David B. Miller, MD, FRCPC, Karin H. Humphries, DSc

A new way to evaluate randomized controlled trials? New approach does more harm than good

The Therapeutics Initiative uses a "total serious adverse events" rate to evaluate RCTs, but this flawed approach can lead to invalid conclusions and needless concern on the part of physicians and patients.

ABSTRACT: The Therapeutics Initiative of British Columbia published "Do statins have a role in primary prevention?" suggesting that statin medications may be harmful in patients without a previous myocardial infarction. This conclusion was drawn by evaluating serious adverse events in two randomized, placebo-controlled studies—AFCAPS/TexCAPS and PROSPER. The Therapeutics Initiative authors believe that total serious adverse events, an endpoint that combines safety and efficacy data, is the best way to evaluate clinical trials. We demonstrate in a variety of ways that this approach to evaluating randomized controlled trials can lead to invalid conclusions and can cause needless concern, if not alarm, on the part of physicians and patients.

A recent publication of the Therapeutics Initiative (TI) of British Columbia, "Do statins have a role in primary prevention?" suggested that statin medications (properly called HMG-CoA reductase inhibitors) may do more harm than good in patients without a previous myocardial infarction.[1] This conclusion was drawn by evaluating serious adverse events (SAEs) in two randomized, placebo-controlled studies—AFCAPS/TexCAPS[2] and PROSPER.[3]

While the authors agree that statins demonstrated a statistically significant reduction in the studies' primary endpoints and selected secondary endpoints, they argue that these efficacy endpoints are less relevant than the "total SAE" rate, or safety endpoint. The authors believe that a similar total SAE rate in the placebo and treatment arms of the studies demonstrates that "statins have not been shown to provide an overall health benefit in primary prevention trials."[1] Furthermore, because the total SAE rates are similar in both arms, despite reductions in the secondary efficacy endpoints, there is "the possibility that unrecognized serious adverse events are increased by statin therapy."[1]

Using the TI publication as an example, we will show how this approach to evaluating randomized controlled trials (RCTs) can lead to invalid conclusions and can cause needless concern, if not alarm, on the part of physicians and patients.[4-7] Several aspects of this approach to evaluating RCTs deserve mention. First, there are inherent problems when meta-analyses selectively focus on a small number of trials and secondary rather than primary endpoints. Second, proposing that total SAE rates (a combination of safety and efficacy endpoints) is the most appropriate method to evaluate randomized clinical trials ignores the intent of a clinical trial and could lead to an under-assessment of adverse events. We will demonstrate that when this type of analysis is used, any trial that demonstrates superior efficacy against a

Dr Miller is an endocrinologist in Victoria, BC, and head of Endocrinology for the Vancouver Island Health Authority, South Islands. Dr Humphries is an epidemiologist/statistician and assistant professor, Department of Medicine, Division of Cardiology and Centre for Health Evaluation and Outcome Sciences at St. Paul's Hospital in Vancouver, BC.

background of equivalent safety will be open to the charge of being ineffective and having unrecognized serious adverse events.

Discussion

Primary versus secondary endpoints

PROSPER was designed and powered to evaluate the efficacy of pravastatin for the prevention of coronary heart disease (CHD) death, non-fatal MI, and fatal or nonfatal stroke.[3] In AFCAPS/TexCAPS lovastatin was used to reduce fatal or nonfatal MI, sudden cardiac death, or unstable angina.[2] The TI evaluation of these two trials, on the other hand, focuses on total mortality, for which neither study was powered, and MI or stroke, which were secondary endpoints. The absolute risk reduction in PROSPER for the primary endpoint was 2.1% (P=.014), while in AFCAPS/Tex-CAPS it was 4.1% (P<.001). The secondary endpoint chosen in the TI meta-analysis, MI and stroke, had an absolute risk reduction of 1.8%. This is still statistically significant but underestimates the full benefit demonstrated in these trials.

Reporting of serious adverse events

The main premise of the alternative approach to RCT evaluation is that because the SAE rates in the statin and placebo arms in these two trials were the same, despite a reduction in MI and stroke in the statin arm, the statin arm must have incurred "unrecognized SAEs."[1] This unconventional approach of combining efficacy (the clinical endpoint the trial is designed to assess) with safety outcomes and then subtracting the efficacy outcomes is problematic both with respect to how SAEs are reported and defined and with respect to methodological and statistical validity.

How are adverse events defined? The definitions for SAEs are quite uniform. The International Conference on Harmonisation (ICH) defines an SAE as, "any untoward medical occurrence that results in death, is life-threatening, requires inpatient hospitalization, or results in persistent or significant disability/incapacity."[8] The United States Federal Drug Administration (USFDA) defines an SAE as "any experience that is fatal or life-threatening, is permanently disabling, requires hospitalization, or is a congenital anomaly, cancer, or overdose."[9] All SAEs must be reported to the appropriate regulatory agency irrespective of whether the SAE—like death or acute myocardial infarction—is also an efficacy endpoint for a particular trial.

For the purposes of reporting trial results in a journal, the reader is interested in all adverse events, not just those defined as serious. The reader is interested in determining whether the efficacy of the intervention was achieved with acceptable safety by comparing the adverse events in the treatment arm with those in the placebo arm. Unfortunately, the definition of adverse events varies among pharmaceutical companies, research organizations, and journals. The ICH Good Clinical Practice (GCP) guideline defines an adverse event as any unintended and unfavorable sign, symptom, or disease temporally associated with the use of a medical product, whether or not it is considered related to the product.[8] Northington defines an adverse event as any unfavorable change in the structure (signs), function (symptoms), or chemistry (laboratory data) of the body temporally associated with participation in the clinical trial, irrespective of the believed relationship to the study drug.[10] This definition would thus include intercurrent illness or injury,

clinically significant results from laboratory tests or other medical procedures, and clinically significant findings uncovered during a physical examination.

Adverse events should always be reported when a trial is presented in a peer-reviewed journal. The ICH GCP Statistical Guideline[11] devotes an entire section to the appropriate analysis of adverse events, separate from a section on the analysis of primary and secondary endpoints. Similarly, the Consolidated Standards of Reporting Trials (CONSORT)[12] discusses the presentation and reporting of efficacy endpoints as distinct from the presentation of adverse events. Neither of these internationally recognized groups recommends combined reporting of efficacy and safety events.

Why report efficacy and safety endpoints separately? One of the cardinal rules of statistical inference in clinical trials is the assumption of independence of the observations.[13] This is relevant in the assessment and reporting of both efficacy and safety events. Multiple events for an individual are unlikely to be independent, and in a clinical trial the patient is the unit of observation. Data are presented as the proportion of patients experiencing a given endpoint. For a composite efficacy endpoint (like CHD death, MI, or stroke), a patient with both an MI and a CHD death contributes only one count to the efficacy endpoint. In addition, if the patient experienced two adverse events (like an elevation in liver enzymes and hospitalization for appendicitis), the patient also contributes one count to the SAE total. This is different from regulatory requirements where both events would be counted, because regulatory bodies are interested in the actual number and nature of adverse events observed.

If one adopts the "total SAE" approach, the patient who experienced each of these four outcomes would count only once. Thus, the total SAE approach can underestimate the true magnitude of a drug's adverse effects. Also in the total SAE analysis, a transient elevation in liver enzymes would be equivalent to a disabling stroke; most physicians and patients would reject this inference.

By assuming that the SAE rate includes all efficacy endpoints, and that subtracting the efficacy endpoints from the total SAE rate can derive "other SAE," any drug that demonstrates statistical efficacy with safety equivalent to placebo will have a higher other-SAE rate. In fact the greater the efficacy, the higher the other-SAE rate will be and the more likely such a trial will be found to demonstrate "unrecognized serious adverse events." This is illustrated in **Table 1**, where a study has a conventionally reported 20% SAE rate in each arm of the study.

Of the two studies cited in the TI report, Professor James Shepherd, the principal investigator of the PROSPER study, has confirmed that the primary outcomes in his study were not included in the SAE rate quoted in the publication (personal written communication with DBM, 2003). For AFCAPS/TexCAPS, the number of serious adverse events is reported, rather than proportions, making assessment on a per patient basis impossible.[2]

Even if both PROSPER and AFCAPS/TexCAPS had reported an SAE rate that comprised both efficacy and safety endpoints, the statement that "unrecognized serious adverse events are increased by statin therapy" cannot be supported. We will demonstrate how adding a relatively large number of SAEs, with high variance, to a much lower primary event rate, with lower variance, allows a statistically significant difference to be lost.

Table 1. Calculation of "other SAE."

		Treatment	Placebo	Absolute risk reduction
Trial A	Primary endpoint	8.0%	9.8%	1.8%
	SAE	20%	20%	0%
	Other SAE	12%	10.2%	−1.8%*
Trial B	Primary endpoint	5.0%	10.2%	5.2%
	SAE	20%	20%	0%
	Other SAE	15%	9.8%	−5.2%*

* (–) suggests harm

Table 2. Adding SAE to the primary outcome.

Trial	Outcome	Treatment	Placebo	Relative risk (95% CI)
1	MI + stroke	8.0%	9.8%	0.82 (0.73-0.92)
2	Total SAE	18.0%	19.8%	0.91 (0.83-0.99)
3	Total SAE	28.0%	29.8%	0.94 (0.88-1.02)
4	Total SAE	38.0%	39.8%	0.95 (0.89-1.03)

CI=confidence interval

In **Table 2**, Trial 1 duplicates the MI and stroke outcome of the AFCAPS/TexCAPS and PROSPER meta-analysis. In Trial 2, a very modest 10% rate of other SAEs is added to the efficacy outcome in both groups, producing a total-SAE rate. Note that the relative risk reduction has been reduced from 18% to 9%, even though the absolute risk reduction has remained at 1.8%. Also note that the 95% confidence interval is now approaching unity and that statistical significance has almost been lost. In Trial 3 and Trial 4, the other SAE have been set at 20% and 30% respectively. In these two analyses, a non-statistically significant reduction in total SAE is seen. In the AFCAPS/TexCAPS and PROSPER meta-analysis, the other SAE rate was even higher: 34% to 36%. Therefore, where the other SAE rate is high and the absolute risk of the primary outcome is low, the total SAEs are almost guaranteed to produce a non-statisti-cally significant difference. A useful treatment might be rejected.

Why not just look at mortality? Many studies are not powered to investigate differences in disease-specific mortality: in the case of statins, cardiovascular mortality. The studies are usually powered to find a difference in combined endpoints of morbidity such as nonfatal MI and/or stroke. Nonfatal events are more common than cardiovascular death and are valid, clinical endpoints worth preventing. Finding a difference in cardiovascular death would simply require a larger and/or longer trial. In these studies the lack of a statistically significant reduction in mortality could be an example of a type II or β error—failing to find a difference where one truly exists. Examining all-cause mortality has the same problem as examining total SAE. If cause-specific mortality is low and other (non-drug-related) mortality is high, and equal in

the treatment and placebo arms of a study, small differences in cause-specific mortality will be diluted and no statistically significant difference in total mortality will be found. This has occurred in many statin studies in secondary prevention—statistically significant reductions in cardiovascular morbidity and/or mortality have been seen without reductions in all-cause mortality.

Conclusion

Authors affiliated with the Therapeutics Initiative have proposed using the total SAE rate as a better way to evaluate clinical trials.[14] Through the *Therapeutics Letter*,[1,15] this approach has been presented to physicians in British Columbia and through newspapers[4-7] to patients throughout Canada. The aim of providing more information on adverse events is clearly laudable, especially in light of recent findings related to COX-2 inhibitors. But the proposed approach will likely underestimate the adverse event rates and thus have the opposite effect. Only by rigorously adopting the reporting format of the CONSORT group, in which efficacy and safety endpoints are reported separately, can one critically evaluate the results of an RCT.

There are many ways to evaluate and interpret clinical trials of pharmaceutical agents. One can look at statistical and clinical significance. One can look at primary outcomes, secondary outcomes, and adverse events. One can look at relative risk reductions, absolute risk reductions, and numbers needed to treat.[16] One should look at adverse events but one shouldn't look at the total SAE rate, as many effective treatments will be discarded as useless, or even harmful, and perversely, treatments with high adverse event rates will likely be underestimated using this approach.

Statin treatment is useful in many settings, both in primary and secondary prevention; though where the risk of an event is low, as in the primary prevention of cardiovascular disease, even an efficacious treatment will have a small absolute risk reduction and a high number needed to treat.[16] The safety of statins has been established in dozens of major clinical trials, which have included tens of thousands of patients studied for up to 8 years.[17] The unconventional evaluation of statin safety and efficacy proposed by the TI is methodologically and statistically unsound and will ultimately do more harm than good.

Competing interests
None declared.

References

1. Therapeutics Initiative. Do statins have a role in primary prevention? Therapeutics Letter 2003;48. www.ti.ubc.ca/pages/letter48.htm (accessed 27 January 2005).
2. Downs JR, Clearfield M, Weis S, et al. Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: Results of AFCAPS/TexCAPS. Air Force/Texas Coronary Atherosclerosis Prevention Study. JAMA 1998;279:1615-1622.
3. Shepherd J, Blauw GJ, Murphy MB, et al. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): A randomised controlled trial. Lancet 2002;360:1623-1630.
4. Taylor P. Unclogging the heart debate: Do cholesterol drugs help or harm? Globe and Mail. 24 January 2004;F4.
5. Munro M. Cholesterol pills carry risk, UBC group says: "Fine print" raised doubts: Finding applies only to preventive use. National Post. 16 September 2003;A1.
6. Munro M. Friend or foe?: First we were told statins could save lives. Now some say they could be putting lives at risk. National Post. 16 September 2003;A15.
7. Munro M. Cholesterol pill's side effects worry UBC drug specialists. Victoria Times Colonist. 16 September 2003;A4.
8. ICH Steering Committee. International Conference on Harmonisation Tripartite Guideline for Good Clinical Practice. 1997. Minister of Public Works and Government Services Canada. Ottawa, ON. Cat #H42-2/67-11-1997IN.
9. Chow S-C, Liu J. Safety Assessment. In: Design and Analysis of Clinical Trials Concepts and Methodologies. New York: John Wiley & Sons Inc.; 1998.
10. Northington B. A review of issues in the collection and reporting of adverse events. Biopharmaceut Rep 1996;4:1-5.
11. ICH Steering Committee. International Conference on Harmonization Tripartite Guideline for Good Clinical Practice Process. Statistical Principles for Clinical Trials. 2003. Minister of Public Works and Government Services Canada. Ottawa, ON. Cat #H49-171/2003E.
12. Moher D, Schulz KF, Altman DG. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001;357:1191-1194.
13. Bolton S. Independence and statistical inference in clinical trial designs: A tutorial review. J Clin Pharmacol 1998;38:408-412.
14. Wright JM, Puil L, Bassett CL. Analysis of serious adverse events. Lipid-lowering therapy revisited. Can Fam Physician 2002;48:486-499, 92-95.
15. Therapeutics Initiative. Serious adverse event analysis: Lipid-lowering therapy revisited. Therapeutics Letter 2001;42. www.ti.ubc.ca/pages/letter42.htm (accessed 27 January 2005).
16. Miller D. Secondary prevention for ischemic heart disease. Relative numbers needed to treat with different therapies. Arch Intern Med 1997;157:2045-2052.
17. Gotto AM Jr, Pownall H. Chapter 9. In: Manual of Lipid Disorders. Lippincott, Williams & Wilkins; 2003:169-198. **BCMJ**