

# The agonising negative trend in monitoring of clinical trials

David L DeMets, Stuart J Pocock, Desmond G Julian

Randomised clinical trials are undertaken in the hope of showing positive benefits of a new treatment, but on occasion quite the opposite trend can occur, if the interim data suggest possible negative (harmful) effects of a new treatment. The handling of such emerging negative trends is among the most complicated and ethically challenging scenarios in monitoring clinical trials through repeated interim analyses. Statistical methods are helpful to detect the point of no likely beneficial effect, and the point that separates neutral results from harmful results. However, in practice the decision whether (and exactly when) to stop such a trial involves a complex of other issues that depends on the context of the disease, the treatment being assessed, and the current practice of medicine. Owing to this complexity, an independent Data and Safety Monitoring Board (DSMB) is best suited to deal with such a situation. Prediction of whether a negative trend will emerge in any trial is not possible. Negative trends were not anticipated in the cardiovascular trials and the trials of lung-cancer prevention described here. In the light of these experiences, all trials and their DSMBs should consider ahead of time the possibility of unexpectedly harmful results, and should document appropriately the statistical guidelines and the decision-making process required to cope with such undesirable events.

## Introduction

Randomised clinical trials for evaluation of new treatments must be monitored carefully during their conduct for early evidence of treatment benefit or harm. The most common practice is to have an independent data and safety monitoring board (DSMB) and to document the decision process to terminate early. While such decisions are generally complex, our paper focuses on the difficulty caused when negative or harmful trends begin to emerge. All medical care and research is committed to first doing no harm, but reactions to emerging negative trends depend on the disease, the status of the treatment being assessed, prior expectations, and background scientific evidence supportive of potential benefit. The DSMB must be responsible for current and future patients in the trial, and also for all patients who might otherwise use the intervention being tested. We assess the issues arising from emerging negative trends by taking several cardiovascular and cancer trials as examples. These issues, however, are not unique to these examples.

## Methods and principles

In the USA, DSMBs were initially recommended by a National Heart Lung and Blood Institute (NHLBI) task force<sup>1</sup> while the Coronary Drug Project<sup>2</sup> was being designed. The DSMB is an external, independent committee that gives recommendations to the trial sponsor and investigators' leadership about many aspects of trial progress. Such DSMBs are now widely used for

public-sponsored trials (eg, US National Institutes of Health and UK Medical Research Council) and as industry-sponsored trials.<sup>3,4</sup> Most DSMBs are responsible for following the accumulating primary, secondary, and safety outcome data. Generally, they are the only group reviewing the outcome data by treatment assignment, since sponsors, investigators, and regulatory agencies remain unaware of treatment allocation. DSMBs assume a major responsibility for the safety of patients in the trial, and for the conduct of the trial. If interim data show that a treatment benefit is emerging, or that harm exists, the DSMB may recommend early trial termination or protocol modification.<sup>3-5</sup>

The DSMB's interactions with the sponsor and investigators are important to delineate,<sup>3,6</sup> especially for trials with harmful negative trends. Since the DSMB's primary responsibility is to protect the patients' interests in the trial, it must maintain independence from both sponsor and investigators. Thus, DSMB members should not be representatives of either sponsor or investigators. DSMB meetings can have preliminary open sessions to allow appropriate dialogue with sponsor and investigator representatives. However, outcome and safety data by treatment should be reviewed in a closed session by DSMB members only.<sup>3,6</sup> Remuneration of DSMB members should not be excessive, but at usual consultant rates, to avoid financial conflicts. DSMB members should also have no financial investments in the sponsor.

Statistical monitoring procedures assist the DSMB in repeatedly inspecting outcome data while controlling for false claims for treatment benefit or harm.<sup>7-10</sup> Although these statistical "stopping rules" are useful, they are not absolute rules in fact, but are objective guidelines to aid DSMB decisions.<sup>2-5</sup> Ethics require that trials continue no longer than necessary to assess the role of a new treatment, but repeated interim analysis of accumulating data increases the risk of false claims of treatment effect, positive or negative. For example, with five interim analyses each having a nominal 5% p value for stopping, the actual false-positive rate would be nearly 19%. This

*Lancet* 1999; **354**: 1983-88

Department of Biostatistics, University of Wisconsin-Madison, USA (D L DeMets PhD); Medical Statistics Unit, London School of Hygiene and Tropical Medicine (S J Pocock PhD); and University of Newcastle upon Tyne (D G Julian MD); UK

Correspondence to: Dr David L DeMets, Department of Biostatistics and Medical Informatics, University of Wisconsin, 600 Highland Avenue, Room K6/446-4675, Madison, WI 53792, USA (e-mail: demets@biostat.wisc.edu)

dilemma is true whether or not one uses frequentist or Bayesian statistical analysis. However, such statistical methods give the DSMB guidelines for interpreting interim results, since they are appropriately cautious but not inflexible.

The most frequently used statistical monitoring approach is the group sequential design,<sup>7-9</sup> which requires more extreme results for early termination than conventional criteria such as  $p < 0.05$ . How extreme these results are depends on the specific method, but the approach of O'Brien-Fleming,<sup>9</sup> as implemented by Lan and DeMets,<sup>10</sup> is widely used. The approach requires very strong early evidence of a treatment difference (ie, a very small  $p$  value), and slackens that criterion as the trial progresses, nearly returning to the boundary of conventional significance level at the final analysis. A similar approach, recommended by Peto and colleagues<sup>7</sup> requires a treatment difference of 3 SE for early stopping. These methods appeal because they are conservative early when results are more unstable, yield only very slight conservatism in the final scheduled analysis, require negligible increase in sample size to allow for repeated looks at data, and allow for unscheduled analyses.

Jennison and Turnbull<sup>11</sup> proposed a similar approach based on repeated CIs, with intervals made wider than the conventional 95% CIs. Repeated CIs are particularly useful in assessment of "equivalency" of two treatments and of negative trends. If the interim CI is wholly below some pre-declared minimum required benefit, then one may stop the trial early, even if the negative trend is not significant.<sup>5,11</sup>

Symmetrical boundaries may be preferred for comparison of two proven therapies. However, it is generally appropriate to have less stringent statistical criteria for negative trends than for superiority when comparing a new therapy with an established therapy. Such formal asymmetrical boundaries<sup>12</sup> are less commonly used than they should be, but most DSMBs operate with some degree of asymmetry. The recent METoprolol Randomised Intervention Trial in Heart Failure (MERIT-HF)<sup>13</sup> had a very conservative upper boundary for survival benefit of 3.5 SE by logrank test (a nominal  $p$  value of less than 0.001) at all interim analyses, with a critical value of 2.05 SE ( $p < 0.04$ ) at the end of the study. The lower boundary for harm was somewhat less conservative, ranging from -2.6 SE initially to -2.3 SE at the scheduled termination. The rationale for this asymmetry was that  $\beta$ -blocker drugs such as metoprolol, although well-established for use after myocardial infarction, were not previously recommended for patients with heart failure because they might be harmful. In fact, the upper sequential boundary for treatment benefit was exceeded halfway through the trial.

Another stochastic curtailment approach<sup>14</sup> identifies when a negative trend is sufficiently strong that the trial is unlikely to show an eventual benefit of the new treatment. The closer the trial is to its scheduled end, the less time there is to recover from negative trends and become positive. A practical Bayesian approach<sup>15</sup> formally incorporates prior data (or belief) about the new treatment's effect. The more prior evidence to suggest benefit, the less additional positive trial data is required to establish superiority. Also, in the face of prior positive evidence, strong negative data from the continued trial would be needed to recommend early termination to avoid harmful effects.

All these statistical approaches are not strict stopping rules, for there are other factors to consider in early termination decisions.<sup>2,4,5</sup> One main difficult issue is that there are rarely second chances in decision-making. If a trial is erroneously stopped early, the reputation of a promising new treatment may forever be tarnished. If a premature claim of treatment superiority is not subsequently accepted as convincing, then another trial may be required. Even worse, no further trials may be done, leaving a potentially valuable treatment in limbo. However, a DSMB cannot wait until trends become so convincing that no-one would ever challenge them. The conflict over how much evidence is needed to convince the scientific, medical, regulatory, and public-health audiences is where the debate centres.

Other factors affecting reliability of decisions are: treatment group similarity, complete and unbiased evolution of outcome of patients, treatment compliance, and consistency of results across multiple primary and secondary outcome measures.<sup>2</sup> Internal consistency across key subgroups is generally helpful, as is consistency of emerging trends with external information. The quality, completeness, and timeliness of available data is important. Experience suggests that any large comparative randomised trial should develop, before the start of the trial, a reasonably detailed documentation of the statistical methods to be used as guidelines for early termination. Inclusion of such monitoring plans in the protocol or DSMB charter before the start is important.

Many trials have stopped early for reasons of benefit, futility, or harm. Examples of positive benefit include the Betablocker Heart Attack Trial,<sup>16</sup> the Physician's Health Study,<sup>17</sup> and the MERIT-HF study.<sup>13</sup> These trials show that the decision to terminate early for benefit is not straightforward, and requires careful consideration of the above issues. In this paper, however, we want to focus on trials with emerging negative trends, in which the decision process is even more challenging.

### Clinical trials with a negative trend

Our examples of clinical trials with negative trends happen to be placebo-controlled trials, but the principles apply similarly to trials with actively-treated standard comparators.

The CONSENSUS-II trial<sup>18</sup> assessed bolus administration of enalapril of 6090 patients with acute myocardial infarction. As a negative trend in mortality emerged, the DSMB first made a protocol modification and eventually recommended early termination. The 6-month mortality comparison (10.2% mortality for placebo vs 11.0% for enalapril,  $p = 0.26$ ) showed that a beneficial effect could be ruled out. Death due to progressive heart failure was higher for enalapril than for placebo (4.3% of patients vs 3.4%,  $p = 0.06$ ). Since bolus administration of enalapril was not routine, and unlikely to be used if not beneficial, the DSMB argued that no further benefit to patients would be accrued and that there was a distinct possibility of harm. Although continuation would have provided additional scientific data for later analysis, nothing would have been gained from the patient's perspective—a consideration that the DSMB must always make a priority.

A more rapidly emerging negative trend occurred in the Cardiac Arrhythmia Suppression Trial (CAST), which assessed three drugs for suppression of cardiac ventricular arrhythmias, which are associated with an increased risk

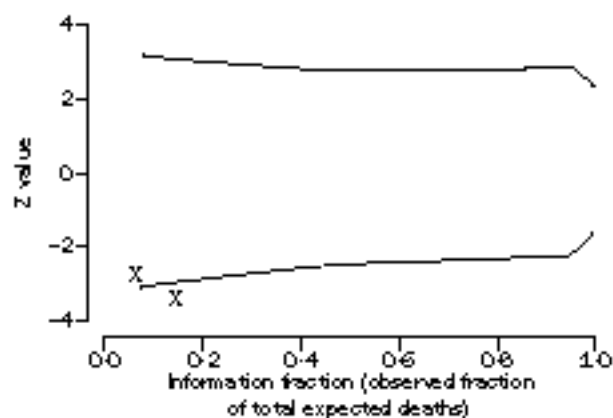


Figure 1: Group sequential boundaries for the CAST trial<sup>19</sup>  
Group sequential boundaries set at two-sided 5% significance. Plotted points=logrank test. Crossing upper boundary=benefit, crossing lower boundary=harm.

of sudden death.<sup>19,20</sup> The hypothesis was that in patients with ventricular arrhythmia after myocardial infarction, suppression of these arrhythmias should reduce the risk of sudden death. The drugs were approved for severe ventricular arrhythmia disorders but also had widespread use in patients with more mild arrhythmias. Thus, many believed that the drugs were beneficial, although this belief had not been shown in randomised trials. Others believed that these drugs may cause a ventricular fibrillation that is difficult to control. In the CAST study,<sup>20</sup> after less than 20% of expected total deaths had occurred, the DSMB recommended early termination due to an excess mortality for patients on two of the drugs (encainide, flecainide). There were 56 deaths (7.7%) among patients on these drugs compared with 22 deaths (3.0%) in the placebo group ( $p=0.0006$ ). Figure 1 shows the percentage of total expected deaths on the horizontal axis and the standardised test statistical group and sequential boundaries on the vertical axis. Observed test statistics that fall outside these bounds show a highly significant treatment difference that is sufficient to consider stopping the trial. The negative trend for the two drugs quickly crossed the lower boundary. The DSMB believed that the results suggesting harm were convincing, and their recommendation for early termination was accepted by the CAST steering committee. The US Food and Drug Administration issued an immediate physician alert. Later, the third drug was also stopped because of excess deaths.<sup>20</sup> Trials such as this must have adequate statistical power to detect important beneficial and harmful effects, because underpowered designs could lead to claims of "no harm".

Fortunately, most negative trends do not emerge with such speed. A more typical pattern is a steady accumulation of harmful evidence, as shown by three heart-failure trials (table 1). The PROMISE trial (Prospective Randomised Milrinone Survival Evaluation)<sup>21</sup> was a large multicentre randomised placebo-controlled trial to assess the effect of milrinone on total mortality in

| Trial                 | Mortality        |                  |
|-----------------------|------------------|------------------|
|                       | Drug (n)         | Placebo (n)      |
| PROMISE <sup>21</sup> | 29.9% (168/561)  | 24.1% (127/527)  |
| PROFILE <sup>22</sup> | 20.6% (214/1170) | 15.4% (181/1175) |
| VEST <sup>23</sup>    | 22.9% (292/1275) | 18.9% (242/1283) |

Table 1: Final mortality results in three heart-failure trials with negative trends

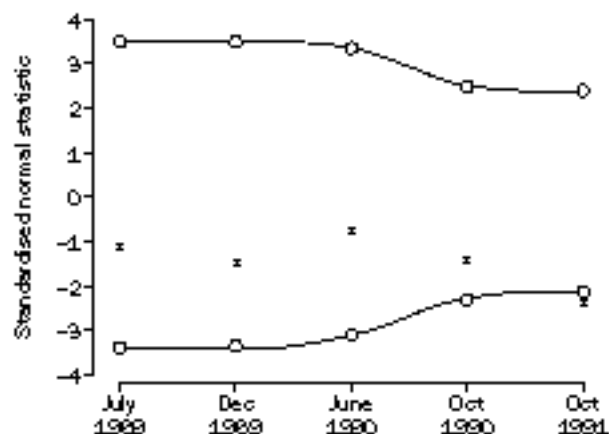


Figure 2: Group sequential boundaries for the PROMISE trial<sup>20</sup>  
Horizontal axis=information fraction (observed fraction of total expected deaths). Group sequential boundaries set at two-sided 5% significance. Plotted points=logrank test. Crossing upper boundary=benefit, crossing lower boundary=harm.

patients with moderate to advanced heart failure. O'Brien-Fleming type group sequential symmetrical boundaries<sup>9,10</sup> were used (figure 2). Interim logrank tests for survival steadily approached the lower boundary for harm. The DSMB had to assess the other advantages of the drug against the potential mortality risk. A neutral result would be important to know about, since milrinone improves heart function and exercise ability in such patients. Patients might even accept a possible higher risk of mortality if daily life were improved significantly while alive. Thus, the DSMB waited for the data to mature—a difficult time in which it seemed unlikely that PROMISE would show a mortality benefit. If milrinone had no such positive benefits, PROMISE would not have continued for as long. Ultimately, the negative mortality trends (168 deaths for milrinone, 127 deaths for placebo,  $p=0.038$ ) crossed the lower boundary and the DSMB recommended trial termination. A similar experience arose in the PROFILE (PROspective Randomised Flosequinone Longevity Evaluation) study of another heart-failure drug, flosequinone.<sup>22</sup>

A more complex scenario occurred in the VESnarinone Trial (VEST),<sup>23</sup> which assessed the drug vesnarinone in patients with advanced heart failure. Vesnarinone improved cardiac function, and a previous small placebo-controlled trial suggested that the 60 mg dose reduced mortality by over 50%.<sup>24</sup> If true, this result would be a major advance for high-risk patients. However, in the earlier trial the 120 mg dose of vesnarinone was terminated because of excess mortality (16 deaths on 120 mg, six on placebo,  $p=0.01$ ). By contrast, the continued part of the trial observed 33 deaths on placebo and 13 deaths in the group given a 60 mg dose. As a result, the VEST was required by the US Food and Drug Administration. Enthusiasm for the drug ran high among patients and cardiologists, and recruitment for VEST was rapid, with 3833 patients randomly assigned placebo, or low dose (30 mg) or higher dose (60 mg) vesnarinone. The DSMB used O'Brien-Fleming group sequential boundaries<sup>9,10</sup> with the total  $p=0.05$  overall type 1 error divided between the two comparisons ( $p=0.025$  for each), low dose versus placebo and higher dose versus placebo. The consequent boundaries are of the same shape but somewhat more extreme (conservative) than figure 2. The trial was designed to continue until 232 deaths occurred on placebo.

| Trial                | Lung-cancer incidence (%) |                    |
|----------------------|---------------------------|--------------------|
|                      | $\beta$ -carotene (n)     | Placebo (n)        |
| ATBC <sup>25</sup>   | 3.26% (474/14 560)        | 2.76% (402/14 573) |
| CARET* <sup>26</sup> | 2.37% (223/9420)          | 1.86% (165/8894)   |

\*Numerators calculated from reported relative risk of 1.28.

Table 2: **Results for two cancer prevention trials of  $\beta$ -carotene**

Unfortunately, a negative trend in mortality began to emerge that steadily gained strength. The point was passed at which it was extremely unlikely that VEST would show a positive mortality benefit, but the negative mortality trend was not yet near the lower boundary and was contradicted by the previous trial. Bayesian methods<sup>15</sup> were helpful to address the latter issue. Termination when a positive effect was unlikely would not have resolved the question of the role of vesnarinone in heart failure. As the negative trends continued, the DSMB increased the frequency of interim analyses. In fact, the target number of placebo deaths and the crossing of the lower boundary occurred simultaneously for the 60 mg group. At the end, VEST observed 242 deaths on placebo, 268 in the 30 mg group, and 292 in the 60 mg group. Comparison of the 60 mg group versus placebo was significant ( $p=0.023$ ), after adjustment for the repeated interim analyses. The DSMB decision to keep going had led to a more definitive result.

A series of trials to test the role of  $\beta$ -carotene in the prevention of cancer, especially lung cancer, provide another example of emerging negative trends but in relatively healthy populations (table 2). The Finnish AlphaTocopherol Beta Carotene (ATBC) trial assessed 28 000 male heavy smokers who were relatively healthy but at high risk of lung cancer.<sup>25</sup> Observational studies implied that  $\beta$ -carotene could reduce risk of lung cancer and other cancers. Incidence of lung cancer was the primary outcome in ATBC, with lung-cancer mortality as a key secondary outcome. Several years of follow-up were required to obtain sufficient statistical power. As the trial progressed, a slight negative trend emerged, contrary to established belief about  $\beta$ -carotene. The DSMB knew that these worrisome trends could not be definitive until the final follow-up radiography for endpoint ascertainment was completed. When that task was completed at the scheduled end, the negative mortality trends became stronger for lung-cancer death (302 on  $\beta$ -carotene, 262 on placebo,  $p=0.1$ ), and other cancer outcomes were also consistent. Many scientists were not convinced by these unexpected results.

Shortly after ATBC, the CARET trial<sup>26</sup> ( $\beta$ -Carotene And Retinol Efficacy Trial) showed similar results. The CARET population was a combination of smokers and asbestos workers, both at high risk of lung cancer. The DSMB noted a negative trend, with a relative risk for 388 new cases of lung cancer of 1.28 ( $p=0.02$ ) in favour of placebo, and 1.17 ( $p=0.02$ ) for total mortality. Although these results did not meet CARET prespecified stopping criteria, after some deliberation the DSMB recommended early termination of treatment to the executive committee, with continued follow-up. CARET was terminated 21 months early. Although CARET results by themselves were not overwhelming, they matched the ATBC results, making the combined evidence compelling.

The Prevention of Atherosclerotic Complications with Ketanserin (PACK) study<sup>27</sup> compared ketanserin with placebo in 3899 patients with intermittent claudication. The DSMB noted a harmful trend in patients who were also taking potassium-losing diuretics. In this subgroup,

there were 35 deaths on ketanserin and 15 deaths on placebo (a highly significant excess) but there was no such trend in the remaining trial patients. Although decisions based on subgroups must always be made with great care, in this case the DSMB made a difficult decision to discontinue the trial for patients on potassium-losing diuretics, so that such patients would not be placed at potential risk.

Commonly, a negative trend diminishes as additional data become available. The HERS (Heart and Estrogen/Progestin Replacement Study) trial<sup>28</sup> assessed the effect of hormone-replacement therapy compared with placebo on cardiovascular mortality and morbidity in 2700 women after myocardial infarction. After 1 year of follow-up, the DSMB observed a raised relative risk (1.8) of coronary heart disease in the group on hormone-replacement therapy. However, by 4 years' follow-up the risk of coronary heart disease was similar in each group (relative risk 0.99 [95% CI 0.8–1.2]). Mortality followed a similar pattern. The early negative trend against hormone-replacement therapy, while worrisome, was not sufficiently convincing to terminate the trial early. Such a decision would have led to a dramatically different conclusion about the use of hormone-replacement therapy. In their final report,<sup>28</sup> the HERS investigators concluded that because of the early possible risk, women should not start hormone therapy for secondary prevention of heart disease, but it may be appropriate to continue for those already on therapy.

A complete reversal of an early negative trend for mortality occurred in the ISIS-I (International Study of Infarct Survival) trial<sup>29</sup> of atenolol versus placebo in myocardial infarction. Early on, the DSMB observed a negative mortality trend of 2 SD ( $p=0.05$ ). Although this negative trend was just nominally significant, the finding did not take into account the effect of repeated statistical testing and the instability of early trends based on small numbers. The DSMB did not recommend early termination. The longer-term follow-up in 16 000 patients showed a significant and clinically important 15% reduction in mortality in favour of atenolol. Reversals of this degree are not common, but examples such as ISIS-I remind us to be cautious and not to overreact to emerging negative trends.

## Discussion

Most DSMBs are very cautious in recommending early termination of a clinical trial because of interim evidence for treatment benefit. Their deliberations are generally exhaustive, detailed, and prolonged. While the assessment of early beneficial trends may be complex, the emergence of a negative trend is even more complex for the DSMB. The examples described above provide some insight into this challenge. The ethical dilemma for the DSMB hinges on not continuing any trial with a negative trend for longer than necessary, while also needing to resolve clearly whether a new treatment may be beneficial, neutral, or actually harmful. Whether the DSMB needs to go beyond simply ruling out possible benefit to resolve the "neutrality versus harm" issue depends on several factors. If the new treatment is in the initial stages of assessment, a negative trend that rules out a clinically important benefit might well be enough for recommendation of early termination. Patients in those trials could go on to other promising treatments. By contrast, if the treatment is already in widespread or enthusiastic use, then determination of whether or not it is actually inferior

becomes the more important decision. In addition, determination of whether a new treatment is neutral or harmful for a primary outcome (eg, mortality) is even more important if there are secondary benefits such as symptom relief or improved quality of life.

Any trial's monitoring process should always be prepared for the unexpected, such as trends going in the wrong direction. This result is sufficiently common that trials should be always designed as two-sided and monitored with two-sided boundaries. This strategy also avoids one-sided significance testing in claims of benefit—an unacceptable practice that is too easy to achieve and open to abuse. However, the statistical stopping boundaries and the DSMB's attitudes to trial termination do not need to be symmetrical. That is, DSMBs generally do not require the same level of evidence to stop a trial for harm as they do to stop for benefit, although boundaries still need to be sufficiently conservative to control the type I error in either direction. If it is too easy to reject the null hypothesis of no treatment effect and to claim harm, potentially promising drugs with other useful attributes might be dismissed too readily.

Probably the most useful method for monitoring a negative trend is to calculate repeated CIs, choosing more extreme levels of confidence (>95%) to take account of the repeated looks at accumulating data.<sup>5,11</sup> A treatment can be declared harmful if the repeated CIs are wholly on the wrong side of no effect. The repeated CIs can also be used to rule out positive benefits. If no clinically important treatment benefit falls within the repeated CIs, then the trial shows that the treatment is not of interest and termination should be considered.

Previous trials sometimes suggest a positive treatment benefit, but are not convincing on their own, as shown by the vesnarinone example. Bayesian methods can be used to assess whether a current negative trend is sufficiently strong to overwhelm previous information, and these methods provide a means to assess all the evidence. A positive external result can provide some basis for caution in termination of a current negative trend, whereas a negative external result may greatly enhance the evidence, as was the case in the ATBC trial's effect on decision-making in the CARET trial.

Subgroup analyses are difficult to interpret even at the end of a trial.<sup>30</sup> Excessive probing of the data can show subgroups with nominally significant negative trends even when the overall result is neutral or even positive, but these are very likely to be spurious findings. However, subgroup analysis becomes a special challenge in the presence of an overall negative trend. A negative trend could be truly confined to one particular subgroup at high risk from a new treatment, say because of drug interactions. A negative trend in a subgroup may also be due to chance, or may eventually emerge across the entire treatment group after collection of more data. Statistically, it is important to assess whether the subgroup finding really differs from the rest of the patients, by doing an appropriate statistical test of interaction. This test can generally assuage anxieties about a worrisome subgroup. The combination of subgroup analysis and repeated testing of accumulating data has a high risk of generating spurious subgroup findings (whether for benefit or harm), which, though apparently interesting, need very cautious interpretation. In these cases, results should be interpreted in light of biological knowledge. Only in the most exceptional circumstances

would we recommend terminating all or part of a trial on the basis of a subgroup analysis.

## Concluding recommendations

Emerging negative trends are especially challenging for any DSMB. They may continue, sometimes rapidly, to show a harmful effect, may return to neutrality, or may occasionally reverse to become beneficial. The ethical and scientific dilemma for a DSMB is not to react too quickly and not to allow a harmful trend to continue beyond what is ethically acceptable. We offer these recommendations to future DSMBs and investigators and sponsors of trials:

- In planning, all randomised trial collaborations should be prepared for results with trends in either direction, no matter how unlikely a negative harmful trend may seem at the outset.
- Early trends with small numbers of patients and events should be interpreted cautiously, since such results are generally unstable and any decisions based on them could be incorrect.
- Early trial termination should require more extreme evidence than conventional 5% significance, and various statistical methods for data monitoring can provide useful guidelines but not absolute rules in that evaluation.
- A greater degree of evidence is generally required to declare a positive beneficial trend than to declare a negative harmful trend, thus creating asymmetrical boundaries for statistical comparisons.
- The degree of asymmetry in the statistical criteria depends on the current role and use of the experimental therapy. A new therapy not yet in general use requires a less extreme negative trend to declare a harmful effect than does a therapy already in practice or with other background evidence of benefit.
- In monitoring of emerging negative trends, a DSMB must balance the ethical ramifications of stopping a trial too soon (and thus having little effect) against waiting too long so that more patients than necessary become exposed to a harmful treatment. Dealing with emerging negative trends requires wise judgment based on all the evidence, and an experienced, independent DSMB is best suited to this purpose.
- Once a DSMB has made recommendations to terminate early for a negative harmful trend, it must collaborate efficiently and responsibly with the sponsor and investigators about the consequences.
- It is ethically and scientifically imperative to present and publish harmful negative results as quickly as possible, so that all patients currently in the trial, all future patients, and investigators may benefit from these results, no matter how disappointing they are.
- It would be nice if there were fewer trials with negative trends, and that requires wiser judgement by trialists in realistically “backing winners” in the pursuit of clinical progress.

## References

- 1 Heart Special Project Committee. Organization, review, and administration of cooperative studies (Greenberg Report): a report from the Heart Special Project Committee to the National Advisory Heart Council, May 1967. *Controlled Clin Trials* 1988; **9**: 137–48.
- 2 Coronary Drug Project Research Group. Practical aspects of decision making in clinical trials: the Coronary Drug Project as a case study. *Controlled Clin Trials* 1981; **1**: 363–76.
- 3 Fleming TR, DeMets DL. Monitoring of clinical trials: issues and recommendations. *Controlled Clin Trials* 1993; **14**: 183–97.

- 4 Schwartz PJ, Julian DG, Bigger JT, et al, and the Task Force of the Working Group on Arrhythmias of the European Society of Cardiology. The early termination of clinical trials: causes, consequences, and control—with special reference to trials in the field of arrhythmias and sudden death. *Circulation* 1994; **89**: 2892–907.
- 5 Pocock SJ. When to stop a clinical trial. *BMJ* 1992; **305**: 235–40.
- 6 DeMets DL. Principles of data and safety monitoring boards. In: Hennekens CH, ed. *Clinical trials in cardiovascular disease*. Philadelphia: WB Saunders, 1999: 31–42.
- 7 Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observations of each patient: I: introduction and design. *Br J Cancer* 1976; **34**: 585–612.
- 8 Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977; **64**: 191–99.
- 9 O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979; **35**: 549–56.
- 10 DeMets DL, Lan KKG. Interim analyses: the alpha spending function approach. *Stat Med* 1994; **13**: 1341–52.
- 11 Jennison C, Turnbull BW. Interim analyses: repeated confidence interval approach. *J R Stat Soc* 1989; **51**: 305–61.
- 12 DeMets DL, Ware JH. Asymmetric group sequential boundaries for monitoring clinical trials. *Biometrika* 1982; **69**: 661–63.
- 13 MERIT-HF Study Group. Effect of metoprolol CR/XL in chronic heart failure: metoprolol CR/XL randomised intervention trial in congestive heart failure (MERIT-HR). *Lancet* 1999; **353**: 2001–07.
- 14 Lan KKG, Wittes J. The B-value: a tool for monitoring data. *Biometrics* 1988; **44**: 579–85.
- 15 Freedman LS, Spiegelhalter DJ, Parmar MKB. The what, why, and how of Bayesian clinical trials monitoring. *Stat Med* 1994; **13**: 1371–83.
- 16 Beta-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.
- 17 Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med* 1989; **321**: 129–35.
- 18 Swedberg K, Held P, Kjekshus J, Rasmussen K, Ryden L, Wedel H. Effect of the early administration of enalapril on mortality in patients with acute myocardial infarction: results of the Cooperative New Scandinavian Enalapril Survival Study II (Consensus II). *N Engl J Med* 1992; **327**: 678–84.
- 19 Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med* 1989; **312**: 406–12.
- 20 The Cardiac Arrhythmia Suppression Trial II Investigators. Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. *N Engl J Med* 1992; **327**: 227–33.
- 21 Packer M, Carver JR, Rodehoffer RJ, et al, for the PROMISE Study Research Group. Effect of oral milrinone on mortality in severe chronic heart failure. *N Engl J Med* 1991; **325**: 1468–75.
- 22 Packer M, Rouleau J, Swedberg K, et al, and the PROFILE Investigators. Effect of flosequinone on survival in chronic heart failure: preliminary results of the PROFILE study. *Circulation* 1993; **88** (suppl I): I–30I (abstr).
- 23 Cohn JN, Goldstein SO, Greenberg BH, et al, for the Vesnarinone Trial Investigators. A dose-dependent increase in mortality with vesnarinone among patients with severe heart failure. *N Engl J Med* 1998; **339**: 1810–16.
- 24 Feldman AM, Bristow MR, Parmley WW, et al. Effects of vesnarinone on morbidity and mortality in patients with heart failure. *N Engl J Med* 1993; **329**: 149–55.
- 25 The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N Engl J Med* 1994; **330**: 1029–35.
- 26 Omenn GS, Goodman GE, Thornquist MD, et al. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *N Engl J Med* 1996; **334**: 1150–55.
- 27 Prevention of Atherosclerotic Complications with Ketanserin Trial Group. Prevention of atherosclerotic complication: controlled trial of ketanserin. *BMJ* 1989; **298**: 424–30.
- 28 Hulley S, Grady D, Bush T, et al, for the Heart and Estrogen/Progestin Replacement Study (HERS) Research Group. Randomised trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA* 1998; **280**: 605–13.
- 29 First International Study of Infarct Survival (ISIS) Collaborative Study Group. Randomised trial of intravenous atenolol among 16,027 cases of suspected acute myocardial infarction: ISIS-1. *Lancet* 1986; **ii**: 57–66.
- 30 Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991; **266**: 93–98.