 2011;11:6.
4. Carlo WA, Finer NN, Walsh MC, et al. Target ranges of oxygen saturation in extremely preterm infants. N Engl J Med 2010;362:1959-69.
5. Vaucher YE, Peralta-Carcelen M, Finer NN, et al. Neurodevelopmental outcomes in the early CPAP and pulse oximetry trial. N Engl J Med 2012;367:2495-504.
6. Schmidt B, Whyte RK, Asztalos EV, et al. Effects of targeting higher vs lower arterial oxygen saturations on death or disability in extremely preterm infants: a randomized clinical trial. JAMA 2013; 309:2111-20.
7. Darlow BA, Marschner SL, Donoghoe M, et al. Randomized controlled trial of oxygen saturation targets in very preterm infants: two year outcomes. J Pediatr 2014;165(1):30-5.e2.
8. Stenson BJ, Tarnow-Mordi WO, Darlow BA, et al. Oxygen saturation and outcomes in preterm infants. N Engl J Med 2013;368:2094-104.
9. Tin W, Milligan DW, Pennefather P, Hey E. Pulse oximetry, severe retinopathy, and outcome at one year in babies of less than 28 weeks gestation. Arch Dis Child Fetal Neonatal Ed 2001;84:F106-10.
10. Johnston ED, Boyle B, Juszczak E, King A, Brocklehurst P, Stenson BJ. Oxygen targeting in preterm infants using the Masimo SET Radical pulse oximeter. Arch Dis Child Fetal Neonatal Ed 2011;96: F429-33.
11. Stenson B, Brocklehurst P, Tarnow-Mordi W. Increased 36-week survival with high oxygen saturation target in extremely preterm infants. N Engl J Med 2011; 364:1680-2.
12. Johnson S, Moore T, Marlow N. Using the Bayley-III to assess neurodevelop-

mental delay: which cut-off should be used? Pediatr Res 2014;75:670-4.
13. Palisano RJ, Hanna SE, Rosenbaum PL, et al. Validation of a model of gross motor function for children with cerebral palsy. Phys Ther 2000;80:974-85.
14. Johnson S, Marlow N, Wolke D, et al. Validation of a parent report measure of cognitive development in very preterm infants. Dev Med Child Neurol 2004;46:389-97.
15. Schmidt B, Roberts RS, Whyte RK, et al. Impact of study oximeter masking algorithm on titration of oxygen therapy in the Canadian Oxygen Trial. J Pediatr 2014; 165(4):666-71.e2.
16. Lakshminrusimha S, Manja V, Mathew B, Suresh GK. Oxygen targeting in preterm infants: a physiological interpretation. J Perinatol 2015;35:8-15.
17. Schmidt B, Whyte RK, Roberts RS. Trade-off between lower or higher oxygen saturations for extremely preterm infants: the first Benefits of Oxygen Saturation Targeting (BOOST) II trial reports its primary outcome. J Pediatr 2014; 165:6-8.
18. Bassler D, Briel M, Montori VM, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. JAMA 2010;303:1180-7.
19. Schou IM, Marschner IC. Meta-analysis of clinical trials with early stopping: an investigation of potential bias. Stat Med 2013;32:4859-74.
20. Askie LM, Darlow BA, Davis PG, et al. Effects of targeting higher versus lower arterial oxygen saturations on death or disability in preterm infants (protocol). Cochrane Database Syst Rev 2014;(7): CD011190.
21. Rosychuk RJ, Hudson-Mason A, Eklund D, Lacaze-Masmonteil T. Discrepancies between arterial oxygen saturation and functional oxygen saturation measured with pulse oximetry in very preterm infants. Neonatology 2012;101:14-9.
22. Quine D, Stenson BJ. Arterial oxygen tension (Pao2) values in infants <29 weeks of gestation at currently targeted saturations. Arch Dis Child Fetal Neonatal Ed 2009;94:F51-3.
23. Jones JG, Lockwood GG, Fung N, et al. Influence of pulmonary factors on pulse oximeter saturation in pre-term infants. Arch Dis Child Fetal Neonatal Ed 2015 November 24 (Epub ahead of print).
24. 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;105:55-63.
25. Manja V, Lakshminrusimha S, Cook DJ. Oxygen saturation target range for extremely preterm infants: a systematic review and meta-analysis. JAMA Pediatr 2015;169:332-40.
26. Field D, Bajuk B, Manktelow BN, et al. Geographically based investigation of the influence of very-preterm births on routine mortality statistics from the UK and Australia. Arch Dis Child Fetal Neonatal Ed 2008;93:F212-6.
27. Draper ES, Zeitlin J, Fenton AC, et al. Investigating the variations in survival rates for very preterm infants in 10 European regions: the MOSAIC birth cohort. Arch Dis Child Fetal Neonatal Ed 2009; 94:F158-63.
28. Sola A, Golombek SG, Montes Bueno MT, et al. Safe oxygen saturation targeting and monitoring in preterm infants: can we avoid hypoxia and hyperoxia? Acta Paediatr 2014;103:1009-18.
29. Tarnow-Mordi W, Cruz M, Morris J. Design and conduct of a large obstetric or neonatal randomized controlled trial. Semin Fetal Neonatal Med 2015;20:389-402.