

ENDGAMES

STATISTICAL QUESTION

Meta-analyses: heterogeneity and subgroup analysisPhilip Sedgwick *reader in medical statistics and medical education*

Centre for Medical and Healthcare Education, St George's, University of London, Tooting, London, UK

Researchers undertook a meta-analysis to evaluate the effectiveness of comprehensive geriatric assessment in hospital for older adults admitted as an emergency. They included randomised controlled trials that compared comprehensive geriatric assessment with usual care. Comprehensive geriatric assessment is a multidimensional interdisciplinary diagnostic process used to determine the medical, psychological, and functional capabilities of a frail elderly person so as to develop a coordinated and integrated plan for treatment and long term follow-up. Usual care usually involved admission to a general medical ward setting under the care of a non-specialist. Twenty two trials were identified, evaluating 10 315 participants in six countries.¹

The primary outcome was “living at home” at the end of the scheduled follow-up period. This outcome was reported by 18 trials evaluating 7062 participants. The median follow-up was 12 months (range six weeks to 12 months). The test of heterogeneity for these trials gave $\chi^2=28.49$, $df=17$, $P=0.04$, $I^2=40\%$. The total overall estimate indicated that the odds of a patient living at home at the end of scheduled follow-up were significantly higher in those patients who had undergone comprehensive geriatric assessment than in those who received usual care (odds ratio=1.16 (95% confidence interval 1.05 to 1.28; $P=0.003$)).

Subgroup analysis was undertaken, based on the type of model of comprehensive geriatric assessment performed. Two broad types of model were identified: assessment in designated wards by a coordinated specialist team; and assessment by mobile teams wherever the patient was admitted. The test of heterogeneity for “ward” gave $\chi^2=17.66$, $df=13$, $P=0.17$, $I^2=26\%$ while that for “team” gave $\chi^2=1.86$, $df=3$, $P=0.60$, $I^2=0\%$.

The subtotal estimate for “ward” indicated that comprehensive geriatric assessment was significantly more likely to result in patients being in their own homes at the end of scheduled follow-up than was usual care (odds ratio 1.22 (1.1 to 1.35; $P<0.001$)). However, when comprehensive geriatric assessment was undertaken by mobile teams its effects were inconclusive in comparison with usual care (odds ratio 0.75 (0.55 to 1.01; $P=0.06$)). The test for subgroup differences gave $\chi^2=9.06$, $df=1$, $P=0.003$, $I^2=89\%$.

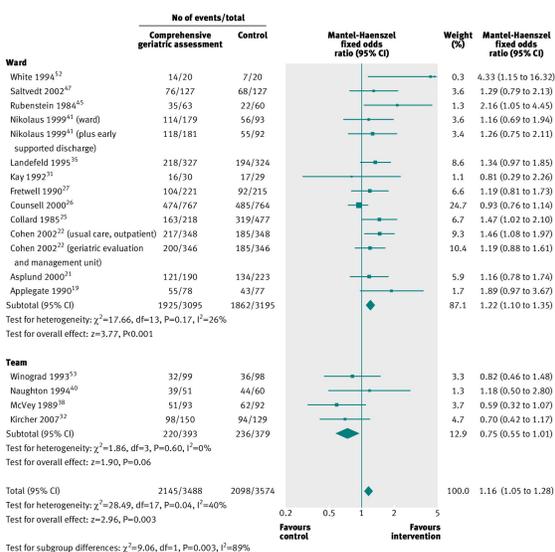
Which of the following statements, if any, are true?

- It can be inferred that homogeneity existed between the sample estimates across all trials.
- Homogeneity existed between the sample estimates in both subgroups of “ward” and “team.”
- It can be inferred that the effect of treatment on the primary outcome was different in the subgroups of wards and teams on the basis of the statistical significance in the subgroups
- A significant interaction existed between the subgroups of “ward” and “team” in the primary outcome.

Answers

Statements *b* and *d* are true, whereas *a* and *c* are false.

The aim of the meta-analysis was to combine the sample estimates of the population parameter of the odds ratio of living at home for comprehensive geriatric assessment when compared with usual care. The forest plot for the meta-analysis is shown (figure).



Forest plot of odds ratios for older adults living at home at end of scheduled follow-up (median 12 months), comparing comprehensive geriatric assessment with usual care

The total overall effect was calculated for all trials, regardless of whether comprehensive geriatric assessment occurred in designated wards or was undertaken by mobile teams. It was essential that the meta-analysis incorporated a statistical test of heterogeneity to assess the extent of variation between the sample estimates across all trials. The most popular tests for statistical heterogeneity are Cochran's Q and Higgins's I². Cochran's Q is the more traditional test and is based on the χ^2 test. It is carried out in a similar way to traditional statistical hypothesis testing, with a null hypothesis and an alternative hypothesis. The null hypothesis states that homogeneity existed between the sample estimates of the population parameter across the trials; any variation between them was no more than expected when taking samples from the same population—that is, the variation between them was minimal and a result of sampling error. The alternative hypothesis states that heterogeneity existed between the sample estimates.

Cochran's Q test may not always accurately detect heterogeneity in sample estimates. Because of this, Higgins's I² statistic is often used as well. Higgins's I² represents the percentage of variation between the sample estimates that is due to heterogeneity rather than to sampling error. It can take values from 0% to 100%, with 0% indicating that statistical homogeneity exists. It has been suggested that the adjectives low, moderate, and high (heterogeneity) be assigned to I² values of 25%, 50%, and 75%. Significant heterogeneity is typically considered to be present if I² is 50% or more.

In the above meta-analysis, the P value for Cochran's Q test and Higgins's I² for the test of heterogeneity across all sample estimates are displayed towards the bottom of the forest plot in the line "Test for heterogeneity: $\chi^2=28.49$, $df=17$, $P=0.04$, $I^2=40\%$ ". The P value of 0.04 meant that the null hypothesis was rejected in favour of the alternative at the 5% critical level of significance. Higgins's I² statistic suggested low to moderate heterogeneity. It was concluded that statistical heterogeneity existed between the sample estimates (a is false).

A subgroup analysis was done to explore this heterogeneity. This analysis was based on the model of comprehensive geriatric assessment used—that is, designated wards and mobile teams. Each subgroup analysis still required a test of heterogeneity,

and these are shown in the figure below the list of studies in each respective subgroup. That for the "ward" subgroup is $\chi^2=17.66$, $df=13$, $P=0.17$, $I^2=26\%$, while that for "team" is $\chi^2=1.86$, $df=3$, $P=0.60$, $I^2=0\%$. Therefore, homogeneity existed between the sample estimates in both subgroups (b is true).

The result of the test of heterogeneity influenced how the total estimate in each subgroup was obtained. The presence of homogeneity indicated that fixed effects methods should be used to derive the subtotal of the treatment effect. In the presence of heterogeneity so called random effects methods would have been used. A random effects meta-analysis would have produced a wider confidence interval for the subtotal effect than a fixed effects meta-analysis, resulting in a less accurate subtotal effect size.

The subgroup analyses indicated that patients who underwent comprehensive geriatric assessment in designated wards were significantly more likely than those who received usual care to be in their own homes at the end of scheduled follow-up (odds ratio 1.22 (1.1 to 1.35; $P<0.001$)). However, when comprehensive geriatric assessment undertaken by mobile teams was compared with usual care the result was inconclusive (odds ratio 0.75 (0.55 to 1.01); $P=0.06$).

The test of the treatment effect of comprehensive geriatric assessment compared with usual care was significant for the "ward" subgroup but not for the "team" subgroup. However, it would be wrong to infer that the effect of treatment on the primary outcome was different in the subgroups of wards and teams on the basis of the significance in the subgroups (c is false); the correct way to proceed would be to compare directly the magnitude of the treatment effects in the subgroups. Furthermore, the inference that homogeneity existed between sample estimates in each subgroup does not necessarily indicate that the model of assessment (ward or team) explained the heterogeneity observed between sample estimates across all trials, as described above. In particular, the numbers of trials and of participants may be too small for subgroup analyses to have adequate statistical power, whether to demonstrate significance of treatment effect or heterogeneity.

Treatment effects in subgroups should be compared by a test of interaction rather than by comparison of significance through P values. The test of interaction investigates whether the effect of intervention (comprehensive geriatric assessment compared with usual care) in the primary outcome varied between the subgroups. Interaction is sometimes referred to as effect modification. In a meta-analysis the test of interaction is undertaken using Cochran's Q test and Higgins' I². The test statistics compare the subtotal estimates between the subgroups. This is in contrast to above, where Cochran's Q and Higgins's I² were used to compare the sample estimates of the treatment effect across all the trials.

For the test of interaction, Cochran's Q provides a test of the null hypothesis that homogeneity existed between the subgroup estimates of the population parameter; any variation between them was no more than expected when sampling subgroups within the same population—that is, the variation between them was minimal and a result of sampling error. Higgins's I² measures the proportion of total variation in subgroup estimates that is due to heterogeneity rather than to sampling error. The test of interaction for the above meta-analysis is presented at the bottom of the forest plot in the line with the title "Test for subgroup differences: $\chi^2=9.06$, $df=1$, $P=0.003$, $I^2=89\%$ ". Cochran's Q test was significant at the 5% level of significance, while Higgins's I² was greater than 50%. Therefore, both Cochran's Q and Higgins's I² indicated that a significant

interaction existed between the subtotal estimates for the subgroups (d is true). It can be concluded that the subgroups estimated different population parameters.

Competing interests: None declared.

- 1 Ellis G, Whitehead MA, Robinson D, O'Neill D, Langhorne P. Comprehensive geriatric assessment for older adults admitted to hospital: meta-analysis of randomised controlled trials. *BMJ* 2011;343:d6553.

Cite this as: *BMJ* 2013;346:f4040

© BMJ Publishing Group Ltd 2013