Selected topics & Case Studies of Inappropriately Interpreted Clinical Trials

Table of contents

Case studies of inappropriately interpreted clinical trials

Introduction: There is a great deal of variability in the information the public reads in the media regarding medical clinical trials. An occasional, though infrequent, cause for this variability is the result of the incorrect application of statistical methods. A more frequent source of error is that the author's conclusions often do not derive directly from the data generated by the study. Even though the data obtained maybe valid, the conclusions are inaccurate and not a reasonable reflection of the data. Though the mechanics of the statistical computations are usually performed correctly, fundamental errors in the use of statistics occur and can lead to incorrect conclusions. (The problem of inaccurate reporting of information by the media is beyond the scope of this discussion.)

The following are some examples of the misuse of statistics, as well as the problem of drawing inappropriate conclusions, exhibited in studies from the medical literature.

The **first example** involves a potential Type II error in evaluating the complications of a medical procedure. (The study has an inadequate sample size to assess for the statistical significance of an infrequent, but important complication.)

The **second example** involves the complete misunderstanding of valid statistical techniques. A group of researchers divide one of their study groups on the basis of a certain variable. They then analyze whether there are statistically significant differences between the two groups in regards to the very same variable that they used to initially separate the groups.

The **third example** is of an inappropriate extrapolation of a study's findings to a wider group of people. Oat bran was studied in a small group of healthy individuals who already had a relatively low fat diet. This is a group where oat bran would have the least effect on cholesterol. The authors incorrectly thought that because they saw little effect on cholesterol in this small group, that oat bran was without significant effect. **This example also examines how the conclusions of a meta-analysis can be biased**.

The **fourth example** illustrates the hazards of inappropriate subgroup analysis.

Example 1

This study by Block et al[1] concerned the evaluation of the safety of patients undergoing outpatient cardiac catheterization.

Some studies have multiple errors. In this one study, the authors miscounted their trial's primary endpoints, included patients that did not belong in the trial, then incorrectly analyzed the data with a potential Type II statistical error, and finally failed to appreciate the implications of their miscounted data

Elective heart catheterization in stable patients is usually associated with a major complication in only 1 out of 1000 patients. This study reported that 3 of 192 outpatients (1 out of 64) in the study experienced a heart attack with elective outpatient cardiac catheterization.(Catheters are guided up the artery and x-ray dye is placed in the arteries of the heart without any intervention such a balloon dilation.) Only 1 patient of 189 patients with inpatient cardiac catheterization in this study had a heart attack**. If a heart attack occurred in 1 out of 64 stable patients undergoing elective diagnostic procedures (rather than 1 out of 1000), then that program warrants investigation rather than commendation as safe.**

Because this difference was not statistically significant, the authors incorrectly concluded their trial demonstrated outpatient catheterization is safe. (Though outpatient catheterization is safe, the data in this study trended in the opposite direction.) The authors did note that because of the small sample size they could not exclude a small increase in complication rates. Nevertheless, their data did not suggest outpatient catheterization was safe. As it turned out, the seemingly high complication rate with outpatient catheterization in this study was a result of the authors miscounting data points as well as not following their own trial protocol.

Analogy of incorrectly declaring a procedure safe.

An example that parallels this study would be if a trial was performed to evaluate whether drawing blood with needles that have not been fully sterilized resulted in complications. If 2 out of 100 patients develop an infection and died with suboptimally sterilized needles, versus 0 out of 100 patients in a group using fully sterilized needles, this would not be statistically significant. **However, it would be a Type II statistical error to state that this data leads to the conclusion that suboptimally sterilized needles are safe because statistical significance was not reached.**

Type II Statistical Error

A Type II statistical error can potentially occur when the numbers being studied are too small to reliably determine that no significant difference exists between the groups being studied. A significant difference, at times may exist, but not show statistical significance. Larger numbers of patients or subjects are required to prove that no difference exists compared to the smaller numbers that may be required to prove that a difference is present. A Type II statistical error occurs if there is truly a difference present, but the numbers are too small to show that a statistically significant difference occurred. **Though heart catheterizations are safe to do as an outpatient, this article does not demonstrate that at all.**

Subsequent explanation by the authors of the study of reason for the high frequency of complications.

After this study was published, a letter to the editor was written which was published in the New England Journal of Medicine questioning the investigators about the details of the patients with myocardial infarctions since it was such an unusual occurrence (and not reassuring) that 1 out of 64 reportedly low risk patients undergoing elective outpatient heart catheterization developed a heart attack in this study.

The authors answered: They had mistakenly counted patients having a heart attack twice rather than once. This occurred even though having a heart attack as a complication of the procedure was the most important primary trial data endpoint and they only counted a total of 4 heart attacks**. Furthermore, contrary to their description of the patients as being low risk outpatient cardiac catheterization, they accidentally included at least one patient in the outpatient group that had a much higher risk procedure, an elective balloon angioplasty who went on to have a heart attack.** These errors explained the high frequency of heart attacks that occurred in the outpatient group.

Hence, in this one study, errors of both interpretation and execution of the trial were made.

1. A Prospective Randomized Trial of Outpatient vs. Inpatient Cardiac Catheterization; Block PC et al, New England Journal of Medicine. 1988 319: 1251-5

Example 2:

Analysis of the Flawed Statistical Methods and Conclusions Found in the Article by Geltman et al.

In the September 1990 Journal of American College of Cardiology there is a study performed by Geltman[1] el al which evaluated myocardial perfusion in patients with angina who had angiographically normal coronary arteries. **The fundamental framework from which they performed the statistical analysis of the data had no validity.**

The study consisted of a control group and a chest pain group. The authors subdivided the chest pain group into those patients having a low myocardial perfusion reserve and those patients without a low myocardial perfusion reserve. This resulted in "three groups" 1) chest pain patients with a low myocardial perfusion reserve, 2) chest pain patients with a normal or high myocardial perfusion reserve, and 3) the undivided control group.

The authors then inappropriately statistically compared the three groups in regard to myocardial perfusion reserve, which is the same value that was used as the basis for selectively subdividing the chest pain group. Naturally, the chest pain group selected for a low myocardial perfusion reserve would have a statistically different lower value compared to the undivided control group. The authors then went and compared the groups for three other values. This included maximal myocardial blood flow and resting myocardial perfusion which are all related to myocardial perfusion reserve. This was not statistically valid.

Analogy explaining fundamental statistical error that was made:

A way to make this fundamental statistical error more understandable is to note the following analogy. A study examines two groups of people, one group from the north and one group from the south. The north group was then divided into the tall northerners and the short northerners on the basis of height. The researchers then inappropriately compared the tall north group to the entire undivided south group.

Naturally, the tall north subgroup who were selected on the basis of being tall would on the average be taller than the undivided total south group. (Similarly, the tall north group would also be likely to have a statistically greater value for a foot to waist measurement or waist to head measurement than the average value for the undivided south group.) **However, one could not say that because of this comparison, that there is a unique subgroup that exists in the north group of tall people that doesn't exist in the south group. This is fundamental error that the authors of this study made**. It was perhaps less obvious because the measurements they used involved PET scanners and cardiac measurements, but that doesn't make this inappropriate statistical analysis any less wrong than this example of people divided on the basis of height.

.The statistical analysis in the Geltman study did not reliably support the contention that angina in patients with normal coronary arteries is attributable to abnormalities of perfusion at rest, maximal myocardial perfusion, or myocardial perfusion reserve. **The invalid statistical analysis that was performed obscured the true implications of the data**

1. Geltman EM, et al. Increased myocardial perfusion at rest and diminished perfusion reserve in patients with angina and angiographically normal coronary arteries. J AM Coll Cardiol 1990; 16:586-95

Example 3: The Death of the oat bran fad. (Murdered by a poorly conceived study.)

The oat bran fad passed with the publication of a flawed study[1] in a prominent medical journal, The New England Journal of Medicine.

This article received wide media attention. In 1990, many people were making and eating oat bran muffins to lower cholesterol. The fad aspect of oat bran ended when an article was published claiming to indicate that oat bran did not significantly lower cholesterol. The article was a flawed study.

In this study, 20 volunteers, many of them young healthy dieticians with low cholesterol values at baseline, were evaluated to determine the effect of ingesting oat bran compared to wheat bran. The study was flawed in studying a very small group of individuals with preexisting low cholesterol values and generalizing the results to the population at large.

The New England Journal of Medicine article was a seriously flawed study.

1. The primary problem with this study is that there were insufficient numbers of participants to exclude a cholesterol lowering effect. There were only 20 participants. **To prove an effect is not present, a large number of participants need to be studied. A trial demonstrating a positive effect generally requires fewer participants than a trial trying to reliably prove that no difference exists.**

2. Cholesterol lowering drug interventions tend to have a lesser magnitude effect on patients who have a very low cholesterol intake and start out with low cholesterol levels. The individuals in the trial had a very low fat and low cholesterol intake compared to the rest of the U.S. population, and had low cholesterol levels and LDL cholesterol at the time of initiation of the study. Many of these subjects for this trial were dietitians with a baseline low fat diet different from the average older individual at high risk for heart disease.

Letters written pointing out the flaws in this study:

Subsequent letters to The New England Journal of Medicine (NEJM) noted the inadequacy of the study. A letter[2] by the Robenoubonoffs, M.D. and R.D. suggests, **"The authors have managed to confuse the American public further with a poorly designed and underpowered trial that draws erroneous conclusions."** A letter[3] by James Burrous, M.D. published in the same issue of The New England Journal of Medicine asked**, "What does a small … study of dietitians with desirable cholesterol levels … tell us about the population of the American public at risk?"**

A later different study documents the cholesterol lowering effect of oat bran:

A subsequent study was performed on 84 middle aged men and women who were placed on metabolic diets comparing wheat bran to oat bran. This study showed a statistically significant 5% reduction in total cholesterol for oat bran beyond what was found for wheat bran.

Multiple studies were later combined and analyzed to further assess the effects of oat bran on cholesterol:

There were at 2 subsequent meta-analyses concerning oat bran:

One meta-analysis[4] concluded that oat bran modestly reduced cholesterol.

This meta-analysis included 10 trials for analysis of oat bran. **A statistically significant decrease in total cholesterol with oat bran was found in the combined trials. There was an average reduction of 5mg/dL (0.13mmol/L) in cholesterol**.
Larger reduction in cholesterol with oat bran was found to occur if the initial cholesterol level was higher than 229 mg/dL, particularly when the dose of oat bran was 3g or more.
The authors' conclusions were that "This analysis supports the hypothesis that incorporating oat products into the diet causes a modest reduction in blood cholesterol level."

There was a second meta-analysis[5] which was done by the senior author of the initial NEJM article which presented the conclusions of this meta-analysis in a way to be as consistent as possible with the initial flawed study.

A subsequent meta-analysis of the effects of oat bran on cholesterol levels was written a group that included the senior author of the widely publicized study that suggested oat bran was without significant effect on cholesterol levels

There is a tendency for any person to desire to confirm prior opinions or stances in the literature rather than refuting their own prior results.. **The conclusions of this meta-analysis were presented in such a way as to minimize any contradiction to the initial flawed study published in 1990.**

The conclusions of a meta-analysis, like any other study, can be presented with a particular slant. The results of this study can be viewed as a cup half full or half empty depending on the bias of the author.

Negative viewpoint expressed by the authors

The authors of this meta-analysis[5] conclude that "increasing soluble fiber can make only a small contribution to dietary therapy to lower cholesterol."

Ingestion of 3g of soluble oat fiber in this meta-analysis resulted in a decrease of .13mmol/L in total cholesterol LDL cholesterol. The

authors note that soluble fiber from 3 bowls (instant packages?) (28g) of oatmeal is required to achieve even a 3 g soluble fiber intake. (Other forms of oats make 3 g of soluble fiber seem much more easily achieved as noted later.)

Positive viewpoint for the data from the same meta-analysis

An equally slanted positive alternative statement of the conclusions for the identical data presented in this same meta-analysis would be the following:
This meta-analysis indicated that 3g of soluble oat fiber can result in a decrease cholesterol of .13mmol/L (5mg/dL), and 6 g of soluble fiber can result in a decrease of .26mmol/L (10mg/dL) in cholesterol.
This study suggests that an intake of 3g of oat soluble fiber can result in a 2% reduction in cholesterol. A 2% reduction in cholesterol has been estimated to correlate to a 4% reduction in cardiovascular disease. An intake of 6g of soluble fiber can result in a 4% reduction in cholesterol which is estimated to result in an 8% reduction in cardiovascular disease. This would be a significant public health benefit.

(Additionally, it could be noted that a standard 40g serving of oatmeal, Quaker Oats Old Fashioned Oatmeal, contains 2g of soluble fiber and 40g standard serving of Quaker Oat Bran hot cereal contains 3g of soluble fiber/serving.)

Similar to any other type of study, the conclusions of a meta-analysis are potentially subject to bias.

1. Swain JF, Rouse,IL, Curley CB, Sacks FM. Comparison of the effects of oat bran and low-fiber wheat on serum lipoprotein levels and blood pressure. N Engl J Med 1990; 322:147-52.

2. Roubenoff RA, Roubenoff R. Letter to the Editor, Oat Bran and Serum Cholesterol. N Engl J Med 1990; 320:1746-47.

3. Burris, J. Letter to the Editor, Oat Bran and Serum Cholesterol . N Engl J Med 1990; 320:1748.

4. Oat Products and Lipid Lowering. A Meta-analysis. Ripsin CM, Keenan J, Van Horn L, et al. JAMA 1992; 267:3317-25.

5. Brown L, Rosner B, Lillett W, Sacks F. Cholesterol-lowering effects of dietary fiber: a meta-analysis. Am J Clin Nutr 111; 69:30-42

Example 4: The Hazards of Subgroup Analysis

Dr. Peter Sleight has made insightful comments in regards to the limitations of subgroup analysis.[1,]

Inappropriate subgroup analysis can lead to ludicrous results. Dr. Sleight and colleagues did a subgroup analysis of patient outcome by astrological signs to illustrate the potential limitations in reliability of subgroup analysis. This subgroup analysis suggested that the treatment was quite effective and statistically significant for all patients except those born under the sign of Gemini or Libra.

The ISIS-2 trial

The ISIS-2 (Second International Study of Infarct Survival) trial[2] was a very large randomized clinical trial performed, by many physicians including Dr. Sleight. This study showed that both aspirin and a clot dissolving medication had important and statistically significant benefits for patients having a heart attack.

The beneficial effect of aspirin for patients having a myocardial infarction (heart attack) was very substantial and equal to the effect of streptokinase (a clot dissolving medication) in this large trial of 17,000 patients. Both were life saving medications.

A subgroup analysis was performed by the ISIS trial investigators to illustrate the potential hazards of subgroup analysis **The subgroup analysis suggested that Gemini and Libra had an adverse effect (NS), rather than a beneficial effect with aspirin, which was not a true relation. These patients would benefit from aspirin to an equal degree as the rest of the group.** Subgroup analysis, at times, can lead to findings that are incorrect.

The ISIS-2 investigators reported that "subdivision of the patients in ISIS-2 with respect to their astrological birth sign appears to indicate that for persons born under Gemini or Libra, there was a slightly adverse effect of aspirin on mortality (9% increase in risk, NS), while for patients born under all other astrological signs there was a striking beneficial effect (28% reduction in risk: 2p <0.00001.)"

What is the reliability of a finding for a small subgroup of a trial that unexpectedly has a different outcome from the rest of the group?

In particular, if a given therapy has a highly significant and strongly beneficial effect for the group as a whole, a subgroup analysis that results in the unexpected finding of certain subgroups that do not have benefit is frequently incorrect**. In fact, it is more likely that the unexpected subgroup finding that runs counter to the group finding is simply not valid.**

Validity of subgroup analysis

The view of this website is that subgroup analysis can be quite useful. The validity tends to be inversely proportional to the number of subgroups which are analyzed. **A study is not immune to an incorrect subgroup analysis outcome simply because the subgroup was prespecified, particularly if there were a large number of prespecified subgroup analyses**.

(If 20 subgroup analyses are prespecified, then it is expected that one of these subgroup analyses may show a false result for a P=.05 probability relationship.) **Part of the benefit of a prespecified subgroup analysis is that there are necessarily fewer such analyses than the almost unlimited number of ways to subdivide the data in a post hoc analysis after the trial results have been obtained.**

For an enlightening look at potential limitations of subgroup analysis, the following articles are recommended:

*1. Debate: Subgroup analyses in clinical trials: fun to look at- but don't believe them! Peter Sleight. Current Control Trial Cardiovasc Med. 2000 1(1): 25-27.

* 2. ISIS-2 (Second International Study of Infarct Survival). Lancet 1988: ii: 349-360 (pages of interest 356-357)

Consolidated list of referenced articles for case studies:

Geltman EM, et al. Increased myocardial perfusion at rest and diminished perfusion reserve in patients with angina and angiographically normal coronary arteries. J AM Coll Cardiol 1990; 16:586-95

A Prospective Randomized Trial of Outpatient vs. Inpatient Cardiac Catheterization; Block PC et al, New England Journal of Medicine. 1988 319: 1251-5

Swain JF, Rouse,IL, Curley CB, Sacks FM. Comparison of the effects of oat bran and low-fiber wheat on serum lipoprotein levels and blood pressure. N Engl J Med 1990; 322:147-52.

Roubenoff RA, Roubenoff R. Letter to the Editor, Oat Bran and Serum Cholesterol. N Engl J Med 1990; 320:1746-47.

Burris, J. Letter to the Editor, Oat Bran and Serum Cholesterol . N Engl J Med 1990; 320:1748.

Oat Products and Lipid Lowering. A Meta-analysis. Ripsin CM, Keenan J, Van Horn L, et al. JAMA 1992; 267:3317-25.

Brown L, Rosner B, Lillett W, Sacks F. Cholesterol-lowering effects of dietary fiber: a meta-analysis. Am J Clin Nutr 111; 69:30-42

Debate: Subgroup analyses in clinical trials: fun to look at- but don't believe them! Peter Sleight. Current Control Trial Cardiovasc Med. 2000 1(1): 25-27.

ISIS-2 (Second International Study of Infarct Survival). Lancet 1988: ii: 349-360 (pages of interest 356-357)

LIMITATIONS OF META-ANALYSES

The assumption that a meta-analysis always represents the final and accurate viewpoint in an area of research is not warranted.

A meta-analysis combines similar trials in order to obtain a larger number of patients to improve the evaluation of whether statistically reliable differences exist between comparison groups. Meta-analyses are by no means perfect. On some occasions, a large clinical trial is subsequently performed evaluating the the same clinical question with an outcome quite different from the initial meta-analysis. Discrepancies between meta-analyses and subsequent large randomized clinical trials are documented in literature (Le Lorier, et al, NEJM, 1997; 337:536-42).

A very large randomized clinical trial is the most reliable way of obtaining reproducible results. This means that if the same trial protocol was repeated in another study with a similar patient population using a sufficient number of patients, the same trial results would be expected to occur. (However, even a very large trial does not guarantee that the specific treatment protocol that is being studied was constructed optimally. Nor does a very large trial provide any guarantee that the study's authors make appropriate and conservative conclusions in regards to the study data.)

The more similar the trials are that are being added together, the more likely the meta-analysis will result in valid conclusions. The most straightforward trials to add together for a meta-analysis are comparisons of a single drug to placebo or a single drug to another single drug where that study has been repeated in the same fashion. The more similar the trials are to one another with the same patient population, the more reliable the meta-analysis. The addition of study protocols that are significantly different from one another tends to make a meta-analysis less reliable.

Since a meta-analysis is a summation of trials, it is only as good as the trials that are included that go into making up the meta-analysis. If a very large trial is poorly done and is part of a meta-analysis, the results of the meta-analysis are inherently limited by the impact of that trial.

The conclusions made in a meta-analysis by the authors are subject to the same potential for bias as the smallest of clinical studies. The authors of the meta-analysis must assess the limitations of their analysis and decide what conclusions to state. In addition, they need to determine how broadly their conclusions can be applied and to what patient groups. **Conservative conclusions derived directly from the data with a realistic assessment of the limitations of the study are optimal, but are by no means universal. A meta-analysis is particularly subject to biased conclusions and minimization of limitations of the meta-analysis when it is created by advocates of a controversial opinion regarding the same topic the meta-analysis is addressing.**

A meta-analysis has a number of areas of potential for bias (which is usually unintentional). The potential biases of a meta-analysis include:

1. Inclusion/exclusion criteria used, including decisions which trials are thought to be sufficient in similarity to be included.
2. The quality of the trials included for meta-analysis
3. Methods used to perform the meta-analysis.
4. The conclusions which are reached.
5. Statements by the authors regarding the reliability of their particular meta-analysis.
6. Declarations of broad applicability for the conclusions that are derived in a particular meta-analysis.

Meta-analyses can be quite useful and beneficial for the analysis of similar trials. However, an assumption that every meta-analysis represents the final and accurate viewpoint on an area of research is unwarranted.

NETWORK META-ANALYSIS

There is a type of meta-analysis called network meta-analysis that is potentially more subject to erroneous conclusions than a routine meta-analysis. A network meta-analysis adds an additional variable to a meta-analysis. Rather than simply summing up trials that have evaluated the same treatment compared to placebo (or compared to an identical medication), different treatments are compared by inference. (If A is better than B, and B is equal to C, then A is better than C.)

The problem with network analysis in regards to a meta-analysis, is that a network meta-analysis tends only to be valid for very similar studies. Since network meta-analysis combines multiple studies, there is even more potential for combining studies that are not adequately similar. The quality of some recent network analyses for the hypertensive literature highlights the problems of this type of analysis.

STRENGTHS AND POTENTIAL WEAKNESSES OF LARGE RANDOMIZED CLINICAL TRIALS.

Very large randomized trials can have significant problems in trial construction and interpretation.

The very large randomized clinical trial has many significant advantages. A major strength is that there are sufficient numbers of patients involved so that the results of the trial are reproducible if the same trial is repeated in the same patient population. In addition, a very large trial is usually sufficiently powered for statistically valid prespecified subgroup analyses to be performed. A very large trial greatly increases the probability that repetition of that very same trial would result in the same outcome.

However, a very large trial does not guarantee that the trial protocol is optimal or that the trial results are broadly applicable or that the conclusions reached are valid.

Does a very large randomized clinical trial guarantee an inherently good trial?

The quality of a very large clinical trial is dependent on having a well constructed treatment protocol to the same degree that a smaller trial is dependent on this issue. If one or both of the treatment arms of a clinical trial has serious limitations in regards to representing optimal therapy for a particular strategy being evaluated, then the clinical applicability of the results to patient care is less directly meaningful. **Hence, a very large trial is dependent on the quality of the treatment protocol in regards to the clinical applicability of the trial for physicians in managing and treating similar patients.**

Validity of Conclusions of Large Randomized Clinical Trials

The conclusions that the authors of a very large randomized trial derive from their results are subject to

the same potential for author bias as the conclusions that are made in a small study. The authors of both types of studies give an opinion on the strengths and limitations of the study protocol as well as a statement regarding how broadly applicable the trial results are thought to be. The authors can state conclusions that are a direct reflection of the trial results or make conclusions not truly warranted by their study. **Though there is an admirable tradition of reaching conclusions that conservatively reflect the data in very large clinical trials, this practice of reaching conclusions that are directly derived from the data is not uniformly practiced, even in very large randomized trials**.

A TALE OF TWO LARGE TRIALS

One trial done quite well and the other suboptimally interpreted.

The following compares two very large randomized trials, one done very well, the **Heart Protection Study**[1], and one interpreted quite suboptimally, the **ALLHAT Trial**[2].

Very large randomized clinical trials effectively deal with the issue of having adequate numbers for statistical significance. A very large randomized trial ensures reproducibility of results if the trial was repeated in the same fashion.

A large trial, however, does not guarantee that the treatment protocols being studied are good ones or broadly applicable, nor does it guarantee that valid conclusions regarding the trial's outcome are made by the authors.

Heart Protection Study (An Excellent Study)

The Heart Protection Study1 was a trial of 20,536 patients. The patient population consisted of individuals at substantial risk for developing a future coronary artery disease event because of preexisting coronary disease or very high risk factors. The trial treated all patients with the same dose of cholesterol lowering medicine (simvastatin - Zocor 40 mg) regardless of whether the patient's pretreatment cholesterol was high, moderate, or low. **This landmark trial reliably showed through the strength of its large numbers and clinically relevant treatment protocol that the benefit of a statin medication for this group of patients was the same regardless of the initial level serum cholesterol. The results of this trial will favorably change the treatment of patients at high risk for heart disease.** The fundamental understanding of physicians of whether a baseline cholesterol level is pertinent in regards to initiation of therapy for this population has been effectively and appropriately changed.

However, even in this excellent trial, it would have been possible for the authors to extend their conclusions beyond what the trial results warranted. The authors could have inappropriately concluded that their trial results applied only to the particular cholesterol drug tested, simvastin. Their unfounded conclusion would then have been that only simvastin should be used to treat patients with coronary heart disease who have a normal or low pretreatment cholesterol levels**. In addition, the authors could have inappropriately concluded in their report's formal conclusions that their trial data proves that measuring cholesterol levels in this type of patients is not useful since all these patients should be treated with the cholesterol lowering medicine regardless.** Though individual authors of this study may have thought the trial raises the question of whether follow-up blood tests are really needed, they did not inappropriately extend their conclusions beyond what the trial directly tested.

Hence, even an excellent trial such as the Heart Protection Study could be substantially diminished if the authors of that trial had chosen to make inappropriate conclusions and if the editors of the journal that published the paper did not require those conclusions to be revised prior to publication. Fortunately, the Heart Protection Study had stellar investigators both in the formulation of their trial protocol and in the conclusions which they directly derived from the data.

ALLHAT trial (A large trial, suboptimally interpreted)

The ALLHAT trial on the other hand, is a poster child for conclusions extending beyond the data in a large randomized trial. The ALLHAT trial failings result primarily from the overextended and inappropriate conclusions. **The trial treatment protocol specified specific blood pressure treatment regimens which had serious limitations in regards to substantially differing from the way hypertensive patients are currently treated by physicians. If the trial had been conservatively interpreted by the trial's authors, these limitations would have been noted and some conclusions deriving directly**

from the trial would have been formulated. Instead, the authors in their 2001 report made a sweeping statement that their trial results indicated that a diuretic should be the first drug used for the treatment of hypertension.

There are a number **of major limitations in regards to the ALLHAT trial results. Despite the fact that the primary end point was identical for the three treatment strategies, the ALLHAT authors inappropriately stated that their trial data proved that diuretic therapy is the preferred drug in initiating treatment for hypertension. In fact, what the ALLHAT trial reliably showed was the outcome for a particular combination of drugs used in this patient population.**
(See www.improvingmedicalstatistics.com/ALLHAT.htm for an extended critique of the ALLHAT trial.)

The advantage of this being a very large trial is that if this trial protocol was repeated for these particular suboptimal combinations of medication for this particular patient population, the same results would occur. Repeating the ALLHAT trial with a similar 100,000 patients rather than 30,000 patients would not make the specific blood pressure combinations studied more broadly applicable, though it would be expected to have the same patient outcome. A very large randomized trial only ensures reproducibility in that the same trial would have similar results if repeated in the same fashion.

Hence, very large randomized clinical trials effectively deal with the issue of having adequate numbers for statistical significance. A large trial, however, does not guarantee that the treatment protocols being studied are good ones, nor does it guarantee that correct conclusions regarding the trial's outcome are made by the authors.

1. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20536 high-risk individuals: a randomised placebo-controlled trial. Lancet 2002; 360: 7-22.

2. ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs. diuretic: the antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). JAMA 2002; 288: 2981-2997.

BEWARE OF META-ANALYSES BEARING FALSE GIFTS.

Meta-analyses performed by strong advocates of a particular position in an ongoing controversy are at higher risk for bias.

A meta-analysis is subject to a set of potential problems and pitfalls just as a clinical trial is subject to potential problems. It is well documented that the conclusions of a meta-analysis (a summation of multiple smaller trials) can be shown to differ from a subsequent, large, more definitive, randomized clinical trial.

The initial hurdle for doing a clinically meaningful meta-analysis is the criteria for how similar the studies must be in order to be included in the meta-analysis. The more similar the studies that are combined, the more valid the meta-analysis. If the meta-analysis combines multiple trials for similar patient populations where a placebo is compared to the same drug; or compares identical treatment regimens, then the meta-analysis results and conclusions are more likely to be in concordance with reality. A less restrictive requirement of similarities between the studies allows more trials to be entered into the meta-analysis which makes it easier to reach statistical significance. However, this actually decreases the reliability of the conclusions of the meta-analysis.

The interpretation of a meta-analysis is potentially subject to an author's bias by what inclusion and exclusion criteria is selected, the type of statistical evaluation performed, decisions made on how to deal with disparities between the trials, and how the subsequent results are presented. Whether the conclusions of a meta-analysis are broad reaching or limited can be affected by the inherent bias that the author of the meta-analysis brings to the study.

Human nature dictates that each of us tends to find it

more satisfying to confirm a previously held opinion, particularly a published opinion, rather than create an analysis that refutes our own prior conclusions. Hence, interpretive bias is even more likely to occur when a meta-analysis is conducted by an author with a strong particular viewpoint in an area of controversy. When the meta-analysis is conducted by a strong advocate of a particular position, it is more likely to be biased in concordance with the author's previously advocated opinion.

A meta-analysis[1] was subsequently published after the ALLHAT trial publication[2] by the some of the same authors who were involved in formulating ALLHAT's inappropriate conclusions. The authors of this meta-analysis tried to bolster their contention that the ALLHAT trial demonstrated that a diuretic drug should be the initial drug used for the treatment of hypertension. **The overly broad conclusions of this meta-analysis do not appropriately reflect the differences in blood pressure between the diuretic led therapy vs. the other therapies studied.**
(For a detailed critique of the ALLHAT related meta-analysis, see:
www.improvingmedicalstatistics.com/allhat_metfinal.htm)

Two separate meta-analyses analyzed the effects of oat bran and other soluble fibers on cholesterol levels.[3,4] A prior flawed individual study incorrectly stated that oat bran does not significantly lower cholesterol.[5] A subsequent meta-analysis written by the senior author of that study was interpreted in a manner to minimize any incongruity with the prior initial incorrect study. A separate meta-analysis by a different group concluded that oat bran modestly reduced cholesterol.[4] (For a more detailed examination of how the conclusions of the meta-analysis in this case can be slanted pro or con, see pages 10 and 11.)

A meta-analysis is not some infallible, final, arbitrator of a clinical question. This is particularly the case when there are significant differences between the trials being combined and when different patient populations are being studied. More reliable is a single, very large, well done

study comparing treatment options consistent with the modern management of patients. This is particularly the case when those results are interpreted in a conservative manner where the conclusions are directly derived from the clinical results. (The Heart Protection Study is an example of an optimally conducted and interpreted large clinical trial and is in contrast to a suboptimally interpreted large trial, the ALLHAT trial.)

1. Psaty B, Lumley T, Furberg C, et al. Health outcomes associated with various antihypertensive therapies used as first-line agents, a network meta-analysis. JAMA 2003; 289: 2534-2544

2. ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs. diuretic: the antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). JAMA 2002; 288: 2981-2997

3. Ripsin CM, Keenan J, Van Horn L, et al. JAMA 1992; 267:3317-25. Oat Products and Lipid Lowering. A Meta-analysis.

4. Brown L, Rosner B, Lillett W, Sacks F. Cholesterol-lowering effects of dietary fiber: a meta-analysis. Am J Clin Nutr 111; 69:30-42

5. Swain JF, Rouse,IL, Curley CB, Sacks F. Comparison of the effects of oat bran and low-fiber wheat on serum lipoprotein levels and blood pressure. N Engl J Med 1990; 322:147-52.