LETTERS TO THE EDITOR

Combining or not combining published results in the presence of heterogeneity?

Dear Sir:

In their updated systematic review and meta-analysis focusing on the effects of n–3 polyunsaturated fatty acids (PUFAs) on depressed mood, Appleton et al (1) reported results supporting some beneficial effects of n–3 PUFAs according on both fixed- and random-effects models [pooled standardized differences in means or effect sizes of 0.10 ($P = 0.01$) and 0.26 ($P < 0.01$), respectively]. These results were to be interpreted with caution given the considerable heterogeneity present between studies ($I^2 = 65\%$, $P < 0.01$) and their corresponding low clinically beneficial effect. According to the authors, the existence of asymmetry in the funnel plot suggested that the likely source of heterogeneity was attributable to publication bias. On the basis of this, they mentioned that the greatest weight should therefore be given to fixed-effects models and referred to the works of Sterne et al (2), Deeks et al (3), and Egger et al (4). In their conclusion, greater emphasis was put on the effects of n–3 PUFAs in individuals with diagnosed depressive illness because of an effect size of 0.41, but substantial heterogeneity remained in this subset of studies ($I^2 = 71\%$, $P < 0.01$).

Publication bias was defined as a systematic error related to the selective publication of studies. In the search strategy of this systematic review, there is no mention of the consideration of results published in conference abstracts or proceedings. These data are a significant source of gray literature, and failure to identify them might affect the results of a systematic review (5). About two-thirds of results from abstracts describing randomized or controlled trials are published in full (6).

Publication bias was investigated by using a funnel plot presenting evidence of asymmetry, which was linked to publication bias. A meta-regression analysis was performed showing a link between study effect size and study size. However, though the presence of funnel plot asymmetry can suggest the presence of publication bias, meta-regression results are expected even when there is no publication bias. A more appropriate formal test of funnel plot asymmetry consists of regressing the standardized effect size on the inverse of the SE. Under the null hypothesis of symmetry, the intercept of this regression should not be different from $0$ (7). We would have liked to present the results of such an analysis; unfortunately study SEs were missing from the article.

As mentioned, funnel plot asymmetry may be explained by publication bias but also by several other sources including poor methodologic quality of studies, true heterogeneity, data irregularities, artifactual due to poor choice of effect measure, chance, and choice of axis in funnel plot representation (8, 9). The funnel plot presented by Appleton et al (1) shows the precision (1/SE) as the measure of study size on the vertical axis. However, this plot is not funnel-shaped, and the 95% CI lines are curved. Sterne and Egger (9) suggest that the proper funnel plot for publication bias detection should use the SE on the vertical axis. The 95% CI lines also allow judgment of the presence of heterogeneity, giving further clues to the possible link between publication bias and heterogeneity. Despite this, further investigation is required to assure that publication bias is the only cause of heterogeneity.

Deeks et al (3) mention that a thoughtful consideration of whether it is appropriate or not to combine all of the studies in a meta-analysis is a key step in a systematic review. Throughout the analyses, considerable heterogeneity was observed, and careful interpretation of the results was made. According to Appleton et al (1), results on the basis of fixed-effects models were favored because of publication bias because random-effects models give relatively more weight to smaller studies than the fixed-effect models and lead to wider CIs. However, the use of random-effects models has frequently been advocated if there is heterogeneity between study results (4). In either case, in the presence of publication bias, results will most likely be inappropriate. It is then important to perform a thorough sensitivity analysis to explain the different sources of the underlying heterogeneity. Appleton et al (1) presented 3 meta-regression analyses that could have been part of such an investigation. However, in all cases heterogeneity remained elevated, which suggests that there may be more than one source of heterogeneity at play. Unfortunately, as is often the case in meta-analyses, the relatively small number of studies does not allow a complete multiple meta-regression analysis to be performed.

In summary, Appleton et al (1) were right to declare that the overall statistic produced by using the evidence of the effects of n–3 PUFAs on depressed mood may be erroneous given the heterogeneity between the studies. The same remark should stand for the evidence in individuals diagnosed with a depressive disorder. The only true result in this meta-analysis is that there is no evidence of

REFERENCE


any benefit of n–3 PUFAs on individuals with no psychiatric diagnoses, with or without depressed mood (both P values >0.05 and I² = 0%).

Neither author reported a conflict of interest.

Danielle Laurin
Pierre-Hugues Carmichael

Unité de Recherche sur le Vieillissement
Centre de Recherche FRSQ du CHA
Laval University
Quebec
Canada
E-mail: danielle.laurin@pha.ulaval.ca

REFERENCES


Reply to D Laurin and P-H Carmichael

Dear Sir:

We thank Laurin and Carmichael for highlighting the questionable value of combining studies in the presence of heterogeneity. We agree with Laurin and Carmichael that the interpretation of any meta-analysis conclusion is considerably compromised by the presence of heterogeneity and agree that the only robust conclusion that can be drawn from our work (1) is the absence of evidence of any benefit of n–3 polyunsaturated fatty acids (n–3 PUFAs) in individuals with no psychiatric diagnosis. A possible effect in individuals with diagnosed depressive illness is suggested from the analyses as a result of the combined effect size, but, as we previously stated, this should be treated with caution given the heterogeneity found in this analysis (1), and as Laurin and Carmichael suggest, this finding is far from definitive.

We agree that the heterogeneity found in these analyses is likely to result from many factors, and that investigation of all these factors is important (2, 3). Although we investigated 3 possible sources of heterogeneity, the data are not currently available in this area to investigate other possible sources. Our analyses suggest differences between studies as a result of study size, severity of depressive symptoms, and psychiatric diagnosis (1), but additional possible sources of heterogeneity include the variety of methodologic forms of bias mentioned by Laurin and Carmichael (2, 3) and those as a result of the specific data investigated: the demographic, dietary, and psychiatric history of participants; the type, dose, and duration of n–3 PUFAs used for supplementation; the placebo comparison; measures used; and trial quality (4). We accept the limitations of a review on the basis of only the published literature identified by the databases used (2).

Our suggestion of publication bias is confirmed by results from the tests for funnel plot asymmetry and publication bias by both Begg and Mazumdar (5) and Egger et al (6) [Begg’s test: z (continuity corrected) = 4.90; P < 0.01; Egger’s test: bias regression coefficient = 2.31; 95% CI: 1.53, 3.09; P < 0.01] and is clearly shown in Begg’s funnel plot with pseudo 95% CI limits including hypothesized unpublished studies in Figure 1. The funnel plot, which was published as part of our review (1), was formatted to allow easier comparison with the funnel plot published in a previous review (7). This comparison shows that there is more evidence now available and some apparent reduction in bias over this time period. Nonetheless, publication bias clearly remains, and although this bias may explain some of the heterogeneity, the existence of additional sources of heterogeneity seems highly likely (2–4). This demonstration of bias in our meta-analysis highlights the value of formal methodologically sound attempts to synthesize studies in this area (2) and the need for increased evidence in the public domain (2). Our emphasis on the results of fixed-effects models over those of random-effects models is also justified given the publication bias (8), although results from both random- and fixed-effects models have been provided throughout the article and differ only minimally.

The value of attempting to combine studies in the presence of heterogeneity is thus possibly less concerned with the pooled estimate achieved and more concerned with the explicit demonstration of this heterogeneity and the need for caution in interpretation. Others have combined studies investigating the impact of n–3 PUFAs on depressed mood/depression to report positive findings without highlighting this need for caution (9). In summary, we would argue

FIGURE 1. Begg’s funnel plot with pseudo 95% CI limits showing current published studies (circles) and hypothesized unpublished studies (squares).