Pain relieving interventions for retinopathy of prematurity: Systematic review and network meta-analysis

Disher, T.a, Cameron, C.b, Cathcart, K.a, Campbell-Yeo, M.a,c

**Affiliations:** aDalhousie University, School of Nursing, Halifax, NS; bCornerstone Research Group, Burlington, ON; cIWK Health Centre, Halifax, NS

**Address correspondence to:** Dr. Marsha Campbell-Yeo, Associate Professor, Dalhousie University School of Nursing, Forrest Building, 6299 South St, Halifax, NS, Canada, B3H 1X7, [marsha.campbell-yeo@dal.ca](mailto:marsha.campbell-yeo@dal.ca), 902-494-4283

**Financial Disclosure:** Disher is supported by a Canadian Institutes for Health Research (CIHR) Masters Scholarship, Killam Scholarship, Dalhousie Faculty of Graduate Studies Research and Innovation Award, Nova Scotia Health Research Foundation Scotia Scholars Doctoral award, and the Dalhousie School of Nursing; Electa Maclennan Memorial Scholarship; Campbell-Yeo is supported by a CIHR New Investigator salary award, and Canadian Child Health Clinician Scientist Program Career Development Award.

**Conflict of Interest:** All authors have indicated they have no conflicts of interest to declare.

**Abbreviations:**to be added later.

**Table of Contents Summary:** to be added later.

**What’s Known on This Subject:** The use of an exclusive human milk fortifier is associated with improvements in rates of NEC, BPD, ROP, and sepsis. Previous estimates considering the cost of NEC alone suggested that a cost savings of $8000 for infant could be expected.

**What This Study Adds:** To be added

###### page break

# Abstract

**Context:** Despite substantial research effort to identify effective pain relieving interventions, there has been no systematic review that compares all available interventions.

**Methods:** Systematic review and network meta-analysis using NetMetaXL.

**Data sources:** Cochrane CENTRAL, MEDLINE, Embase, Web of Science,Clinicaltrials.gov, and the WHO ICTRP (2004-Feb 2017).

**Study Selection:** Interventions intended to provide pain relief and could include pharmacological, non-pharmacological, multisensory, or procedural modifications (e.g. the use of wide field digital retina imaging) were included.

**Data Extraction:** Abstract and title screen, full-text screening, and data extraction, were conducted independently by two reviewers. The primary outcome was pain in during the exam period (pain reactivity) with pain following the exam (pain recovery) and adverse events as secondary outcomes.

**Results:** Seventeen studies (n = 1187) investigating nine interventions assessed pain reactivity phase. Only multisensory interventions combined with topical anesthetic showed statistically significant improvement over topical anesthetic alone (Random effect MD = 2.96, 95% Credible Interval = -5.23 to -0.62). Eleven studies (n = 630) assessed pain recovery with multisensory interventions showing a statistically significant decrease in pain (Random effect MD = -4.79, 95% Credible Interval = -8.81 to 0.79). No differences were detected in adverse events.

**Limitations:** We found evidence that vague priors artificially inflated network heterogeneity. Additional heterogeneity within nodes could not be explained through sensitivity analyses.

**Conclusions:** Recommendations for best treatments for ROP eye exams is complicated by significant heterogeneity. Use of informative priors may improve precision but significant unexplained heterogeneity is likely to remain. Multisensory interventions are likely optimal, but results should be interpreted with caution.

###### page break

# Background

Preterm neonates, those born before 37 weeks gestational age, are at higher risk for pain exposures, which have been associated with numerous short and long-term sequalae including altered cortical development and changes in response to later pain1,2. Retinopathy of prematurity (RoP) is a potentially serious disease that arises from the immature vasculature of the preterm retina3. If left untreated, RoP can result in blindness. Current guidelines recommend that infants born less than 30 weeks receive eye serial eye exams until their retina reach maturity3. Standard practice for eye exams involves indirect ophthalmic examines which require eyelid retraction and scleral pressure3. This procedure is widely recognized as being painful, with neonates showing both immediate pain behaviours and prolonged physiological arousal4.

To reduce the pain associated with RoP eye examination, researchers have identified several pharmacological, non-pharmacological, and procedural modification interventions4. While this is a positive development, the plurality of approaches makes a direct comparison of all interventions unfeasible without a large multi-centre trial. As a result, despite the topic being the subject of at least three recent reviews4–6, it has not been possible to provide a statistically derived estimate of the most effective treatment. The purpose of this systematic review will be to combine all existing randomized trials of pain-relieving interventions for RoP exams using network meta-analysis in order to allow for comparison of direct and indirect evidence.

# Objectives

To conduct a systematic review and network meta-analysis comparing all available pain-relieving strategies for RoP eye examinations.

# Methods

## Protocol and registration

## Study Design

We conducted a systematic review with network meta-analysis using a Bayesian approach in WinBugs7. A pre-specificed protocol was followed (PROSPERO 2017:CRD42017058231).

## Search Strategy and Selection Criteria

A database search was conducted and February, 2017. The search strategy was developed in partnership with library professional and included searches of the Cochrane Library Central Registry of Controlled Trials (1966-present), MEDLINE (1946-present), Embase (1974-present), and Web of Science (1900-present). A complete search strategy for MEDLINE can be found in the supplemental material (Appendix A). Eligible trials designs included randomized clinical trials comparing at least two pain-relieving strategies for ROP eye exams conducted in preterm neonates.

## Study Selection and Data Extraction

Parallel group and cross-over designs were included. Eligible interventions included those that were intended to provide pain relief and could include pharmacological (e.g. sucrose, glucose, nitric oxide, topical anesthetic), non-pharmacological (e.g. facilitated tucking, bundling, holding, skin-to-skin contact, non-nutritive sucking), combined interventions, or procedural modifications (e.g. the use of wide field digital retina imaging). Abstract and title screen, full-text screening, and data extraction, were conducted independently by two reviewers using Covidence8. All conflicts were resolved by reviewers and, if necessary, consultation with a third reviewer. Data was extracted using standardized forms.

## Outcomes

The primary outcome was pain score as measured by validated pain assessment tools. As there are over 50 tools in the literature, we made a decision apriori to convert all tools to a common scale (the premature infant pain profile) as recommended by the OMERACT group9. Since these transformations introduced a bias, we conducted a priori sensitivity analysis excluding these trials from the analysis. A challenge commonly faced during meta-analysis of neonatal pain trial is the large number of time points typically measured (e.g. 30, 60, 90, and 120s after the procedure)10. Following the approach outlined in Pillai Riddell et al’s11 Cochrane review of nonpharmacologic pain relieving interventions in neonates, we selected one timepoint measured during the procedure (pain reactivity), and the first timepoint after completion of the procedure (pain recovery). Secondary outcomes included adverse events defined as apnea (Sp02 less than 80%) or bradycardia (heart rate 60bpm or less).

## Quality Assessment – Risk of Bias

Critical appraisal was conducted using the Cochrane risk of bias tool for randomized controlled trials12. Two reviewers assessed each study, with conflicts resolved through consultation or, if required, consultation with a third reviewer.

## Statistical Analysis

Relevant clinical and study design characteristics were compared between eligible trials in order to assess acceptability to synthesis. These included infant postmenstrual age at time of procedure, birthweight, use of speculum and scleral depression during procedure, and infant positioning (e.g. swaddled or contained). Pairwise and network meta-analysis was conducted using WinBugs through the freely available program NetMetaXL13. Results of continuous scales were expressed in mean difference and accompanied with their 95% credible intervals (CrI). Adverse events were expressed as odds ratios (ORs). In order to provide a coherent probabilistic ranking of treatments, surface under the cumulative ranking (SUCRA) curves were developed14. Heterogeneity was assessed through he standard deviation of the random effect distribution, using previously established benchmarks which identify values of 0.1 to 0.49 as reasonable, 0.5-0.9 as moderate, and greater than 1 as potentially indicative of serious heterogeneity15.

Assessment of inconsistency within the network (e.g. agreement between direct and indirect evidence) was assessed by comparing the deviance information criteria (DIC) between a consistency and inconsistency model16. Lower DIC indicates better fit, with differences of 3-5 indicating better fit. Cross-over trials. Cross-over trials can lead to unit of analysis issues if correlations between treatment and control are unaccounted for17. If treated as if the data were generated by parallel design, the precision of the trial will be underestimated and will receive less weighting in the analysis. The Cochrane handbook outlines several methods for addressing this issue, including calculation of corrected standard errors from paired statistical tests17. Sensitivity analyses were conducted excluding cross-over trials.

### Skewed data.

Many authors report that pain scales often produce skewed data, which may have important implications for pooled analyses. When results were reported as median and IQR, we assessed whether the assumption of symmetry was supported. In the absence of signs of skewness, we used scores as reported, treated medians as the mean and imputed the standard deviation using method outlined by the Agency for Healthcare Research and Quality (AHRQ) 18. Sensitivity analyses were conducted excluding these trials.

### Outcomes reported using multiple scales.

There are over fifty scales validated for the measurement of pain in preterm neonates10. While the standardized mean difference is often used in this context, it has several important limitations9. Suggestions of cut points for small, moderate, and large effects are potentially valuable although not directly relevant or interpretable for clinicians9. The SMD can also lead to bias in estimate in that trials conducted with more homogenous populations will have smaller SDs and thus larger estimated effects9. To allow for combination of multiple scales, we scaled scores to the premature infant profile to allow for a single clinically relevant outcome19.

# Results

## Search results

The database search returned 797 citations after removal of duplicates, of which 20 met all inclusion criteria (figure 1). All included trials were published in peer-reviewed journals.

## Study characteristics

Most trials were parallel randomized controlled trials20–34, with four studies35–38 reported as randomized crossover trials and one study reported as a multi-site randomized crossover trial39. Interventions included no treatment21,35, anesthetic eye drops20–22,25,29,31–35,37–39, non-nutritive sucking (NNS) alone36, sweet taste alone28, sweet taste combined with anesthetic eye drops20,22,26,29,33,37, NNS combined with anesthetic eye drops20,23,24,27,30,36, sweet taste combined with anesthetic eye drops and non-nutritive sucking20,23,24,27,30, sweet taste combined with non-nutritive sucking and nitrous oxide26, expressed breast milk combined with anesthetic eye drops31, sensorial saturation combined with anesthetic eye drops34, acetaminophen combined with anesthetic eye drops3225,33, and morphine combined with anesthetic eyedrops32. All but three included studies were two-arm trials, with one four-arm20 and two three-arm included32,33. All but two studies used binocular indirect ophthalmoscopy for the procedure, with one study examining pain reaction to the use of wide field digital retinal imaging38, an one study that used two methods at the different sites39. One study did not use a speculum to hold the infant’s eyelids together29, and in two studies it was unclear whether a speculum was used33,39. Scleral depression was used in all but two studies38,39, but was unclear in an additional four25,29,40. There was a mix of eligible eye exams, with six using only the first eye exam20,22,23,25,28,30, and the remainder allowing the use of any eye exam as the procedure. When reported, the average post-menstrual age at time of procedure was similar across studies, as was birthweight (supplementary tables 1 to 3). Reporting of infant positioning during the procedure was mixed, with seven studies stating that infants were swaddled24,26,27,30,31,36,37, and the remainder stating infants were contained. There is mixed evidence regarding whether either approach offers better baseline pain control11, but there were no observable stable differences between these studies.

Based on consultation with clinicians and pain researchers, we elected to lump together the following interventions: Sucrose and glucose were combined in a node describing sweet tasting solutions, combinations of more than two interventions whose mode of action is through a different sensory input were combined in a multisensory category. A full list of study characteristics, actual comparisons and the treatment groups lumped together and their individual results can be found in supplementary tables one to three.

## Risk of bias within studies

Studies assessing easily blindable interventions (e.g. sweet taste, oral acetaminophen) were generally considered to be at an overall low risk of bias (supplementary table 4). Details of sequence generation and allocation concealment were unclear in most studies. Studies that tested interventions which were difficult or impossible to blind (e.g. sensorial saturation) generally scored as high risk of bias on outcome assessors.

## Publication Bias

We intended to use funnel plots to investigate signs of publication bias, although no comparisons had sufficient studies12. We identified several trial registries indicating trials that are or should realistically be complete without an identifiable publication of results in abstract or manuscript form41–46. One of these was a trial assessing efficacy of acetaminophen which was stopped early because the intervention showed no effect43. None of the contacts listed responded to e-mails.

## Pain reactivity

Seventeen studies (n = 1187) investigating nine interventions reported results of a validated pain assessment scale during the pain reactivity phase (Fig 2A). Five of eleven nodes included at least two studies, and comparisons between pair-wise and nma estimates suggested that the heterogeneity parameter was being influenced by the vague prior that was placed on it (Table 1). As a crude estimate of the influence of the vague prior we duplicated pairwise analyses with frequentist methods and found clinically significant (greater than 2 points) differences in width of confidence intervals of the random effect model (supplementary table 5). The greatest impact was observed within single study connections (e.g. WFDRI vs topical anesthetic). This may disguise a statistically significant difference between topical anesthetic and no treatment which disappears in the Bayesian pairwise and nma results. We considered the use of a fixed effect model, but found it to be a poor fit for the data (posterior deviance = 79.49 on 19 data points, DIC 109.9 vs 59.4), so the random effect model was used. The most common comparator was anesthetic drops alone (13 studies, n = 458) and thus it was set as the reference treatment. Generally, adding additional interventions resulted in incrementally larger differences in treatment effect (Fig 3). There was evidence of considerable network heterogeneity, although the true value is disguised by residual effect of vague priors (supplementary table 4). Sensitivity analyses were conducted excluding trials with large residual deviance contribution (> 1.5), crossover trials, and studies in which imputed means or scaled scores were used (Table 1). Estimates were similar with regards to direction and magnitude in all scenarios. Estimates from the random effect consistency and inconsistency models were similar, and deviance information criteria was slightly higher in the inconsistency model (59.4 vs 58.9) suggesting it was not a good fit to the data.

## Pain recovery

Eleven studies assessing nine interventions (n = 630, Fig 2B) were included in the analysis. The most common interventions were anesthetic eye drops (8 studies, n = 211) followed by multisensory interventions combined with anesthetic drops (4 studies, n = 91). A single study of sweet taste alone found a substantial effect leading to it being associated with the largest point estimate compared against anesthetic drops alone (-4.99, 95% CrI = -11.10 to 1.03). As in the case of the reactivity analyses, large amounts of network heterogeneity resulted in wide credible intervals that prevented any intervention except anesthetic drops combined with a multi-sensory intervention from reaching statistical significance (Table 2). With the exception of sweet taste being considered the most-likely best intervention, ranking based on sucra generally followed the trend that increased number of interventions was associated with increased pain relieving effects (Figure 4). Sensitivity analyses were conducted removing studies with scaled scores or imputed means, cross-over designs, and residual deviance greater than 1.5 (Table 2). Direction and magnitude of effect did not meaningfully change for any comparison. The inconsistency model for this comparison only offered a slight improvement in DIC (0.13 points lower), suggesting that the consistency model was a better fit for the data. As in the reactivity analysis, we found evidence that vague priors were exhibiting substantial influence over the precision of the pooled results (supplementary table. 6).

## Adverse events

Few studies reported frequency of adverse events, with two distinct networks preventing a complete comparison of the include treatments. The first network consisted of two studies comparing anesthetic drops, anesthetic drops in addition to acetaminophen, and anesthetic drops with sweet taste (n = 160, Fig. 2D) . Adverse events in all cases were bradycardia, but there was a large unexplained in the baseline risk of events between studies. The second network consisted of two studies comparing anesthetic drops with NNS against anesthetic drops with a multisensory intervention (n = 104, Fig 2C). All arms in all included studies had at least one event. Because all nodes were single study connections, a fixed effect model was used. In both networks there were no statistically significant differences found between interventions (Fig 5).

# Discussion

To our knowledge, this is the first attempt at a network meta-analyses of interventions intended to decrease procedural pain in preterm neonates. While most comparisons failed to reach statistical significance, probabilistic results support the hypothesis that engaging more sensory systems likely results in improved pain relief. The exception is in the recovery period where sweet taste was the optimal intervention, although it should be considered that this was the result of a single trial. Point estimates were robust to sensitivity analyses, which provides some degree of confidence in their findings although all results should be interpreted with caution.

Interpretation of the findings is complicated by profound heterogeneity across all networks. The source of heterogeneity appeared to be residual influence of vague priors in combination with within comparison heterogeneity. As the model we used derives a single heterogeneity term for the entire network, the result is that only a single intervention appeared to offer a statistically improvement over anesthetic eye drops despite generally large estimates of effect size. This is especially problematic considering pairwise meta-analyses results showed that acetaminophen, and multisensory interventions combined with topical anesthetic, were superior to topical anesthetic alone. No sensitivity analysis could meaningfully reduce heterogeneity, which was driven in all cases by disparate results in the current studies. Investigation of potential sources of heterogeneity was further limited by incomplete procedure reporting which made it difficult to know the exact timing of the recording of the pain response, and potentially important procedural modifiers (e.g. presence or absence of eyelid speculum or scleral depression). While these factors may explain some additional heterogeneity, it should be noted that this appears to be a consistent problem faced in meta-analysis pain-relieving interventions in neonates. Two recently updated Cochrane systematic reviews assessed nonpharmacologic11 and skin-to-skin contact10 as interventions for reducing pain from commonly performed painful procedures in preterm and term neonates. In both cases, moderate the high heterogeneity was commonly cited a reason for downgrading the level of evidence from combined analysis. The recently updated review of sucrose for painful procedures did not experience the same difficulty, however most analyses were limited to single studies47. No reviews have had success in identifying explanations for this heterogeneity, and thus it is unclear whether it is the result of methodological or clinical heterogeneity. It is possible that this heterogeneity in results at least partially explains the diversity of treatments investigated and used in practice.

An unexpected challenge in interpreting these results arises from the need to compare interventions which are amenable to blinding with those that generally are not. For example, it is relatively simple to blind sterile water against sweet tasting solutions, but not possible to do so when comparing anesthetic drops alone to a multisensory intervention (e.g. NIDCAP, or sensorial saturation). As a result, it is possible that estimates of multisensory interventions are biased which may make them appear to be superior to sweet taste alone when in fact they are not. This may have clinical implications, as multisensory interventions are more resource intensive to implement.

While using cut scores to determine mild, moderate, or severe pain should generally be avoided, it is important to comment on the relative inability of any intervention to meaningfully reduce average raw scores. For example, when assessing pain during eye exam 12 of the 17 studies reported PIPP scores of 11 or greater in both groups with six of those documenting scores considered to reflect severe pain. This is placed in comparison to the same interventions used to reduce pain from vaccination, heel lance, or venepuncture where scores in intervention groups are routinely lying within four and six points on the same scale10,11,47,48. Strong conclusions from these findings are not warranted, however it does draw our attention to two points of consideration: Whether existing pain scales valid for assessment of pain from RoP exams, and should we be giving greater consideration to more powerful interventions? The PIPP was the most used pain scale included in this review, however in the most recent review of validation and reliability authors have not repored validation studies for RoP exams47. This is of relevance because the exam itself includes direct manipulation of the face including eyelid speculum and a bright light, which makes up three domains (and a possible 9 points) on the PIPP. The lowest raw score on the PIPP came from a unit that did not use the eyelid speculum29, and while it is tempting to suggest this as an approach to relieve pain it is difficult to say whether this is an artifact created by the scale not being valid for these procedures. Others have suggested that persistently high raw pain scores suggest that stronger analgesics should be investigated (e.g opiates). One ongoing clinical trial will investigate the use of morphine for pain reduction during eye exam and use the PIPP to assess pain. While this will undoubtedly provide important information, it will need to be interpreted within the possibility that the tool itself is not a valid measure of pain perse.

# Conclusions

Substantial network heterogeneity limits confidence in findings and may disguise real differences in treatment effects observed in pairwise meta-analysis. Current evidence suggests that the optimal treatments are likely to be anesthetic eye drops in combination with either sweet taste or multisensory interventions although the evidence is of very low quality. Estimates for all interventions include potentially clinically meaningful benefit or harm, introducing substantial uncertainty in recommendations for practice and future research. Assumptions about the validity of pain scores for ROP exams may need to be revisited. Future studies should be careful to report exactly when pain scores were assessed, and provide all relevant details of the procedure including use of speculum and scleral depression, and infant positioning.