ONE MORE STUDY – HOW DOES A NEW STUDY AFFECT THE RESULTS OF A PUBLISHED META-ANALYSIS?

Dorothy Mackerras

Introduction

The news media often report the results of the latest scientific study without referencing any previous research. This is potentially quite misleading. Conclusions about the strength and direction of a postulated relationship must be based on the totality of the evidence, not the latest study.

Meta-analysis is a technique for combining the results of a number of studies. However, they are done at discrete points in time.

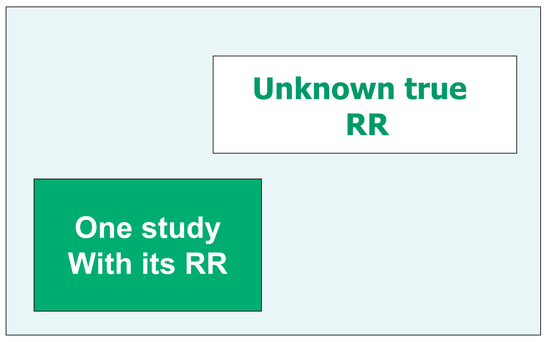
This poster illustrates a simple method for obtaining a quick "gut feel" as to whether the latest study importantly alters the results of a previously compiled set of literature (e.g. a meta-analysis). More sophisticated methods are also available.

The example uses randomised controlled trials examining the effect of folic acid plus other B vitamins on recurrent heart disease. These trials were done to test the hypothesis that increasing folic acid intake would decrease homocysteine levels which would decrease heart disease rates.

Interpreting study results

A single study is one of many possible studies that could be done to test a hypothesis. It may or may not give a good estimate of the true unknown relationship that is being investigated (Fig. 1) As the number of studies testing the same hypothesis increases, our estimate as to the true effect improves.

Fig. 1: What is the relationship between the findings of one study and the unknown, true result?



What is a 95% confidence interval?

The relative risk/risk ratio (RR) for any study is the best estimate from that study of the true unknown result.

→ The result of any particular study may not be a good estimate of the true unknown result.

The results from any study are consistent with a range of true underlying effects: this range is described by a confidence interval, such as a 95% CI.

The 95% CI tells you that

→ if the underlying model is correct and there is no bias, then over an unlimited number of replications of the study, 95% of the replications will have a 95 % CI that contains the true unknown value

→ and in 5% of the studies, the 95% CI will not contain the unknown true value

Fig. 2 shows the results from 3 hypothetical studies

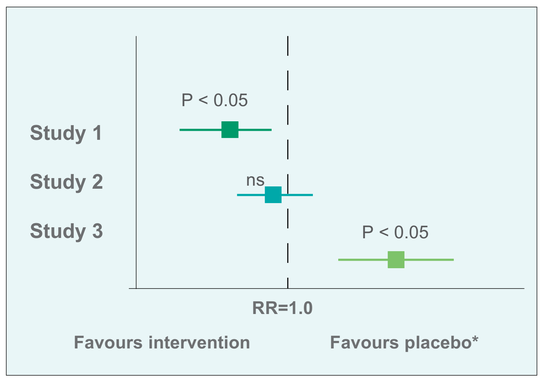
→ Study 1 is significant, Study 2 is not significant but their confidence intervals overlap – and so it would not be correct to say that they disagree with each other. The overlap in CI starts to suggest where the true unknown RR lies.

→ Why is Study 3 different? (It is significant but with the opposite effect from Study 1)

- Is it one of the 5% of study replications that has a 95% CI which does not contain the unknown true value?
- Or does it contain the true unknown RR and are Studies 1 & 2 the outliers?
- Or is there an important methodological difference that might explain the discrepancy – e.g. a different dose, a different form of vitamin (e.g. retinol vs beta-carotene), a different length of follow-up, a different measurement method for the outcome etc etc.
- Exploring these differences could be very enlightening. In this case, there would be important "heterogeneity" in the overall analysis, and this is a warning that gaps in understanding.

→ Clearly more good studies are needed.

Fig. 2: Results from 3 hypothetical studies – RR (square) & 95% confidence intervals (line).



* NB for studies where an increase is a good outcome (e.g. in survival after cancer) an RR>1 means that the intervention is favourable

What about statistical significance?

Statistical significance is commonly misunderstood.

→ Statistical significant result is not is not a guide to the value of the study result

→ Studies which are statistically significant are not "better" or more useful studies than those with non-significant results.

Statistical significance describes the probability that a RR may be different from 1.0 (the null value).

If there is truly no relationship (RR=1.0) then this result cannot statistically significant.

The results of many studies testing the same hypothesis will lie around the true, unknown, result (assuming the studies are well conducted).

→ If the true, unknown, result is RR=1 (no effect), then approximately half the studies will find a RR <1 and half RR>1.

→ This does not mean that there is disagreement among the studies, even if some of them are large enough to have statistically significant results

Illustration

Davey Smith and Ebrahim produced a Forrest plot of studies of folic acid (with vitamins B6 and B12) on recurrence of heart disease outcomes in late 2005. The black print on Fig. 3 shows the plot from this paper.

→ The RR and its 95% CI interval are shown for each study and the combined overall effect. (In this example, RR < 1 indicates that the vitamins reduced heart disease, RR=1 means no effect, and RR >1 means that vitamins increased heart disease)

→ Only Lange had statistically significant results, which indicated an adverse effect of vitamins

→ The squares for each study show how they contribute to the weighted average; the largest studies (with the narrowest 95% CI) have the greatest influence on the combined result.

The funnelling effect showing that the largest trials home in on the true RR is clear in this diagram.

→ The smaller trials, even the statistically significant trial by Lange are random noise around the true result illustrated by the large trials.

→ The diamond at the bottom shows the weighted average result for all studies combined – RR=1.05 or a 5% increase. The RR is small, not significant, and was interpreted as "no effect of vitamins" (although some might suggest a possible small adverse effect might exist).

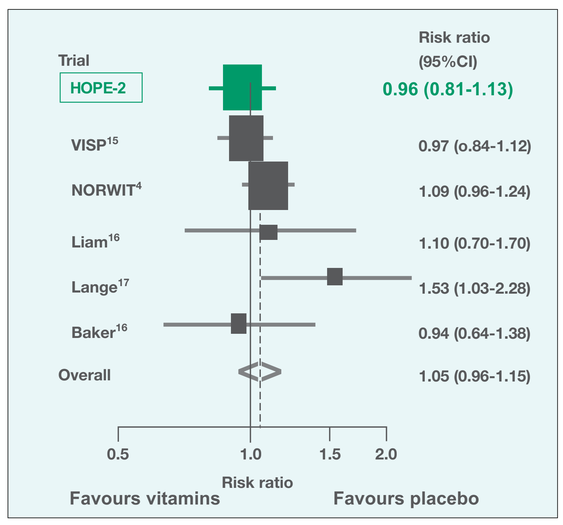
In early 2006, the results of another large randomised controlled trial (5522 subjects followed for 5 years) were published. This was the HOPE-2 trial (plotted in blue on Figure 3).

→ It found a non-significant favourable effect of folic acid; its CI overlaps CIs of the previous large studies and the previous combined effect (the diamond).

The effect of the HOPE-2 study is clear – it will pull the diamond back towards 1.0, the no effect value. Our previous conclusion is strengthened: increasing folic acid, B6 and B12 intake has no effect on recurrent heart disease incidence. The possibility of an adverse effect has essentially disappeared.

→ Note that if the new study had been small but significant adverse effect (e.g. like Lange's result) it would have had virtually no effect on the location of the diamond. Once large studies have been done, new small studies have little influence on the combined estimate.

Fig. 2: Randomised trials of folic acid and B vitamins on heart disease outcomes. Original plot from Davey Smith & Ebrahim, 2005 with more recent HOPE-2 trial added



Discussion

From the above, it is clear why statements such as "one study indicated an adverse effect" is not a good summary of the totality of the evidence.

Any technique is open to misuse. Summaries of the literature are more robust if

→ they clearly define the question being examined

→ found all the papers including those not in English

→ describe how quality criteria were applied: e.g. standards for good randomisation, measurement instruments, age & sex of subjects, extent of controlling for confounding that is required.

→ assessed whether publication bias might exist in the body of literature available for examination.

Studies should only be combined if they address the same question. Good description of the above factors makes it easier to determine whether the latest study should be added to a previous meta-analysis or not.

References

Davey Smith & Ebrahim. Folate supplementation and cardiovascular disease. Lancet 2005:366:1679-81.

Lonn et al. Homocysteine lowering with folic acid and B vitamins in vascular disease. NEJM 2006:354:1567-77

Australia
New Zealand

PO Box 7186 Canberra BC ACT 2610
PO Box 10559 The Terrace Wellington 6036

www.foodstandards.gov.au
www.foodstandards.govt.nz

FOOD STANDARDS Australia New Zealand
Te Mana Kounga Kai – Ahitereiria me Aotearoa