Department of Health Sciences

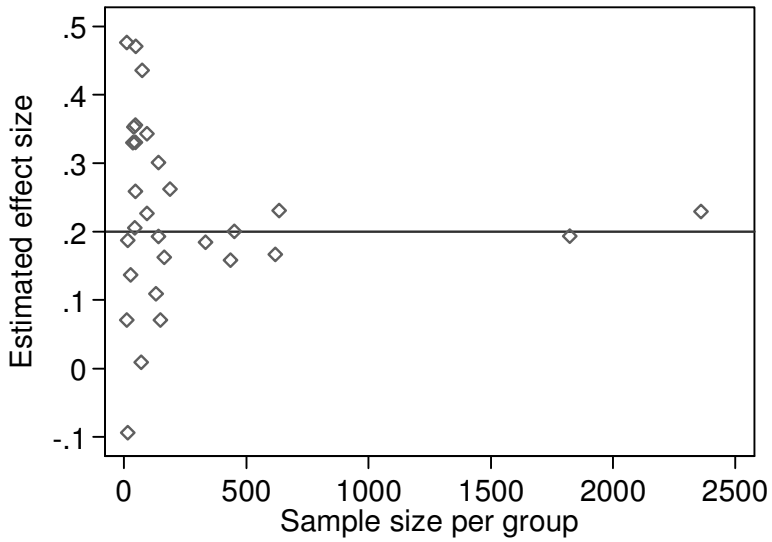
M.Sc. Module: Systematic Reviews

Suggested answers to exercise on meta-analysis: multivitamin supplements

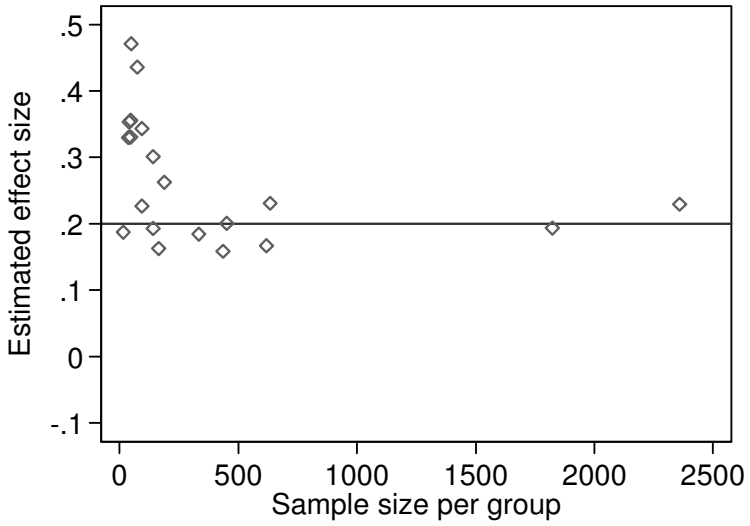
1. *How is the research question framed in terms of population/condition, intervention, control, and outcome?* The population is elderly people, but no definition, e.g. of minimum age, is given. The condition is prevention of infections in elderly people, which infections is not specified so we could presume they include any and all, nor is it specified whether the elderly people should be otherwise healthy. This aspect of the question is not well specified. The intervention is multivitamin and mineral supplements, no dosage or frequency being specified. Single vitamins or minerals are specifically excluded. The control treatment is placebo. The outcomes are incidence of infections and duration of infections which occur. Hence the main problem is the definition of the patient population.

2. *Why did the authors use so many data-bases in their search?* No database is able to index all research. We therefore use a large number of databases to try to get as many of the relevant studies as possible. The same studies may be retrieved from many different databases, but this does not matter. The aim is not miss studies.

3. *The authors say that they had 'intended to use funnel plots to assess the possibility of publication bias, the relatively small number of studies reporting each outcome precluded such an assessment.' What is a funnel plot and how can it assess publication bias? Why did the small number of studies lead the authors not to do this?* Publication bias is when some studies are not published because their results are not significant or not in the desired direction. Such trials are often small. A funnel plot is used to see whether the estimated treatment effect is related size of the study. We plot the difference between the treatments against either the size of the study, the meta-analysis weight, the standard error of the difference, or some other related quantity. The difference which may be a difference between means, a difference/standard deviation (effect size), log odds ratio, etc. For example, a funnel plot may look like this:



If there are small, not-significant trials missing, it might look like this:



The plot is no longer symmetrical about the horizontal effect size line, because some of the small, not-significant trials were not published. If there are very few trials, as in this paper, there will be too few to see a pattern, so the authors could not do this.

4. *In Table 1 the Jadad score is given. What is the Jadad score and how can we interpret the scores in Table 1?* The Jadad score is an attempt to assess the quality of a reported trial. It gives a score between 0 and 5, depending on whether the trial is randomised and blinded, whether randomisation and blinding are appropriate, and whether dropouts and withdrawals are well described. It partly assesses the quality of trial design (randomisation and blinding) and partly the quality of reporting (withdrawals). The scores are all 2, 3, or 4. The trials ranged from mediocre to good, but none were of the best quality.

5. *What sort of graph is shown in Figures 2, 3, and 4? What do the squares, horizontal lines, and diamond shape represent?* These are examples of a forest plot, so called because when treatment difference is on the vertical axis it is said to look like trees in a forest. The graph shows the results of several studies and the meta-analysis combined estimate for all of them. The squares represent the effect estimates from the individual studies, the horizontal lines the confidence intervals for these estimates. The diamond represents the meta-analysis estimate. The horizontal width shows the confidence interval and the point with greatest vertical depth is the point estimate.

6. *Why is the scale in Figure 2 a natural scale and in Figure 3 a logarithmic scale?* Figure 2 shows the difference in mean number of days with infection. A difference of 10 days in one direction would be represented by the same distance as a difference of 10 days in one direction. Also, these data would be analysed on the natural scale and so the confidence intervals will be symmetrical on this scale. Figure 3 shows a difference in terms of an odds ratio. A difference measured by a odds ratio = 2 in one direction would be the same as an odds ratio of ½ in the other direction. A logarithmic scale represents these odds ratios by equal distances. Also, odds ratios are analysed using their logarithms and the confidence interval for the log odds ratio is symmetrical. Hence on a logarithmic scale the confidence interval for the odds ratio is symmetrical.

7. *The authors say 'We used random effects models to perform meta-analyses if the heterogeneity between studies was estimated to be greater than zero; otherwise we used the model reduced to a fixed effect model.' (Quantative (sic) data synthesis.) What is the difference between random and fixed effects models?* In a fixed effects model, we assume that all the studies estimate the same effect. The only reason they differ is random variation between the trial subjects. The treatment effect is a constant. We use only the sampling variation within the trials. In a random effects model, we assume that the trials are estimating different treatment effects, which vary from trial to trial and so are not a constant but have a variance. We use the sampling variation within the trials and the sampling variation between trials. We assume that the trials are a sample from a population of possible of trials where the treatment effect varies. They must be a representative or random sample, which is a strong assumption.

   If the treatment effect really is the same in all trials, i.e. no heterogeneity, the fixed effects model is more powerful and easier. We get a correct pooled estimate, confidence interval, and P-value. The random effects model is less powerful because P values are larger and confidence intervals are wider.

   When heterogeneity exists, a fixed effects model gives a pooled estimate which may give too much weight to large studies, a confidence interval which is too narrow, and a P-value which is too small. A random effects model may give us a different (but correct) pooled estimate with a different interpretation, a wider confidence interval, and a larger P-value.

8. *The authors say that 'Although the direction of results [in Figure 2] is consistent, studies are heterogeneous (the I^2 statistic, which indicates the proportion of variability in the weighted mean differences attributable to heterogeneity, is estimated to be 97.3%, which is considered (very) large)'. What is 'heterogeneity' and how does the I^2 statistic measure it?* 'Heterogeneity' means that studies do not estimate the same treatment effect. We may have clinical heterogeneity, where studies differ in terms of the patient population, interventions, outcomes, or design, or statistical heterogeneity, random variation in treatment effects. We can test for heterogeneity using a chi-squared test and check for it visually by forest and Galbraith plots.

   The chi-squared test provides a test of significance for heterogeneity, but it does not measure it. The I^2 statistic is defined from the chi-squared heterogeneity statistic X^2

$$I^2 = 100 \times \frac{X^2 - d}{X^2}$$

   where d is the degrees of freedom. If I^2 is negative we set it to zero.

   If there is no heterogeneity the expected value of the chi-squared statistic is equal to its degrees of freedom. Hence I^2 is the percentage of the chi-squared statistic which is not explained by the variation within the trials. We can interpret $I^2 = 0\%$ as meaning no heterogeneity, $I^2 = 25\%$ as low heterogeneity, $I^2 = 50\%$ as moderate heterogeneity, and $I^2 = 75\%$ as high heterogeneity. Hence $I^2 = 97.3$ is very high heterogeneity, clearly visible in the forest plot.

9. *The diamond shape in Figure 2 is much wider horizontally than the horizontal lines and the diamond shapes in Figures 3 and 4 are much narrower horizontally than the horizontal lines. Why is this?* In Figure 2, the meta-analysis was done using a random effects model because there was a very large amount of heterogeneity (I^2 = 97.3%). The confidence intervals for the individual trials, which are represented by the horizontal lines, use only the variation between patients within the trials. The confidence interval for the meta-analysis estimate, represented by the diamond shape, uses both the variation between patients within the trials and the variation between the trials. As the latter is much larger than the former, the total variance used in calculating this confidence interval is much larger than that used for any of the trials and the interval is wider. In Figures 3 and 4, the meta-analyses were done using fixed effects models because there was very little heterogeneity (I^2 = 0% and not given). The confidence interval for the meta-analysis estimate uses only the variation between patients within the trials, just as for the individual trial estimates. As the meta-analysis estimate has more patients than any individual trial and uses the same variance, this confidence interval is narrower than that for the trials. This is usually the case in meta-analysis. **(*N.B. The situation in Figure 2 is very unusual, random effect models usually give narrower confidence intervals for meta-analysis estimates than for the trials, though wider than for fixed effects models.*)**

10. *The authors state their conclusions in the abstract as follows: 'The evidence for routine use of multivitamin and mineral supplements to reduce infections in elderly people is weak and conflicting. Study results are heterogeneous, and this is partially confounded by outcome measure.' What in their results leads them to these conclusions and do you agree?* They say that results are conflicting because the trials produce a fairly large and highly significant reduction in duration of infections but small and non-significant differences in incidence of infections. They say that study results are heterogeneous because the trial results for duration of infection are highly variable with a huge I^2 statistic. They say results are partially confounded by outcome measure because they get different results for duration than they do for incidence. They are right up to a point, but we would not necessarily expect the same treatment to prevent infection and to alleviate its effects. The results are quite consistent with vitamins having no effect on the incidence of infections but acting to ameliorate infections once they have been acquired.