



Published in final edited form as:

N Engl J Med. 2008 September 25; 359(13): 1400–1402. doi:10.1056/NEJMe0807372.

Identifying and Addressing Safety Signals in Clinical Trials

Thomas R. Fleming, Ph.D.

From the Department of Biostatistics, University of Washington, Seattle.

Interventions for the prevention or treatment of disease that are based on our understanding of the pathobiologic features of the illness can provide benefit in how a patient feels or functions or in whether the patient survives. Such benefits are generally achieved through on-target biologic effects of the intervention. However, there are numerous recent cases in which it has been established or strongly suggested that off-target effects of such interventions have adversely altered their risk–benefit profile. For example, in patients with rheumatoid arthritis or osteoarthritis, the cyclooxygenase-2 (COX-2) inhibitors provide an analgesic benefit with a reduced risk of gastrointestinal side effects, as compared with nonselective nonsteroidal antiinflammatory drugs, yet they increase the risk of death from cardiovascular causes, myocardial infarction, or stroke.

Frequently, such off-target effects are initially recognized through exploratory analyses of data from randomized clinical trials. However, there are important reasons why results from exploratory analyses should be viewed with caution.¹ The P values from a plethora of exploratory analyses of efficacy and safety are far less interpretable than are P values from a prespecified primary analysis of the prespecified primary end point. Furthermore, an impressive estimated effect, whether it be a benefit on an exploratory efficacy measure or a harm on an exploratory safety measure, is likely to be exaggerated owing to the regression-to-the-mean effect. Such bias occurs because there are both true signal and random noise in every estimate of treatment effect, and when many analyses are conducted, the results that appear to be most extreme tend to be at least partially due to random overestimates of the true effect.²

For example, results of preclinical studies or exploratory analyses in clinical trials may suggest a signal of excess risk or specific off-target effects that could have a meaningful effect on the risk–benefit profile of an intervention. In most instances, these results should be considered hypothesis-generating and should be addressed in confirmatory trials. Three criteria, stated in terms of questions, should be considered in assessing the reliability of exploratory safety analyses. First, is it statistically unlikely that such events can be explained by chance? For example, when three cases of progressive multifocal leukoencephalopathy were identified in randomized trials evaluating natalizumab (Tysabri) in patients with multiple sclerosis and Crohn's disease, the finding represented an increase in the event rate by a factor of 1000, as compared with the rate among historical control subjects. Second, is the safety risk biologically plausible? And third, can one identify independent, prospectively obtained data to confirm the finding?

For illustration, we can look at interventions to slow the progression of aortic-valve stenosis. In an article in this issue of the *Journal*, Rossebø et al.³ describe the results of the Simvastatin and Ezetimibe in Aortic Stenosis (SEAS) trial (ClinicalTrials.gov number, NCT00092677), which has provided a signal that therapy with a combination of simvastatin and ezetimibe may increase the incidence of cancer and the risk of cancer-related death. In the simvastatin–

ezetimibe group, as compared with the placebo group, there was an increase in the incidence of cancer (101 vs. 65 patients; hazard ratio, 1.55; 95% confidence interval [CI], 1.13 to 2.12) and in cancer-related deaths (37 vs. 20 deaths; hazard ratio, 1.78; 95% CI, 1.03 to 3.11). Ezetimibe blocks the absorption of phytosterols and other phytonutrients that are linked to protection against cancer, which provides some biologic plausibility that the drug could have an effect on the growth of cancer cells.^{4–6} On the basis of the three above-mentioned criteria, this exploratory finding in the SEAS trial is a signal that requires independent confirmation, ideally through prospective clinical trials.

The primary goal of these confirmatory trials is not to determine whether there is significant evidence to rule out the hypothesis of no increased risk, with the absence of such significant evidence considered to be a “positive result.” Trials designed in this manner could be conducted with poor quality and could provide a relatively small number of clinical end points. Such trials could easily fail to detect clinically meaningful adverse effects. Declaring that a treatment is “safe” in such a setting is tantamount to accepting that absence of evidence is evidence of absence.

Safety trials should be designed to rule out a level of increased risk that would be clinically unacceptable in the context of the level of benefit that is provided by the intervention.⁷ For illustration, the PRECISION (Prospective Randomized Evaluation of Celecoxib Integrated Safety versus Ibuprofen or Naproxen) trial (NCT00346216) has been designed to rule out the hypothesis that a COX-2 agent, celecoxib, is associated with an increase of 33% in the rate of death from cardiovascular causes, myocardial infarction, or stroke in patients with rheumatoid arthritis or osteoarthritis, as compared with naproxen.⁷ To rule out this hypothesis with a probability of 90% when in truth there is no increase in risk, using the Cox regression analysis with a one-sided false positive error rate of 0.025, investigators would need a trial of sufficient size and duration that 508 patients would have the primary safety end point. This might require the enrollment of approximately 7000 patients per randomized group, with average follow-up of 2 to 3 years. Figure 1 shows several possible outcomes in such a trial and the interpretation of the results.

For such a safety trial to provide interpretable evidence regarding ruling out harm, several performance standards must be met. These include timely enrollment of the target population in which excess risk from the experimental regimen is most plausible, the enrollment of patients who are at sufficiently high risk for the targeted number of primary safety end points to be achieved, adherence to an experimental regimen at a level that matches the best that would be achievable in a real-world setting, minimization of the number of control subjects who cross over to the experimental group, and long-term retention of nearly all patients who undergo randomization.

As described by Peto et al.⁸ in this issue of the *Journal*, two ongoing trials — IMPROVE-IT (the Improved Reduction of Outcomes: Vytorian Efficacy International Trial) (NCT00202878), in which simvastatin plus ezetimibe is compared with simvastatin plus placebo, and the SHARP (Study of Heart and Renal Protection) trial (NCT00125593), in which simvastatin plus ezetimibe is compared with placebo alone — provide the opportunity for independent confirmation of the signal from the SEAS trial that simvastatin–ezetimibe may increase the incidence of cancer and the risk of cancer-related death. To avoid the regression-to-the-mean effect, a meta-analysis of confirmatory trials should be conducted that does not include the hypothesis-generating trial, SEAS. Regarding the incidence of cancer, the analysis of the IMPROVE-IT and SHARP trials showed that incident cancers had been diagnosed in 313 patients in the combined active-treatment groups as compared with 326 patients in the combined control groups (hazard ratio, 0.96; 95% CI, 0.82 to 1.12). Hence, the hypothesis that simvastatin–ezetimibe increases the relative risk of cancer by more than 12% can be ruled out.

Given the possibility that simvastatin–ezetimibe could have an effect on the rate of cancer-related deaths through a cancer-promoter mechanism, separate analyses should be performed for the end point of cancer-related mortality. Unfortunately, these two trials indicate an increase in the number of cancer-related deaths associated with active treatment, as compared with control therapy (97 vs. 72 deaths; approximate hazard ratio, 1.34; 95% CI, 0.98 to 1.84), indicating an increase in risk of 34%. Given this confidence interval, investigators cannot exclude a relative increase of as much as 84% in the risk of cancer-related death associated with the use of simvastatin–ezetimibe. Hence, there are clinically important increases in the risk of cancer-related death that are not ruled out by these data.

The analyses of the IMPROVE-IT and SHARP trials raise additional important concerns. The interim nature of the data from these trials is problematic. As is frequently discussed,^{2,9} there is a serious risk of misinterpreting interim data as well as disturbing the integrity of ongoing trials through the release of interim data. Hence, access to interim data from ongoing trials such as IMPROVE-IT and SHARP should be restricted to data monitoring committees. Furthermore, without full access to peer-reviewed summaries of these data, it is not possible to determine the extent to which the two trials meet the list of above-mentioned performance standards for safety trials.

Additional data are needed to adequately address the signal that simvastatin–ezetimibe is associated with an increased risk of death from cancer. Such data should be provided by completed randomized trials that have been prospectively designed and conducted to meet the performance standards for safety trials. Such confirmation is especially important in the case of agents, such as ezetimibe, for which there are safety signals of major illness or death and evidence of efficacy that is limited to documented effects on a biomarker.

References

1. O'Neill RT. Secondary endpoints cannot be validly analyzed if the primary endpoint does not demonstrate clear statistical significance. *Control Clin Trials* 1997;18:550–556. [PubMed: 9408717]
2. Fleming TR, Sharples K, McCall J, Moore A, Rodgers A, Stewart R. Maintaining confidentiality of interim data to enhance trial integrity and credibility. *Clin Trials* 2008;5:157–167. [PubMed: 18375654]
3. Rossebø AB, Pedersen TR, Boman K, et al. Intensive lipid lowering with simvastatin and ezetimibe in aortic stenosis. *N Engl J Med* 2008;359:1343–1356. [PubMed: 18765433]
4. Bradford PG, Awad AB. Phytosterols as anticancer compounds. *Mol Nutr Food Res* 2007;51:161–170. [PubMed: 17266177]
5. Assmann G, Kannerbert F, Ramey DR, Musliner TA, Gutkin SW, Veltri EP. Effects of ezetimibe, simvastatin, atorvastatin, and ezetimibe-statin therapies on non-cholesterol sterols in patients with primary hypercholesterolemia. *Curr Med Res Opin* 2008;24:249–259. [PubMed: 18053317]
6. Imanaka H, Koide H, Shimizu S, et al. Chemoprevention of tumor metastasis by liposomal β -sitosterol intake. *Biol Pharm Bull* 2008;31:400–404. [PubMed: 18310900]
7. Fleming TR. Current issues in non-inferiority trials. *Stat Med* 2008;27:317–332. [PubMed: 17340597]
8. Peto R, Emberson J, Landray M, et al. Analyses of cancer data from three ezetimibe trials. *N Engl J Med* 2008;359:1357–1366. [PubMed: 18765432]
9. Ellenberg, SS.; Fleming, TR.; DeMets, DL. Data monitoring committees in clinical trials: a practical perspective. Chichester, England: John Wiley; 2002.

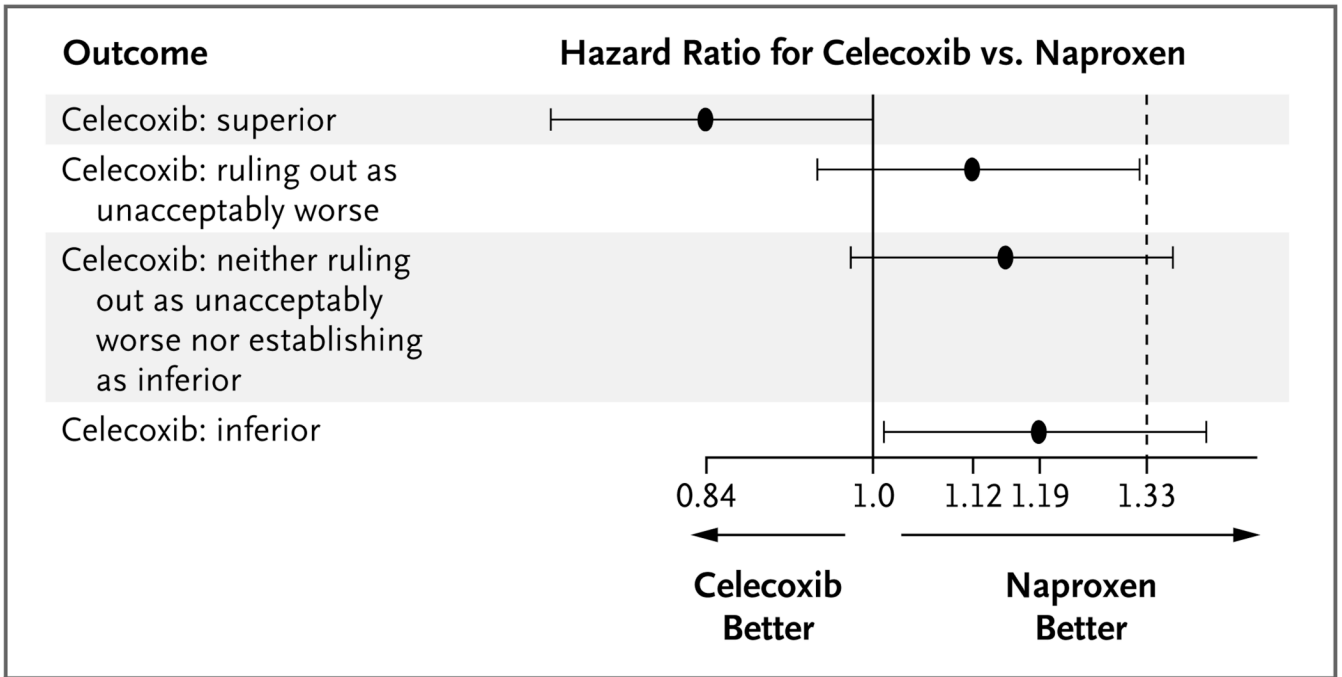


Figure 1. Hazard Ratios for the Composite Safety End Point of Death from Cardiovascular Causes, Myocardial Infarction, or Stroke in the Celecoxib Group, as Compared with the Naproxen Group, According to Possible Outcomes in an Ongoing Trial

The graph shows four possible outcomes in the PRECISION (Prospective Randomized Evaluation of Celecoxib Integrated Safety versus Ibuprofen or Naproxen) trial, which is being conducted to assess whether investigators could rule out an unacceptable increase in safety risks associated with celecoxib, as compared with naproxen. The black circles represent the point estimates, and the horizontal lines the 95% confidence intervals, for the risk of the composite safety end point with celecoxib, as compared with naproxen. In essence, if the entire confidence interval for the hazard ratio lies to the left of 1.33, the trial is “positive” in that investigators can rule out the possibility that the risk of the composite safety end point is 33% greater in the celecoxib group than in the naproxen group. If the entire confidence interval lies to the left of 1.0, celecoxib is superior to naproxen with regard to the risk of the composite safety end point, and if the entire confidence interval lies to the right of 1.0, celecoxib is inferior to naproxen. Data are from Fleming.⁷