



# Systematic Reviews in Neonatal Respiratory Care: Are Some Conclusions Misleading?

Andres Maturana<sup>1,2\*</sup>, Fernando Moya<sup>3</sup> and Steven M. Donn<sup>4</sup>

<sup>1</sup> Neonatology, Clinica Alemana, Department of Pediatrics, Santiago, Chile, <sup>2</sup> Faculty of Medicine, Centro de Desarrollo Educativo, Universidad del Desarrollo, Santiago, Chile, <sup>3</sup> Betty Cameron Children's Hospital, Coastal Carolina Neonatology, Coastal Children's Services, PLLC, Wilmington, NC, United States, <sup>4</sup> Division of Neonatal-Perinatal Medicine, C.S. Mott Children's Hospital, Michigan Medicine, Ann Arbor, MI, United States

An increasing amount of information is currently available in neonatal respiratory care. Systematic reviews are an important tool for clinical decision-making. The challenge is to combine studies that address a specific clinical question and have similar characteristics in terms of populations, interventions, comparators, and outcomes, so that their combined results provide a more precise estimate of the effect that can be validly extrapolated into clinical practice. The concept of heterogeneity is reviewed, emphasizing that it should be considered in a wider perspective and not just as a mere statistical test. A case is made of how well-designed studies of the neonatal respiratory literature, when equivocally combined, can provide very precise but potentially biased results. Systematic reviews in this field and others should be rigorously peer-reviewed before publication to avoid misleading readers to potentially biased conclusions.

**Keywords:** neonatal respiratory care, meta-analysis, systematic reviews, clinical decision-making, infant-newborn

## OPEN ACCESS

### Edited by:

Saadet Arsan,  
Ankara University, Turkey

### Reviewed by:

Hannes Sallmon,  
Charité Medical University of  
Berlin, Germany

Daniel Vijlbrief,  
University Medical Center  
Utrecht, Netherlands

### \*Correspondence:

Andres Maturana  
amaturana@alemana.cl

### Specialty section:

This article was submitted to  
Neonatology,  
a section of the journal  
Frontiers in Pediatrics

**Received:** 05 November 2019

**Accepted:** 09 January 2020

**Published:** 31 January 2020

### Citation:

Maturana A, Moya F and Donn SM  
(2020) Systematic Reviews in  
Neonatal Respiratory Care: Are Some  
Conclusions Misleading?  
Front. Pediatr. 8:7.  
doi: 10.3389/fped.2020.00007

## INTRODUCTION

We are currently confronted with an overwhelming amount of information in all medical disciplines, and neonatal care is no exception (1, 2). A systematic review of the current literature can provide information that may be combined, thus increasing statistical power and providing a quantitative estimate of the effect in a meta-analysis (3). Although systematic reviews addressing a specific clinical question can help clinicians appraise in a summarized format all or most of the existing research pertaining to that topic and aid in bedside decision-making, they have recognized limitations (4, 5). Clinicians are sometimes confronted with systematic reviews that claim results based on combining studies that differ in substantial ways and therefore yield conclusions that are very difficult to interpret (6). Most of us would agree that almost any respiratory outcome in premature infants could be significantly influenced by antenatal steroid exposure and gestational age. Nevertheless, systematic reviews combining study populations with significant differences in these relevant variables have been published (Table 2).

The purpose of this review is to raise awareness of the importance of adequately appraising systematic reviews, using examples from the neonatal respiratory literature that, in our view, can sometimes lead to misleading conclusions. Table 1 summarizes the definitions of terms that will be used.

**TABLE 1** | Terms used in this review.

Term	Definition (*)
Systematic Review	The identification, selection, appraisal, and summary of primary studies that address a focused clinical question using methods to reduce the likelihood of bias.
Meta-Analysis	A statistical technique for quantitatively combining the results of multiple studies that measure the same outcome into a pooled or summary estimate.
Heterogeneity	Differences among individual studies included in a systematic review. These differences can refer to study characteristics or study results.
$I^2$ Statistic	The $I^2$ statistic is a test of heterogeneity. The results range from 0 to a 100% indicating no heterogeneity to high heterogeneity, respectively.
Bias	Systematic deviation from the truth because of a feature of the design or conduct of a research study. This can skew the outcome in a certain direction.
Selection Bias	Occurs when the population that is selected for a study is not representative of the general population addressed by the question the study intends to answer. This has the consequence that study results although not necessarily biased may not be applicable to the general population.

\*Definitions adapted from *Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice, Third Edition (7)*.

## THE CONCEPT OF HETEROGENEITY

A systematic review summarizes the existing research that addresses a specific clinical question in a systematic and reproducible way. For the purpose of this review, we will refer to systematic reviews addressing the effect of therapeutic interventions in randomized clinical trials. In some cases, the studies found in the review process can be combined using meta-analysis, so as to provide a single more precise estimate of the effect (3). This entails some assumptions about the studies included in the analysis. First, the magnitude and direction of the treatment effect across the different studies should be relatively similar and that there are no significant variations in the results that could be explained by relevant differences among the studies. The studies should be combined only if they lack significant bias, if they answer the same specific question, if they include similar populations, and if they attempt to compare similar interventions and measure equivalent outcomes, so that a pooled effect of the results from individual studies yields a more precise and representative estimate of the treatment effect (6). The challenge is how much difference (heterogeneity) we are willing to tolerate in these parameters among the different studies without compromising the confidence of the pooled estimate. The usual approach to this conundrum is to evaluate heterogeneity in a statistical manner. Any of the tests used for this purpose are only providing information about differences between study results and telling us how likely the differences in individual trial results are from chance alone (9). A frequently used test for evaluating heterogeneity is the  $I^2$  statistic that estimates the heterogeneity as the magnitude of variability. It is easily interpreted as the percentage of heterogeneity in the point estimates from individual studies. When it approaches

0%, the reader can be relatively confident that any differences between the individual point estimates of the included studies is explained merely by chance and, therefore, the summary estimate of the treatment effect is credible. When this percentage approaches 100% the probability that only chance explains these differences is substantially less likely and, therefore, a summary effect is more difficult to interpret (10). The problem is that sometimes we can be confronted with differences in study design that make any pooled estimate of the effect difficult to interpret or even meaningless, and are not necessarily detected by any statistical test for heterogeneity. Therefore, heterogeneity between studies in a meta-analysis needs to be examined as much more than a simple statistical test, and clearly, one more relevant issue when critically appraising a systematic review.

## HETEROGENEITY IN INCLUDED POPULATIONS, INTERVENTIONS, CONTROL GROUPS, AND OUTCOMES

If we are considering therapeutic interventions, a certain homogeneity in the populations included in the different studies considered in a systematic review can be a very relevant issue. We should not feel comfortable drawing any conclusions from a meta-analysis within a systematic review that combines studies including populations that differ in characteristics that could potentially influence the magnitude or direction in the effect of the intervention being studied.

A systematic review by Ferguson et al. addressing the question of interventions to improve rates of successful extubation in preterm infants can help exemplify this point (8). If we review the comparison between high flow nasal cannula and nasal continuous positive airway pressure (CPAP) on the outcome respiratory failure, three studies are included in this analysis (Table 2) (11–13). As an example, the populations in the study by Yoder include more mature infants (>28 weeks) and with a significantly lower percentage of antenatal steroid receipt (<35%) than the other two included studies, and these are two well-recognized prognostic factors for respiratory failure. Fortunately, in this case we are alerted by an  $I^2$  of 55%, suggesting that chance does not adequately explain the variability between the point estimates. Regrettably, this is not always the case.

An intervention will have an effect that will reflect a magnitude and a direction. Evidently, this is dependent upon the comparative intervention. It would not be correct to claim a certain magnitude of effect of a certain intervention if it is being compared to anything different than the standard of care for the control group, since this could potentially overestimate the real effect of the intervention. It would not make much sense to combine studies that have different comparators in a meta-analysis. A recently published systematic review by Wu et al. addresses the outcomes of surfactant administration in a minimally invasive way (via thin endotracheal catheter) to spontaneously breathing infants (14). In this review, four studies are included for the outcome of requiring mechanical ventilation within the first 72 h of life (15–18). The trial by

**TABLE 2 |** Heterogeneity in populations included in the Meta-Analysis by Ferguson et al. (8).

Trial	Mean gestational Age (treated/controls)	% of Antenatal Steroids (treated/controls)	% of Caffeine use (treated/controls)
Yoder et al. (11)	33.5 ± 3.6/33.2 ± 3.2	38/32	27.0/30.0
Manley et al. (12)	27.7 ± 2.1/27.5 ± 1.9	93.4/94.7	99.3/98.0
Collins et al. (13)	27.9 ± 1.9/27.6 ± 1.9	88.0/89.0	100.0/100.0

Göpel compared a less invasively administered surfactant (LISA) to intubation and rescue surfactant via endotracheal tube in the control group, while Kanmas and Bao compared LISA with the Intubation-Surfactant-Extubate (INSURE) procedure in the control group. In these studies, specific criteria for respiratory failure were defined in the protocols. The included study by Kribs used LISA and compared this to surfactant administration with mechanical ventilation. In this last case, indications for mechanical ventilation were defined by protocol for the control group and, in fact, only one infant was not mechanically ventilated. For this analysis, the  $I^2$  statistic shows 0% heterogeneity, suggesting that the summary point estimate is not biased by any relevant differences between the studies. Nevertheless, it is obvious that these studies are completely different and probably should not have been combined for this outcome.

When evaluating the impact of an intervention on a specific outcome across different studies, an important assumption is that the outcome in each of the studies was similarly defined, so as to render the combined effect in a meta-analysis interpretable. This is particularly relevant when considering physician-driven outcomes, which are those that depend upon the treating physician and therefore rely on how every protocol in each study defined the criteria for this outcome. An example of such an outcome in neonatal practice is nasal CPAP failure or intubation for mechanical ventilation. We can expect differences in clinical practice among different centers and even within a single center among different clinicians. When one performs a systematic review, one forgoes the ability to conduct logistic regression analysis using center effect as a variable. The problem arises when we try to interpret combined results of studies that have, for instance, significant differences in the criteria for intubation, especially if it is not defined *a priori* in the various studies included in the systematic review.

Another example of this is the recently published review by Conte that addresses the comparison of high flow nasal cannula and nasal CPAP as the initial strategy to treat RDS in preterm infants (19). In this review, six studies are included in the analysis for the outcome of respiratory failure, but only five of them contribute with outcomes (11, 20–23). If we look at the  $I^2$  statistic, it shows that there is relatively little heterogeneity (17%) within the included studies for this outcome and, therefore, we should be fairly confident in interpreting this summary estimate of the treatment effect. Unfortunately, this statistic can only detect the mathematical heterogeneity in the individual point estimates of the effect but will not reflect relevant differences within the studies. In this example, three of the studies (20, 22, 23)

have intubation thresholds utilizing an  $FiO_2$  of 0.4, whereas Nair and Karna (21) and Yoder et al. (11) have significantly higher thresholds for intubation (0.6 and 0.7, respectively). These differences will evidently bias the results toward a lower difference between the groups for this outcome, since fewer patients will meet the threshold. If we exclude these two studies, the analysis yields a significantly greater magnitude in the point estimate against using high flow nasal cannula as the initial support strategy (1.72 vs. 1.57).

## LIMITATIONS IN GENERALIZABILITY

When examining the conclusions of any trial, including those conducted under high standards, they can only provide an answer to a clinical question that generally is fairly specific (primary outcome), and applicable to the population studied. Good examples of this paradigm are those studies that compared CPAP at or soon after birth vs. intubation with or without surfactant administration. For instance, the COIN trial enrolled preterm infants of a minimum gestational age of 25 weeks or more, who were spontaneously breathing at 5 min of life (24). Therefore, their findings *do not* apply to all infants born at 25 weeks or more, but obviously more to those who were in apparently better status immediately after delivery. Furthermore, their findings *do not* apply at all to preterm infants below 25 weeks. In fact, in the systematic review of Schmolzer et al. comparing CPAP to intubation (usually plus surfactant), only one trial enrolled infants <25 weeks' gestation (SUPPORT) (25, 26). In this large trial, essentially all extremely preterm infants for whom informed consent had been obtained antenatally were enrolled. This is an important difference compared to the other trials included in this systematic review, where a more select population of preterm infants was enrolled. The critical nature of this potential source of bias is clearly demonstrated by Rich et al. who reported outcomes of all infants that were eligible for the SUPPORT trial but *were not enrolled* (27). Undoubtedly, essentially all meaningful outcomes were worse among those infants, signaling a clear selection bias, albeit smaller than in other trials of this systematic review.

## A PLAUSIBLE EXPLANATION FOR STATISTICAL ASSOCIATIONS

When interpreting the pooled results of a systematic review, we should not accept the results without considering some logical explanation behind them. An example of this point can

be made in relation to a recently published systematic review by King and colleagues (28). In this review, two interfaces to deliver nasal CPAP were compared and a total of seven studies met the inclusion criteria; however, only six of them were considered for the outcomes of nasal CPAP failure and bronchopulmonary dysplasia (BPD) (29–33). When we look at the pooled results for nasal CPAP failure within 72 h after initiation, we see a marginally significant result in favor of nasal mask vs. binasal prongs (Risk ratio 0.72, 95% CI 0.53–0.97) without considerable heterogeneity ( $I^2$  of 16%). What is more promising is the fact that there is a significant difference again in favor of the nasal mask interface with a reduction in moderate to severe BPD, this time with moderate heterogeneity ( $I^2$  of 30%). Nevertheless, if we try to find a plausible explanation for this difference based on better effectiveness and less failure with the nasal mask, the results do not support this. The study by Say et al. is the major contributor to the difference observed in moderate to severe BPD, but it shows no difference in the failure rate between the compared nasal CPAP interfaces (33). This strongly suggests that this observed association probably occurred by chance and is not related to the intervention.

## CONCLUSIONS

Systematic reviews are in great demand and remain a significant contribution for clinical decision-making and effectively provide an updated and informative perspective of the current state of the literature in a specific topic but their results should be interpreted with care. The Cochrane Library, which in many ways has set the standards for systematic reviews in therapeutic interventions, has not been always able to keep the published reviews updated with sufficient promptness,

thus creating a valid space for alternate versions of already published topics.

We have shown how well-designed studies can be equivocally combined in a meta-analysis and lead to biased summary point estimates of the effect. Heterogeneity among studies is a potential source of bias and may not always be detected by statistical tests. The latter aim to detect variability between study results but cannot detect relevant differences in design that could result in a meaningless conclusion from the combination of very different studies. This problem should be better described in the existing literature. Publication requirements for systematic reviews should be strengthened, following currently existing guidelines and undergo a rigorous peer-review process that considers some of the issues discussed previously. Clinicians should definitely be more aware of potential sources of bias when reading published systematic reviews to avoid being misled by only interpreting their conclusions.

## AUTHOR CONTRIBUTIONS

AM: substantial contributions to the conception, analysis, and interpretation of the work. Drafting and revising the manuscript critically for important intellectual content. Agrees to be accountable for all aspects for all aspects of the work in ensuring that questions related to the accuracy of any part are appropriately resolved. FM and SD: substantial contribution to the analysis and interpretation of the work. Drafting and revising the manuscript critically for important intellectual content.

## FUNDING

Funding for the presented work and publishing fees was provided by Clinica Alemana.

## REFERENCES

- Hall A, Walton G. Information overload within the health care system: a literature review. *Health Inf Libr J.* (2004) 21:102–8. doi: 10.1111/j.1471-1842.2004.00506.x
- Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med.* (2010) 7:e1000326. doi: 10.1371/journal.pmed.1000326
- Glass GV. Primary, secondary, and meta-analysis of research. *Educ Res.* (1976) 5:3–8. doi: 10.3102/0013189X005010003
- Lavis JN. How can we support the use of systematic reviews in policymaking? *PLoS Med.* (2009) 6:e1000141. doi: 10.1371/journal.pmed.1000141
- Murthy L, Shepperd S, Clarke MJ, Garner SE, Lavis JN, Perrier L, et al. Interventions to improve the use of systematic reviews in decision-making by health system managers, policy makers and clinicians. *Cochrane Database Syst Rev.* (2012) CD009401. doi: 10.1002/14651858.CD009401.pub2
- Eysenck HJ. Meta-analysis or best-evidence synthesis? *J Eval Clin Pract.* (1995) 1:29–36. doi: 10.1111/j.1365-2753.1995.tb00005.x
- Guyatt G, Rennie D, Meade MO, Cook DJ. *Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice.* Third Edition. New York, NY: McGraw-Hill Education. (2015).
- Ferguson KN, Roberts CT, Manley BJ, Davis PG. Interventions to improve rates of successful extubation in preterm infants: a systematic review and meta-analysis. *JAMA Pediatr.* (2017) 171:165–74. doi: 10.1001/jamapediatrics.2016.3015
- Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statist Med.* (1999) 18:2693–708.
- Borenstein M, Higgins JPT, Hedges LV, Rothstein HR. Basics of meta-analysis:  $I^2$  is not an absolute measure of heterogeneity. *Res Synth Methods.* (2017) 8:5–18. doi: 10.1002/jrsm.1230
- Yoder BA, Stoddard RA, Li M, King J, Dirnberger DR, Abbasi S. Heated, Humidified high-flow nasal cannula versus nasal CPAP for respiratory support in neonates. *Pediatrics.* (2013) 131:e1482–90. doi: 10.1542/peds.2012-2742
- Manley BJ, Owen LS, Doyle LW, Andersen CC, Cartwright DW, Pritchard MA, et al. High-flow nasal cannulae in very preterm infants after extubation. *N Engl J Med.* (2013) 369:1425–33. doi: 10.1056/NEJMoa1300071
- Collins CL, Holberton JR, Barfield C, Davis PG. A randomized controlled trial to compare heated humidified high-flow nasal cannulae with nasal continuous positive airway pressure postextubation in premature infants. *J Pediatr.* (2013) 162:949–54. doi: 10.1016/j.jpeds.2012.11.016
- Wu W, Shi Y, Li F, Wen Z, Liu H. Surfactant administration via a thin endotracheal catheter during spontaneous breathing in preterm infants: surfactant administration via a thin endotracheal catheter. *Pediatr Pulmonol.* (2017) 52:844–54. doi: 10.1002/ppul.23651
- Göpel W, Kribs A, Ziegler A, Laux R, Hoehn T, Wieg C, et al. Avoidance of mechanical ventilation by surfactant treatment of spontaneously breathing preterm infants (AMV): an open-label, randomized, controlled trial. *Lancet.* (2011) 378:1627–34. doi: 10.1016/S0140-6736(11)60986-0

16. Bao Y, Zhang G, Wu M, Ma L, Zhu J. A pilot study of less invasive surfactant administration in very preterm infants in a Chinese tertiary center. *BMC Pediatr.* (2015) 15:21–27. doi: 10.1186/s12887-015-0342-7
17. Kanmaz HG, Erdeve O, Canpolat FE, Mutlu B, Dilmen U. Surfactant administration via thin catheter during spontaneous breathing: randomized controlled trial. *Pediatrics.* (2013) 131:e502–9. doi: 10.1542/peds.2012-0603
18. Kribs A, Roll C, Göpel W, Wieg C, Groneck P, Laux R, et al. Nonintubated surfactant application vs conventional therapy in extremely preterm infants: a randomized clinical trial. *JAMA Pediatr.* (2015) 169:723–30. doi: 10.1001/jamapediatrics.2015.0504
19. Conte F, Orfeo L, Gizzi C, Massenzi L, Fasola S. Rapid systematic review shows that using a high-flow nasal cannula is inferior to nasal continuous positive airway pressure as first-line support in preterm neonates. *Acta Paediatr.* (2018) 107:1684–96. doi: 10.1111/apa.14396
20. Lavizzari A, Veneroni C, Colnaghi M, Ciuffini F, Zannin E, Fumagalli M, et al. Respiratory mechanics during NCPAP and HHFNC at equal distending pressures. *Arch Dis Child - Fetal Neonatal Ed.* (2014) 99:F315–20. doi: 10.1136/archdischild-2013-305855
21. Nair, G, Karna, P. Comparison of the effects of vapotherm and nasal CPAP in respiratory distress in preterm infants. *Pediatr Acad Soc Annu Meet.* (2005) 57:2054.
22. Roberts CT, Owen LS, Manley BJ, Frøisland DH, Donath SM, Dalziel KM, et al. nasal high-flow therapy for primary respiratory support in preterm infants. *N Engl J Med.* (2016) 375:1142–51. doi: 10.1056/NEJMoa1603694
23. Shin J, Park K, Lee EH, Choi BM. Humidified high flow nasal cannula versus nasal continuous positive airway pressure as an initial respiratory support in preterm infants with respiratory distress: a randomized, controlled non-inferiority trial. *J Korean Med Sci.* (2017) 32:650–55. doi: 10.3346/jkms.2017.32.4.655
24. Morley CJ, Davis PG, Doyle LW, Brion LP, Hascoet JM, Carlin Jb et al. Nasal CPAP or intubation at birth for very preterm infants. *N Engl J Med.* (2008) 358:700–8. doi: 10.1056/NEJMoa072788
25. Schmolzer GM, Kumar M, Pichler G, Aziz K, O'Reilly M, Cheung PY. Non-invasive versus invasive respiratory support in preterm infants at birth: systematic review and meta-analysis. *BMJ.* 347:f5980. doi: 10.1136/bmj.f5980
26. SUPPORT Study Group. Early CPAP versus surfactant in extremely preterm infants. *N Engl J Med.* (2010) 362:1970–9. doi: 10.1056/NEJMoa0911783
27. Rich W, Finer NN, Gantz MG, Newman NS, Hensman AM, Hale EC, et al. Enrollment of extremely low birth weight infants in a clinical research study may not be representative. *Pediatrics.* (2012) 129:480–4. doi: 10.1542/peds.2011-2121
28. King BC, Gandhi BB, Jackson A, Katakam L, Pammi M, Suresh G. Mask versus prongs for nasal continuous positive airway pressure in preterm infants: a systematic review and meta-analysis. *Neonatology.* (2019) 4:1–15. doi: 10.1159/000496462
29. Bashir T, Murki S, Kiran S, Reddy VK, Oleti TP. 'Nasal mask' in comparison with 'nasal prongs' or 'rotation of nasal mask with nasal prongs' reduce the incidence of nasal injury in preterm neonates supported on nasal continuous positive airway pressure (nCPAP): a randomized controlled trial. *PLoS ONE.* (2019) 14:e0211476. doi: 10.1371/journal.pone.0211476
30. Chandrasekaran A, Thukral A, Jeeva Sankar M, Agarwal R, Paul VK, Deorari AK. Nasal masks or binasal prongs for delivering continuous positive airway pressure in preterm neonates—a randomised trial. *Eur J Pediatr.* (2017) 176:379–86. doi: 10.1007/s00431-017-2851-x
31. Goel S, Mondkar J, Panchal H, Hegde D, Utture A, Manerkar S. Nasal mask versus nasal prongs for delivering nasal continuous positive airway pressure in preterm infants with respiratory distress: a randomized controlled trial. *Indian Pediatr.* (2015) 52:1035–40. doi: 10.1007/s13312-015-0769-9
32. Kieran EA, Twomey AR, Molloy EJ, Murphy JFA, O'Donnell CPF. Randomized trial of prongs or mask for nasal continuous positive airway pressure in preterm infants. *Pediatrics.* (2012) 130:e1170–6. doi: 10.1542/peds.2011-3548
33. Say B, Kanmaz Kutman HG, Oguz SS, Oncel MY, Arayici S, Canpolat FE, et al. Binasal prong versus nasal mask for applying CPAP to preterm infants: a randomized controlled trial. *Neonatology.* (2016) 109:258–64. doi: 10.1159/000443263

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Maturana, Moya and Donn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.