Treating Individuals 2

Subgroup analysis in randomised controlled trials: importance, indications, and interpretation

Lancet 2005; 365: 176–86

Peter M Rothwell

Stroke Prevention Research Unit, University Department of Clinical Neurology, Radcliffe Infirmary, Oxford OX2 6HE, UK (P M Rothwell FRCP)

peter.rothwell@clneuro.ox.ac.uk

Large pragmatic trials provide the most reliable data about the effects of treatments, but should be designed, analysed, and reported to enable the most effective use of treatments in routine practice. Subgroup analyses are important if there are potentially large differences between groups in the risk of a poor outcome with or without treatment, if there is potential heterogeneity of treatment effect in relation to pathophysiology, if there are practical questions about when to treat, or if there are doubts about benefit in specific groups, such as elderly people, which are leading to potentially inappropriate undertreatment. Analyses must be predefined, carefully justified, and limited to a few clinically important questions, and post-hoc observations should be treated with scepticism irrespective of their statistical significance. If important subgroup effects are anticipated, trials should either be powered to detect them reliably or pooled analyses of several trials should be undertaken. Formal rules for the planning, analysis, and reporting of subgroup analyses are proposed.

Introduction

> "The essence of tragedy has been described as the destructive collision of two sets of protagonists, both of whom are correct. The statisticians are right in denouncing subgroups that are formed post hoc from exercises in pure data dredging. The clinicians are also right, however, in insisting that a subgroup is respectable and worthwhile when established a priori from pathophysiological principles."
>
> A R Feinstein, 1998[1]

Randomised controlled trials (RCTs) and systematic reviews are the most reliable methods of determining the effects of treatments.[2–5] However, when trials were first developed for use in agriculture, researchers were presumably concerned about the effect of interventions on the overall size and quality of the crop rather than on the wellbeing of any individual plant. Clinicians have to make decisions about individuals, and how best to use results of RCTs and systematic reviews to do this has generated considerable debate.[6–22] Unfortunately, this debate has polarised, with statisticians and predominantly non-clinical (or non-practising) epidemiologists warning of the dangers of subgroup analysis and other attempts to target treatment, and clinicians warning of the dangers of applying the overall results of large trials to individual patients without consideration of pathophysiology or other determinants of individual response. This rift, described by Feinstein as a "clinicostatistical tragedy",[1] has been widened by some of the more enthusiastic proclamations on the extent to which the overall results of trials can properly inform decisions at the bedside or in the clinic.[23–25]

The results of small explanatory trials with well-defined eligibility criteria should be easy to apply, but generalisability is often undermined by highly selective recruitment, resulting in trial populations that are unrepresentative even of the few patients in routine practice who fit the eligibility criteria.[26] Recruitment of a higher proportion of eligible patients is a major strength of large pragmatic trials, but deliberately broad and sometimes ill-defined entry criteria mean that the overall result can be difficult to apply to particular groups,[27] and that subgroup analyses are necessary if heterogeneity of treatment effect is likely to occur. Yet despite the adverse effects on patient care that can result from misinterpreted or inappropriate subgroup analyses (table 1), there are no reviews or guidelines on the clinical indications for subgroup analysis and no consensus on the implications for trial design, analysis, and interpretation of subgroup effects, and the CONSORT statement on reporting of trials includes only a few lines on subgroup analysis.[28] This article discusses arguments for and against subgroup analyses, the clinical situations in which they can be useful, and rules for their performance and interpretation. Illustrative examples are taken mainly from treatments

Observation	Refutation
Aspirin is ineffective in secondary prevention of stroke in women[29,30]	31
Antihypertensive treatment for primary prevention is ineffective in women[32,33]	34
Antihypertensive treatment is ineffective or harmful in elderly people[35]	36
Angiotensin-converting enzyme inhibitors do not reduce mortality and hospital admission in patients with heart failure who are also taking aspirin[37]	38
β blockers are ineffective after acute myocardial infarction in elderly people,[39] and in patients with inferior myocardial infarction[41]	40
Thrombolysis is ineffective >6 hours after acute myocardial infarction[42]	43
Thrombolysis for acute myocardial infarction is ineffective or harmful in patients with a previous myocardial infarction[42]	44
Tamoxifen citrate is ineffective in women with breast cancer aged <50 years[45]	46
Benefit from carotid endarterectomy for symptomatic stenosis is reduced in patients taking only low-dose aspirin due to an increased operative risk[47]	48
Amlodipine reduces mortality in patients with chronic heart failure due to non-ischaemic cardiomyopathy but not in patients with ischaemic cardiomyopathy[49]	50

Table 1: **Examples of subgroup analyses that have shown apparently clinically important heterogeneity of treatment effect which has subsequently been shown to be false**

For personal use. Only reproduce with permission from Elsevier Ltd

for cerebrovascular or cardiovascular disease but the principles are relevant to all areas of medicine and surgery.

Arguments against subgroup analysis

> " . . . it would be unfortunate if desire for the perfect (ie, knowledge of exactly who will benefit from treatment) were to become the enemy of the possible (ie, knowledge of the direction and approximate size of the effects of treatment of wide categories of patient)."
>
> S Yusuf et al, 1984[4]

The main argument against subgroup analysis is that qualitative heterogeneity of relative treatment effect (defined as the treatment effect being in different directions in different groups of patients, ie, benefit in one subgroup and harm in another) is very rare.[2–5] However, this observation is much less reassuring than it seems. First, it automatically excludes most treatments because they do not have a substantial risk of harm and can only be effective or ineffective. Yet use of an ineffective treatment can be highly detrimental if this prevents the use of a more effective alternative or if adverse effects impair quality of life. Second, the

***Panel 1:* Rules of subgroup analysis: a proposed guideline for design, analysis, interpretation, and reporting**

Trial design

- Subgroups analyses should be defined before starting the trial and should be limited to a small number of clinically important questions.
- Expert clinical input into the design of subgroup analyses is needed to ensure that all relevant baseline clinical and other data are recorded.
- The direction and magnitude of anticipated subgroup effects should be stated at the outset.
- The exact definitions and categories of the subgroup variables should be defined explicitly at the outset in order to avoid post hoc data-dependent variable or category definitions. For continuous or hierarchical variables the cut-off points for analysis should be predefined.
- Stratification of randomisation by important subgroup variables should be considered.
- If important subgroup-treatment effect interactions are anticipated, trials should ideally be powered to detect them reliably.
- Trial stopping rules should take into account anticipated subgroup-treatment effect interactions and not simply the overall effect of treatment.
- If relative treatment effect is likely to be related to baseline risk, the analysis plan should include a stratification of the results by predicted risk. The risk score or model should be selected in advance so that the relevant baseline data can be recorded.

Analysis and reporting

- The above design issues should be reported in the methods section along with details of how and why subgroups were selected.
- Significance of the effect of treatment in individual subgroups should not be reported; rates of false negative and false positive results are extremely high. The only reliable statistical approach is to test for a subgroup-treatment effect interaction.
- All subgroup analyses that were done should be reported—ie, not only the number of subgroup variables but also the number of different outcomes analysed by subgroup, different lengths of follow-up etc.
- Significance of pre hoc subgroup-treatment effect interactions should be adjusted when multiple subgroup analyses are done.
- Subgroup analyses should be reported as absolute risk reductions and relative risk reductions. Where relevant the statistical significance of differences in absolute risk reductions should be tested.
- Ideally, only one outcome should be studied and this should usually be the primary trial outcome, irrespective of whether this is one outcome or a clinically important composite outcome.
- Comparability of treatment groups for prognostic factors should be checked within subgroups.
- If multiple subgroup-treatment effect interactions are identified, further analysis is needed to check whether their effects are independent.

Interpretation

- Reports of the significance of the effect of treatment in individual subgroups should be ignored, especially reports of lack of benefit in a particular subgroup in a trial in which there is overall benefit, unless there is a significant subgroup treatment effect interaction
- Genuine unanticipated subgroup-treatment effect interactions are rare (assuming that expert clinical opinion was sought in order to pre-define potentially important subgroups) and so apparent interactions that are discovered post hoc should be interpreted with caution.
  No test of significance is reliable in this situation.
- Pre hoc subgroup analyses are not intrinsically valid and should still be interpreted with caution. The false positive rate for tests of subgroup-treatment effect interaction when no true interaction exists is 5% per subgroup.
- The best test of validity of subgroup-treatment effect interactions is their reproducibility in other trials.
- Few trials are powered to detect subgroup effects and so the false negative rate for tests of subgroup-treatment effect interaction when a true interaction exists will usually be high.

For personal use. Only reproduce with permission from Elsevier Ltd

observation refers only to so-called unanticipated heterogeneity.[2–5] As outlined below, there are many examples in which qualitative heterogeneity of relative treatment effect has been correctly anticipated. Third, the observation only applies to single outcome events; it is argued that subgroup analyses based on composite outcomes are inappropriate.[2–5,51] However, since qualitative heterogeneity of relative treatment effect is only possible for treatments that have a risk of harm, and such treatments almost always need a composite outcome to express the balance of both risk and benefit, qualitative heterogeneity as defined will inevitably be rare—a Catch-22, in fact.

There are several other arguments against attempts to target treatment. First, it is said that clinicians already tend to undertreat patients,[52] and we should not risk effective treatments being further restricted. However, one of the main purposes of subgroup analysis is to extend the use of treatments to subgroups that are not currently treated in routine practice. Subgroup analyses in epidemiological studies and trials often show that benefit from treatment is likely to be more universal than expected and that current indications for treatments in routine clinical practice are inappropriately narrow, as is now clear, for example, with treatment thresholds for blood pressure lowering or lipid lowering.[53,54] Second, it is argued that subgroup analyses are almost always underpowered,[55–60] but this is simply an argument for larger trials and for meta-analysis of individual patient data. Third, it has also been argued that false positive subgroup effects might be more common than genuine heterogeneity,[2–5,55–60] and these false observations might harm patients—"subgroups kill people."[61] Subgroup analyses have certainly led to mistaken clinical recommendations (table 1), but these analyses would not have satisfied the rules suggested in panel 1. Moreover, not doing subgroup analysis can also be harmful. Properly powered subgroup analyses most commonly show that relative treatment effect is consistent across subgroups and, or, that treatments should be used more extensively than is currently the case.[53,62,63] Without such evidence, unfounded clinical concerns about possible heterogeneity or inappropriately narrow indications for treatment would reduce the use of effective treatments in routine practice.[26] Not doing subgroup analyses has very probably killed more people.

***Panel 2:* The four main clinical indications for subgroup analysis**

Potential heterogeneity of treatment effect related to risk

- Differences in risks of treatment
- Differences in risk without treatment

Potential heterogeneity of treatment effect related to pathophysiology

- Multiple pathologies underlying a clinical syndrome
- Differences in the biological response to a single pathology
- Genetic variation

Clinically important questions related to the practical application of treatment

- Does benefit differ with severity of disease?
- Does benefit differ with stage in the natural history of disease?
- Is benefit related to the timing of treatment after a clinical event?
- Is benefit dependent on comorbidity?

Underuse of treatment in routine clinical practice due to uncertainty about benefit

- Underuse of treatment in specific groups of patients eg, elderly people
- Confinement of treatment according a narrow range of values of a relevant physiological variable—eg, treatment thresholds for cholesterol level or blood pressure

Situations in which subgroup analyses should be considered

> "The tragedy of excluding cogent pathophysiologic subgroup analyses merely because they happen to be subgroups will occur if statisticians do not know the distinction, and if clinicians who do know it remain mute, inarticulate or intimidated."
>
> A R Feinstein, 1998[1]

Subgroup analyses should be predefined and carefully justified. Feinstein and others have emphasised the need for determination of pathophysiological heterogeneity, but there are three other indications for subgroup analysis (panel 2), each of which are discussed below, which are probably more important.

Heterogeneity related to risk

Clinically important heterogeneity of treatment effect is common when different groups of patients have very different absolute risks with or without treatment. The need for reliable data about risks and benefits in subgroups and individuals is greatest for potentially harmful interventions, such as warfarin or carotid endarterectomy, which are of overall benefit but that kill or disable a proportion of patients. However, evidence-based guidelines usually recommend these treatments in all cases similar to those in the relevant RCTs.[64–66] In considering this approach, it is useful to draw an analogy with the criminal justice system. Suppose that research showed that individuals charged by the police with specific crimes were usually guilty. Few would argue that they should therefore be sentenced without trial. Automatic sentencing would, on average, do more good than harm, with most criminals correctly convicted, but any avoidable miscarriages of justice are widely regarded as unacceptable. In contrast, relatively high rates of

For personal use. Only reproduce with permission from Elsevier Ltd

treatment-related death or disability (miscarriages of treatment) are tolerated by the medical scientific community precisely because, on average, treatment will do more good than harm. In both situations systems need to be in place to avoid doing harm. Yet the contrast between the effort that is put into the defence of the accused in order to avoid wrongful conviction and the very limited efforts of the medical scientific community to identify patients at high risk of harm is obvious. Admittedly, determination of guilt in a criminal trial is based on knowledge of past events, which can often be established with certainty, whereas probable benefit or harm from medical treatment depends on future events, which are usually less certain. However, the probable balance of risk and benefit in individual patients can be predicted to some extent with subgroup analysis and risk models, as has been shown, for example, with carotid endarterectomy.[67–70] In view of the fact that treatment complications are now a leading cause of death in developed countries,[71] effort is needed to more effectively target potentially harmful interventions.

Differences in the risk of a poor outcome without treatment can also lead to clinically important heterogeneity of treatment effect. Trial populations are often skewed in terms of control group risk, with a few individuals contributing much of the observed risk,[72] and treatment may be ineffective or harmful in the low risk majority. In vascular medicine, this is the case with endarterectomy for symptomatic carotid stenosis,[69] anticoagulation for uncomplicated non-valvular atrial fibrillation,[73] coronary artery bypass grafting,[74] and anti-arrhythmic drugs after myocardial infarction.[75] Clinically important heterogeneity of relative treatment effect by baseline risk has also been shown for blood pressure lowering,[76] aspirin,[77] and lipid lowering[78] in primary prevention of vascular disease, and in treatment of acute coronary syndromes with clopidogrel,[79] and with enoxaparin versus unfractionated heparin.[80,81] There are many similar examples in other areas of medicine,[82,83] and this issue is the subject of the next article in this series.

Pathophysiological heterogeneity

Differences between groups of patients in underlying pathology, biology, or genetics can each lead to clinically important heterogeneity of treatment effects. Examples will probably be identified more frequently as our understanding of the molecular mechanisms of disease is enhanced.

Multiple underlying pathologies

Clinicians often have to treat patients with ill-defined clinical syndromes, which probably have many underlying pathologies, rather than one disease. Primary generalised epilepsy is a typical example in which treatment effects differ between patients, probably because of the different underlying molecular pathologies. In vascular disease, clinically important heterogeneity of treatment effect in relation to underlying pathology is seen with thrombolysis for acute ischaemic stroke,[84,85] with aspirin in primary prevention of vascular disease (in which benefit may be largely confined to men with elevated levels of C-reactive protein,[86] probably indicating underlying atherosclerosis), and with blood pressure-lowering in secondary prevention of transient ischaemic attack and stroke, in which guidelines suggest that all patients be treated.[87–89] However, there is clinical concern about patients with carotid stenosis or occlusion in whom cerebral perfusion is often severely impaired.[90,91] Table 2 shows stroke risk by systolic blood pressure in patients with and without flow-limiting (⩾70%) carotid stenosis who were randomly assigned to medical treatment in RCTs of endarterectomy.[92] Major increases in stroke risk were noted in patients with flow-limiting stenosis, but only if systolic blood pressure <150 mm Hg: 5-year risk in patients with bilateral (⩾70%) stenosis was 64·3% versus 24·2% (p=0·002) at higher blood pressures. This difference in risk was absent in patients who had been randomly assigned to endarterectomy (13·4% *vs* 18·3%, p=0·6), suggesting a causal effect and indicating that aggressive blood pressure-lowering would very probably be harmful in patients with bilateral severe carotid disease in whom endarterectomy was not possible.

	Systolic blood pressure (mm Hg)			
	<130	130–149	150–169	>170
Stenosis group				
Bilateral <70%	1	1	1	1
Unilateral ⩾70%	1·90 (1·24–2·89) p=0·02	1·18 (0·92–1·51) p=0·30	1·27 (0·99–1·64) p=0·13	1·64 (1·15–2·33) p=0·03
Bilateral ⩾70%	5·97 (2·43–14·68) p<0·001	2·54 (1·47–4·39) p=0·001	0·97 (0·4–2·35) p=0·95	1·13 (0·50–2·54) p=0·77

The hazard ratios are derived from a Cox proportional hazards model stratified by trial and adjusted for age, sex and previous coronary heart disease. Patients with bilateral <70% stenosis are allocated a hazard of 1. ⩾70% stenosis is only consistently associated with an increase in the risk of stroke at lower levels of systolic blood pressure.

Table 2: **Hazard ratios (95% CI) for risk of stroke in patients categorised according to severity of carotid disease within pre-defined blood pressure groups[92]**

Biological heterogeneity

Subgroup analyses can also be useful when there are predictable differences in the biological response to the underlying disease. For example, perioperative administration of antilymphocyte antibodies reduces rejection in cadaveric renal transplantation by 30%,[93,94] but is expensive and has serious adverse effects. Clinical concern that benefit might depend on pre-existing immune sensitisation prompted a meta-analysis of individual patient data from five RCTs. As predicted, treatment was highly effective in sensitised patients (hazard ratio for allograft failure at 5 years=0·20, 95% CI=0·09–0·47) but was ineffective in the remaining 85% (0·97, 0·71–1·32).[94] The subgroup-treatment effect interaction was significant (p=0·009)—ie, the effect of treatment was significantly different between the subgroups. A similar pre-specified immunological subgroup analysis in a large trial of

For personal use. Only reproduce with permission from Elsevier Ltd

roxithromycin versus placebo after coronary angioplasty showed that treatment reduced restenosis and the need for revascularisation if the titre of *Chlamydia pneumoniae* antibody was high but was ineffective or harmful if the titre was low (interaction p=0·006).[95]

Genetic heterogeneity

Individuals respond differently to some drugs and this tendency can be inherited.[96,97] Genotype is an important determinant of both the response to treatment and the susceptibility to adverse reactions for a wide range of drugs.[98,99] For example, response to chemotherapy is dependent on gene expression in both colon cancer[100] and breast cancer,[101] and HDL cholesterol response to oestrogen replacement therapy is highly dependent on sequence variants in the gene encoding oestrogen receptor α.[102] In each of these cases, significant subgroup-treatment effect interactions have been reported. There is also great interest in the effects of genetics on the response to treatment in patients with HIV-1.[103] Subgroup analyses based on genotype have particular methodological problems since many genotypes may be studied and analyses will often be post hoc.

Heterogeneity related to practical application

Many of the arguments used against subgroup analyses misinterpret their main function. The main potential of subgroup analysis is not in the identification of groups that differ in their response to treatment for reasons of pathophysiology, but is in answering practical questions about how treatments should be used most effectively, such as at what stage of the disease is treatment most effective, how soon after a clinical event is treatment sufficiently safe or most effective, or how are the risks and benefits related to comorbidity? Subgroup analyses related to questions of the practical application of interventions can be vital to effective clinical practice.

Severity or stage of disease

Treatment effects often depend on severity of disease. In primary prevention of vascular disease, a pooled analysis of RCTs of pravastatin showed that the relative risk reduction with treatment increased with baseline LDL cholesterol (interaction p=0·01): relative risk reduction=3% in the lowest quintile and 29% in the two highest quintiles.[104] In stroke medicine, carotid endarterectomy is highly effective for ⩾70% recently symptomatic stenosis, modestly effective for 50–69% stenosis, but harmful for <50% stenosis (interaction p<0·0001).[105] In cardiology, thrombolysis for acute myocardial infarction is ineffective or harmful in patients with ST segment depression, but highly beneficial in patients with ST elevation (interaction p<0·01),[106] and early invasive treatment of unstable angina is of no benefit in patients with only minor ST segment change but of major

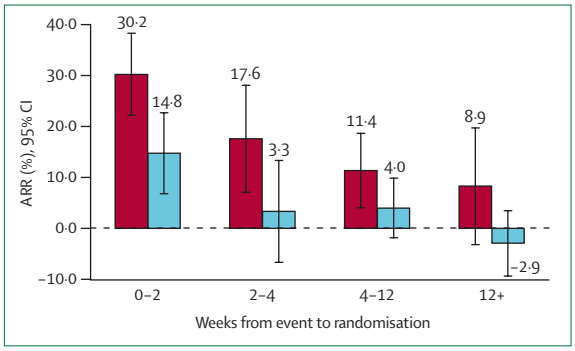

Figure 1: **Effect of carotid endarterectomy in patients with 50–69% and ⩾70% symptomatic stenosis in relation to time from last symptomatic ischaemic event to randomisation[70]**

Numbers above bars indicate actual absolute risk reduction. Vertical bars are 95% CIs. ARR=absolute risk reduction.

benefit in patients with more marked changes (interaction p=0·006).[107] The stage of disease can also determine the effect of treatment of non-vascular disease, as is seen in people with cancer,[108,109] or HIV/AIDS.[110–112]

Timing of treatment and comorbidity

Effect of treatment is often critically dependent on timing, as shown in figure 1, for benefit from endarterectomy for recently symptomatic carotid stenosis. The risk of a stroke is very high during the first few days and weeks after a transient ischaemic attack,[113] especially in patients with carotid stenosis,[114] but falls rapidly with time, as therefore does benefit from endarterectomy.[70] Similar time-dependence has been shown for benefit from thrombolysis for both acute myocardial infarction[106] and acute ischaemic stroke.[115]

Treatment effects may also depend on comorbidity. For example, angiotensin-converting enzyme inhibitors and angiotensin II receptor blocking drugs are harmful in patients with renovascular disease but highly beneficial in other hypertensive patients.[116] Benefit from diltiazem after myocardial infarction may depend on the presence of heart failure because of the negative chronotropic and inotropic effects of the drug.[117]

Underuse of treatment in specific groups

Treatments that are effective in trials are often underused in specific groups of patients in routine practice. For example, statins were not used in elderly people for many years until the drugs were proved highly effective by subgroup analysis in the Heart Protection Study.[53] Proof of some benefit by subgroup analysis was also needed to counter underuse in elderly patients of thrombolysis for acute myocardial infarction in elderly people,[106] and similar underuse of endarterectomy for symptomatic carotid stenosis.[70] In each case, treatment had already been shown to be highly effective overall. Use of treatment in routine clinical practice is also often inappropriately limited to patients with measurements of

For personal use. Only reproduce with permission from Elsevier Ltd

Day of birth	Events/patients: Surgical	Medical	ARR (%)	95% CI	p value
Sunday	7/56	6/41	3·1	−11·3 to 17·5	0·34
Monday	4/66	10/44	16·7	3·0 to 30·3	0·008
Tuesday	8/76	6/28	10·5	−6·9 to 27·9	0·12
Wednesday	8/67	13/47	18·3	2·3 to 34·2	0·01
Thursday	9/75	9/36	12·8	−3·8 to 29·4	0·07
Friday	1/56	6/37	15·1	2·3 to 27·9	0·01
Saturday	6/51	8/41	9·5	−6·6 to 25·6	0·12
Total	43/447	58/274	12·3	6·5 to 18·1	<0·001

Heterogeneity: p=0·83

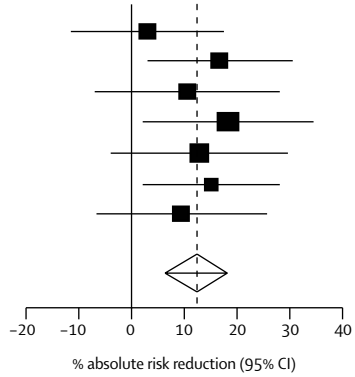

Figure 2: **Effect of carotid endarterectomy in patients with ⩾70% symptomatic stenosis in ECST[126] according to day of week on which patients were born**

physiological parameters above specific arbitrary cut-off points, such as treatment thresholds for blood pressure and total cholesterol in prevention of vascular disease. There is increasing evidence from subgroup analysis in large trials that such thresholds are inappropriate.[53,87] Proof of the generalisability of benefit is a major function of subgroup analysis. However, such analyses should be sufficiently powered to detect benefit, and pooled analyses of multiple trials will often be needed for subgroups such as elderly people who are commonly under-represented in trials.[26]

Estimation and interpretation of subgroup effects

> "Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise."
>
> J W Tukey, 1962[118]

Multiplicity, post hoc analyses, and publication bias

In one trial of β blockers after myocardial infarction,[119] 146 subgroup analyses were done,[120] several of which showed apparent differences in the effect of treatment. However, none of the differences were confirmed by subsequent studies.[40] Pocock reviewed 50 trials published in major journals in 1997 and noted that 70% reported a median of four subgroup analyses,[55] which was little changed from 10 years previously.[121] The reliability of these subgroups depends to a great extent on whether they were predefined and how many other analyses were done but not reported. Selective reporting of post hoc subgroup observations, which are generated by the data rather than tested by them, is analogous to placing a bet on a horse after watching the race. There is certainly evidence of selective reporting of significant analyses,[122–124] but this is difficult to judge when assessing an individual trial. The only solution is for a small number of potentially important subgroups to be pre-defined in the trial protocol, along with their anticipated directions. Post hoc observations are not automatically invalid (many medical discoveries have been fortuitous), but they should be regarded as unreliable unless they can be replicated.

Statistical significance

Subgroup analyses can be wrong in two ways. First, they can falsely indicate that treatment is beneficial in a particular subgroup when the trial shows no overall effect—the situation in which subgroup analyses are most commonly done.[56,57] Simulations of RCTs powered to determine the overall effect of treatment suggest that false subgroup effects will be noted by chance in 7%–21% of analyses depending on other factors.[58] More commonly (in 41%–66% of simulated subgroups) simulations can falsely indicate that there is no treatment effect in a particular subgroup when the trial shows benefit overall.[58] Benefit is most likely to be absent in small subgroups, which probably explains the recurrent and usually mistaken finding that treatments are ineffective in women[29,32,125] and in elderly people,[32,35] who tend to be under-represented in RCTs.[26] The correct analysis is not the significance of the treatment effect in one subgroup or the other, but whether the effect differed significantly between the subgroups—the test of subgroup-treatment effect interaction. For example, although endarterectomy for severe stenosis in the European Carotid Surgery Trial (ECST)[126] was only significantly beneficial in patients born on specific days of the week (figure 2), this was, of course, due to chance and there was no subgroup-treatment effect interaction (p=0·83). Data from simulation studies have shown that tests of subgroup-treatment effect interaction are reliable, with a false positive rate of 5% at p<0·05, which is robust to differences in the size of subgroups, the number of categories, and to continuous data.[58] However, although testing of subgroup-treatment effect interactions is widely recommended,[51,55–57,121] Pocock's review showed that 37% of RCTs reported only p values for treatment effect within subgroups and only 43% reported tests of interaction.[55]

For personal use. Only reproduce with permission from Elsevier Ltd

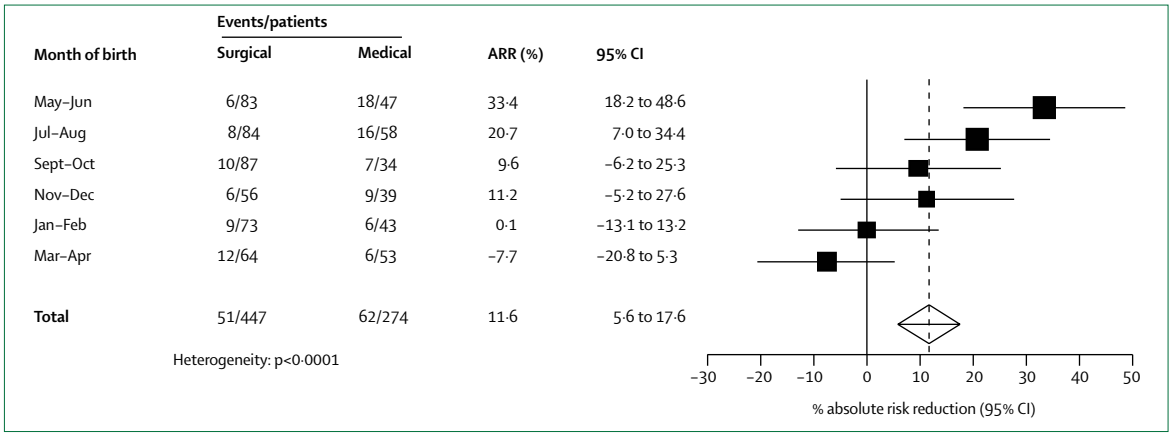

Figure 3: **Effect of carotid endarterectomy in patients with ⩾70% symptomatic stenosis in ECST[126] according to month of birth in six 2 month periods**

Chance

The effect of chance on subgroup analyses is usually illustrated with the ISIS-2 trial example (aspirin *vs* placebo in acute myocardial infarction), in which aspirin was ineffective in patients born under the star signs of Libra and Gemini (150 deaths on aspirin *vs* 147 on placebo, 2p=0·5), but was beneficial in the remainder (654 deaths on aspirin *vs* 869 on placebo, 2p<<0·0001).[3–5] The significance of this subgroup treatment effect interaction has never been reported, but it seems to be p=0·01 (Breslow Day test). However, Libra and Gemini are not adjacent on the Zodiac and merely splitting a trial of an effective treatment into 12 subgroups and comparing the two subgroups with the least evidence of benefit with the remainder will almost inevitably produce substantial heterogeneity. A more appropriate test of the subgroup-treatment effect interaction across the 12 separate birth signs would undoubtedly be non-significant in ISIS-2. However, highly significant interactions can occur by chance. Figure 3 shows the effect of endarterectomy for severe carotid stenosis by month of birth in the ECST (interaction: p<0·001 across the 12 months). The remarkable trend in benefit (p<0·0000001), highest in patients born in May (absolute risk reduction=37·5%, 16·3–58·7) or June (29·7%, 8·1–51·1) and falling smoothly to possible harm in March (–7·2%, –22·3 to 7·9) and April (–10·5%, –32·8 to 11·8), would have been very difficult to ignore if it had been in relation to age, blood pressure, or some other plausible variable, illustrating the unreliability of unanticipated subgroup effects. One of the most damaging unanticipated subgroup interactions (p=0·003) was the observation in the Canadian Cooperative Study Group trial[29] that aspirin was effective in preventing stroke and death in men (RR=0·52, p<0·005) but not in women (1·42, p=0·35). Women were undertreated for at least a decade before subsequent trials and overviews suggested benefit.

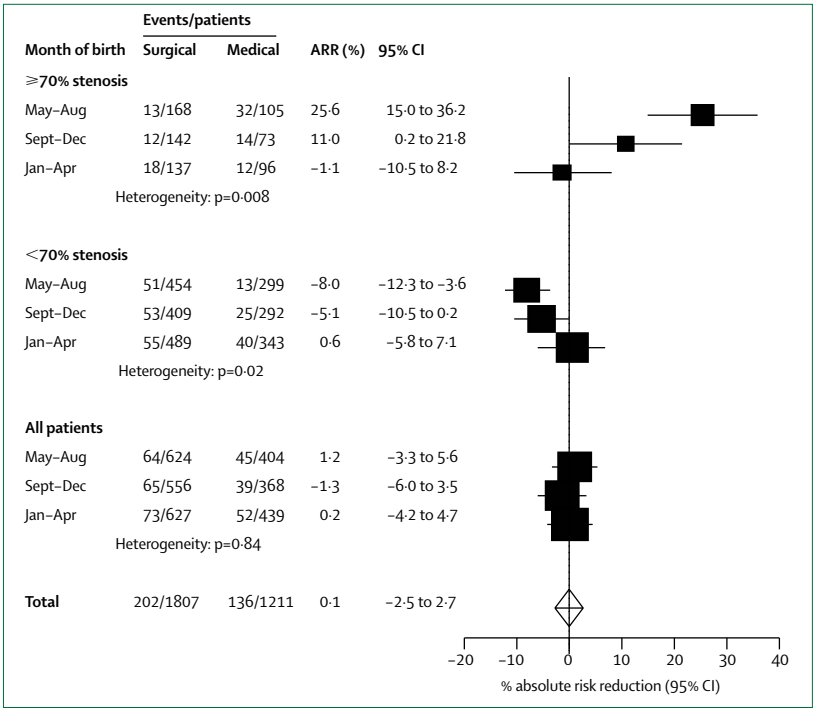

Figure 4: **Effect of carotid endarterectomy in ECST[126] according to month of birth in three 4 month periods in patients with ⩾70%, <70%, and all degrees of symptomatic stenosis**

Replication

The best test of the validity of subgroup analyses is not significance but replication. For example, month of birth interaction was not replicated in ECST patients with <70% stenosis (figure 4) or in other trials, whereas the effect of the timing of surgery on benefit from endarterectomy was present in the two different stenosis groups (figure 1) and in an independent trial (figure 5). For post-hoc analyses, replication is absolutely essential irrespective of plausibility or significance. For example, a rigorous pooled analysis of RCTs of tamoxifen citrate in breast cancer showed that treatment was ineffective in women aged <50 years (mainly pre-menopausal) but very effective in older women (mainly post-menopausal).[45] The interaction was

For personal use. Only reproduce with permission from Elsevier Ltd

highly significant (p<0·0001), but was not replicated in subsequent trials. Similarly, a large RCT of a calcium antagonist in chronic heart failure showed no reduction in mortality in patients with ischaemic cardiomyopathy (RR=1·04, 0·83-1·29) but major benefit (0·64, 0·37-0·79) in patients with non-ischaemic cardiomyopathy (interaction p=0·004).[49] This effect was also manifest as a difference between patients with and without angina (RR=1·09, 0·84-1·42 *vs* 0·59, 0·44-0·81, p=0·002). However, the direction of these interactions was opposite to that expected, and a subsequent trial failed to confirm the benefit in non-ischaemic cardiomyopathy.[50]

Anticipated subgroup-treatment effect interactions that are underpowered but reproducible are more reliable than unanticipated interactions no matter how significant. For example, although an early RCT of coronary artery bypass grafting suggesting that survival benefit was mainly confined to patients with left main coronary artery disease or three-vessel disease had only a few hundred patients,[127] the observation was biologically plausible and was reproduced in a subsequent trial.[128] However, it was not until 20 years later that a pooled analysis of seven RCTs had sufficient power to demonstrate a significant interaction.[74]

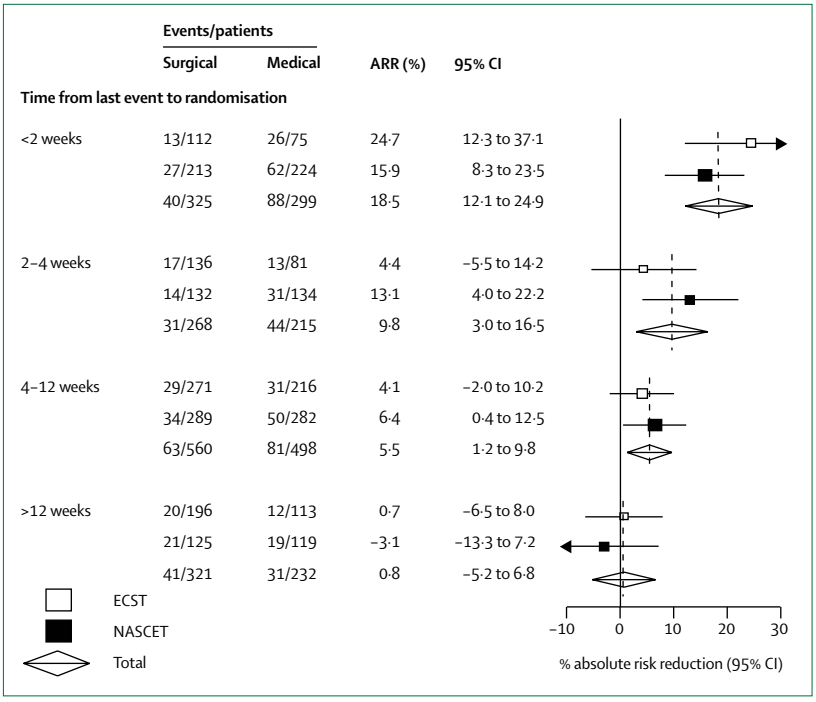

Figure 5: **Effect of carotid endarterectomy in patients with 50–99% symptomatic stenosis in relation to time from last symptomatic ischaemic event to randomisation**
Data taken from ECST and North American Symptomatic Carotid Endarterectomy Trial.[70]

Power, choice of outcome, and prognostic imbalance

If trials are powered to determine the overall effect of treatment, virtually all subgroup analyses will be underpowered. If a genuine subgroup-treatment effect interaction exists, the chance of a false negative result with a formal test of interaction will therefore be far greater than the 5% false positive rate in a trial in which no true interaction exists. The ability of formal tests of interaction to correctly identify subgroup effects also depends on the size of the interaction relative to the overall treatment effect. For example, if a trial has 80% power to detect the overall effect of treatment (not uncommon), reliable detection of an interaction of the same magnitude as the overall effect (ie, potentially clinically important) would need a four-fold greater sample size.[58]

Rules for stopping trials are usually based on the demonstration of an overall effect of treatment and take no account of the need for data about the effect of treatment in subgroups. If potentially important subgroup effects are anticipated, then there should be separate stopping rules for different subgroups so that adequate numbers of patients are recruited into each. This approach was used very effectively in the trials of endarterectomy for symptomatic carotid stenosis in which there was independent stopping and reporting in different subgroups of patients according to the degree of carotid stenosis.[70]

The choice and number of outcomes studied will affect the validity of subgroup analyses. Outcomes that are not directly affected by treatment, as is often the case for all-cause mortality, should be avoided. For example, in the Medical Research Council trial of blood pressure lowering in mild hypertension, treatment reduced all-cause mortality in men but increased mortality in women.[33] However, the excess mortality in women on active treatment was due entirely to non-cardiovascular deaths and treatment reduced the risk of stroke in both sexes. Ideally, only one outcome should be studied and this should usually be the primary outcome for the trial, irrespective of whether it is one or a composite outcome. For example, sex had no effect on benefit from carotid endarterectomy for 50%–69% symptomatic stenosis in the pooled RCTs if the analysis was based on the risk of ipsilateral ischaemic stroke only (ie, the outcome that surgery prevents), but there was harm in women and benefit in men (interaction p=0·008) if the clinically relevant composite outcome of ipsilateral ischaemic stroke plus operative stroke or death (the primary outcome in the trials) was considered.[70] The overall effect of surgery depends on the balance of two outcomes which have different mechanisms and risk factors. Women had a lower risk of stroke on medical treatment than men (hazard ratio=0·79, 0·64-0·97, p=0·03) but a higher operative risk (1·50, 1·14-1·97, p=0·004).[70] In this situation there is an argument for modelling risk and benefit separately but patients and clinicians still need to know what the overall effect of treatment on the composite outcome is.

In large trials, randomisation ensures that the prognosis in the different treatment groups is similar at baseline, but this cannot be assumed in subgroups

For personal use. Only reproduce with permission from Elsevier Ltd

unless randomisation was appropriately stratified.[60] It is important to check that differences in treatment effect between subgroups are not attributable to baseline imbalances between the treatment arms, although the power of testing for balance between treated and control arms in subgroups will usually be low. It is also important to understand that subgroup effects are not necessarily independent. For example, in the pooled analysis of RCTs of carotid endarterectomy, benefit from surgery across all degrees of carotid stenosis was similar in patients with ocular ischaemic events versus hemispheric transient ischaemic attack or stroke.[70] However, the mean degree of carotid stenosis was higher in patients with retinal events (54% *vs* 41%, p<0·0001), which would increase the likelihood of benefit from surgery. Surgery was more effective in patients with hemispheric than ocular events when subgroup analyses were stratified by degree of stenosis.[70]

Conclusions

Large randomised controlled trials with broad eligibility criteria and high inclusion rates provide the most reliable data about the effects of treatments, but these should be designed, analysed, and reported in a way that allows clinicians to use the results as effectively as possible in routine practice. Subgroup analyses can be useful if there are widely differing risks of a poor outcome with or without treatment between specific groups, if there are important differences in pathophysiology that might influence the effect of treatment, if there is uncertainty about when to treat, or if there is undertreatment in specific groups in routine clinical practice. Clinical concerns about heterogeneity of treatment effects will often be unfounded, but if they are not addressed they will restrict the use of treatment in routine practice. A limited number of clinically important analyses must be carefully predefined and justified, and post-hoc observations should be treated with scepticism irrespective of their significance. Adherence to the guidelines for planning, analysis, and reporting of subgroup analyses proposed in panel 1 would increase reliability.

Conflict of interest statement

I declare that I have no conflict of interest.

References

1 Feinstein AR. The problem of cogent subgroups: A clinicostatistical tragedy. *J Clin Epidemiol* 1998; **51:** 297–99.
2 Peto R. Statistics of cancer trials. In Halan KE ed. Treatment of Cancer. London: Chapman & Hall, 1981.
3 Collins R, Peto R, Gray R, Parish S. Large-scale randomised evidence: trials and overviews. In Weatherall DJ, Ledingham JGG, Warrell DA, eds. Oxford Textbook of Medicine Oxford. Oxford University Press, 1996: 21–32.
4 Yusef S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984; **3:** 409–22.
5 Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. *Lancet* 2001; **357:** 373–80.
6 Bailey KR. Generalising the results of randomised clinical trials. *Control Clin Trials* 1994; **15:** 15–23.
7 Davey Smith G, Egger M, Phillips AN. Beyond the grand mean. *BMJ* 1997; **315:** 1610–14.
8 Wittes RE. Problems in the medical interpretation of overviews. *Stat Med* 1987; **6:** 269–76.
9 Lau J, Ionnidis JPA, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet* 1998; **351:** 123–27.
10 Ionnidis JPA, Lau J. Uncontrolled pearls, controlled evidence, meta-analysis and the individual patient. *J Clin Epidemiol* 1998; **51:** 709–11.
11 Chalmers I. A patient's attitude to the use of research evidence for guiding individual choices and decisions in healthcare. *Clin Risk* 2000; **6:** 227–30.
12 Sullivan FM, MacNaughton RJ. Evidence in consultations: interpreted and individualised. *Lancet* 1996; **348:** 941–43.
13 Senn S. Applying results of randomised trials to patients. N of 1 trials are needed. *BMJ* 1998; **317:** 537–38.
14 Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992; **116:** 78–84.
15 Black D. The limitations of evidence. *J R Coll Physicians Lond* 1998; **32:** 23–26
16 Hampton JR. Size isn't everything. *Stat Med* 2002; **21:** 2807–14.
17 Caplan LR. Evidence based medicine: concerns of a clinical neurologist. *J Neurol Neurosurg Psychiatry* 2001; **71:** 569–74.
18 Evans JG. Evidence-based and evidence-biased medicine. *Age Ageing* 1995; **24:** 461–63.
19 Swales JD. Evidence-based medicine and hypertension. *J Hypertens* 1999; **17:** 1511–16.
20 Feinstein AR, Horwitz RI. Problems in the "evidence" of "evidence-based medicine". *Am J Med* 1997; **103:** 529–35.
21 Fahey T. Applying the results of clinical trials to patients in general practice: perceived problems, strengths, assumptions, and challenges for the future. *Br J Gen Pract* 1998; **48:** 1173–78.
22 Mant D. Can randomised trials inform clinical decisions about individual patients? *Lancet* 1999; **353:** 743–46.
23 Sackett DL, Straus SE. Finding and applying evidence during clinical rounds: the "evidence cart". JAMA 1998; **280:** 1336–38.
24 Sackett DL. Applying overviews and meta-analyses at the bedside. *J Clin Epidemiol* 1995; **48:** 61–70.
25 Ellis J, Mulligan I, Rowe J, Sackett DL. Inpatient general medicine is evidence based. *Lancet* 1995; **346:** 407–10.
26 Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?". *Lancet* 2005; 365: 82–93.
27 Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet* 1995; **345:** 1616–19.
28 Altman DG, Schulz KF, Moher D, et al, for the CONSORT Group. The revised CONSORT statement for reporting randomised trials: explanation and elaboration. *Ann Intern Med* 2001; **134:** 663–94.
29 The Canadian Cooperative Study Group. A randomised trial of aspirin and sulfinpyrazone in threatened stroke. *N Engl J Med* 1978; **299:** 53–59.
30 Fields WS, Lemak NA, Frankowski RF, Hardy RJ. Controlled trial of aspirin in cerebral ischaemia. *Stroke* 1977; **8:** 301–14.
31 Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy. I: Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. *BMJ* 1994; **308:** 81–106.
32 Anastos K, Charney P, Charon RA, et al. Hypertension in women: what is really known? *Ann Intern Med* 1991; **115:** 287–93.
33 Medical Research Council Working Party. MRC trial of treatment of mild hypertension: principal results. *BMJ* 1985; **291:** 97–104.
34 Gueyffier F, Boutitie F, Boissel JP, et al. Effect of antihypertensive drug treatment on cardiovascular outcomes in men and women. *Ann Intern Med* 1997; **126:** 761–67.
35 Amery A, Birkenhager W, Brixko P, et al. Influence of antihypertensive drug treatment on morbidity and mortality in patients over the age of 60 years. European Working Party on High blood pressure in the Elderly (EWPHE) results: sub-group analysis on entry stratification. *J Hypertens Suppl* 1986; **4:** S642–47.
36 Gueyffier F, Bulpitt C, Boissel JP, et al. Antihypertensive drugs in very old people: a subgroup meta-analysis of randomised controlled trials. *Lancet* 1999; **353:** 793–96.
37 Cleland JGF, Bulpitt CJ, Falk RH, et al. Is aspirin safe for patients with heart failure? *Br Heart J* 1995; **74:** 215–19.

For personal use. Only reproduce with permission from Elsevier Ltd

38 Flather MD, Yusuf S, Kober L, et al. Long-term ACE-inhibitor therapy in patients with heart failure or left-ventricular dysfunction: a systematic overview of data from individual patients. *Lancet* 2000; **355:** 1575–81.

39 Anderson MP, Bechsgaard P, Frederiksen J, et al. Effects of alprenolol on mortality among patients with definite or suspected acute myocardial infarction: preliminary results. *Lancet* 1979; **2:** 865–68.

40 Yusuf S, Peto R, Lewis J, Collins R, Sleight P. β blockade during and after acute myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis* 1985; **27:** 335–71.

41 Multicenter International Study: Supplemental report: reduction in mortality after myocardial infarction with long-term β-adrenoreceptor blockade. *BMJ* 1977; **2:** 419–21.

42 Gruppo Italiano per lo Studio della Streptochinasi nell'Infarcto Myocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet* 1986; **1:** 397–402.

43 ISIS-2 Collaborative Group. Randomised trial of IV streptokinase, oral aspirin, both or, neither among 17 187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988; **2:** 349–60.

44 Fibrinolytic Therapy Trialists' (FTT) Collaborative Group. 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:** 311–22.

45 Early Breast Cancer Trialists' Collaborative Group. Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials among 28,896 women. *N Engl J Med* 1988; **319:** 1681–92.

46 Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer (Cochrane Review). *Cochrane Database Syst Rev* 2001; **1:** CD000486.

47 Barnett HJ, Taylor DW, Eliasziw M, et al. Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis. North American Symptomatic Endarterectomy Carotid Trial Collaborators. *N Engl J Med* 1998; **339:** 1415–25.

48 Taylor DW, Barnett HJ, Haynes RB, et al. Low-dose and high-dose acetylsalicylic acid for patients undergoing carotid endarterectomy: a randomised controlled trial. ASA and Carotid Endarterectomy (ACE) Trial Collaborators. *Lancet* 1999; **353:** 2179–84.

49 Packer M, O'Connor CM, Ghali JK, et al. Effect of amlodipine on morbidity and mortality in severe chronic heart failure. *N Engl J Med* 1996; **335:** 1107–14.

50 Wijeysundera HC, Hansen MS, Stanton E, et al. Neurohormones and oxidative stress in nonischemic cardiomyopathy: relationship to survival and the effect of treatment with amlodipine. *Am Heart J* 2003; **146:** 291–97.

51 Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomised clinical trials. *JAMA* 1991; **266:** 93–98.

52 Rose G. High-risk and population strategies of prevention: ethical considerations. *Ann Med* 1989; **21:** 409–13.

53 Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet* 2002; **360:** 7–22.

54 Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 2002; **360:** 1903–13

55 Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; **355:** 1064–69.

56 Stallones RA. The use and abuse of subgroup analysis in epidemiological research. *Prev Med* 1987; **16:** 183–94.

57 Pocock SJ, Hughes MD. Estimation issues in clinical trials and overviews. *Stat Med* 1990; **9:** 657–71.

58 Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess* 2001 **5:** 1–56.

59 Moreira ED, Stein Z, Susser E. Reporting on methods of subgroup analysis in clinical trials: a survery of four scientific journals. *Braz J Med Biol Res* 2001; **34:** 1441–46.

60 Cui L, Hung HNJ, Wang SJ, Tsong Y. Issues related to subgroup analysis in clinical trials. *J Biopharm Stat* 2002; **12:** 347–58.

61 van Gijn J. Extrapolation of trial data into practice: where is the limit? *Cerebrovasc Dis* 1995; **5:** 159–62.

62 International Stroke Trial Collaborative Group. The International Stroke Trial (IST): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke. *Lancet* 1997; **349:** 1569–81.

63 Fourth International Study of Infarct Survival Collaborative Group. ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58 050 patients with suspected acute myocardial infarction. *Lancet* 1995; **345:** 669–85.

64 Intercollegiate Stroke Working Party. National Clinical Guidelines for Stroke 2nd edition. London: Royal College of Physicians of London, 2000.

65 Biller J, Feinberg WM, Castaldo JE, et al. Guidelines for carotid endarterectomy: a statement for healthcare professionals from a Special Writing Group of the Stroke Council, American Heart Association. *Stroke* 1998; **29:** 554–62.

66 Pearson TA, Blair SN, Daniels SR, et al. AHA guidelines for primary prevention of cardiovascular disease and stroke: 2002 update: consensus panel guide to comprehensive risk reduction for adult patients without coronary or other atherosclerotic vascular diseases. *Circulation* 2002; **106:** 388–91.

67 Rothwell PM, Slattery J, Warlow CP. A systematic comparison of the risks of stroke and death due to carotid endarterectomy for symptomatic and asymptomatic stenosis. *Stroke* 1996; **27:** 266–69.

68 Bond R, Rerkasem K, Rothwell PM. A systematic review of the risks of carotid endarterectomy in relation to the clinical indication and the timing of surgery. *Stroke* 2003; **34:** 2290–301.

69 Rothwell PM, Warlow CP, on behalf of the ECST. Prediction of benefit from carotid endarterectomy in individual patients: a risk modelling study. *Lancet* 1999; **353:** 2105–10.

70 Rothwell PM, Eliasziw M, Gutnikov SA, et al. Effect of endarterectomy for recently symptomatic carotid stenosis in relation to clinical subgroups and the timing of surgery. *Lancet* 2004; **363:** 915–24.

71 Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalised patients. *JAMA* 1998; **279:** 1200–05.

72 Ionnidis JPA, Lau J. The impact of high-risk patients on the results of clinical trials. *J Clin Epidemiol* 1997; **50:** 1089–98.

73 Laupacis A, Boysen G, Connolly S, et al. Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Analysis of pooled data from five randomised controlled trials. *Arch Intern Med* 1994; **154:** 1449–57.

74 Yusuf S, Zucker D, Peduzzi P, et al. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists' Collaboration. *Lancet* 1994; **344:** 563–70.

75 Boissel JP, Collet JP, Lievre M, Girard P. An effect model for the assessment of drug benefit: example of antiarrhythmic drugs in postmyocardial infarction patients. *J Cardiovasc Pharmacol* 1993; **22:** 356–63

76 Li W, Gueyffier F, Boissel JP, Girard P, Boutitie F, Cucherat M. Identification and prediction of responders to a therapy: a model and its preliminary application to actual data. *Arch Mal Coeur Vaiss* 1998; **91:** 1059–63. [In French].

77 Sanmuganathan PS, Ghahramani P, Jackson PR, Wallis EJ, Ramsay LE. Aspirin for primary prevention of coronary heart disease: safety and absolute benefit related to coronary risk derived from meta-analysis of randomised trials. *Heart* 2001; **85:** 265–71.

78 West of Scotland Coronary Prevention Group. West of Scotland Coronary Prevention Study: identification of high-risk groups and comparison with other cardiovascular intervention trials. *Lancet* 1996; **348:** 1339–42.

79 Bundaj A, Yusuf S, Mehta SR, et al. Benefit of clopidogrel in patients with acute coronary syndromes without ST-segment elevation in various risk groups. *Circulation* 2002; **106:** 1622–26.

80 Antman EM, Cohen M, Bernink PJ, et al. The TIMI risk score for unstable angina/non-ST elevation MI. A method for prognostication and therapeutic decision making. *JAMA* 2000; **284:** 835–42.

For personal use. Only reproduce with permission from Elsevier Ltd

81 Cohen M, Demers C, Gurfinkel EP, et al. A comparison of low molecular weight heparin with unfractionated heparin for unstable coronary artery disease. *N Engl J Med* 1997; **337:** 447–52.

82 Pagliaro L, D'Amico G, Soronson TIA, et al. Prevention of bleeding in cirrhosis. *Ann Intern Med* 1992; **117:** 59–70.

83 International Study of Unruptured Intracranial Aneurysms Investigators. Unruptured intracranial aneurysms—risks of rupture and risks of surgical intervention. *N Engl J Med* 1998; **339:** 1725–33.

84 Alder SJ, Moody AR, Martel AL, et al. Limitations of clinical diagnosis in acute stroke. *Lancet* 1999; **354:** 1523.

85 Hacke W, Brott T, Caplan L, et al. Thrombolysis in acute ischemic stroke: controlled trials and clinical experience. *Neurology* 1999; **53** (suppl 4)**:** S3–14.

86 Ridker PM, Cushman M, Stampfer MJ, Tracy RP, Hennekens CH. Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *N Engl J Med* 1997; **336:** 973–79.

87 PROGRESS Collaborative Group. Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6105 individuals with previous stroke or transient ischaemic attack. *Lancet* 2001; **358:** 1033–41.

88 No authors listed. National clinical guidelines for stroke: a concise update. *Clin Med* 2002; **2:** 231–33.

89 McAlister FA, Zarnke KB, Campbell NR, et al. The 2001 Canadian recommendations for the management of hypertension: Part two-Therapy. *Can J Cardiol* 2002; **18:** 625–41

90 Van der Grond J, Balm R, Kappelle J, Eikelboom BC, Mali WP. Cerebral metabolism of patients with stenosis or occlusion of the internal carotid artery. *Stroke* 1995; **26:** 822–28.

91 Grubb RL, Derdeyn CP, Fritsch SM, et al. Importance of hemodynamic factors in the prognosis of symptomatic carotid occlusion. *JAMA* 1998; **280:** 1055–60.

92 Rothwell PM, Howard SC, Spence D. Relationship between blood pressure and stroke risk in patients with symptomatic carotid occlusive disease. *Stroke* 2003; **34:** 2583–90.

93 Szczech LA, Berlin JA, Aradhye S, Grossman RA, Feldman HI. Effect of anti-lymphocyte induction therapy on renal allograft survival: a meta-analysis. *J Am Soc Nephrol* 1997; **8:** 1771–77.

94 Szczech LA, Berlin JA, Feldman HI. The effect of antilymphocyte induction therapy on renal allograft survival. A meta-analysis individual patient-level data. Anti-lymphocyte antibody induction therapy study group. *Ann Intern Med 1998*; **128:** 817–26.

95 Neumann F, Kastrati A, Miethke T, et al. Treatment of Chlamydia pneumoniae infection with roxithromycin and effect on neointima proliferation after coronary stent placement (ISAR-3): a randomised, double-blind, placebo-controlled trial. *Lancet* 2001; **357:** 2085–89.

96 Kalow W, Gunn DR. Some statistical data on atypical cholinesterase of human serum. *Ann Hum Genet* 1959; **23:** 239–50.

97 Price Evans DA, Manley KA, McKusick VA. Genetic control of isoniazid metabolism in man. *BMJ* 1960; **2:** 485–91.

98 Weinshilboum R. Inheritance and drug response. *N Engl J Med* 2003; **348:** 529–37.

99 Philips KA, Veenstra DL, Oren E, Lee JK, Sadee W. Potential role of pharmacogenomics in reducing adverse drug reactions. *JAMA* 2001; **286:** 2270–79.

100 Barratt PL, Seymour MT, Stenning SP, et al. DNA markers predicting benefit from adjuvant flourouracil in patients with colon cancer: a molecular study. *Lancet* 2002; **360:** 1381–91.

101 Chang JC, Wooten EC, Tsimelzon A, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* 2003; **362:** 362–69.

102 Herrington DM, Howard TD, Hawkins GA, et al. Estrogen-receptor polymorphisms and effects of estrogen replacement on high-density lipoprotein cholesterol in women with coronary disease. *N Engl J Med* 2002; **346:** 967–74.

103 Telenti A, Aubert V, Spertini F. Individualising HIV treatment—pharmacogenetics and immunogenetics. *Lancet* 2002; **359:** 722–23.

104 Sacks FM, Tonkin AM, Shepherd J, et al. Effect of pravastatin on coronary disease events in subgroups defined by coronary risk factors. *Circulation* 2000; **102:** 1893–900.

105 Rothwell PM, Eliasziw M, Gutnikov SA, et al. Pooled analysis of individual patient data from randomised controlled trials of endarterectomy for symptomatic carotid stenosis. *Lancet* 2003; **361:** 107–16.

106 Fibrinolytic Therapy Trialists' (FTT) Collaborative Group. Indications for fibrinolytic thereapy 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:** 311–22.

107 Holmvang L, Clemmensen P, Lindahl B, et al. Quantitative analysis of the admission electrocardiogram identifies patients with unstable coronary artery disease who benefit the most from early invasive treatment. *J Am Coll Cardiol* 2003; **41:** 905–15.

108 PORT Meta-analysis Trialists Group. Postoperative radiotherapy in non-small-cell lung cancer: systematic review and meta-analysis of individual patient data from nine randomised controlled trials. *Lancet* 1998; **352:** 257–63.

109 Gale RP, Horowitz MM. How best to analyse new strategies in bone marrow transplantation. *Bone Marrow Transplant* 1990; **6:** 357–59.

110 Fischl MA, Richman DD, Grieco MH, et al. The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. A double-blind, placebo-controlled trial. *N Engl J Med* 1987; **317:** 185–91.

111 Concorde Coordinating Committee. Concorde: MRC/ANRS randomised double-blind controlled trial of immediate and deferred zidovudine in symptom-free HIV infection. *Lancet* 1994; **343:** 871–81.

112 Egger M, Neaton JD, Phillips AN, Davey Smith G. Concorde trial of immediate versus deferred zidovudine. *Lancet* 1994; **343:** 1355.

113 Coull A, Lovett JK, Rothwell PM, on behalf of the Oxford Vascular Study. Early risk of stroke after a TIA or minor stroke in a population-based incidence study. *BMJ* 2004; **328:** 326.

114 Lovett JK, Coull A, Rothwell PM. Early risk of recurrent stroke by aetiological subtype: implications for stroke prevention. *Neurology* 2004; **62:** 569–73.

115 Wardlaw JM, Warlow CP, Counsell C. Systematic review of evidence on thrombolytic therapy for acute ischaemic stroke. *Lancet* 1997; **350:** 607–14.

116 Brown MJ. Matching the right drug to the right patient in essential hypertension. *Heart* 2001; **86:** 113–20.

117 Multicenter Diltiazem Postinfarction Trial Research Group. The effect of diltiazem on mortality and reinfarction after myocardial infarction. *N Engl J Med* 1988; **319:** 385–92.

118 Tukey JW. The future of data analysis. *Ann Math Stat* 1962; **33:** 13–14.

119 β-blocker Heart Attack Trial Research Group. A randomized trial of propranolol in patients with acute myocardial infarction. I. Mortality results. *JAMA* 1982; **247:** 1707–14.

120 Furberg CD, Byington RP. What do subgroup analyses reveal about differential response to β-blocker therapy? The β-Blocker Heart Attack Trial experience. *Circulation* 1983; **67:** I98–I110.

121 Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* 1987; **317:** 426–32.

122 Tannock IF. False positive results in clinical trials: multiple significance tests and the problem of unreported comparisons. *J Natl Cancer Inst* 1996; **88:** 206–07.

123 Gelber RD, Goldhirsch A. Interpretation of results from subset analyses within overviews of randomised clinical trials. *Stat Med* 1987; **6:** 371–88.

124 Hahn S, Williamson PR, Hutton L, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analysis within studies. *Stat Med* 2000; **19:** 3325–36.

125 Frasure-Smith N, Lesperance F, Prince RH, et al. Randomised trial of home-based psychological nursing intervention for patients recovering from myocardial infarction. *Lancet* 1997; **350:** 473–79.

126 European Carotid Surgery Trialists' Collaborative Group. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: final results of the MRC European Carotid Surgery Trial (ECST). *Lancet* 1998; **351:** 1379–87.

127 Takaro T, Hultgren HN, Lipton MJ, Detre KM. The VA cooperative randomised study of surgery for coronary arterial occlusive disease. II. Subgroup with significant left main lesions. *Circulation* 1976; **54** (6 suppl)**:** III107–17.

128 European Coronary Surgery Study Group. Long-term results of a prospective randomised study of coronary artery bypass surgery in stable angina patients. *Lancet* 1982; **2:** 1173–80.

For personal use. Only reproduce with permission from Elsevier Ltd