

Use and Abuse of the Controlled Clinical Trial

Hasan Yazici, M.D.

There has been much discussion of late that confidence in medicine is being eroded. No less an authority than Elias A. Zerhouni, M.D., Director of the National Institutes of Health (NIH), Bethesda, Maryland, has recently said, "Forty percent of science news relates to health and medicine," noting as well, "and we are seeing a gradual erosion of public trust."¹ The main thrust of this discussion is the ever profiteering drug industry, the ever loosening professional standards of the physicians, and the standing and expectations from the state health authorities, i.e., the Federal Drug Administration (FDA), among others. The most common scenario is the new "blockbuster" drug, introduced initially as *the* breakthrough. After enjoying the market for several years, the drug turns out to be not much better and, in fact, is observed to be more toxic than what had been available when it was first launched. In all this discussion, with the inevitable "saucy" representation in the news media, I think there is one important "paragraph" that has been forgotten, the highlighting of which is my particular aim with this communication.

Although some would disagree,² the randomized controlled trial (RCT) is the appropriate golden measure to scientifically judge the claimed efficacy of a new remedy. It should come as no surprise, with some historical review, that the first RCT designed to test the efficacy of streptomycin in treating tuberculosis was conducted shortly after World War II in 1948.² During the times that led to the first RCT, scientific research, in general, underwent an important change in methodology. The inductive reasoning that had dominated Western science, practically since the Renaissance, gave way to deductive reasoning, with its "falsifiable hypotheses." These were the times of, among others, Einstein, Medawar,

and Feynman. The dominant philosopher of this new era was, of course, Karl Popper, who, in brief, argued that the main activity of the scientist was to falsify a hypothesis, and it was only after the failure of a genuine effort at falsification that a scientific truth could be established.³

I propose that the RCT is a nearly perfect example of the science of the Popperian deductive era. In the "true to the art" RCT, we start with the hypothesis that a proposed new treatment A is better than the old treatment B or better than doing nothing, treatment P (P for placebo), in managing disease D. To this end, we take a group of patients with disease D and randomize them to treatments A, B or P, with neither patients nor physicians knowing who received what. We perform power calculations in the beginning to account for the likelihood of chance if we see no effect in the end and significance calculations at the end to see the degree to which the efficacy we observed would have been due to sheer chance rather than a true effect. Only if this genuine effort to falsify the hypothesis or "ourselves" turns out to be not successful, can we say that A is better than B or P in managing disease D.

One can also justifiably say that our scientific endeavor at falsification does not end even after our trial is over. When the scientific paper is written and submitted to a reputable journal, the chances are quite good that our peer reviewers will be among our competitors in the field. They too will often not spare their share of vicious criticism, falsification if you will. Finally, at the end, when the paper sees the light of publication, then our successful and unrefuted hypothesis becomes most vulnerable. It is now open to members of the community of scientists to falsify.

The majority of RCTs published in our best scientific journals, at first glance, would seem to satisfy the requirements of an RCT as a scientific endeavor. Randomization, blindness, placebo, power calculation, and scrutiny for statistical significance are present. On the other hand, a closer

Hasan Yazici, M.D., is from the Department of Medicine, Division of Rheumatology, University of Istanbul, Turkey.

Correspondence: Hasan Yazici, M.D., Sefa Sokak, Sen Apt 17/7, Kadikoy, Istanbul 81310, Turkey.

Table 1 Weaknesses of RCTs

-
- Short duration
 - Limited number of patients enrolled
 - Multiple drugs
 - Big differences between the patients enrolled in a trial and seen in real practice
-

look can bring inherent problems to the surface. The concept of equipoise is rather simple.⁴ If a scientific experiment is valid in genuinely testing a hypothesis and if one conducts an adequate number of experiments, about one-half of these experiments will verify and the other half will falsify the hypothesis being tested. This being the case, we all know that almost all of the abstracts related to RCTs in medical meetings report only positive⁴ results regarding the efficacy of any drug tested as a new medication. The same is true for the vast majority of original papers reporting RCTs in medical journals.

One obvious reason for this state of affairs is the so-called publication bias,⁴ where investigators like to submit and editors like to publish positive findings. Another reason, not much appreciated, is that when a new drug hits Phase III testing, the phase of the properly powered RCT, after Phases I and II—we have already learned a considerable amount about the value of the medication at-hand through animal and some human experimentation. In other words, we are pretty certain that that the new drug will also go through Phase III testing with flying colors if it did well during the initial two. Moreover, almost always, the prerequisite of a Phase III RCT is successful performance in the previous phases. This truism has even led some dedicated and critical students of quantitative, evidence-based medicine to propose that we, perhaps, do not need RCTs after all.⁴

One cannot but admit that there is solid reason in this nihilistic view, as I briefly attempted to explain, above. I propose, however, that doing away with the RCT can only be justified in a situation where the new drug is being tested against placebo or a decidedly inferior competitor. On the other hand, an at least as equally important issue regarding any new drug is whether that drug is superior in efficacy when compared with the current best available at-hand. For this purpose, I strongly maintain that the RCT is the best tool we have available, at least for the time being.

As with every measurement tool, the RCT also has some weaknesses (Table 1). Of the points listed in Table 1, I propose the real Achilles heel of the RCT is the last bulleted item: differences between trial subjects and patients presenting in actual practices. All too often, the patients enrolled in clinical trials have disease severity higher than that seen in daily practice, as was noted in TNF alpha antagonists for the treatment of rheumatoid arthritis.^{5,6} The demography of the study group is also usually different, with under representation of especially the elderly. Finally, the impoverished, the uncooperative, the drug abuser, the homeless, in short, the

Table 2 Problems with Extension Studies

-
- Equipoise
 - Informed consent
 - Power calculations
 - Intent to treat
 - Delaying useful treatment
 - Changing of authors
 - No new information
-

misfit is seldom included in the RCT, thus potentially limiting the external validity (generalizability) of the trial results. It is important not to forget that the four main headings listed as weaknesses in Table 1 hold equally true for both efficacy and toxicity (harm) outcomes from an RCT.

It is interesting to note that recent awareness⁷ of these weaknesses has led some investigators to give more and more credence to open, uncontrolled, or “real life” data to judge efficacy.⁸ Parallel to this is, again, a trend to give more credence to the RCT as a tool to look for adverse events. For example, in 2004, after the worldwide withdrawal of VIOXX[®] (rofecoxib), Peter S. Kim, Ph.D., President of Merck Research Laboratories, promptly announced that “Merck has always believed that prospective, randomized, controlled clinical trials are the best way to evaluate the safety of medicines.”⁹ Kim was, of course, right in the sense that a common condition such as myocardial infarction was found to be higher among the among the group receiving the test drug and this led to halting the APPROVE [Adenomatous Polyp Prevention on VIOXX] trial and, eventually, to the withdrawal of rofecoxib from the market. What he did not emphasize was the fact that this was a serendipitous finding. Surely, that particular trial was not designed to look for this adverse event. In fact, this contention can be generalized to *any* RCT as a scientific tool, based on deductive, hypothesis falsifying methodology. Thus, in the case of rofecoxib, Popperian science would have required enrollment of people already more prone to the event of myocardial infarction (MI) for allocation to either placebo or the rofecoxib arm for the development of MI, in that the hypothesis to be tested – to be falsified – is that use of rofecoxib is not associated with more MI events than placebo. Or, similarly, for a new coxib thought to be less harmful on the gastrointestinal tract than the best other remedy on the market, the scientific methodology would require enrollment of people who already have peptic ulcer disease. It is obvious that such an undertaking would truly be unethical. It is highly improbable, more likely not possible, that a subject would sign a consent form that stated an (let alone *the sole*) objective of the exercise was to search for harm.

This is not to say that we should not turn our eyes away from adverse events during formal hypothesis testing for efficacy. This experience may come both from the first and the second phases of formal drug development, or at times, from experience with the same drug in other diseases. There

will also be a good chance that we will, indeed, learn more about adverse events of the drug in question after a well conducted RCT. However, we should expect that any such information gained will be, by definition, *serendipitous*.

In other words, I strongly suggest that the RCT is a tool that has been developed to test efficacy, and the use of this tool to search for harm does not have a sound scientific basis. To look for adverse events is a never ending process during the life history of a drug. It is, of course, sound reasoning to formalize this process in the post-marketing period, using schemes like MedWatch (FDA safety information and adverse event reporting program),¹⁰ among others. In short, we must maintain the realization that we can never perform a true scientific experiment to search for harm.

The so-called extension study is another activity that I, among others,¹¹ consider detrimental to the very essence of the RCT. The usual scenario is the continuation of a RCT, during which the efficacy of the tested drug has already been shown. The general rationale is to achieve a better understanding about both the efficacy and the harm of the drug. In some such extension studies, blindness is also claimed to be maintained. Table 2 lists some of the “sins” of such extension studies. As listed in Table 2, apart from the lack of equipoise discussed above, such studies usually lack the basic requirements of a RCT, such as the intention to treat principle, in that only those that responded favorably in the first place continue with the extension. There are ethical problems like withholding an efficacious treatment from the control arm. What is more, such studies usually do not provide additional information. Finally, the greatest transgression, I propose, of an extension study is its being “over-proving” or “over killing,” rather than a falsifying exercise, which I described as the backbone of good science in establishing proof.

I once finished a letter to an editor about a particular RCT¹² by stating, “Sometimes, however, it is not the con-

trolled clinical trial, but its interpretation that needs to be put on trial.” I now believe that I should have included the word “implementation” side-by-side with “interpretation” in this statement.

References

1. Kramer BS, Wilentz J, Alexander D. Getting it right: Being smarter about clinical trials. *PLoS Medicine* 2006;3:756-61.
2. Coulter HL. *The Controlled Clinical Trial: An Analysis*. Washington, DC: Center for Empirical Medicine, 1991.
3. Magee B. Scientific method the traditional view and Popper's view. *In: Popper*. Fontana: London, 1982, pp. 18-34.
4. Fries JF, Krishnan E. Equipoise, design bias, and randomized controlled trials: The elusive ethics of new drug development. *Arthritis Res Ther*. 2004;6:250-5.
5. Sokka T, Pincus T. Most patients receiving routine care for rheumatoid arthritis in 2001 did not meet inclusion criteria for most recent clinical trials or American College of Rheumatology criteria for remission. *J Rheumatol*. 2003;30:1138-46.
6. Gogus F, Yazici Y, Yazici H. Inclusion criteria as widely used for rheumatoid arthritis clinical trials: Patient eligibility in a Turkish cohort. *Clin Exp Rheumatol*. 2005;23:681-4.
7. Cleophas TJ, Zwinderman AH. Limitations of randomized clinical trials, proposed alternative designs. *Clin Chem Lab Med*. 2000;38:1217-23.
8. Pincus T. Guidelines for monitoring of methotrexate: “Evidence-based medicine” outside of clinical trials. *Arthritis Rheum*. 2003;48:2706-9.
9. Anonymous. Merck Announces Voluntary Worldwide Withdrawal of VIOXX®. http://www.vioxx.com/rofecoxib/vioxx/consumer/voluntary_withdrawal_statement.jsp. Accessed February 6, 2007.
10. Kessler DA. MEDWatch: The new FDA medical products reporting program. *Am J Hosp Pharm*. 1993;50:1151-2.
11. Taylor WJ, Weatherhall M. What are open-label extension studies for? *J Rheumatol*. 2006;33:642-3.
12. Yazici H. Is the STAR trial really a safety trial? *J Rheumatol*. 2004;31:2528.