Limitations of Randomized Clinical Trials

John B. Kostis, MD, DPhil*, and Jeanne M. Dobrzynski, BA

Randomized clinical trials are essential for determining the efficacy of interventions, but have limitations. The types of limitations discussed in this review may be grouped in 11 categories including incorrect statistical inference, low internal or external validity, misinterpretation of the difference between frequentist and Bayesian statistical approaches, publication bias, that healthy persons with a given condition participate in clinical trials although they are not representative of the population as a whole, the rather short duration (3 to 5 years) that does not give correct estimates of the lifetime effects of the interventions or the legacy effect when participants who receive active therapy derive residual benefit after the end of the study when all participants receive active medication and the tension between the generalizability of the evidence versus the reliability of the findings of different types of clinical trials and the difficulty in applying the findings of randomized clinical trials to individual patients. These limitations are described and illustrated by examples and figures from the literature. In conclusion, this review will be useful to clinical trialists, clinical trial participants and regulatory agents. **(Am J Cardiol 2020;129:109−115)**

Randomized clinical trials are essential in determining efficacy of pharmacologic and other interventions and together with meta-analyses are at the apex of studies providing reliable information on these issues. Blinding of randomization and treatment allocation of the patient and the investigators (double-blind studies) are common and important. Prospective Randomized Open Blinded End-point design is used in trials where adjustment of the study medications by the research staff is needed and the event adjudicators are blinded. The majority of therapeutic interventions are supported by the findings of randomized trials, and approval of pharmaceuticals in primary and secondary preventive treatments involving drugs are based on randomized clinical trials. However, randomized clinical trials have important limitations. In this review, we describe the important limitations of clinical trials that may affect their application in clinical and research settings. They include general limitations that affect all clinical trials such as statistical inference and p values and specific limitations including the applicability of clinical trials to individual patients and their internal and external validity.

Statistical Inference—Type I and Type II Errors

Statistical inference is a process used to report the overall findings of randomized clinical trials and other research. It is based on comparison of sample data derived from 2 or more distributions. A distribution is an arrangement that shows the frequency of the values of a group (variable). It is a listing of the number of times that a variable takes each of its possible values. Distributions may be presented graphically as a histogram or a curve. In statistical analyses a null hypothesis is set when a statement, usually the opposite of the assumed relation, is stated. There are 2 types of errors in comparing distributions: type I and type II. A type I error occurs when the null hypothesis is rejected when in reality it is true. Ordinarily, the probability of committing a type I error is equal to the a priori set value for significance for the hypothesis being tested, usually 0.05. However, this probability may be higher when the sample of items included in the study is not random and therefore does not represent the intended population to be studied leading to a false-negative decision. The probability of committing a type II error is equal to 1 minus the statistical power of the test and can be decreased by increasing the sample size. In many instances type I and type II errors, especially the latter lead to misinterpretation of the findings of randomized clinical trials. Thus, a type I error in comparing samples from 2 (or more) distributions may lead to a false-negative decision whereas a type II error may lead to a false-positive decision.

An example of a type I error is the concluding statement of the Aspirin Myocardial Infarction Study (AMIS). AMIS was a National Heart, Lung, and Blood Institute-sponsored, multicenter, randomized, double-blind, placebo-controlled trial designed to test whether the regular administration of aspirin in persons who had experienced 1 or more myocardial infarctions would result in lower rate of mortality during a 3-year follow-up. This study was performed with extreme care not to bias the results including the intent-to-treat approach to the events, adjudication of mortality and nonfatal events blinded to the medication allocation by investigators who were blinded. In addition, adherence to blinded medication was examined by pill count, platelet aggregation testing, and urine test for salicylates performed in a blinded manner. However, the final statement of AMIS was "based on AMIS results aspirin is not recommended for the routine use in patients who have survived a myocardial infarction".[1] In retrospect, this statement is wrong and AMIS represents a false-negative study since aspirin was better than placebo in the Persantine-Aspirin Reinfarction Study.[2] In addition, a wealth of evidence supports the use of aspirin in the secondary prevention of atherosclerotic

Cardiovascular Institute, Rutgers Robert Wood Johnson Medical School, New Brunswick, New Jersey. Manuscript received February 27, 2020; revised manuscript received and accepted May 5, 2020.

Funding: None.

See page 115 for disclosure information.

*Corresponding author: Tel: (732) 235-7685; fax: (732) 235-7039.

E-mail address: kostis@rwjms.rutgers.edu (J.B. Kostis).

0002-9149/
https://doi.org/10.1016/j.amjcard.2020.05.011

cardiovascular disease.[3] Also, current American College of Cardiology/American Heart Association guidelines recommend lifetime use of aspirin in all patients with coronary artery disease.[4] The false-negative findings of AMIS probably represent a chance finding.

An example of a type II error was the erroneous impression that statins did not decrease cardiovascular morbid and mortal events in women. Previous studies did not show statistically significant benefits in women or a significant interaction by gender especially in primary prevention probably because of underrepresentation of women in these trials. This implied lack of benefit of statins in women, that is, a type II error. This error was corrected by increasing the aggregate number studied and the statistical power of the relevant previous studies by performing a meta-analysis. Eighteen randomized clinical trials with sex specific outcomes including 141,235 men and 40,275 women with 21,468 cardiovascular events were included in the meta-analysis. The analysis indicated a statistically significant benefit of statins in both sexes in both primary and secondary prevention. The benefit was quantitatively similar in women and men (odds ratio [OR] 0.81, 95% confidence interval [CI] 0.75 to 0.89; $p < 0.0001$, and OR 0.77, 95% CI 0.71 to 0.83, $p < 0.0001$, respectively) and there was no significant interaction by gender (p for interaction = 0.45).[5]

Internal and External Validity

The internal and external validity of a given clinical trial are examined in order to estimate their utility in practice. To avoid bias and assure internal validity, randomized controlled trials are conducted using strict protocols that may compromise their external validity. Detailed inclusion and exclusion criteria result in increased internal validity whereas unbalanced co-interventions, crossovers, and patients lost to follow-up decrease external validity. Tessa Kennedy-Martin et al reported on 52 studies (20 in cardiology, 17 in mental health, and 15 in oncology). They concluded that some trials were conducted in highly selected patients that were not representative of real-world populations who had lower risk profiles. They recommend a thoughtful approach to clinical evidence where the trade-offs between internal and external validity are considered in a balanced manner.[6] Another approach is to perform real-world pragmatic trials. Pragmatic trials do not have the drawbacks of clinical trials described above where the participants are different from the general population. Benefits of pragmatic trials are their high generalizability and external validity, although, they may have variability in the quality of the data, crossover and lack of standardization.

The P Value

Fisher and Mackenzie developed the statistical theory underlying comparative clinical trials when they studied the effect of different potato varieties to manure.[7] They reported that there was no significant variation in the response of different varieties to manure with respect to weight of potatoes per plot. Fisher then developed the concept of the "p value" that compares the findings of a study to a hypothesis, the "null hypothesis." The p value is a function of the observed sample results (in this case the yields of the fields used by Fisher) assuming that the sample used by R.A. Fisher was representative of all similar fields that were used for testing the null hypothesis. A threshold p value (usually 0.05) is stated before the study and when the p value is equal to or higher than this level the observed data are not compatible with the null hypothesis which is rejected. Therefore, the p value does not prove or reject a hypothesis; rather it describes the findings of the study.

P values depend upon both the magnitude of association and the sample size. As a result, the p value comparing 2 distributions decreases as the number of elements (N) increases. If the magnitude of effect is small and clinically unimportant, the p value can be ''significant'' if the sample size is large. Conversely, an effect can be large, but fail to meet the $p < 0.05$ criterion if the sample size is small. There is a temptation to embark on ''fishing expeditions'' in which investigators test many possible associations. When many possible associations are examined using a threshold p value of ≤ 0.05, the probability of finding at least one that meets the critical point increases with the number of times the hypothesis is tested. The Bonferroni correction is used to adjust for the error resulting from multiple testing. The Bonferroni corrected threshold becomes smaller and smaller according to the number of tests performed on the same hypothesis. So, when the p value stated before the experiment is 0.05 with repeated tests, the threshold p value becomes smaller and smaller according to the number of tests, for example, if 3 tests are done it is 0.05 divided by 3 (i.e., $p = 0.016$). The corrected p value is used to compensate for type I errors (see above type I and type II errors). The p values and the corrected p values are used so that "we are not frequently wrong." In addition, statistical significance does not account for the effects of bias and confounding.

In March 2016, the American Statistical Association released a statement on Statistical Significance and p values stating the following 6 principles on the proper use and interpretation of the p value.[8] (1) P values can indicate how incompatible the data are with a specified statistical model; (2) p values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone; (3) scientific conclusions and business or policy decisions should not be based only on whether a p value passes a specific threshold; (4) proper inference requires full reporting and transparency; (5) a p value, or statistical significance, does not measure the size of an effect or the importance of a result; and (6) by itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

The increased quantification of scientific research and a proliferation of large, complex data sets has expanded the scope for statistics and the importance of appropriately chosen techniques, properly conducted analyses, and correct interpretation. The p value was never intended to be a substitute for scientific reasoning. Well-reasoned statistical arguments contain much more than the value of a single number and whether that number exceeds an arbitrary threshold. The American Statistical Association statement was intended to steer research into a "post $p < 0.05$ era." Nevertheless, it appears the p value has become a gatekeeper for whether work is published.

Frequentist and Bayesian Approaches to Statistical Inference

The American Statistical Association statement on p values emphasizes estimation over testing, including confidence, credibility, or prediction intervals; Bayesian methods, likelihood ratios (Bayes Factors) and other approaches such as decision-theoretic modeling. These approaches rely on further assumptions, but they may more directly address the size of an effect (and its associated uncertainty) or whether the hypothesis is correct.[8] Decision analysis describes how decisions should be made using a set of tested tools for framing the questions, structuring decision problems, quantifying uncertainty and preferences, and discovering factors in a model that are critical for the decision. In frequentist statistics data are presented using descriptive statistics such as the mean, variance, standard deviation, median, interquartile range. Also, samples are compared using statistical inference that allows one to make general statements by rejecting or not rejecting hypotheses as described above. In light of misuses and misconceptions concerning p values, statisticians often supplement or even replace p values with other approaches. These include methods "that emphasize estimation over testing such as confidence, credibility, or prediction intervals; Bayesian methods; alternative measures of evidence such as likelihood ratios or Bayes factors; and other approaches such as decision-theoretic modeling and false discovery rates. In Bayesian analysis the probability before an analysis (Figure 1, top) is modified by new evidence resulting in revision of the previous beliefs to the posterior beliefs.

An example of the difference between the frequentist and Bayesian approaches in analyzing the same data of randomized controlled trials pertains to the question on whether statins should be used in persons older than 75 without history of atherosclerotic cardiovascular disease, that is, for primary prevention. Statins have been successfully used to reduce the risk of cardiovascular disease for more than 30 years and they have become one of the most proven pharmacologic interventions for atherosclerotic cardiovascular disease. However, current American College of Cardiology/American Heart Association clinical guidelines do not recommend use of statins for primary prevention in persons >75 years of age because there is not sufficient information to prove a benefit and because of uncertainty regarding drug safety, drug to drug interactions, and polypharmacy in this age group. A meta-analysis of 35 randomized clinical trials examined the effects of statins compared

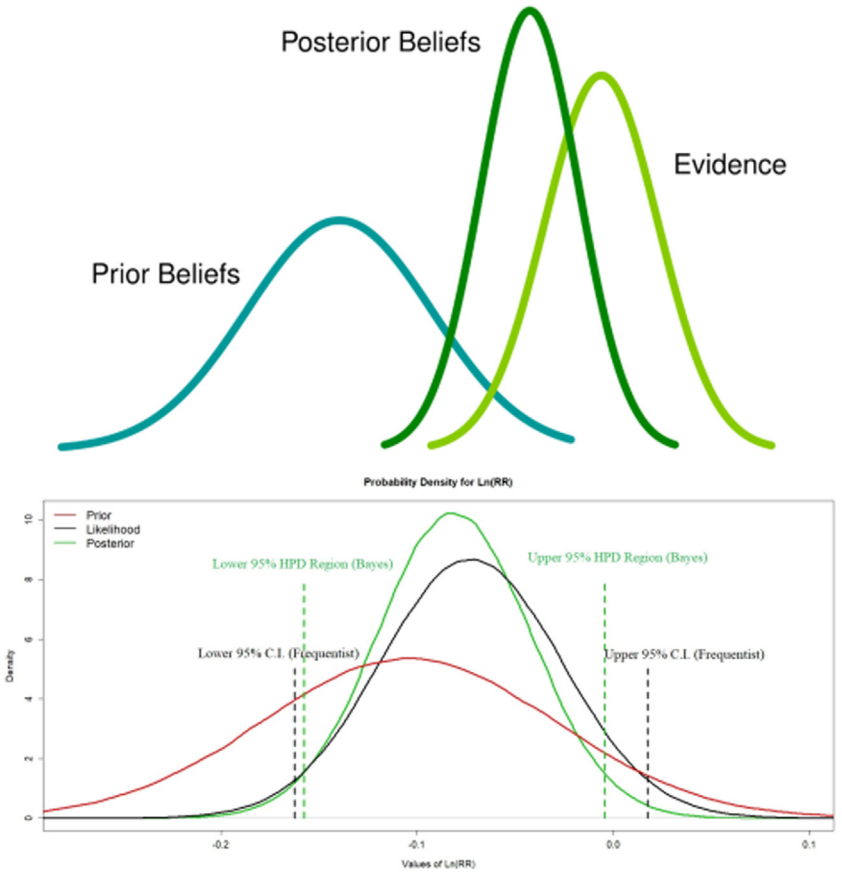

Figure 1. (*Top*) In Bayesian analysis the probability before an analysis (previous beliefs) is modified by new evidence resulting in revision of the previous beliefs to the posterior beliefs.
(Bottom) For all-cause death frequentist analysis indicated a nonsignificant effect for treating patients older than 75 with statins. However, using Bayesian analysis, a significant benefit in terms of all-cause death is observed. Red line is the previous distribution. Black line shows data in studies on patients older than 75. Green line is posterior distribution after consideration of the previous distribution and the data available on patients older than 75 using Bayes' rule. HPD = highest Posterior Density.

with placebo or usual care on all-cause mortality by both frequentist and Bayesian approaches. In this study, the frequentist analysis did not show a statistically significant difference between cases (on statins) and controls ($p = 0.16$), whereas the Bayesian analysis indicated a definite, statistically significant ($p = 0.03$) and clinically relevant benefit of statin treatment for primary prevention in persons >75 years of age (Figure 1, bottom).[9]

In addition to the frequentist, Bayesian and decision-theoretic modeling approaches, a "false discovery rate" approach may be used when Bonferroni correction is not appropriate because the p values are too many or correlated.[10] This approach was used in Trial of Non-pharmacologic Interventions in the Elderly where there were so many analyses performed that statistically significant results were likely to occur by chance. A randomized dataset where the genotype (21 polymorphisms) of each participant was assigned a phenotype (all other variables including blood pressure) at random was created. Since in this randomized dataset there was no relationship between genotype and phenotype, any statistically significant associations happened by chance because of the large number of analyses performed. The number of real associations is equal to the number of observed associations in the original dataset minus the number of false discoveries at a particular p value.[11]

Publication Bias and Regression to the Truth

These 2 types of bias are corrected by regression to the mean when more publications become available. The practice of scientific journals to publish reports with statistically significant findings leads to the file-drawer problem where research with statistically significant outcomes are much more likely to get published, whereas other work that might be just as scientifically important is never published. It also leads to practices called by such names as "p-hacking" and the difficulty in applying the findings of randomized clinical trials to individual patients that emphasize the search for small p values over other statistical and scientific reasoning. An additional problem is that compliance with the Food and Drug Administration Amendments Act of 2007 by ClinicalTrials.gov to report the results is poor with only 40.9% (95% CI 39.4 to 42.2) complying within the 1-year deadline.[12]

The opposite problem of publication bias is the phenomenon of "regression to the truth." Early trials report very significant usually positive findings whereas the average effect is observed when many reports are published. Regression to the truth derives from the biological concept of regression to the mean, whereby random fluctuations in a biological variable occur over time, such that the true value of the variable is approached with repeated measurements.[13]

Clinical Trial Participants Are Healthier Than the Average Person With the Condition Under Study

Clinical trial participants are usually healthier than the average population suffering with the condition under study because they are health conscious with better life-style. Also, the physical and laboratory examinations dictated by the research protocol may discover "incidentalomas," that is, important clinical conditions that would not have been discovered during usual healthcare, but are detected because of the research protocol. This does not apply to patients with cancer and other malignancies where a clinical trial is the last resort after standard therapies have been unsuccessful. The Systolic Hypertension in the Elderly Program (SHEP) is an example of a clinical trial where the participants were healthier than the average person with systolic hypertension.

Compared with an exact age and gender matched cohort, SHEP participants had markedly higher overall life expectancy ($p < 0.0001$) and greater chance of reaching the ages of 80 (81.3% vs 57.6%), and 90 (30.5% vs 22.0%; Figure 2). The high survival was attributed to the strict protocol of SHEP that only included participants who were alive at the age of 60 (average age 72) and healthy with relatively low rate of smoking and diabetes and without history of neoplasia or other serious disease. The benefit of randomization to the active group was much smaller than the difference between the randomized groups and actuarial controls.

Estimation in Gains in Life Expectancy Using Randomized Clinical Trials and the Legacy Effect

The usual duration of randomized controlled clinical trials is approximately 3 to 5 years although hypercholesterolemia and other risk factors for atherosclerotic cardiovascular disease are lifelong conditions. The long-term follow-up of SHEP provides an opportunity to examine the long-term effects of a therapeutic intervention, that is, chlorthalidone-based stepped care therapy compared with placebo.

At the 22-year follow-up, participants in the active treatment group lived 145 more days (95% CI 23 to 260, $p = 0.012$) than those in the placebo group (Figure 2).[14]

The "legacy effect" is the persistence of beneficial effects of pharmacologic interventions after the end of the randomized phase in an open label follow-up trial when all participants, those initially randomized to active therapy and those randomized to placebo, receive active therapy. The legacy effect is important in providing information on the long-term benefits or harms of treatment in patients who continue active therapy and those who discontinue therapy. Most of the information on the legacy effect is derived from retrospective analyses of randomized clinical trials in patients with hypercholesterolemia, diabetes and hypertension.

The legacy effect of lipid-lowering agents was shown in a meta-analysis of 8 randomized trials. The benefit observed in the randomized period of the studies, a decrease in all-cause mortality (0.84, 95% CI 0.76 to 0.93; $p = 0.0006$), was also observed during the follow-up phase (OR 0.90, 95% CI 0.84 to 0.97, $p = 0.0035$). A similar pattern was observed in cardiovascular mortality (0.72, 95% CI 0.63 to 0.82; $p = 0.0001$) in the randomized phase and 0.82 95% CI 0.73 to 0.93 $p = 0.0014$ in the follow-up phase when all participants, that is, those initially randomized to active therapy and those initially randomized to placebo, received active therapy.[15] Similar legacy effects were observed with niacin in the Coronary Drug Project.[16]

Generalizability of Evidence Versus Reliability of Findings

An important consideration in applying the findings of clinical trials and other research designs in practice is the fact that as the reliability of the evidence increases the

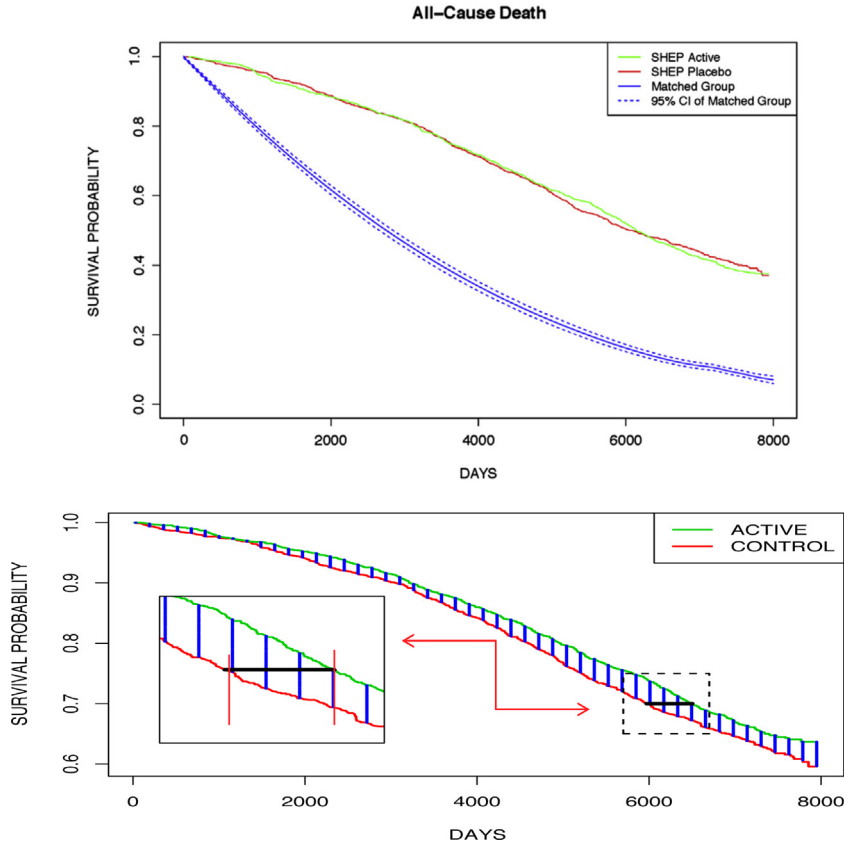

Figure 2. *(Top)* SHEP participants had markedly higher overall life expectancy ($p < 0.0001$) and greater chance of reaching the ages of 80 (81.3% vs 57.6%), and 90 (30.5% vs 22.0%) compared with an exact age and gender matched cohort.
(Bottom) The long-term effect of chlorthalidone-based stepped care therapy compared with placebo. At the 22-year follow-up, participants in the active treatment group lived 145 more days (95% CI 23 to 260, $p = 0.012$) than those in the placebo group.

generalizability decreases. For example, studies such as case reports and case-controlled studies are easy to apply directly to the individuals participating in the studies. However, such studies cannot be generalized to large groups of patients. On the contrary, meta-analyses and clinical trials generalize their findings to the group of patients studied whereas it is difficult to apply the findings to individual patients.

Bias in Choosing the Question to be Studied, Type of Controls and Inadequate Methodologies Pertaining to Randomization and Blinding

Randomized clinical trials may include many types of bias some of which are highlighted in this review. An important somewhat subtle bias pertains to equipoise and the chance of observing a beneficial effect in trials sponsored by for-profit organizations. Als-Nielsen and associates, examining 370 randomized drug trials that were included in published meta-analyses, observed that trials funded by for-profit organizations were more positive. After adjustment for treatment effect and double-blinding, conclusions were significantly more likely to recommend the experimental drug as treatment of choice in trials funded by for-profit organizations compared with trials funded by non-profit organizations (OR 5.3, 95% CI 2.0 to 14.4).[17] Djulbegovic et al reported that in trials supported only or in part by commercial organizations, new treatments were significantly favored over the standard treatment (74% vs 26%, $p = 0.004$). In this study, the violations may be due to inferior comparators.[18] Also, in studies of children, the parents may not be in equipoise and prefer a newer treatment.[19] Examining 234 unique meta-analyses containing 1,973 trials, Savovic et al reported that study design characteristics were associated with exaggerated estimates of intervention effects reporting subjectively assessed outcomes such as inadequate generation of randomization sequence, allocation concealment, and lack of double blinding.[20]

Presentation of RRR Appears to Magnify the Benefit of Randomized Clinical Trials

The results of randomized clinical trials may be presented by emphasizing different quantitative aspects of the results. Relative risk reduction is calculated as the risk (number of events divided by the number treated) in the control group minus the risk in the treatment group divided by the risk in the control group. It is the percentage decrease in risk. Many reports emphasize the relative risk of the findings because usually it is numerically large. Figure 3 shows the effect of treating patients with isolated systolic hypertension with chlorthalidone-based stepped-care therapy compared with placebo in preventing hospitalizations for heart failure. A significantly lower rate of heart failure admissions occurred in the treatment group. The relative risk reduction was 47% but because the risks in the

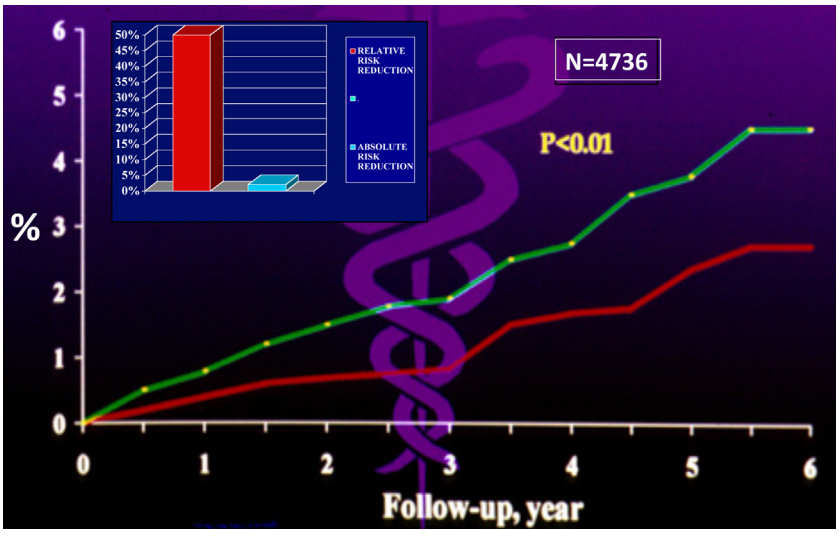

Figure 3. Presentation of relative risk reduction and absolute risk reduction of heart failure hospitalizations at the 6-year follow-up of SHEP participants randomized to active therapy (red) or placebo (green). The relative risk reduction is 47% (red bar in the insert) and the absolute risk reduction is 2.10% (blue bar in the insert).

treatment and control group were low (2.33% and 4.43%, respectively), the absolute risk reduction is low (2.10%). A different way to describe the true magnitude of the effect is to report the Number Needed to Treat to prevent one event (NNT). The NNT is the reciprocal of the absolute risk reduction and in this case 48. It is important to report RRR, ARR, NNT and their confidence intervals.[21]

The Difficulty in Applying the Findings of RCTs to Individual Patients

A difficulty in applying the results of clinical trials to individual patients is emphasized by the fact that many patients have individual characteristics that are not common in the outcomes whereas a positive or negative effect is observed when all patients are considered as a whole. The concept of the p value developed by Fisher and Mackenzie is described including fields with manure and control fields where there was no manure, "placebo fields" (Figure 4). Fisher and Mackenzie used the weight of potato yield per unit area as the dependent variable. Although fields treated with manure produced a higher weight yield than the "placebo fields," not all potatoes in a given yield are the same as shown in the figure where some potatoes are smaller, and in a clinical sense patients are worse off in spite of the presence of an overall benefit of a clinical trial.[7]

Figure 4. The figure shows that some potatoes are small and therefore the manure had an adverse effect on them. Although fields treated with manure produced a higher weight yield than the "placebo fields," not all potatoes in a given yield are the same as shown where some potatoes are smaller. In a clinical sense, patients are worse off in spite of the presence of an overall benefit of a clinical trial.

Conclusion

Randomized clinical trials are essential in determining efficacy of pharmacologic and other interventions and are necessary for the approval by regulatory authorities. However, randomized clinical trials have important limitations that are discussed in this review. Statistical inference is difficult because inference in general is challenging and does not describe the difficulty in estimating the findings of a given trial throughout the life span of the participants, and that the source of funding influences the outcomes. The American Statistical Association statement on p values states that what is needed is a more nuanced approach to interpreting, communicating, and using the results of statistical methods in research.

Registries, real-world observational data, pragmatic studies, and nonrandomized studies provide complementary evidence to clinical trials in order to generate a more complete picture of the current knowledge of a given issue.[22,23] Also, a recent publication in the *American Journal of Cardiology* has recommended that the findings of a Bayesian analysis indicating the use of statins for primary prevention in persons older than 75 years be considered by guideline committees.[9]

Disclosures

All authors report no conflicts of interest.

1. Aspirin Myocardial Infarction Study Research Group. A randomized, controlled trial of aspirin in persons recovered from myocardial infarction. *JAMA* 1980;243:661–669.
2. The Persantine-Aspirin Reinfarction Study Research Group. Persantine and aspirin in coronary heart disease. *Circulation* 1980;62:449–461.
3. Hennekens CH, Dyken ML, Fuster V. Aspirin as a therapeutic agent in cardiovascular disease. A statement for healthcare professionals from the American Heart Association. *Circulation* 1997;96:2751–2753.
4. 2016 ACC/AHA Guideline Focused Update on Duration of Dual Antiplatelet Therapy in Patients With Coronary Artery Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines: An Update of the 2011 ACCF/AHA/SCAI Guideline for Percutaneous Coronary Intervention, 2011 ACCF/AHA Guideline for Coronary Artery Bypass Graft Surgery, 2012 ACC/AHA/ACP/AATS/PCNA/SCAI/STS Guideline for the Diagnosis and Management of Patients With Stable Ischemic Heart Disease, 2013 ACCF/AHA Guideline for the Management of ST-Elevation Myocardial Infarction, 2014 AHA/ACC Guideline for the Management of Patients With Non-ST-Elevation Acute Coronary Syndromes, and 2014 ACC/AHA Guideline on Perioperative Cardiovascular Evaluation and Management of Patients Undergoing Noncardiac Surgery. *Circulation* 2016;134:e123–e155.
5. Kostis WJ, Cheng JQ, Dobrzynski JM, Cabrera J, Kostis JB. Meta-analysis of statin effects in women versus men. *JACC* 2012;59:572–582.
6. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials* 2015;16:495.
7. Fisher RA, Mackenzie WA. Studies in crop variation. II. The manorial response of different potato varieties. *J Agric Sci* 1923;13:311–320.
8. Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. *Am Statistician* 2016;70(2):129–133.
9. Kostis JB, Giakoumis M, Zinonos S, Cabrera J, Kostis WJ. Meta-analysis of usefulness of treatment of hypercholesterolemia with statins for primary prevention in patients older than 75 years. *Am J Cardiol* 2020;125:1154–1157.
10. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med* 1990;9:811–818.
11. Kostis WJ, Cabrera J, Hooper WC, Whelton PK, Espeland MA, Cosgrove NM, Cheng JQ, Deng Y, De Staerck C, Pyle M, Maruthur N, Reyes I, Anderson CA, Liu J, Kostis JB. Relationships between selected gene polymorphisms and blood pressure sensitivity to weight loss in elderly persons with hypertension. *Hypertension* 2013;61:857–863.
12. DeVito NJ, Bacon S, Goldacre B. Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: a cohort study. *Lancet* 2020;395:361–369.
13. Krum H, Tonkin A. Why do phase III trials of promising heart failure drugs often fail? The contribution of "regression to the truth". *J Cardiac Fail* 2003;9:364–367.
14. Kostis WJ, Cabrera J, Messerli FH, Cheng JQ, Sedjro JE, Cosgrove NM, Swerdel JN, Deng Y, Davis R, Kostis JB. Competing cardiovascular and noncardiovascular risks and longevity in the systolic hypertension in the elderly program. *Am J Cardiol* 2014;113:676–681.
15. Kostis WJ, Moreyra AE, Cheng JQ, Dobrzynski J, Kostis JB. Continuation of mortality reduction after the end of randomized therapy in clinical trials of lipid-lowering therapy. *J Clin Lipidol* 2011;5:97–104.
16. Canner PL, Berge KG, Wenger NK, Stamler J, Friedman L, Prineas RJ, Friedewald W. Fifteen year mortality in Coronary Drug Project patients: long-term benefits with niacin. *J Am Coll Cardiol* 1986;8:1245–1255.
17. Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials. A reflection of treatment effects or adverse events? *JAMA* 2003;290:921–928.
18. Djulbegovic B, Lacevic M, Cantor A, Fields KK, Bennett CL, Adams JR. The uncertainty principle and industry-sponsored research. *Lancet* 2000;356:635–638.
19. Hissink Muller PCE, Yildiz B, Allaart CF, Brinkman DMC, van Rossum M, van Suijlekom-Smit LWA, van den Berg JM, ten Cate R, de Vries MC. Participation in a single-blinded pediatric therapeutic strategy study for juvenile idiopathic arthritis: are parents and patient-participants in equipoise? *BMC Med Ethics* 2018;19:96.
20. Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, Als-Nielsen B, Balk EM, Gluud C, Gluud LL, Ioannidis JP, Schulz KF, Beynon R, Welton NJ, Wood L, Moher D, Deeks JJ, Sterne JA. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;157:429–438.
21. Kostis JB, Davis BR, Cutler J, Grimm RH Jr, Berge KG, Cohen JD, Lacy CR, Perry HM Jr, Blaufox MD, Wassertheil-Smoller S, black HR, Schron E, Berkson DM, Curb JD, Smith WM, McDonald R, Applegate WB. SHEP Cooperative Research Group. Prevention of heart failure by antihypertensive drug treatment in older persons with isolated systolic hypertension. *JAMA* 1997;278:212–216.
22. Ezzat VA, Lee V, Ahsan S, Chow AW, Segal O, Rowland E, Lowe MD, Lambiase PD. A systematic review of ICD complications in randomised controlled trials versus registries: is our 'real-world' data an underestimation? *Open Heart* 2015;2:e000198.
23. Ford I, Norrie J. Pragmatic trials. *N Engl J Med* 2016;375:454–463.