Are Clinical Trials in Coronary Heart Disease Oversold or Undersold?

KATHERINE M. DETRE, M.D., DR. PH., JAMES WARE, PH. D., AND NATHAN MANTEL, M.A.

SUMMARY Randomized clinical trials constitute the formal experiments in therapeutics. Many such trials in coronary heart disease have terminated inconclusively or in controversy. In this editorial, we analyze some of the methodologic issues that may lead to controversy; the main reason for the low success rate may lie in insufficient understanding of the complex biology of the disease and in failure to select the appropriate models for therapy. We argue that these difficulties only strengthen the need for the rigorous experimental approach to the evaluation of therapies for coronary heart disease.

In his presidential address to the section of Experimental Medicine and Therapeutics of the Royal Society of Medicine in 1949, Professor Pickering described therapeutics as “a branch of medicine that, by its very nature, should be experimental.” Although controlled clinical trials, which are the formal experiments of therapeutics, have gained acceptance in cardiology over the past decade or so, many in the profession are not convinced of the value of such trials.

Among people with experience in large-scale clinical trials, a consensus has emerged on what constitutes methodologic excellence. The criteria include strict definitions of eligibility, detailed procedures for treatment and follow-up, randomized treatment assignment to avoid selection bias and to assure comparability, and unbiased clinical and statistical treatment evaluation. The design should dictate the main statistical comparisons of the study and should, if possible, spell out subgroup analyses that are expected to be of major interest.

Early chemotherapeutic trials of tuberculosis quickly and successfully resolved questions of effective treatment regimens for tuberculosis with minimal complications and side effects. Why have many of the recent coronary heart disease (CHD) trials finished inconclusively and controversially?

The model for CHD is very complex and poorly understood. The therapies on trial tend to be much less specific than chemotherapies and their expected success is usually contingent on the validity of many assumptions, e.g., lowering cholesterol as a CHD intervention, when we only know that cholesterol is a risk factor in the development of CHD. Thus, this type of clinical trial would be handicapped even under conditions of methodologic excellence. In addition, because these trials are based on low event rates, such as mortality, they must go on for years at multiple locations. In time, operational problems inevitably develop with noncompliance and incompleteness of follow up. During the course of the study, reports by others may appear about new subgroups, and on changes in treatment techniques. When such studies are published, methodologic impurities and “outmoded techniques” become the focus of controversies. One may lose sight of the fact that the streptomycin-tuberculosis trials and the Veterans Administration hypertension trial arrived at definite conclusions not because their methods were superior, but because of the large treatment response. Although excellent methods usually accompany good science, they certainly do not guarantee conclusive results. When treatment effects are modest, it is not unusual to find variable or inconclusive study outcomes.

The case in point is illustrated by the aspirin trials in post-myocardial infarction patients. Five of six large studies demonstrated a small beneficial effect on total mortality, whereas the effect in the largest trial was just the reverse. The studies were not statistically significant individually or collectively. Using results of those same studies, Peto reported definite prophylactic value for aspirin, achieving this by focusing his investigations on vascular deaths, including heart disease and stroke. Significance emerged after pooling the data from all six studies.

One of the classic causes for controversy is multiplicity — the analyses of many subgroups or many end points or both. Multiple comparisons are usually associated with large clinical trials, which may yield barely significant overall results. Having invested a great deal of work, time and funding, investigators seek to learn all they can from the study by conducting many comparisons, which increases the chances of false-positive findings. Reduction of vascular deaths in the aspirin example, being only one of the many possible subgroup results, does not carry the same weight as reduction of total mortality would, even if it can be argued that this outcome relates more closely to a scientific hypothesis about the action of aspirin. The multiplicity of tests and the knowledge that counts of vascular deaths are far less reliable and more subject to bias than counts of all deaths weaken the finding.

A similar problem in a much more severe form emerged in the recent Anturane Reinforcement Trial, where it was again hypothesized that deaths and

From the University of Pittsburgh, Graduate School of Public Health, Department of Epidemiology, Pittsburgh, Pennsylvania.

Supported in part by USPHS grant Ca-15686 and by grant RD-102 from the American Cancer Society.

Received November 25, 1980; accepted January 6, 1981.

Circulation 64, No. 4, 1981.
myocardial infarction rates would be reduced by means of the platelet mechanism. However, the effect revealed by the analysis was a reduction of sudden deaths early after myocardial infarction. Can this finding be accepted? The controversies in this study center on three issues: multiplicity; the decision rules to include deaths that occur in a specific time frame relative to drug intake, while excluding deaths outside that time frame; and questions concerning the validity of the “cause of death” classification. Naturally, if nonsudden deaths were misclassified as sudden, conclusions on sudden deaths would be immediately invalidated. However, the issue regarding exclusions could only shed doubts but would not necessarily invalidate the findings. Sackett suggested the need to develop alternatives to analyses based on all events befalling all subjects at any time after randomization. Because the main findings, however, did not bear out prior scientific expectations, the unhypothesized subgroup result on sudden deaths carried less weight because of the multiplicity aspect alone, even though post hoc, an excellent scientific case could be made in its support.

In the midst of doubts about the value of clinical trials in CHD, Lee et al. produced an example for clinicians of how in a simulated randomized clinical trial statistical significance can be obtained, in the absence of any true difference, through large numbers of subgroup analyses. However, such chance findings from multiple comparisons are known in most biologic investigations, not only in clinical trials. We disagree with the recommendation by the authors that the use of statistical adjustment and clinical judgment over the findings are the remedies to overcome these problems, and alternatively wish to emphasize certain points not suggested by them.

Statistical analysis should be sound and complete, including adjustment for confounding should it occur through imbalance of risk factors between the treatment groups. Having done that, however, false security may result, if one ignores that the procedure itself may introduce bias through selection among the several risk factors that could be used in the adjustment. Further, what is not mentioned is that randomization alone provides the best chance for comparability of the unknown risk factors, even if they appear to outweigh the known ones.

However, we fear that weighting the results by “what his training and experience could lead him to reasonably expect” will often not be helpful to the clinician. For example, the crucial subgroup in the simulated trial with the false-positive result consisted of patients with three-vessel disease, abnormal left ventricular contraction and no history of congestive heart failure. What in one’s training would warn us not to accept this subgroup result or, for that matter, an opposite one that might have occurred in those with a history of congestive heart failure? Experience has not shown clinical judgment to be a particularly sensitive measure of scientific validity in therapeutics.

Because therapy is the bread and butter of clinicians and because its science is clinical trials, we suggest that clinicians need to engage in and know much more about randomized clinical trials, including how to design, conduct and participate in them as well as how to evaluate trial results. Medical school would be a good place to start this experience. Clinicians should be able to identify with their own science and not depend on experts to tell them what is right or wrong. Good science in randomized clinical trials, as in other investigations, involves simple principles, such as incorporating prior knowledge into the hypotheses and directing analyses toward evaluation of those hypotheses. Other questions of interest should also be explored by subgroup analyses. Depending on the significance level attaching to the subgroup findings, such findings may occasionally be accepted as demonstrated, though more frequently they will serve only as leads for further studies.

Scientists work years at the bench scoring many failures or inconclusive results for each successful proof of a hypotheses. When the biology is complex, this is expected and prompts us to do more research until the breakthrough occurs. Ideally, this would be the time to think of application of the results to the clinic. However, patients do not have the option of postponing their CHD. The need for therapy is immediate, and randomized trials provide the most ethical and scientific way to proceed in such situations. In the light of vague disease and treatment modalities, inconclusive results can often be expected. Difficulties in experimental therapeutics, just as in basic research, do not make alternative and less scientific methods acceptable or even more feasible. On the contrary, even greater rigor in methods and more emphasis on controlled studies is required when the design is weak because of limitations in substantive knowledge.

References
Are clinical trials in coronary heart disease oversold or undersold?
K M Detre, J Ware and N Mantel

Circulation. 1981;64:667-668
doi: 10.1161/01.CIR.64.4.667

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1981 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on
the World Wide Web at:
http://circ.ahajournals.org/content/64/4/667

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally
published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the
Editorial Office. Once the online version of the published article for which permission is being requested is
located, click Request Permissions in the middle column of the Web page under Services. Further
information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org//subscriptions/