Critically Appraising Randomized Controlled Trials: Is there Substance in Subgroups?

Ricky D Turgeon, B.Sc.(Pharm.), Pharmacy Resident[1]
[1]*Lower Mainland Pharmacy Services, Vancouver, BC, Canada*

"Half of what you'll learn in medical school will be shown to be either dead wrong or out of date within five years of your graduation; the trouble is that nobody can tell you which half—so the most important thing to learn is how to learn on your own." – David Sackett, father of evidence-based medicine

The evidence-based medicine (EBM) movement, particularly the popularization of randomized controlled trials (RCTs), has facilitated large therapeutic advances. In addition to ascertaining the efficacy and safety of new treatments, RCTs have also led to the discovery of harmful effects of interventions long accepted as beneficial. A recent example of this is the ACCORD trial (1,2). This trial showed that, contrary to conventional beliefs, aggressive control of blood glucose and blood pressure in type 2 diabetics produced increases in mortality and serious adverse events, respectively. Despite the strengths of RCTs, they are rarely without their own limitations that must be carefully considered during their interpretation.

In this PSSJ Workshop, we will discuss one of the aspects of RCTs that often misleads healthcare professionals: Subgroups.

Large clinical trials often include patient populations with a range of characteristics including age, gender, race, comorbidities, severity of illness, etc. Subgroup analysis is a statistical technique that divides trial participants into two or more cohorts based on a certain characteristic, such as pre-existing cardiovascular disease or stage of cancer. The effect of the intervention on each of these subpopulations separately is then assessed. The goal of this approach is to individualize care based on patient factors. However, subgroup analyses are frequently overused to mine for positive results, often leading to spurious conclusions (3). Subgroup analysis also contributes to a problem in clinical trials known as multiplicity. That is, if we test a sufficient number of hypotheses, we are bound to eventually find a positive result purely by chance, also known as a false positive. On the other hand, loss of power from dividing the trial population into increasingly smaller populations can also contribute to falsely concluding that an intervention has no effect in a subgroup, known as a false negative.

The authors of the ISIS-2 trial illustrate how subgroups can mislead us (4,5). In this trial, investigators assessed the efficacy and safety of aspirin and streptokinase, separately or combined, versus placebo in patients with suspected myocardial infarction. Aspirin reduced mortality in the overall population, but among 40 subgroups evaluated it was no better than placebo in those born under the Gemini or Libra astrological signs, those with a prior myocardial infarction, and in diabetics. Many readers would dismiss the first subgroup and contemplate the latter two based on biological plausibility, though it is highly likely that all of these are chance findings due to the sheer number of subgroups observed. Thus, without criteria with which to assess the validity of these subgroups, readers may be compelled to withhold lifesaving aspirin in individuals having a heart attack.

In an attempt to improve the rational interpretation of subgroups, Sun and colleagues proposed 11 criteria to evaluate the credibility of subgroup analyses (Table 1) (6). In this article, we will interpret the results of the female gender subgroup of the Heart Protection Study (HPS) using the Sun et al criteria (7).

Do Women Benefit from Statins? Answers from a Heart Protection Study Subgroup Analysis

Cardiovascular disease is the second leading cause of death in Canadian women (8). Despite this, women who have had a cardiovascular event receive fewer evidence-based interventions than men (9). This may stem from a reluctance to apply findings from RCTs in cardiology to women due to the disproportionate inclusion of males in these studies. The Heart Protection Study was a 5-year double-blind, placebo-controlled randomized trial designed to assess the safety and efficacy of statin therapy in over 20,000 men and women at high-risk of death from cardiovascular disease (7).

During study design, investigators selected seven baseline characteristics (10), including gender, for

Design	HPS (7) female subgroup
1) Is the subgroup variable a characteristic measured at baseline (optimal) or after randomization?	At baseline
2) Is the effect suggested by comparisons within (optimal) rather than between studies?	Within
3) Was the hypothesis specified *a priori*?	Yes (10)
4) Was the direction of the subgroup specified *a priori*?	No
5) Was the subgroup effect one of a small number of hypothesized effects tested?	Seven subgroup pre-specified (10)
Analysis	
6) Does the interaction test suggest a low likelihood that chance explains the apparent subgroup effect?	No
7) Is the significant subgroup effect independent?	Does not apply
Context	
8) Is the size of the subgroup effect large?	Does not apply
9) Is the interaction consistent across studies?	Meta-analysis finds no significant test for interaction (11)
10) Is the interaction consistent across closely related outcomes within the study?	First major vascular event was the only outcome tested for subgroup effect
11) Is there indirect evidence that supports the hypothesized interaction (e.g. biological rationale)?	None presented
Is there credible evidence for a subgroup effect?	It is unlikely that there is a difference in the relative effect of statins on major vascular events between men and women based on this trial

Table 1. Criteria to assess the credibility of subgroup analyses from Sun *et al* (6).

Received: March 8, 2013
Revised: August 3, 2013

which they would conduct subgroup analyses to see if the effect of treatment varied based on these features (positive criteria 1, 2, 3 and 5 in Table 1). They did not, however, specify how they hypothesized the subgroup would influence the treatment's effect (i.e. whether females would derive more or less benefit from statins than males; negative criterion 4).

Before proceeding to interpreting the separate efficacy estimate for the female subgroup, it is first important to evaluate the test for interaction. This statistical test assesses the likelihood that a difference between groups is due to chance. In this case, the p-value for the test for interaction between males and females was 0.76 (far above the traditional 0.10 threshold for statistical significance used for this test), suggesting that any apparent subgroup difference is due to chance (negative criterion 6; obviates criteria 7 and 8). A look beyond this study confirms a lack of a gender difference in a subsequent meta-analysis of trials (criterion 9) (11). Authors did not present subgroup analyses for the different components of the "major vascular event" outcome, such as death, though it is unlikely that there are differences given the uniformity of the overall outcome (criterion 10). Finally, there is no strong biological rationale for a gender difference in statin efficacy (criterion 11). Given all of the above, we can be reasonably certain that statins provide the same relative benefit in women as they do in men.

The validity of a subgroup analysis is not determined simply by achieving a high score on a checklist. Sun et al propose that credibility of a subgroup effect lies on a continuum (6). Due to the high risk of chance findings, I am usually cautious in interpreting subgroups, requiring criteria 1 through 5 to be satisfied, as well as replication in one additional study, or at the very least a statistically significant test for interaction, before I incorporate these findings into my therapeutic decisions. In a case where the subgroup effect is not sufficiently credible, such as in the above Heart Protection Study illustration, I will instead use the overall trial relative effects to estimate the benefits for my patients in a specific subgroup. Thus, the Heart Protection Study's overall relative risk reduction of 24% with statin treatment can be applied to the 17.7% absolute risk of a major vascular event over 5 years in women receiving placebo to derive an absolute risk reduction of 4.2% with statin therapy, which can also be reported as a number needed to treat of 24.

References

1) The Action to Control Cardiovascular Risk in Diabetes Study Group. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med. 2008;358:2545–59.

2) ACCORD Study Group. Effects of intensive blood-pressure control in type 2 diabetes mellitus. N Engl J Med. 2010;362:1575–85.

3) Rothwell PM. Treating individuals 2: subgroup analysis in randomised controlled trials: Importance, indications, and interpretation. Lancet. 2005;365:176–86.

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

5) Horton R. From star signs to trial guidelines. Lancet. 2000;355:1033–4.

6) Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ. 2010;340:c117.

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

8) Leading causes of death, by sex [Internet]. Ottawa (ON): Statistics Canada; 2012 [cited 2013 Mar 8]. Available from: http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/hlth36c-eng.htm

9) Lahoud R, Howe M, Krishnan SM, Zacharias S, Jackson EA. Effect of use of combination evidence-based medical therapy after acute coronary syndromes on long-term outcomes. Am J Cardiol. 2012;109:159–64.

10) MRC/BHF Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol-lowering therapy and of antioxidant vitamin supplementation in a wide range of patients at increased risk of coronary heart disease death: early safety and efficacy experience. Eur Heart J. 1999;20:725–41.

11) Gutierrez J, Ramirez G, Rundek T, Sacco RL. Statin therapy in the prevention of recurrent cardiovascular events: A sex-based meta-analysis. Arch Intern Med. 2012;172:909–19.