GRADING the Oxygen Trials in Preterms

Satyan Lakshminrusimha, MD
Professor of Pediatrics,
Chief, Division of Neonatology,
Director, Center for Developmental Biology of the Lung,
University at Buffalo
Target Ranges of Oxygen Saturation in Extremely Preterm Infants

Oxygen Saturation and Outcomes in Preterm Infants

The BOOST II United Kingdom, Australia, and New Zealand Collaborative Group

Effects of Targeting Higher vs Lower Arterial Oxygen Saturations on Death or Disability in Extremely Preterm Infants
A Randomized Clinical Trial

• SUPPORT, BOOST-II and COT rigorously designed randomized controlled trials
• Optimal oxygen target range for extremely preterm neonates
  • 85-89% vs. 91-95%
• Collaborative design to facilitate eventual meta-analysis
• Approximately 5000 infants enrolled
• Why are we still debating the optimal “dose” of oxygen and “target” for oxygen saturation at HOT TOPICS?
Masking Pulse Oximeters

- SUPPORT alarm setting 96%
- BOOST-II NZ alarm setting 93%
- BOOST-II UK alarm setting 95%
- COT alarm setting at 94%
- SUPPORT alarm setting 84%

Diagram showing displayed SpO2 (%) versus true SpO2 (%) for different oximetry systems.
Calibration with Co-oximetry - Original Algorithm

![Graph showing calibration with Co-oximetry for the Original Algorithm. The graph plots saturation against light ratio. It indicates that as the light ratio increases, saturation decreases significantly.]
Calibration with Co-oximetry - Original Algorithm
Nellcor vs. Masimo (original algorithm)

Proportion of time spent at each saturation value (while receiving supplemental oxygen)

- Masimo (old algorithm)
- Nellcor

Saturation %
Revised Algorithm

![Graph showing the relationship between saturation and light ratio.]
Effect of Revising the Algorithm of Oxygen Saturation Distribution in the BOOST-II Trial

- Better separation of lower target group
- No change in the “tail” distribution < 85%
Effect of Revising the Algorithm of Oxygen Saturation Distribution in the BOOST

**United Kingdom**

**Original Algorithm**

- Hazard ratio, 0.87 (95% CI, 0.66–1.14)  
  - P=0.31 by log-rank test

<table>
<thead>
<tr>
<th>No. at Risk</th>
<th>Lower target</th>
<th>Higher target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower target</td>
<td>629</td>
<td>578</td>
</tr>
<tr>
<td>Higher target</td>
<td>630</td>
<td>567</td>
</tr>
</tbody>
</table>

**Revised Algorithm**

- Hazard ratio, 1.65 (95% CI, 1.26–2.16)  
  - P<0.001 by log-rank test

<table>
<thead>
<tr>
<th>No. at Risk</th>
<th>Lower target</th>
<th>Higher target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower target</td>
<td>593</td>
<td>528</td>
</tr>
<tr>
<td>Higher target</td>
<td>590</td>
<td>543</td>
</tr>
</tbody>
</table>
BOOST-II Effect of Change in Algorithm

Original algorithm

Revised algorithm

Difference in number of deaths

Favors higher SpO₂

Favors lower SpO₂

Total number of deaths
<table>
<thead>
<tr>
<th>Study</th>
<th>Pulse ox Algorithm</th>
<th>Death at discharge or follow-up</th>
<th>Severe ROP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUPPORT</td>
<td>Original</td>
<td>↑</td>
<td>↓↓↓↓</td>
</tr>
<tr>
<td>BOOST-II</td>
<td>Original</td>
<td>No diff</td>
<td>↓ tendency (p-0.19)</td>
</tr>
<tr>
<td></td>
<td>Revised</td>
<td>↑↑ (stopped early)</td>
<td>↓ tendency (p-0.11)</td>
</tr>
<tr>
<td>BOOST-II</td>
<td>Original + Revised</td>
<td>↑ tendency (p-0.09)</td>
<td>↓ (p-0.045)</td>
</tr>
<tr>
<td>(pooled)</td>
<td>Revised</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COT</td>
<td>Original</td>
<td>No diff (p-0.99)</td>
<td>No diff (0.48)</td>
</tr>
<tr>
<td>COT</td>
<td>Revised</td>
<td>No diff (p-0.41)</td>
<td></td>
</tr>
</tbody>
</table>
2010: In babies receiving oxygen, saturation should be maintained between 85 to 93% (D) →
2013 update: In preterm babies receiving oxygen, saturation target should be between 90 and 95% (B) - Sweet et al Neonatology 2013
Severe ROP

Saugstad et al, Neonatology 2014

<table>
<thead>
<tr>
<th>Study</th>
<th>Relative risk (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support</td>
<td>0.52 (0.37, 0.73)</td>
</tr>
<tr>
<td>BOOST (UK)</td>
<td>0.79 (0.60, 1.06)</td>
</tr>
<tr>
<td>BOOST (AU)</td>
<td>0.76 (0.49, 1.18)</td>
</tr>
<tr>
<td>BOOST (NZ)</td>
<td>0.86 (0.39, 1.89)</td>
</tr>
<tr>
<td>COT</td>
<td>0.95 (0.65, 1.39)</td>
</tr>
<tr>
<td>Overall</td>
<td>0.74 (0.59, 0.92)</td>
</tr>
</tbody>
</table>

Relative risk with the x-axis ranging from 0.25 to 3.
Necrotizing Enterocolitis

<table>
<thead>
<tr>
<th>Study</th>
<th>Relative risk (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support</td>
<td>1.11 (0.82, 1.54)</td>
</tr>
<tr>
<td>BOOST (UK)</td>
<td>1.35 (0.97, 1.89)</td>
</tr>
<tr>
<td>BOOST (AU)</td>
<td>1.24 (0.80, 1.94)</td>
</tr>
<tr>
<td>BOOST (NZ)</td>
<td>1.25 (0.60, 2.59)</td>
</tr>
<tr>
<td>COT</td>
<td>1.38 (0.94, 2.02)</td>
</tr>
<tr>
<td>Overall</td>
<td>1.25 (1.05, 1.49)</td>
</tr>
</tbody>
</table>

Saugstad et al Neonatology 2014
How does a busy Neonatologist read a Journal Article?

**Optimal Oxygenation of Extremely Low Birth Weight Infants: A Meta-Analysis and Systematic Review of the Oxygen Saturation Target Studies**

Ola Didrik Saugstad² Dagfinn Aune³,⁴

*Department of Pediatric Research, Oslo University Hospital, University of Oslo, Oslo, and Department of Public Health and General Practice, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway; Department of Epidemiology and Biostatistics, School of Public Health, Imperial College, London, UK

**Key Words**
- Extremely low birth weight infants
- Mortality
- NEOPROM
- Oxygen saturation
- Retinopathy of prematurity
- Bronchopulmonary dysplasia
- Necrotizing enterocolitis

**Abstract**

**Background:** The optimal oxygen saturation for extremely low birth weight infants in the postnatal period before the delivery process is unknown. Objectives: To summarize and discuss the results of the randomized trials, constituting the NEOPROM (Neonatal Oxygenation Prospective Meta-analysis) collaborative study, examining the effect of the use of high functional oxygen saturation targets in the postnatal period for premature infants with gestational age <28 weeks. **Methods:** A meta-analysis of SUPPORT (Surfactant, Positive Pressure and Pulse Oximetry/Randomized Trial), the three SOCT II (Benefits of Oxygen Saturation Targeting) studies and the COT (Canadian Oxygen Trial) was performed including a total of 4,911 infants randomized to either a low (85-89%) or high (91-95%) functional oxygen saturation (SpO₂) within the first 24 h after birth. **Results:** Relative risks (RR; 95% CI) comparing a low versus a high oxygen saturation target were 1.41 (1.14-1.74) for mortality at discharge or at follow-up, 0.74 (0.59-0.92) for severe retinopathy of prematurity, 0.95 (0.86-1.04) for physiologic bronchopulmonary dysplasia, 1.25 (1.03-1.50) for necrotizing enterocolitis, 1.02 (0.88-1.19) for brain injury, and 1.01 (0.95-1.08) for patent ductus arteriosus. RR > 1 favors a high oxygen saturation. **Conclusion:** Tidal for mortality and necrotizing enterocolitis are significantly increased and severe retinopathy of prematurity significantly reduced in low compared to high oxygen saturation target infants. These are no differences regarding physiologic bronchopulmonary dysplasia, brain injury or patent ductus arteriosus between the groups. Based on these results, it is suggested that functional SpO₂ should be targeted at 90-95% in infants with gestational age <28 weeks until 36 weeks' postmenstrual age. However, there are still several unanswered questions in this field.

**Introduction**

The optimal oxygenation of extremely low birth weight infants has been hidden in the dark. This is not only due to a lack of knowledge of which values to target but as much due to a lack of understanding of the many detrimental effects of oxygen on the immature newborn infant. Retrolental fibroplasia (RLF) was first described in 1942, and by 1954 it was estimated that 10,000 infants had been blinded by this condition [1, 2]. The appearance of RLF in the 1940s taught the neonatal community slowly
Oxygen is Essential for Life: Elusive Optimal \( \text{PaO}_2 / \text{FiO}_2 \)

**Good**

90-95%

**Ugly – NEC**

**Mortality**

**Bad - ROP**

Oxygen therapy for preterm neonates – the elusive optimal target – Synnes, Miller JAMA Pediatrics 2015
Users’ Guides to the Medical Literature
A MANUAL FOR EVIDENCE-BASED CLINICAL PRACTICE
3rd EDITION

Gordon Guyatt, MD
Drummond Rennie, MD
Maureen O. Meade, MD
Deborah J. Cook, MD

McGraw Hill Education
Objective

• To systematically review evidence evaluating the effect of restricted versus liberal oxygen exposure on morbidity and mortality in extremely preterm infants (<28 weeks gestation).

• GRADE the quality of evidence and risk of bias

• Focus on studies performed within the NEOPROM collaboration
GRADE

• Grading of Recommendations Assessment, Development and Evaluation

• A system for rating the quality of evidence and strength of recommendations that is explicit and comprehensive

• Increasingly adopted by guideline organizations

• Confidence in estimates of effect into 4 levels:
  • High, moderate, low or very low

• Recommendations are graded as
  • Strong
  • Weak
The Cochrane risk-of-bias tool was used to assess study quality of included studies.

- Risk of bias assessment may explain heterogeneity in study findings and reflects the confidence in the estimate of effect.
- Risk of bias is categorized as low, unclear or high based on assessment of the following areas:
  - Sequence generation for randomization
  - Concealment of allocation
  - Blinding
  - Completeness of follow up
  - Selective reporting
  - Publication bias
  - Other biases relating to the specific research question being asked
Using GRADE, the quality of body of evidence for each outcome is graded as high, moderate, low or very low. This relates to:

- **Quality** (design and execution) of the studies included (determines risk of bias)
- **Consistency** of results (also referred to as variability or heterogeneity) – this is assessed by variation in the size of effect, overlapping confidence intervals, statistical significance of heterogeneity and $I^2$
- **Directness** of the studies compares the differences in outcomes in the studies (surrogate outcome for example) and those that are important to patients.
- **Precision** based on sample size and width of confidence intervals
- **Publication bias**
## Methods – GRADE Assessment

Confidence assessment criteria based on GRADE

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Quality of Evidence</th>
<th>Lower if</th>
<th>Higher if</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized trial</td>
<td>High</td>
<td>Risk of bias</td>
<td>Large effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 Large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very Serious</td>
<td>+2 Very Large</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>Inconsistency</td>
<td>Dose response</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 Evidence of a gradient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very Serious</td>
<td></td>
</tr>
<tr>
<td>Observational study</td>
<td>Low</td>
<td>Indirectness</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very Serious</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Very low</td>
<td>Imprecision</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very Serious</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Publication bias</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Likely</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very likely</td>
<td></td>
</tr>
<tr>
<td>Outcomes</td>
<td>RR (95% CI)</td>
<td>Participants</td>
<td>Quality of evidence</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------------------</td>
<td>--------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Death/Disability</td>
<td>1.02 (0.94-1.14)</td>
<td>2716</td>
<td>Moderate</td>
</tr>
<tr>
<td>Death – 24 mo</td>
<td>1.16 (0.98-1.37)</td>
<td>2783</td>
<td>Moderate</td>
</tr>
<tr>
<td>Death &lt; discharge</td>
<td>1.18 (1.03-1.36)</td>
<td>3757</td>
<td>Low</td>
</tr>
<tr>
<td>Disability</td>
<td>1.03 (0.73-1.45)</td>
<td>2252</td>
<td>Moderate</td>
</tr>
<tr>
<td>NEC</td>
<td>1.24 (1.05-1.47)</td>
<td>4929</td>
<td>Moderate</td>
</tr>
<tr>
<td>Severe ROP</td>
<td>0.72 (0.5-1.04)</td>
<td>4066</td>
<td>Low</td>
</tr>
</tbody>
</table>
Why is the quality of evidence rated low to moderate?

- Lack of separation between the two groups
- Pulse oximeter algorithm issues
- Significant unexplained heterogeneity with severe ROP
- BOOST-II UK and Australia – stopped early
- Death before discharge was not the primary prespecified outcome
NEOPROM Studies – Goal Separation

< 3% difference
Significant overlap

Potential zone of increased risk of mortality/NEC

Potential zone of increased risk of ROP
Masking Pulse Oximeters

Zone of instability (low SpO2 target arm) - Tendency to increase inspired oxygen

COT alarm setting at 94%

Oximeters for lower target range

COT alarm setting at 86%

Oximeters for higher target range

Zone of instability (high SpO2 target arm) - Tendency to decrease inspired oxygen

Schmidt et al J Peds 2014
Impact of Algorithm Revision

![Graph showing the impact of algorithm revision on displayed pulse oximeter values. The graph compares revised algorithm performance against the protocol recommended values.]

Range recommended by the protocol for bedside providers during the study period.
Impact of Algorithm Revision

Range recommended by the protocol for bedside providers during the study period.
Impact of Algorithm Revision

- Original algorithm
- Revised algorithm - low display
- Original algorithm - high display
- Revised algorithm - high display

Range recommended by the protocol for bedside providers during the study period.
Impact of Algorithm Revision

Range recommended by the protocol for bedside providers during the study period.
NEOPROM Combined Data – No Dose Response Effect

Original Algorithm - Corresponding % on Revised Algorithm
< Left Tails Right SUPPORT vs. COT

Potential zone of increased risk of mortality/NEC

Potential zone of increased risk of ROP

Actual median oxygen saturation level (%)
Less ROP & More NEC with Restricted $O_2$

**Necrotizing Enterocolitis**

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Restricted Oxygen, No.</th>
<th>Liberal Oxygen, No.</th>
<th>Weight, %</th>
<th>Risk Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOOST II trial, 2013</td>
<td>127 (1221)</td>
<td>97 (1217)</td>
<td>44.1</td>
<td>1.30 (1.01-1.68)</td>
</tr>
<tr>
<td>COT, 2013</td>
<td>74 (602)</td>
<td>56 (599)</td>
<td>26.0</td>
<td>1.31 (0.95-1.83)</td>
</tr>
<tr>
<td>SUPPORT, 2010</td>
<td>76 (641)</td>
<td>70 (649)</td>
<td>29.9</td>
<td>1.10 (0.81-1.49)</td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td>277 (2464)</td>
<td>223 (2465)</td>
<td>100.0</td>
<td>1.24 (1.05-1.47)</td>
</tr>
</tbody>
</table>

Heterogeneity: $\tau^2 = 0.00; \chi^2 = 0.88, df = 2 (P = 0.65); I^2 = 0\%$
Test for overall effect: $z = 2.54 (P = 0.001)$

**Retinopathy of Prematurity**

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Restricted Oxygen, No.</th>
<th>Liberal Oxygen, No.</th>
<th>Weight, %</th>
<th>Risk Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOOST II trial, 2013</td>
<td>110 (1035)</td>
<td>141 (1044)</td>
<td>36.6</td>
<td>0.79 (0.62-0.99)</td>
</tr>
<tr>
<td>COT, 2013</td>
<td>64 (500)</td>
<td>66 (503)</td>
<td>32.3</td>
<td>0.98 (0.71-1.34)</td>
</tr>
<tr>
<td>SUPPORT, 2012</td>
<td>41 (475)</td>
<td>91 (509)</td>
<td>31.0</td>
<td>0.48 (0.34-0.68)</td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td>215 (2010)</td>
<td>298 (2056)</td>
<td>100.0</td>
<td>0.72 (0.50-1.04)</td>
</tr>
</tbody>
</table>

Heterogeneity: $\tau^2 = 0.08; \chi^2 = 8.93, df = 2 (P = 0.01); I^2 = 78\%$
Test for overall effect: $z = 1.73 (P = 0.08)$

Tau$^2$ is used for random-effects modeling to assess heterogeneity.
Heterogeneity

- Differences among individual studies included in a systematic review, typically referring to study results.
  - Visual assessment of variability
    - How similar are the point estimates?
    - To what extent do the confidence intervals overlap?
  - Statistical tests evaluating variability
    - Yes-or-no tests for heterogeneity that generate a P value
    - $I^2$ test that quantifies the variability explained by between-study differences in results
I² Statistic

100% Why are we Pooling?
75% Very concerned
50% Getting concerned
25% Little concerned
0% No worries All is well

Substantial heterogeneity

No heterogeneity
Probe site and PRBC transfusions

C. Location of the probe (pre/postductal)
Effect of transfusions - HbA vs. HbF
# Pulse Oximetry in Neonates - Flaws

<table>
<thead>
<tr>
<th>Displayed SpO2</th>
<th>Site of the probe (pre- vs. postductal)</th>
<th>Corresponding SaO2 (+/- 3%)</th>
<th>Corresponding preductal SaO2 (+/- 2 cf. postductal)</th>
<th>Transfusion status (% HbF and P50)</th>
<th>Preductal PaO2 (mmHg)</th>
<th>Hemoglobin (g/dl)</th>
<th>Arterial oxygen content (ml/dl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>88%</td>
<td>86%</td>
<td>86%</td>
<td>Never transfused (HbF - 90% P50 - 18mmHg)</td>
<td>36</td>
<td>8</td>
<td>9.5</td>
</tr>
<tr>
<td>B</td>
<td>88%</td>
<td>90%</td>
<td>92%</td>
<td>Transfused (HbF - 50% P50 - 22mmHg)</td>
<td>52</td>
<td>13</td>
<td>16.4</td>
</tr>
<tr>
<td>C</td>
<td>93%</td>
<td>91%</td>
<td>91%</td>
<td>Never transfused (HbF - 90% P50 - 18mmHg)</td>
<td>43</td>
<td>9</td>
<td>11.3</td>
</tr>
<tr>
<td>D</td>
<td>93%</td>
<td>95%</td>
<td>97%</td>
<td>Multiple transfusions (HbF - 10% P50 - 27mmHg)</td>
<td>95</td>
<td>14</td>
<td>18.8</td>
</tr>
</tbody>
</table>

Variable effects on oxygen delivery and toxicity
Final Comments

- Highlight the importance of assessing the risk of bias and the quality of evidence in studies and of the level of confidence in the estimate of effect of each outcome.
- Well designed and mostly well executed studies, the significant overlap in oxygenation between the intervention and comparator arms makes it difficult to conclude that the results obtained were predominantly due to difference in oxygenation.
- Death before hospital discharge was not a pre-specified outcome in any of these trials. The confidence in this outcome is lower because of this reason.
- Significant heterogeneity for the outcome of ROP resulting in lowering of the level of confidence for the estimate of effect for this outcome.
- Qualitative factors need to be considered before making treatment recommendations based on systematic reviews of available literature.
Conclusions

• Significant physiological and methodological differences in the results of NEOPROM studies

• Inference: Most likely, there is no real difference in mortality between the two groups.
  – Differences are probably related to episodes of hypoxemia (SpO$_2$ <85%) and hyperoxemia (SpO$_2$ >95%)

• There is still significant uncertainty about the optimal target range for oxygen saturation in extremely preterm infants.

• Plan: Wait for AAP COFN Recommendations