Conclusiveness of the Cochrane Neonatal Reviews: A systematic analysis

DROR MANDEL, YOAV LITTNER, FRANCIS B. MIMOUNI & RONIT LUBETZKY

Department of Neonatology, Lis Maternity Hospital, Tel Aviv Sourasky Medical Center, and the Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

Abstract

Aim: To assess the conclusiveness of the Cochrane Neonatal Reviews (CNRs). We tested the hypotheses that: 1) the majority of the reviews is inconclusive; 2) the majority of reviews recognizes the need for further studies; 3) the ability to reach a conclusion is dependent upon both the number of studies and the number of patients. We also aimed to determine whether the conclusiveness of the CNRs was affected by time. Methods: We selected CNRs available in the Cochrane Library in June 2004. The number of randomized clinical trials (RCTs) found, number of RCTs included for analysis, number of patients enrolled, the stated need for further studies, and the conclusiveness of CNRs were recorded. Results: Out of 170 CNRs, 67.7% were conclusive. The average number of articles was similar, but the total number of patients enrolled was three times higher in the conclusive CNRs. The percentage of articles included in conclusive studies was significantly higher than in inconclusive ones. The vast majority of CNRs recognized the need for further studies. The number of studies included correlated significantly with the total number of patients included. The percentage of conclusive CNRs correlated negatively with year of publication. Conclusion: The majority of CNRs is conclusive, but emphasizes the need for further studies. The ability of a CNR to reach a conclusion is affected by the cumulative sample size and by the number of studies performed. The probability of a newer review to be conclusive is lower than that of an older review.

Key Words: Meta-analysis, neonatology, randomised clinical trial

Introduction

The Cochrane Neonatal Review Group (CNRG) is one of the many collaborative review groups registered in the Cochrane Collaboration [1]. The Cochrane systematic reviews are based upon the comprehensive search of relevant randomized clinical trials (RCTs) in order to produce a review of valid and reliable evidence that will be the basis of decision making in the therapeutic arena. Cochrane reviews are of the effects of interventions. The reviews have a fixed structure that incorporates a conclusion section, which summarizes the findings of the reviewers. A recent review by the administrators and editors of the CNRG surveyed the methods and some of the results and characteristics of the reviews available until 2001 [2] and concluded that 1) most reviews are current, and 2) among reviews finding a statistically significant benefit of treatment, the effect size was large, while among reviews finding no evidence of treatment effect, large and potentially important benefits or harms could not be excluded.

We undertook the following study of all the CNRG reviews available until June 2004 in order to assess the conclusiveness of the reviews from the perspective of the practicing neonatologist. By conclusiveness we meant the ability or not to reach the definitive conclusion that the treatment was beneficial or not, as explained later in detail in the methods section of this article. In particular, we aimed to describe the percentage of the reviews that were able to reach a meaningful therapeutic recommendation. Based upon a small survey of each one of this study's authors' opinion and personal experience/bias, we specifically tested the hypotheses that: 1) the majority of the reviews is inconclusive; 2) the majority of reviews recognizes the need for further and better studies; and 3) the ability to reach a conclusion is dependent upon both the number of studies performed and the number of patients enrolled. We also aimed to
determine whether the conclusiveness of the reviews was affected by a time factor, and whether newer reviews would be more conclusive than older ones.

Materials and methods
We selected all the 170 Cochrane reviews available in the Cochrane Library [3] on 1 June 2004, starting from 1997. One author (DM) extracted from the abstract, conclusion, and, if necessary, from the body of the review the following characteristics of each review: number of RCTs found by the reviewers on the topic, number of RCTs included for final analysis, total cumulative number of patients enrolled in the included studies, stated need or not for further and better studies, and type of conclusion reached by the reviewers. We classified the type of conclusion reached by the reviewers into five subcategories: 1) one strategy or drug was found to be better than the alternative; 2) no significant differences were found between the two strategies or drugs; 3) no decision could be reached because the studies were of insufficient quality, as stated by the reviewers; 4) no decision could be reached because there were insufficient data, as stated by the reviewers; 5) no decision could be reached because the studies were found to be “old”, as stated by the reviewers. We defined the reviews from the first two categories as being “conclusive”, and those from the last three as being “inconclusive”. We also recorded whether or not the review’s author(s) stated the need for additional studies.

Minitab version 13.1 (State College, PA, USA) was used for statistical analyses. The Kruskal-Wallis test was used to study the differences between “conclusive” and “inconclusive” studies in terms of number of RCTs available, number of RCTs included for final analysis, total cumulative number of patients enrolled, and stated need or not for further studies. For the purpose of this analysis, only the main outcome was considered, and not the secondary outcomes of each review. Regression analysis was used to study the correlation between number of included studies and cumulative sample size. A p-value of <0.05 was considered significant. We also used linear regression to determine whether the percentage of conclusive reviews (dependent variable) was affected or not by the year of publication (independent variable).

Results
We identified 170 Cochrane Neonatal Reviews (CNRs) [3]. The range, median, and mean number of RCTs examined, number of RCTs included for analysis, total cumulative number of patients enrolled, the stated need or not for further and better studies, and type of conclusion reached by the reviewers are represented in Table I.

In brief, 115 (67.7%) CNRs fulfilled our criteria for being conclusive versus 55 inconclusive (32.3%). Among the 115 “conclusive” studies, in 59 of them one strategy or drug was found to be better than the alternative and in 56 of them no significant differences were found between the two strategies or drugs. Among the 55 studies classified as “inconclusive,” 9 were judged so because the studies were of insufficient quality, 41 because there were insufficient data, and five because the studies were found to be “old.” The average number of articles was similar between the two groups (median 7, range 0–32), but the total cumulative number of patients enrolled was nearly three times higher in the conclusive CNRs (p<0.01) (Table I). A significantly higher percentage of articles was included in conclusive studies as compared to inconclusive ones (64% vs 49%, p<0.001). The vast majority of CNRs, whether conclusive (93%) or inconclusive (98%), emphasized the need for further studies in their conclusions.

<table>
<thead>
<tr>
<th>Table I. Study characteristics of the Cochrane Neonatal Reviews (CNRs).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of articles</strong></td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>mean ± SD</td>
</tr>
<tr>
<td>(median, range)</td>
</tr>
<tr>
<td>Total cumulative number of patients enrolled</td>
</tr>
<tr>
<td>mean ± SD</td>
</tr>
<tr>
<td>Percent of articles fitting criteria of inclusion</td>
</tr>
<tr>
<td>mean ± SD</td>
</tr>
<tr>
<td>(median, range)</td>
</tr>
<tr>
<td>Percent of reviews where further studies are needed</td>
</tr>
<tr>
<td>mean ± SD</td>
</tr>
<tr>
<td>(median, range)</td>
</tr>
</tbody>
</table>
In regression analysis, the number of studies included correlated significantly with the total number of patients included ($R^2 = 0.31\%, p < 0.0001$). In linear regression, the percentage of conclusive reviews correlated strongly and negatively with the year of publication ($R^2 = 0.78\%, p < 0.003$) (Figure 1).

**Discussion**

Contrary to our hypothesis, the majority of CNRs is conclusive: indeed, 67.7% of CNRs fulfilled our criteria for being conclusive, while only 32.3% were inconclusive. Thus, theoretically, the readers of a given Cochrane review are more likely than not to get answers from the review that might help them in the therapeutic arena. This assumes, obviously, that the review was not biased by the usual factors that may bias a meta-analysis, which include, as reviewed recently by Egger et al. [4]: 1) publication bias [4,5], which is the higher chance for an article with positive results to be published than a study with negative results; 2) location bias [4], according to which studies performed in different types of population are not necessarily applicable to a given patient; 3) language bias [4,6], by which clinical trials reported in languages other than English from meta-analyses may reduce the precision of combined estimates of treatments; 4) citation bias [4], which is created when articles with negative results, less likely to be quoted, are missed by the search (we have even recently demonstrated the evidence for an “impact factor bias”; that is, the greater likelihood for a study with positive results to be accepted for publication in a journal of greater impact factor) [7]; and 5) multiple publication bias introduced when a positive study is reported more than once, increasing its chance of being detected in the search [4].

As hypothesized, whether conclusive (93%) or inconclusive (98%), the vast majority of CNRs recognized the need for further studies in their conclusions. The stated need for additional studies when a conclusion can be reached may appear paradoxical. From an ethical standpoint, it is even questionable to perform additional studies (which expose the patients to research risks, without the benefit of receiving the best-studied medicine) when the meta-analysis is conclusive. The request for additional studies is warranted, however, when new alternative treatments have not been sufficiently studied, when long-term side effects of existing therapies are unknown, when there are no data on specific subgroups of infants, or when the best dose to be used for a specific drug is unknown. For instance, the review by Steer et al. [8] comparing the use of caffeine versus theophylline for apnea in preterm infants concludes that “caffeine…is the preferred treatment for apnea in preterm infants” (i.e. a “conclusive” study according to our study criteria). It also indicates that “there is a need for clinical trials with larger numbers of infants born at lower gestational age to demonstrate the effectiveness and safety of caffeine compared to theophylline treatment with respect to clinically important outcomes including safety and long-term effects on neurodevelopmental outcome. The appropriate dose of methylxanthine therapy requires further investigation.” Thus, the above-mentioned review clearly states that specific subpopulations of infants must be studied (small preterm infants) and that appropriate doses must be evaluated, which, in our view, is a justified conclusion.

Figure 1. Correlation between the percentage of conclusive Cochrane Neonatal Reviews (CNRs) and the year of publication.
As hypothesized, the ability to reach a conclusion was dependent upon both the number of studies performed and the number of patients enrolled. A large sample size decreases the risk of type II error, that is the risk of concluding that no difference exists between groups while a true difference exists. One aim of a meta-analysis is to “combine” the data of several studies in order to increase the sample size, and thus to decrease the type II error. It is therefore logical that the total cumulative number of patients enrolled was nearly three times higher in the conclusive CNRs than in the non-conclusive ones.

The number of studies included correlated significantly with the total number of patients included. Thus, we were not able to analyze the relative importance of each variable (total cumulative number of patients and number of studies) upon the ability to reach a conclusion, as such a regression equation would be affected by significant co-linearity. Nevertheless, the median number of studies included in the CNRs was approximately seven, a number very close to the “typical” Cochrane reviews (in other fields of medicine), which were found to comprise six included trials; in contrast, the median total number of patients was noticeably lower (207) in CNRs than in other fields (945 patients) [9]. A potential limitation of our analyses is that the ratio “number of RCTs included in the analysis”/“total number of RCTs found” is influenced by the review author who determines by pre-set criteria which studies will be excluded, which will affect the numerator of this ratio.

The fact that the percentage of conclusive reviews correlated strongly and negatively with the year of publication is of great interest. It means that, contrary to our hypothesis, the probability of a newer review to be conclusive is lower than that of an older review. We can only speculate about this finding. A possible explanation is that, over the years, reviewers use more and more stringent criteria for conclusiveness. Another potential explanation would be that older reviews dealt with issues that have been more extensively studied, and for which more is known, than newer reviews that deal with more controversial topics, the less controversial ones having been taken care of previously.

In summary, we found that the CNRs allow, in most cases, the reader to reach a conclusion applicable to his/her clinical practice; they also emphasize in both conclusive and inconclusive analyses the major points that require additional or better research.

References