Use of Composite Outcomes in Neonatal Trials: An Analysis of the Cochrane Reviews

Nai Ming Lai[a, b]   Amanda Qiao Ying Yap[c]   Hwee Chin Ong[d]   Sheng Xuan Wai[e]
Julie Hsiao Hui Yeo[f]   Charis Yen Ee Koo[g]   Wen Chin Lah[d]   Yin Sear Lim[a]
Colleen Ovelman[h]   Roger F. Soll[h, i]

[a]School of Medicine, Faculty of Health and Medical Sciences, Taylor's University, Subang Jaya, Malaysia; [b]School of Pharmacy, Monash University, Subang Jaya, Malaysia; [c]Hospital Putrajaya, Ministry of Health, Putrajaya, Malaysia; [d]Hospital Tawau, Ministry of Health, Tawau, Malaysia; [e]University of Malaya Medical Centre, Kuala Lumpur, Malaysia; [f]Hospital Sultanah Bahiyah, Ministry of Health, Alor Setar, Malaysia; [g]Hospital Tuanku Jaafar, Ministry of Health, Seremban, Malaysia; [h]Cochrane Neonatal, Burlington, VT, USA; [i]Division of Neonatal-Perinatal Medicine, The University of Vermont Larner College of Medicine, Burlington, VT, USA

Abstract

Introduction: Composite outcomes are used to increase the power of a study by combining event rates. Many composite outcomes in adult clinical trials have components that differ substantially in patient importance, event rate, and effect size, making interpretation challenging. Little is known about the use of composite outcomes in neonatal randomized controlled trials (RCTs). ***Methods:*** We assessed the use of composite outcomes in neonatal RCTs included in Cochrane Neonatal reviews published till November 2017. Two authors reviewed the components of the composite outcomes to compare their patient importance and computed the ratios of effect sizes and event rates between the components, with an a priori threshold of 1.5, indicating a substantial difference. Descriptive statistics were presented. ***Re-*** ***sults:*** We extracted 7,766 outcomes in 2,134 RCTs in 312 systematic reviews. Among them, 55 composite outcomes (0.7%) were identified in 46 RCTs. The vast majority (92.7%) of composite outcomes had 2 components, with death being the most common component (included 51 times [92.7%]). The components in nearly three-quarters of the composite outcomes ($n = 40$ [72.7%]) had different patient importance, while the effect sizes and event rates differed substantially between the components in 27 (49.1%) and 35 (63.6%) outcomes, respectively, with up to 43-fold difference in the event rates observed. ***Conclusions:*** The majority of composite outcomes in neonatal RCTs had different patient importance with contrasting effect sizes and event rates between the components. In patient communication, clinicians should highlight individual components, rather than the composites, with explanation on the relationship between the components, to avoid misleading impression on the effect of the intervention. Future trials should report the estimates of all individual components alongside the composite outcomes presented.

© 2021 S. Karger AG, Basel

Nai Ming Lai
School of Medicine, Faculty of Health and Medical SciencesTaylor's University
Selangor Darul Ehsan
Subang Jaya 47500 (Malaysia)
lainm@doctors.org.uk

Background

Improving medical care has resulted in decreasing mortality and morbidities. Progressively, newer medical interventions have diminishing superiority over existing treatments. To show statistically significant results in clinical trials, increasingly large samples of populations are required [1]. A convincing sample size estimation is a requirement in the approval and funding of a clinical trial. To increase the chance of a statistically significant result with a manageable sample, trialists may resort to using composite outcomes [2].

A composite outcome refers to an outcome with multiple distinct components [3]. The use of composite outcomes increases the likelihood of timely completion of a trial, by increasing the event rates and power of analysis, hence reducing the sample required to achieve a statistically significant result [3]. Additionally, many researchers believe that reporting a composite outcome reduces the amount of information for patients, by reducing the number of endpoints for decision-making, thus simplifying the communication process. Composite outcomes are considered useful in communicating the "net effect" of an intervention, as represented by outcomes such as event-free survival [1].

Composite outcomes are also used to address the issue of competing risks [4]. Competing risks arise when the occurrence of one outcome precludes the occurrence of another, usually less serious outcome. This might lead to an apparent decrease in the incidence of the outcome concerned, creating a false impression on the usefulness of the intervention [4].

Reviews of adult randomized controlled trials (RCTs) have identified the following problems related to inappropriate use and reporting of composite outcomes: poor definition, discordant patient importance, effect sizes and event rates between components, predominant contribution of the component with lesser importance to patient care, and a lack of reporting on the estimates of the individual components [5–8]. These problems may mislead healthcare decision-makers and patients by giving a false impression on the usefulness of the intervention. It has been recommended that composite outcomes should only be reported if the components are comparable in patient importance, effect sizes, and event rates [2, 3]. It is unclear whether composite outcomes in neonatal RCTs were incorporated appropriately based on the aforementioned criteria.

Objectives

We determined the proportion and characteristics of composite outcomes in neonatal RCTs with the following objectives: to assess the proportion of patients who were presented appropriately, based on similarity in patient importance, effect size, and event rates; to assess the proportion with components of truly competing risk; and to assess the number of individual components that were reported alongside the corresponding composite outcomes.

Methods

We conducted an analysis of the RCTs included in the Cochrane Neonatal reviews (up to issue 11 of 2017). We extracted composite outcomes from the "Characteristics of included studies" table. A summary of our methods is provided in the following text, with a more detailed report of the methods provided in online supplement 1 (for all online supplementary material, see www.karger.com/doi/10.1159/000514402).

We determine the comparability of patient importance of the components in independent duplicates and calculated the ratio of the effect sizes (relative risk) and event rates between the components, with a threshold of 1.5 adopted a priori to denote a substantial difference between the components. We also determined whether the components in the composite were appropriately considered as outcomes with competing risk, using principles outlined in previous publications [4, 9]. Additionally, we assessed the number of components that were highlighted individually alongside the composite outcomes, by inspecting the primary outcomes reported by the authors from the Characteristics of Included studies tables, supplemented by an assessment of the full text of the individual studies. We reported our findings using standard descriptive statistics (SPSS 23, Chicago, IL, USA). This study was approved by Cochrane Neonatal and used only published data from Cochrane Neonatal reviews and individual published studies.

Results

We identified a total of 7,766 outcomes in 2,134 RCTs in 312 reviews. There were 55 composite outcomes (0.7%) identified in 46 RCTs. Thirty-two of these RCTs included only preterm infants, 4 included term and late-preterm infants, and 10 included infants of all gestations. There were 2 components in 51 composite outcomes, and in the remaining 4 outcomes, 2 included 3 components and 2 included 4 components. Mortality, measured at various time points, was included most often, in 51 of 55 outcomes. Apart from mortality, there were 14 other different outcome domains, as displayed in online supplement 2. Bronchopulmonary dysplasia or chronic lung disease

Lai/Yap/Ong/Wai/Yeo/Koo/Lah/Lim/
Ovelman/Soll

was most commonly paired with mortality (24 times), followed by neurodevelopmental disability (9 times) and "brain abnormality" on cranial ultrasound (excluding intraventricular hemorrhage) (8 times).

Comparability of Patient Importance, Effect Sizes, and Event Rates

There was fair interrater agreement on the similarity of patient importance between the components (Cohen's kappa: 0.68). Based on our consensus, 27.3% of the composite outcomes (n = 15) had components of similar patient importance, and among these, mortality was a component in 13 outcomes.

Data on effect sizes and overall event rates of the components were reported in 53 outcomes (112 components) in 44 RCTs. Half of the components (27, 49.1%) differed substantially in effect sizes (by up to 9 folds), and two-thirds (35, 63.6%) differed substantially in event rates (by up to 43-folds).

The components had opposite effects in 25 of 55 outcomes (45.5%). In 51 outcomes where mortality was a component, the paired component had an opposite effect to that of mortality in 24 outcomes (47.1%). The overall effect of the composite outcome was opposite to that of mortality in 15 outcomes (29.4%), with 10 showing an overall reduction in the risk, while mortality was actually increased.

Combination on the Basis of Competing Risk

We considered 40 outcomes (72.7%) as having truly competing risk between the components. Of these, components in 11 outcomes (27.5%) were judged to have similar patient importance. On the other hand, among the 15 composite outcomes that were not considered outcomes with competing risks, 10 of them (66.7%) had components with clearly different patient importance.

Proportion of Individual Components Highlighted

A single component was presented alongside the composite outcome in 16 occasions (29.1%); both components were presented in 13 occasions (23.6%), and all 3 components were presented in one occasion (1.8%). Of 51 composite outcomes that included mortality, mortality was highlighted as an individual component 17 times (33.3%). A detailed list of each composite outcome with its components, effect sizes, overall event rates, rating of similarity in patient importance, and rating of appropriateness of the components as outcomes with competing risk is provided in online supplement 3.

Discussion

The proportion of adult RCTs that included at least one composite outcome ranged approximately from 37 [7] to 47% [6]. In comparison, much fewer neonatal RCTs included composite outcomes. However, like adult trials [5, 7, 8], the majority of the composite outcomes in neonatal RCTs included mortality as a component, while the other components included mostly lung, brain, or developmental conditions that were specific to the neonatal populations. Our main finding, similar to reports from studies on adults, is the high proportion of composite outcomes in neonatal RCTs with contrasting patient importance, effect sizes, and event rates.

The overall effect of the composite was opposite to that of the more important component such as mortality in half of the cases, some showing an overall reduction in mortality or morbidity, while mortality was actually increased. While mortality might have been included as a censoring outcome, the overall findings might confuse lay readers whose focus is on mortality. This reaffirms the concerns that composite outcomes might lend itself to injudicious interpretation, making the intervention appear more useful than it actually is, or even disguising its harmful effect on the outcomes with high patient importance such as mortality [4, 5].

Many clinical trials on adult patients in cardiology, endocrinology, and oncology include composite outcomes on the basis of competing risk [4, 10]. In neonatal research, competing risk is perhaps an even greater issue because death in a newborn is usually considered premature in terms of timing. Statistically, a newborn's death distorts the representation of all conditions that manifest or are assessed in the medium or longer term. Therefore, death is widely seen as a "censoring variable" in the evaluation of long-term outcomes for neonates. One common way of addressing the censoring effect of death is to use a composite of death and another outcome. However, the concern of this approach is the potential failure to predict either death or another outcome [11]. A notable example on the use of composite outcomes in neonatology is the NeOProM Collaboration [12]. In this collaboration, data from 5 RCTs comparing higher (91–95%) versus lower (85–89%) oxygen saturation targets in preterm infants were prospectively combined in a meta-analysis. The primary outcome of the meta-analysis was a composite of death or major disability, which included blindness, deafness, and cerebral palsy, assessed at a corrected age of 18–24 months. There was no difference between groups in this composite outcome, but more in-

fants who received oxygen within the lower target range died. This example shows that although the composite outcome was chosen on the appropriate basis, such results might create confusion in decision-making, especially to caregivers who view death differently from all the other outcomes. Commendably, the authors of this study made available the results of all individual components as secondary outcomes to avoid such confusion. Opinions vary widely on how clinical outcomes should be combined based on competing risk. Manja et al. [4] recommended using composite outcomes on this basis if competing risk is plausible and occurs at sufficiently high frequency to affect interpretation of the outcome of interest. Tai et al. [10] and Lubsen and Kirwan [9] favored statistical adjustments on the outcomes concerned over the use of composite outcomes to account for competing risk. In our study, although nearly two-thirds of the composite outcomes were appropriately chosen on the basis of competing risks, there remains a challenge in getting an accurate message to caregivers about the major effect of the intervention because the vast majority had marked discrepancies in patient importance, with no individual components reported alongside the composite outcomes.

Limitations

First, we did not include consumers in judging the comparability of patient importance, especially on the impact of the component outcomes on life as there might be a difference in the perception between a patient or a caregiver and a medically trained person. Next, we relied on the information in the Characteristics of Included Studies table of the Cochrane reviews. Despite the rigor in the development of Cochrane reviews with clear expectations on the quality of the data [13], we could not rule out a possibility that information in the table were inaccurate as we only cross-checked in some but not all the full texts of over 2,000 individual studies, which was practically impossible. Consequently, we might have underestimated the number of composite outcomes reported.

The information presented in the Characteristics of Included Studies table were insufficient for us to determine whether the composite outcomes were designated as primary or secondary outcomes. Because we did not assess the full texts of all included trials, we were unable to determine if the use of composite outcome is related to trial sample size and whether sample size calculation of each trial, had it been conducted, was based on the composite outcome. We therefore could not verify the statement that a major reason for the use of composite outcome is to manage sample size. However, we believe that even if we had examined all full texts, we still would not have acquired sufficient information for this purpose as sample size calculation strategies were often poorly reported in published trials [14, 15]. Due to the limitations on our data, we were unable to evaluate other potentially valuable research questions on the use of composite outcomes, such as its relationship to methodological rigor, including the possibility of reporting bias, especially in the case where individual components were not fully reported.

To determine what constituted a substantial difference in effect sizes and event rates, we adopted a threshold of 1.5 based on the ratio of one figure over another, rather than a series of absolute differences, as adopted by Montori et al. [2]. Our approach increases sensitivity in capturing relatively important differences in the effect sizes/event rates when they were low, at the expense of missing important differences in high effect sizes/event rates, for instance, in relative risks greater than 1.5 and event rates higher than 50%.

Last, despite having approval from Cochrane Neonatal and using only published data, we did not register our study protocol with an established trial registry. This could have affected readers' perception on the credibility of our findings.

Conclusions

A small proportion of neonatal RCTs included composite outcomes, but many had components with discordant patient importance, effect sizes, and event rates. This suggests that while the use of composite outcomes may be statistically justifiable, care should be taken to communicate their effects appropriately. We strongly recommend trialists to report all individual components alongside the composite outcomes in future trials to facilitate accurate communication between clinicians and caregivers.

Statement of Ethics

The paper is exempt from ethical committee approval as it used only published data. The study was approved by Cochrane Neonatal.

Conflict of Interest Statement

All authors have indicated they have no potential conflicts of interest to disclose.

References

1 Ferreira-González I, Permanyer-Miralda G, Busse JW, Bryant DM, Montori VM, Alonso-Coello P, et al. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. J Clin Epidemiol. 2007;60(7):651–62.

2 Montori VM, Permanyer-Miralda G, Ferreira-González I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite end points in clinical trials. BMJ. 2005;330(7491):594.

3 McCoy C. Understanding the use of composite endpoints in clinical trials. West J Emerg Med. 2018;19(4):631–4.

4 Manja V, AlBashir S, Guyatt G. Criteria for use of composite end points for competing risks-a systematic survey of the literature with recommendations. J Clin Epidemiol. 2017;82:4–11.

5 Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. BMJ. 2010;341:c3920.

6 Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? JAMA. 2003;289(19):2554–9.

7 Lim E, Brown A, Helmy A, Mussa S, Altman DG. Composite outcomes in cardiovascular research: a survey of randomized trials. Ann Intern Med. 2008;149(9):612–7.

8 Ferreira-González I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. BMJ. 2007;334(7597):786.

9 Lubsen J, Kirwan BA. Combined endpoints: can we use them? Stat Med. 2002;21(19):2959–70.

10 Tai BC, Wee J, Machin D. Analysis and design of randomised clinical trials involving competing risks endpoints. Trials. 2011;12:127.

11 Crilly CJ, Haneuse S, Litt JS. Predicting the outcomes of preterm neonates beyond the neonatal intensive care unit: what are we missing? Pediatric Res. 2020 May:1–20.

12 Askie LM, Darlow BA, Finer N, Schmidt B, Stenson B, Tarnow-Mordi W, et al. Association between oxygen saturation targeting and death or disability in extremely preterm infants in the neonatal oxygenation prospective meta-analysis collaboration. JAMA. 2018;319(21):2190–201.

13 Churchill RLT, Chandler J, Tovey D, Thomas J, Flemyng E, Higgins JPT. Standards for the reporting of new Cochrane intervention reviews. Methodological expectations of cochrane intervention reviews [Internet]. London: Cochrane; 2019 Oct.

14 Nikolakopoulos S, Roes KC, van der Lee JH, van der Tweel I. Sample size calculations in pediatric clinical trials conducted in an ICU: a systematic review. Trials. 2014;15:274.

15 Abdul Latif L, Daud Amadera JE, Pimentel D, Pimentel T, Fregni F. Sample size calculation in physical medicine and rehabilitation: a systematic review of reporting, characteristics, and results in randomized controlled trials. Arch Phys Med Rehabil. 2011;92(2):306–15.