Seminars in Fetal & Neonatal Medicine 20 (2015) 378–383

Contents lists available at ScienceDirect

Seminars in Fetal & Neonatal Medicine

journal homepage: www.elsevier.com/locate/siny

Review

How to find and how to read articles in neonatology

Keith J. Barrington*

CHU Sainte Justine, Université de Montréal, Montréal, Québec, Canada

Keywords:
Electronic tables of contents
Critical evaluation
Literature review

SUMMARY

Staying abreast of the neonatal literature is an important task. Being aware of new information and knowing how to evaluate its reliability remain essential to be able to provide the most appropriate, evidence-based, therapy to our patients. This article discusses methods for being informed of, and critically reviewing, published research in order to fulfill these tasks without being overwhelmed by the number or complexity of publications.

1. Introduction

The medical literature is enormous; thousands of new articles are published every week. How to stay on top of this literature, to interpret the value of articles published and whether they should materially change the treatment of our patients is an ongoing challenge. Fortunately the field of neonatology is relatively limited, and articles that are important, and should be therapy-changing, are sparse. The question remains, how to ensure that you are alerted when such an article appears, and how to ensure that the data are valid and worthy of a change in practice.

This article will be a list of principles and recommendations based on my own practice and experience. There are other sources of articles which take a step-by-step approach to evaluation of the medical literature. I highly recommend the *JAMA* series on users' guides to the medical literature which are all available on the Johns Hopkins website (www.hopkinsmedicine.org/gim/training/Osler/osler_JAMA_Steps.html). In this single article I will point out some important principles, and provide pointers to other resources.

2. Ongoing surveillance of the literature

I survey the literature every few days to see if anything new, and worthy of significant attention has been published. There are two complementary ways of doing this.

The first is E Table of Contents (ToC): to be automatically informed of new publications, sign up for e-mail reception of ToC from the major journals that include neonatal articles. Some very high quality journals almost never have neonatology (such as *Annals of Internal Medicine*) so don't bother with those. Some occasionally do, and they tend to be either of high quality, or very controversial; these include *JAMA*, the *New England Journal of Medicine* and *The Lancet*; the *BMJ* very rarely publishes neonatal articles, but they have extremely high editorial and peer review standards.

Most of these journals give the option of sending you an e-mail when they have newly accepted articles available online, in addition to the table of contents of the monthly (or weekly) printed version.

The second is saved searches. There are several services that will, for free, send you an e-mail with links to new neonatal articles on a regular basis. One example is Amedeo, which will e-mail each week a filtered list of articles, pulled from a number of journal eToC. They have a 'neonatology' option (www.amedeo.com/medicine/neo.htm), which will filter out many non-neonatal articles from the ToC of the journals. The weekly e-mail can be opened within PubMed (www.pubmed.gov) which then provides links to the original articles.

"My NCBI" is another source, which is also freely available from the PubMed website (www.ncbi.nlm.nih.gov/myncbi). If you sign in to the site, you can create and save searches which will then be performed at regular intervals, and the results e-mailed to you. I currently have four searches, three of which run every week: a search for clinical trials in newborns, a search for studies of hypotension in preterms, another for pain in the newborn, and one which runs monthly for clinical trials of inhaled nitric oxide. Both of these complementary services are 'spam free' apart from an occasional advert for a medical textbook from Amedeo.

In addition to regular searches to remain up to date, the need for searches to answer a specific clinical question is important. The

* Address: CHU Sainte Justine, 3175 Cote Ste Catherine, Montreal, Quebec H3T 1C5, Canada.
E-mail address: Keith.barrington@umontreal.ca.

http://dx.doi.org/10.1016/j.siny.2015.09.005
1744-165X/

search engines of PubMed have improved progressively over the years, and very few articles of importance are not included in the National Library of Medicine Database. I rarely search other databases, and then usually when preparing a systematic review for publication.

For studies of therapy, you can select a filter for "clinical trial" from the web page; this will eliminate large numbers of items from an initial search; other readily available filters which are valuable are "humans" which can be added to "newborn" in order to eliminate veterinary studies and those that only address older subjects.

Of course the majority of the articles that you obtain from any of these searches are not of therapy-changing consequence. Selecting the useful articles can then begin.

3. The title

The title is a major source of information. Does the article address an issue of interest? Does it appear to be a good quality source of information? It may be immediately obvious that the article is not in neonatology, or is a case report of no personal interest, and it can be deleted.

4. The abstract

The abstract can usually be easily accessed; hopefully it will become clear whether there is a concurrent control group, an adequate sample size, and enough details of methodology to know whether further reading is likely to be worthwhile. The following paragraphs detail some considerations for analysis of a published article.

5. The journal in which the article is published

Articles published in the highest profile journals are not all worthy of much consideration; they need to be evaluated using the same standards as articles in other journals. As one example, a randomized controlled trial (RCT) of steroids for the prevention of bronchopulmonary dysplasia was published in the *New England Journal of Medicine* [1], perhaps suggesting ground-breaking findings. However, that study included a total of only 26 patients in three different groups. It showed a significant difference in the primary outcome variable (death from respiratory failure or continued assisted ventilation at 60 days of age) which should have been considered to be an intriguing finding, worthy of an adequately powered trial. Instead the prolonged 42-day dose regime was adopted around the world.

A second principle is that there are many extremely low-profile journals that are very unlikely to publish high-quality articles worthy of consideration. If you read about a RCT that was submitted and published in the "New journal of neonatal–perinatal and pediatric medicine and nursing" – beware! Many such journals are 'predatory', meaning that they provide online publication at a price, with no real peer review.

Between those two extremes, there are many journals which publish worthwhile articles. Most of those which are likely to be important in changing clinical practice are listed in Box 1, which is not intended to be an exclusive list, and may need to be enlarged as time passes. How to evaluate the profile of a journal and the likely quality of the articles published is not simple. I suggest that you ignore impact factors. Impact factors are a simplistic calculation of how many times articles published in a journal are referenced by other publications within a couple of years of publication. But even in very high profile journals, impact factors are only produced by a minority of the published articles, i.e., those which happen to attract a lot of short-term interest. Also, journals which publish a lot of review articles may have high impact factors, as they receive citations in other articles, but they may publish very few articles of therapy-changing importance.

Box 1
Journals with numerous neonatal articles and electronic tables of contents available.

Pediatrics
Journal of Pediatrics
Acta Paediatrica
Neonatology
Journal of Perinatology
Archives of Diseases in Childhood – Fetal and Neonatal Edition
American Journal of Perinatology
Journal of Paediatrics and Child Health
JAMA – Pediatrics
European Journal of Pediatrics

6. Uncontrolled trials are unreliable

Any study which reports the outcome of a group of patients without a control group is suspect. In cardiovascular support, for example, there are numerous examples of reports showing a change in a particular variable, such as renal Doppler findings after a drug infusion that was started [2] in preterm infants on the first day of life. Renal perfusion is known to increase enormously during the first day of life, so without a control group is it unclear whether any changes seen are due to the intervention. Changes after an intervention are of uncertain significance if there is no comparison control group to reveal whether the changes may have occurred in any case – especially, but not exclusively, when the variable being evaluated may change spontaneously or is known to vary in predictable, or unpredictable ways.

Of great importance for uncontrolled trials is the phenomenon known as regression to the mean, whereby extreme observations, when repeated, usually become less extreme. Regression to the mean is a major component of the placebo effect, probably the major component. Placebo effects are important in neonatology as in other areas of medicine. Only high-quality randomized controls can correctly attribute changes after an intervention to that intervention.

Another example in neonatology is a study of using a pungent aroma to treat apnea of prematurity [3], which showed an apparent effect. Although the possibility of this being effective is intriguing, the study had no controls. Apnea of prematurity is a highly variable condition, with major variations in the number of apneas experienced by preterm infants. Infants are likely to be enrolled in apnea trials when they are having more apneas than average. In other words, control infants without any intervention are also likely to show a reduction in apnea, simply with prolonged observation. Randomized controls would have demonstrated whether the effects of the intervention were due to a real impact on apnea, or were a placebo effect.

7. Beware of review articles

All review articles should be based on systematic review and synthesis of the literature. Many review articles and book chapters are selective and unreliable, reflecting the author's own preoccupations and a biased interpretation of the published literature. A review article should therefore clearly state how the reference list

was assembled, which references are RCTs and systematic reviews, and what attempts were made to ensure that the review is balanced and complete.

8. The hierarchy of evidence

The idea of a hierarchy of reliable evidence came about as a result of the evidence-based medicine movement. The need to evaluate and summarize published evidence was recognized, and various pyramids have been drawn and published. At the top of such pyramids is the systematic review, a method for analyzing and summarizing all the reliable evidence from RCTs.

9. Systematic reviews

Several provisos must be considered when determining whether an individual systematic review should be considered to give a reliable answer about the clinical question being asked.

The most important of these are the size and quality of the component RCTs, and the possibility of publication bias. Numerous small RCTs can inflate the apparent benefit (or harm) of an intervention. This is particularly the case when negative trials may not be published, either from failure to submit, or rejection of an article by a journal because a negative trial is perceived as being less interesting.

As one example, a single, large, high-quality RCT was enough to overturn the results of the previous Cochrane review of the effectiveness of intravenous gammaglobulin in the treatment of newborn infants with suspected sepsis. Eight small RCTs prior to the International Neonatal Immunotherapy Study (INIS) had shown a survival benefit; the quality and especially the size of those previous trials was known to be inadequate [4]. A single, very large trial ($n = 3800$) was clearly negative; no benefit of any kind was demonstrated [5]. One could be comfortable that the initial trials had not shown evidence of harm, which was also confirmed by INIS. I think it is unlikely that another trial of that size or quality will ever be performed to address this issue. In my opinion the RCT is a better source of reliable information than the updated Cochrane systematic review which now includes its results [6].

10. Systematic reviews are only reliable if they are built on reliable data

There are many systematic reviews, even in the best sources, such as the Cochrane Database, which are based on one or a few small trials of limited research quality. A systematic review based on a few low-quality small trials may have misleading results, and may be overturned by a single high-quality trial. Systematic reviews are also at risk of bias; to be reliable they should be performed according to a predefined protocol, and any deviations from that protocol should be clearly explained. A series of questions, adapted from Crowther and Cook [7] can help to evaluate the validity (and applicability) of a systematic review (Box 2).

11. Randomized controlled trials

Randomized controlled trials are clearly the bedrock of clinical evidence, and numerous tables have been constructed to evaluate the quality of each trial. A checklist which can be used to evaluate the reliability of both randomized and non-randomized health care interventions has been published and validated [8]. The 27-item checklist is simple to complete and gives a good overview of the methodologic quality of a study. A shorter checklist is freely available and downloadable from the Scottish Intercollegiate Guidelines Network (a revised version of which is included here as Box 3): for each item an indication of "yes", "no" or "uncertain" (or for the final item, "does not apply") can be marked. That network also has other checklists available for other study types (www.sign.ac.uk/methodology/checklists.html). The CONSORT statement includes a checklist of items that should be included in any report in order to be able to assess how the trial was performed [9]. Although many journals have subscribed to the CONSORT standards, RCT reports in those journals are often missing critical items. The CONSORT statement is intended as a checklist for items that authors should include in their study reports, not as a way to evaluate study quality.

Box 2
Question used to evaluate the usefulness of a systematic review.

- Did the systematic review address a focused clinical question?
- Were the criteria used to select articles for inclusion both defined and appropriate?
- Does it seem likely that relevant studies were missed?
- Was the quality of the included studies assessed?
- Were the assessments reproducible?
- Were the study-to-study results consistent?
- How precise were the results of the review?
- Were all clinically important outcomes evaluated?

In assessing the value of the review, it is important to consider the following questions:

- Can the results be applied to my patients, and will the results help me care for my patients?
- Are the benefits worth the harms and costs?

Box 3
Characteristics of a well-conducted randomized controlled trial.

- The study addresses an appropriate and clearly focused question.
- The assignment of subjects to treatment groups is randomized.
- An adequate concealment method is used to mask the group allocation.
- Subjects and investigators are kept 'blind' about treatment allocation.
- The treatment and control groups are similar at the start of the trial.
- The only difference between groups is the treatment under investigation.
- All relevant outcomes are measured in a standard, valid and reliable way.
- A low percentage of the individuals or clusters recruited into each treatment arm of the study dropped out before the study was completed.
- All the subjects are analysed in the groups to which they were randomly allocated (often referred to as intention-to-treat analysis).
- Where the study is carried out at more than one site, results are comparable for all sites.

11.1. Important characteristics to evaluate

11.1.1. Method of enrollment, and the proportions enrolled

In order to know whether the results of an RCT are applicable to your practice it is essential to know the characteristics of the population that was screened, and the sample that was enrolled. Sometimes there are insufficient data to know whether a sample is representative of the types of babies that you see in your practice.

11.1.2. Masking of group allocation

When a patient is entered into the trial, the group to which the patient will be allocated should not be known. Pseudo-randomized trials (group assignment based on date or hospital record number for example) are too easily "gamed" and should be treated with suspicion. Although masking an intervention is not always possible, masking of allocation is always possible; failure to do so should raise suspicion. Description of enrollment practices is sometimes inadequate to be able to evaluate whether the allocation was masked.

11.1.3. Masking of the intervention

Although sometimes presented as the essential component of a randomized trial, "double-blinding" is sometimes not possible or may be impractical. Studies comparing ventilator strategies, for example, over several days or weeks, may require that the clinical staff be aware of the mode of ventilation in order for the study to be safe. Some drugs – muscle relaxants for example – may have effects that are clinically obvious. Many studies of surfactant were performed without masking the intervention, because of the expense involved in having a separate team of individuals available to administer the surfactant (or placebo) behind a screen. Nevertheless the benefits of surfactant were evident, and the size of the benefit was similar between masked and unmasked trials. Although meta-epidemiologic studies have demonstrated that failure to adequately blind an intervention affects the estimates of treatment effect, the impact appears to be much greater for subjective outcomes than objective outcomes, especially when the subjective outcome is measured on a continuous scale [10]. Thus, death, positive blood cultures and other objective outcomes are much less likely to be influenced by knowledge of the assigned group. Outcomes which may be more subjective, such as the need for oxygen at 36 weeks, can be rendered more objective, e.g. by using oxygen withdrawal testing with strict criteria for interpretation [11]. However, other co-interventions may be affected in non-masked trials, and there remains a chance that even objective outcomes may be affected, although there is little direct evidence for this [12].

11.1.4. Balancing of the groups

The best way to ensure that a difference in outcomes after an intervention is due to the intervention is to compare groups which are identical in all other respects. In real life this cannot be achieved, but similarity of groups will be improved with masked truly random allocation, and large samples. Some improvement in the matching of groups will be achieved with the use of stratified group allocation, or by other techniques such as minimization.

11.1.5. Reliable evaluation of outcomes

Important outcomes of RCTs should be reliably assessed. Most neonatal outcomes of interest can be evaluated reliably, but there are important limitations in some of the important outcomes that we rely on. Stage 2 necrotizing enterocolitis, for example, is diagnosed based on clinical signs and the presence of pneumatosis intestinalis on abdominal radiography. However, there is a great deal of inter-observer variability in interpretation of such radiographs. Head ultrasound abnormalities scored using the Papile system are also open to interpretation and subject to the limitations of the system, which lumps together findings with differing long-term implications.

11.1.6. Complete assessment of subjects for the outcomes

Short-term outcomes in neonatology are usually marked by almost complete evaluation of the groups, with little or no loss to follow-up. Once infants are discharged home, the loss to follow-up and the attendant risks to study validity progressively increase. There is no specific threshold beyond which studies can be classed as unreliable, but sensitivity analyses of how loss to follow-up might affect a trial's conclusions should be performed, especially if more than 10% are not evaluated for the longer-term outcomes that are so important in determining the clinical importance of an intervention.

11.1.7. Intention-to-treat analyses

In order to determine real-world effectiveness of an intervention, all subjects allocated to each group should be evaluated according to the groups to which they were assigned. There is a risk of inflating the impact of an intervention by eliminating patients from the analysis even though they completed the trial, or by re-assigning them to the "treatment actually received" group. It is important to not take claims of intention-to-treat analysis at face value: trials may be reported by so-called intention-to-treat analyses which exclude substantial numbers of randomized infants. An RCT comparing two breast-milk fortification powders, for example [13] eliminated about 20% of enrolled patients (who had already commenced the study intervention) from the dataset before performing an "intention-to-treat" analysis. Analyses by intervention actually received, or after elimination of some patients, may be of some interest to determine the potential physiological effects of a treatment, but such analyses can only be considered to be exploratory.

12. Clinically important outcomes

Outcomes that are important to infants and their families should be the primary outcomes of interest for clinical trials.

13. Secondary outcomes and subgroup analyses

The issue of the reliability and importance of secondary outcomes is thorny, but important. It is widely held that secondary outcomes can only be hypothesis generating, not hypothesis confirming. In a strict sense this is true, as a trial is usually designed to answer one question with sufficient power, and with all the procedures aimed at answering that one question. Any data from two groups of patients will have some significant differences between groups if you look hard enough and long enough. Secondary outcomes examined after inspection of the data are particularly suspect. Only those that are pre-specified should be given credence, and then with caution. Secondary differences may be highly "significant" but must be treated with great caution, even if the *P*-value is very small.

The corollary to these considerations is that we must also be sensitive to unexpected differences between groups, especially where toxicity is involved. A study designed to examine one particular outcome which found a major and significant mortality difference, for example, cannot just be ignored because it was a secondary outcome. Interpreting such findings requires an analysis of whether the difference is biologically feasible, whether the trial is otherwise of good quality and unlikely to have introduced bias, and whether there are consistent data from other sources. Many

outcomes are usually examined in systematic reviews including outcomes which may well have been secondary outcomes in the original trials. One of the benefits of a systematic review of several RCTs is that a consistent change in a secondary outcome between several studies is much less likely to be due to chance.

14. Studies of prognosis

Whenever evaluating a study of prognosis in neonatology it is important to consider how important the outcome is, and why prediction of that outcome may be useful. The characteristics of a prognostic test that make it useful depend on the reason for performing the test. So a test designed to determine future surveillance may require a high sensitivity if the surveillance is affordable and harmless, and/or a high specificity if harm is possible or cost is substantial. Prognostic testing in neonatology has often been used in the past to determine which infants should have consideration of withdrawal or withholding of life-sustaining interventions. For such a prognostic test to be worthwhile it should have a very high specificity and a very high positive predictive value, for outcomes which are very important to the infants and their families.

15. Studies of diagnosis

For studies reporting a diagnostic test, comparison with a gold standard is the most important feature to evaluate. For some diagnoses, however, a "gold standard" may be lacking. For example the gold standard for diagnosis of a patent ductus arteriosus (PDA) is the demonstration of ductal flow on echocardiography. However, the definition of a "hemodynamically significant PDA" remains questionable, and there is currently no gold standard to make this diagnosis.

As another example, one vitally important diagnosis in neonatology is that of sepsis. Culture-positive sepsis is defined by the presence of a positive culture, of a normally sterile site, but the elimination of contaminants is a thorny issue. The most frequent contaminants are also the most frequent infecting organisms, that is, the coagulase-negative staphylococci. Differentiating between the two situations, infection and contamination, requires either, in addition to a positive culture, clinical signs consistent with sepsis and no other explanation, or two blood cultures positive with the same organism and taken within a limited time interval. Both methods for eliminating contamination are clearly fraught with difficulties, so there is no gold-standard method for diagnosing coagulase-negative staphylococcal sepsis. Similarly, for the phenomenon known as culture-negative sepsis, there is no gold standard for diagnosis, or even a clear definition. Infants with clinical signs identical to those with sepsis clearly exist – some even develop shock, and may die – but the diagnosis (and treatment) of such infants will remain unclear until an adequate definition is developed.

Whenever a diagnostic test is evaluated, the purpose for trying to make the diagnosis should be considered. The characteristics of the test that are important will vary depending on the clinical purpose. For example, when a new diagnostic test for sepsis is evaluated, there are two possible benefits: either to treat more quickly in cases where the decision may be uncertain, or to reduce treatment of non-infected babies, which can mean shorter courses of treatment if true infection can be eliminated more quickly, or not even starting antibiotics if the test can be applied with a very rapid result.

For a test to be used to avoid treatment in suspected sepsis, the major characteristic required should be a very high negative predictive value. That is, if the test is negative it is safe to avoid treatment. One example of a widely used diagnostic test which does not satisfy such a criterion is the C-reactive protein (CRP). Due to the delays in production in the inflammatory mediators which lead to elevated CRP concentrations, CRP may be quite low at the time of presentation of clinical signs of sepsis. Sepsis will usually lead to an elevated CRP after a few hours, so the use of a negative test to discontinue treatment after several hours is more reasonable. By contrast, the low specificity implies that using an elevated CRP as an indication to continue treatment will often lead to over-treatment [14].

A test which might help to treat more quickly should have a high sensitivity, and be positive either before other clinical signs are definite or simultaneously with them. Heart rate variability indices have been shown to be highly sensitive for diagnosis of sepsis; indeed an RCT demonstrated that using the index led to reduced mortality because of earlier diagnosis of sepsis [15]. A high specificity will prevent over-treatment of non-infected patients. Thus the characteristics of a useful test may depend on the use that should be made of that test.

16. Studies of physiology

Many interesting studies of physiology have eventually led to improvements in management. It is important, however, to guard against making clear recommendations based only on physiologic studies. Interventions which seem physiologically justified may not lead to improved outcomes. To return to the example of sepsis, the rapid production of inflammatory mediators, which are associated with clinical signs and with mortality, has led to a series of trials of agents designed to interrupt their production. The recent succession of articles examining such agents has been universally disappointing in older adults, just as INIS showed no benefit in the newborn of intravenous immunoglobulin [5]. Indeed recent articles have suggested that we should instead be looking to enhance inflammatory responses, rather than inhibit them.

Similarly inotropic/vasopressor agents in septic shock in adults have differing hemodynamic responses, with norepinephrine, for example, leading to greater increases in cardiac output and renal perfusion than dopamine. Comparative clinical trials, however, have shown no difference in outcomes. Prior to those clinical trials a weak guideline recommendation to use norepinephrine rather than dopamine would have been understandable. Now, no such recommendation can be made.

A specific neonatal example is the treatment of hypotension in the extremely preterm. Dopamine increases blood pressure, and some guidelines promote its use for this purpose (e.g., NANN.org). However, increasing blood pressure without any evidence of improved clinical outcomes is inadequate for such recommendations. Further physiologic studies detailing the hemodynamic responses to dopamine are important preliminaries, but in the final analysis they are insufficient to determine whether routine treatment with the agent is warranted.

17. Conclusion

Staying up to date with the neonatal literature is important, and a difficult but not an insurmountable challenge. Ongoing surveillance of the important journals, repeated regular searches using freely available tools, and focussed searches to answer specific questions are complementary approaches. Evaluation of whether or not the published evidence should lead to changes in clinical practice depends on the nature and the quality of the publications. Studies without controls are unreliable, and studies without concurrent randomized controls are always suspect, but may add to or support the results of primary RCTs.

Only by remaining abreast of the literature and interpreting it reasonably can we provide the best evidence-based therapy for our patients.

Conflict of interest statement

None declared.

Funding sources

None.

References

[1] Cummings JJ, D'Eugenio DB, Gross SJ. A controlled trial of dexamethasone in preterm infants at high risk for bronchopulmonary dysplasia. N Engl J Med 1989;320(23):1505–10.
[2] Seri I, Abbasi S, Wood DC, Gerdes JS. Regional hemodynamic effects of dopamine in the sick preterm neonate. J Pediatr 1998;133(6):728–34.
[3] Marlier L, Gaugler C, Messer J. Olfactory stimulation prevents apnea in premature newborns. Pediatrics 2005;115(1):83–8.
[4] Ohlsson A, Lacy J. Intravenous immunoglobulin for suspected or subsequently proven infection in neonates. Cochrane Database Syst Rev 2010;(3):Cd001239.
[5] Brocklehurst P, Farrell B, King A, Juszczak E, Darlow B, Haque K, et al. Treatment of neonatal sepsis with intravenous immune globulin. New Engl J Med 2011;365(13):1201–11.
[6] Ohlsson A, Lacy JB. Intravenous immunoglobulin for suspected or proven infection in neonates. Cochrane Database Syst Rev 2013;7:Cd001239.
[7] Crowther MA, Cook DJ. Trials and tribulations of systematic reviews and meta-analyses. Hematology 2007;2007(1):493–7.
[8] Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Commun Health 1998;52(6):377–84.
[9] Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA 2013;309(8):814–22.
[10] Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008;336(7644):601–5.
[11] Walsh MC, Yao Q, Gettner P, Hale E, Collins M, Hensman A, et al. Impact of a physiologic definition on bronchopulmonary dysplasia rates. Pediatrics 2004;114(5):1305–11.
[12] Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med 2012;157(6):429–38.
[13] Reis BB, Hall RT, Schanler RJ, Berseth CL, Chan G, Ernst JA, et al. Enhanced growth of preterm infants fed a new powdered human milk fortifier: a randomized, controlled trial. Pediatrics 2000;106(3):581–8.
[14] Mukherjee A, Davidson L, Anguvaa L, Duffy DA, Kennea N. NICE neonatal early onset sepsis guidance: greater consistency, but more investigations, and greater length of stay. Archs Dis Childh Fetal Neonatal Ed 2014.
[15] Moorman JR, Carlo WA, Kattwinkel J, Schelonka RL, Porcelli PJ, Navarrete CT, et al. Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: a randomized trial. J Pediatr 2011;159(6):900–6. e1.