Subgroup analysis: application to individual patient decisions

R John Simes, Val J Gebski and Anthony C Keech

CLINICAL TRIALS PROVIDE EVIDENCE of effectiveness of treatments as an average for a group of patients, yet, in clinical medicine, we usually wish to apply these results to individuals. Can we simply apply the overall trial result for each patient, or can the result be tailored to individual patients in some way?

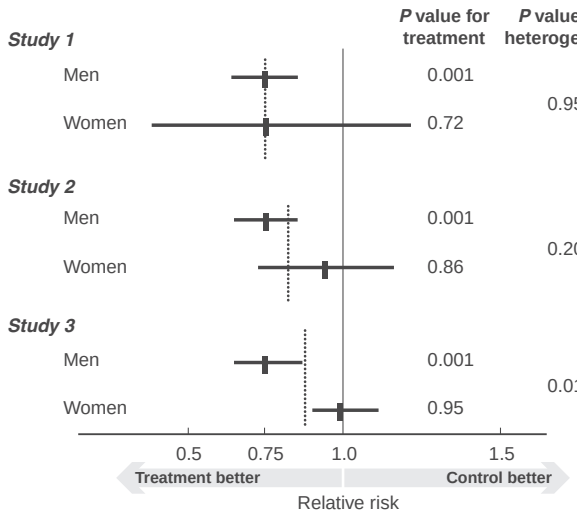
Consider a hypothetical example: a randomised trial comparing treatments A and B shows that treatment A is more effective than B among men ($P < 0.001$), but not among women (not signficant). Does this mean men should receive the new treatment, but women should not?

Box 1 illustrates these results from three different studies. In Study 1, the estimated treatment effect in men and women is the same — a 25% reduction in mortality associated with treatment A — but the much smaller number of women in the study gives rise to wider confidence intervals for this subgroup. In this case, there is no basis to consider that the treatment is any less effective in women (no heterogeneity; ie, non-significant test for interaction). Treatment could be considered effective for any patient regardless of sex.

In Study 2, the treatment effect in women is less than that in men (8% v 25% relative reduction, or 0.92 v 0.75 relative risk, respectively), but the effects in both are still consistent with the overall result of a 20% relative reduction, and there is no evidence of significant heterogeneity between groups (test for interaction, $P = 0.20$). Here, the different results between men and women could be simply due to chance, and it would still be appropriate to apply the overall estimate to both men and women (unless there was additional evidence).[1]

In Study 3, the observed effects for men and women are sufficiently different to suggest that this difference is unlikely to be due to chance (test for interaction, $P = 0.01$), and it is reasonable to conclude that the treatment effect differs between men and women. In this instance, the trial evidence should be considered separately for these subgroups. However, even here, a test of interaction can still give a low P value simply on the play of chance if many subgroups have been evaluated.[2]

1: Treatment effects in subgroups of men and women in three hypothetical trials



Overall relative risk: 0.75 for Study 1; 0.80 for Study 2; 0.87 for Study 3; represented by the vertical dashed line in each case.

A practical approach

Consider applying the overall trial treatment effect to each subgroup

How should we decide in practice whether to consider the treatment effects for these subgroups separately? A practical approach is shown in Box 2. A controlled trial is usually designed with a sample size large enough to show an overall treatment effect, but not necessarily adequate to show significant effects in each subgroup separately. The overall treatment effect is considered the best estimate for each subgroup of patients in the trial (this is sometimes referred to as the effect domination principle[1]). Hence, the treatment-effect results should only be applied differently for different subgroups if there is evidence of heterogeneity (a significant difference between subgroups, sometimes called interaction or treatment-effect modification).[3-7]

In considering whether there is evidence of heterogeneity, it is also worth reviewing the other questions outlined in the checklist for subgroup analyses in the previous article in this series.[1] If there is clear and reliable evidence of heterogeneity, using the treatment effects for each subgroup may be appropriate. However, as these may be unreliable (based on smaller numbers of patients) they may still be considered exploratory and motivate further trials rather than necessarily leading to different treatment guidelines. Further, evidence of heterogeneity may also lead to a search for underlying factors which may be linked to the particular subgroup and provide a more plausible biological explanation for such variation. For example, an apparent difference in treatment effect between men and women may truly relate to differences between these groups in age or smoking status (so-called confounding).

Seek confirmatory evidence

If there is still uncertainty whether differences between the subgroups in treatment effect are real, the following steps should be taken:

NHMRC Clinical Trials Centre, Camperdown, NSW.
R John Simes, MD, SM, FRACP, Director; **Val J Gebski,** BA, MStat, Associate Professor and Principal Research Fellow;
Anthony C Keech, MB BS MScEpid FRACP, Deputy Director.
Reprints will not be available from the authors. Correspondence: Professor R John Simes, NHMRC Clinical Trials Centre, Locked Bag 77, Camperdown, NSW 2050. john@ctc.usyd.edu.au

- seek confirmation from the results of an independent trial, a meta-analysis, or both;
- determine whether the effect is also present for a composite (expanded) endpoint, or surrogate endpoints; and
- establish whether independent evidence exists of a-priori biological plausibility of differences in the treatment effect.

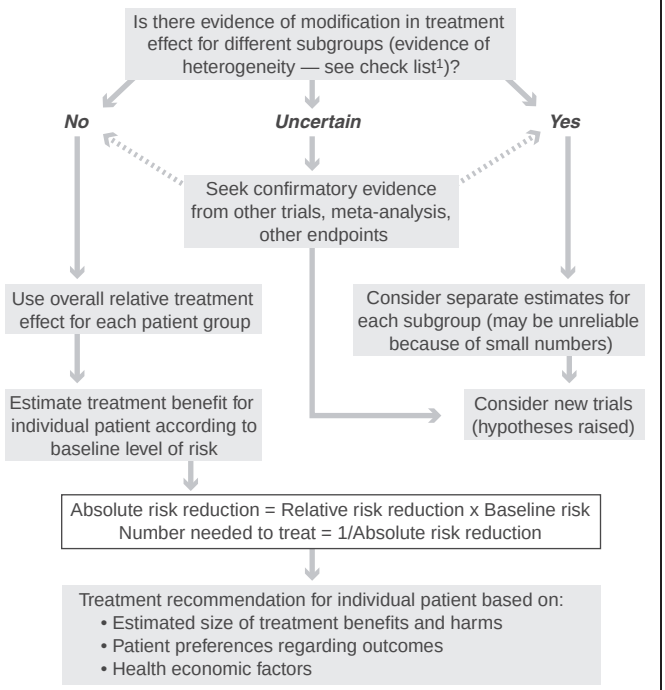
Without a good a-priori rationale for subgroup differences, the overall treatment effect provides a reasonable estimate for each subgroup, unless confirmatory evidence of treatment differences becomes available.

Estimate treatment effect according to baseline risk

If the same (or similar) relative treatment effect applies to different subgroups of patients, then those with a greater baseline risk of an event will derive a larger treatment effect. The absolute risk reduction associated with treatment is simply the absolute baseline risk multiplied by the relative risk reduction (Box 2).[8] For example, for a patient group with a 20% baseline risk, a treatment with a relative risk reduction of 25% would translate into a 20 0.25 = 5% reduction in absolute risk. For a patient group with a 10% baseline risk, this would translate into a 10 0.25 = 2.5% absolute risk reduction.

The number of patients needed to treat (NNT) to avoid one event can be calculated as one divided by the absolute risk reduction (Box 2).[8] This corresponds to $1 \div 0.05 = 20$ NNT for a patient with a baseline risk of 20%, and $1 \div 0.025 = 40$ NNT for a patient with a baseline risk of 10%. Smaller numbers needed to treat will result for patients at higher baseline risk.

2: Interpreting treatment effects in different subgroups within a controlled clinical trial*



* The decision pathway assumes there was a reasonable basis to consider the subgroups to have the same underlying condition.

A practical example

A 75-year-old woman who has had a previous myocardial infarction (MI) presents within 4 hours of symptom onset with suspected acute MI and ST elevation on electrocardiogram (ECG); she is being considered for aspirin and reperfusion therapy. Data from randomised trials of aspirin, thrombolytic therapy and immediate coronary angioplasty are considered. The patient has no known contraindications to these treatments.

Based on evidence from the ISIS-2 trial,[9] there is strong evidence that aspirin reduces the risk of short-term mortality, by about 23%, both overall and within most subgroups examined, including women and patients aged over 70 years. However, in this trial, among patients with a prior MI, no significant treatment benefit was observed. While the treatment effect in this subgroup apparently differed from patients without a prior MI, the interaction may have been a chance finding owing to the many subgroups examined. (For example, the chance of at least one significant result at the 5% level among 20 independent tests is over 50%.[1]) Consequently, the trial evidence still strongly supports the use of aspirin therapy in this patient.

Randomised trials of thrombolytic therapy in the FTT overview[10] have also demonstrated clear evidence of a reduction in mortality from such treatment for patients with acute ST elevation presenting within 12 hours of symptom onset. This overview also suggested diminished effectiveness of such treatment in the elderly (test for trend with older age, $P = 0.01$). However, in this case, much of the heterogeneity could be explained by the fact that older patients more often presented later (after 12 hours) and without the specific diagnosis of ST elevation on ECG. Lack of ST elevation and late presentation to hospital relate directly to the underlying biology and are linked to diminished effects of treatment. Once these confounding factors have been taken into account, there is much less rationale for considering different treatment or withholding thrombolytic therapy simply on the basis of the age of our patient.[11]

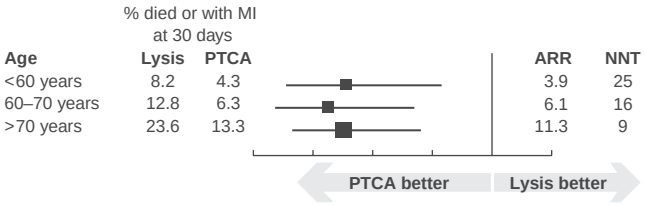
Next, the role of immediate coronary angioplasty in such a patient could be considered. Randomised trials, particularly in specialised centres, have suggested an additional treatment benefit for immediate coronary intervention compared with thrombolysis. An individual patient data overview of earlier randomised trials suggests a relative reduction in death or reinfarction of about 50%, with similar relative effects in each of the subgroups examined.[12] However, the absolute benefits of treatment (absolute risk reductions) were estimated to be much greater in the patients at high baseline risk, particularly those aged over 70 years (see Box 3). Consequently, if treatment with immediate angioplasty is considered appropriate in the particular hospital setting, it would be likely to have greater absolute benefit for this older patient than the average patient.

Finally, the role of long-term treatment in this patient could be considered. Should statin therapy be considered on the basis of the evidence from such trials as the LIPID and CARE studies?[13,14] Both of these had insufficient evidence

3: Absolute risk reduction (ARR) and numbers needed to treat (NNT) in age and sex subgroups

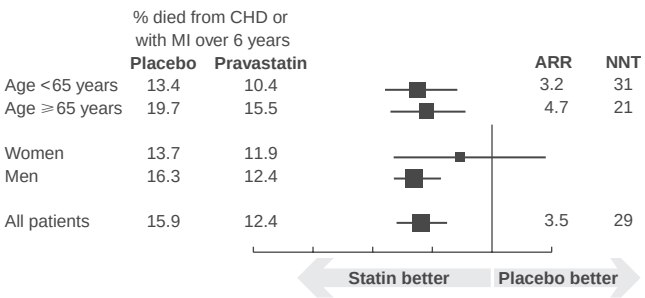
(a) Thrombolysis or percutaneous transluminal coronary angioplasty (PTCA) for acute myocardial infarction (MI)

Results from the PCAT overview[12]

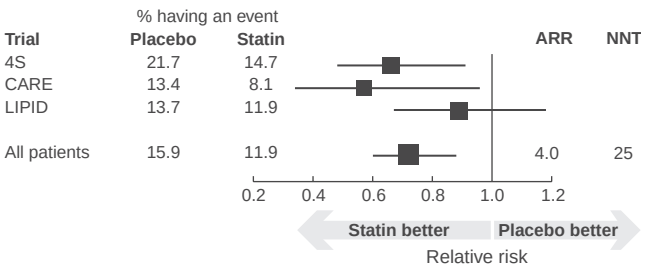


(b) Statin therapy to prevent coronary heart disease (CHD) events

1. Results from the LIPID trial[13,14]



2. Results for women only from the 4S, CARE and LIPID trials combined[14]



In none of these cases is there evidence of treatment effect modification (all *P* values for interaction are non-significant). ARRs and NNTs are derived from the overall relative treatment effect.

to show reductions in mortality with treatment for women separately. In the LIPID trial, older and younger patients had similar relative reductions in events (Box 3), but older patients at higher baseline risk had greater absolute benefit. Fewer women than men were studied in these trials, and yet the results for women were not inconsistent with those for men (Box 3).

The effect of statin therapy for women with prior CHD is illustrated further by the results of the 4S, CARE and LIPID trials.[14] The combined results of these three trials show an overall significant reduction in coronary events; the estimates from the separate trials vary but are still consistent with the overall result. Evidence of a similar relative treatment effect from statin therapy in both women and men has recently been confirmed by the results of the Heart Protection Study.[15]

Finally, for some of these decisions, different recommendations for treatment may still apply even when a similar relative treatment effect seems valid and patients are at the same baseline risk. Circumstances in which different recommendations will be appropriate include:

4: Principles for using subgroup evidence for making decisions about individual patients

- Use the subgroup-specific result only when there is (unconfounded) evidence of interaction and, ideally, confirmatory evidence.
- Use the estimated overall treatment effect if there is no evidence of heterogeneity (no interaction or treatment-effect modification).
- Adjust the size of the treatment benefit (and harm) according to the patient's baseline risk.
- Consider patient preferences regarding each outcome when there are significant trade-offs in benefit and harm.

- Where the importance of different outcomes (of benefit and harm) varies for different patients
- Where patient preference varies for other reasons
- Where there are limitations in applying the trial results in a particular setting, related to such factors as the skill or experience of practitioners and access to technologies.

Principles for applying subgroup analysis to decisions about individual patients are summarised in Box 4. Cautious interpretation of the results of subgroup analyses is generally advisable.

Competing interests

None identified.

References

1. Cook DI, Gebski VJ, Keech AC. Subgroup analysis in clinical trials. *Med J Aust* 2004; 180: 289-291.
2. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001; 357: 1191-1194.
3. Moye LA. Multiple analyses in clinical trials. Fundamentals for investigators. New York: Springer, 2000: 287-289.
4. Altman DG, Matthews JNS. Statistics notes: interaction 1: heterogeneity of effects. *BMJ* 1996; 313: 486.
5. Matthews JNS, Altman DG. Statistics notes: interaction 2: compare effect sizes not P values. *BMJ* 1996; 313: 808.
6. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ* 2003; 326: 219.
7. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of subgroups of patients in randomized clinical trials. *JAMA* 1991; 266: 93-98.
8. Nuovo J, Melnikow J, Chang D. Reporting number needed to treat and absolute risk reduction in randomized controlled trials. *JAMA* 2002; 287: 2813-2814.
9. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17 187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988; ii: 349-360.
10. FTT Collaboration. Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. *Lancet* 1994; 343: 311-322.
11. White H. Hirulog and Early Reperfusion or Occlusion (HERO) 2 Trial Investigators. Thrombin-specific anticoagulation with bivalirudin versus heparin in patients receiving fibrinolytic therapy for acute myocardial infarction: the HERO-2 randomised trial. *Lancet* 2001; 358: 1855-1863.
12. PCAT Collaboration. Primary coronary angioplasty compared with intravenous thrombolytic therapy for acute myocardial infarction: six-month follow-up and analysis of individual patient data from randomized trials. *Am Heart J* 2003; 145: 47-57.
13. Hunt D, Young P, Simes J, et al, for the LIPID investigators. Benefits of pravastatin on cardiovascular events and mortality in older patients with coronary heart disease are equal to or exceed those seen in younger patients. *Ann Intern Med* 2001; 134: 931-940.
14. Hague W, Forder P, Simes J, et al, on behalf of the LIPID Investigators. Effect of pravastatin on cardiovascular events and mortality in 1516 women with coronary heart disease: results from the LIPID study. *Am Heart J* 2003; 145: 643-651.
15. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 4963 people with diabetes: a randomised placebo-controlled trial. *Lancet* 2003; 361: 2005-2016.

(Received 23 Mar 2004, accepted 23 Mar 2004) ❑