Ethical Considerations and Study Design: Discussion of Dr. Heifetz' Presentation

J. E. ALMAN

VA Outpatient Clinic, Boston, Massachusetts 02108

J Dent Res 63(Spec Iss):723-724, May, 1984

It is clear from Dr. Heifetz' presentation that when we invoke ethical considerations, we are looking at the ethics of using human subjects and the steps that must be taken to protect their welfare. In addition, we are considering the ethics implied by our responsibility to the research community – to the dental profession, the dental schools, government agencies, the manufacturers of oral products, to the FDA, and to any other interested parties that I have not mentioned. It is the research community that ultimately judges the credibility of a clinical trial outcome, and we have the ethical responsibility of providing to these judges a true and complete report on the trial.

About a year ago, a group from the Harvard School of Medicine reported in the *New England Journal of Medicine* on a study of the adequacy of reporting the results of medical clinical trials. Since the topics they considered are applicable in most part to dental clinical trials, I make my comments in terms of their list of topics and the way I feel they relate to ethical considerations in dental trials. The list that follows, condensed somewhat from the original list by DerSimonian *et al.* (1982) for purposes of these remarks, provides items that should be reported upon and discussed, to the extent necessary, in the reporting of the clinical trial. Note that, while the items essentially represent our responsibilities to the research community, embedded in them in several ways are our ethical responsibilities for the welfare of the human subjects:

(1) recruitment of subjects and eligibility requirements for admission to trial;
(2) allocation of subjects to treatment;
(3) blindness to treatment by subjects and outcome evaluators;
(4) treatment complications and side-effects;
(5) extent and apparent reasons for loss of subjects to follow-up;
(6) statistical methodology and computer software used in the data analyses; and
(7) power considerations, determination of sample sizes, and levels of differences targeted for detection.

Careful consideration of the reporting on these items can only lead to careful consideration of the same items in the planning stages. Dr. Heifetz has applied this thought to the three topics he discussed in detail. I will go through the list of items with comments on areas he did not cover specifically.

(1) *Admission of subjects to a trial.* – The ethical considerations of informed consent, and informed consent of parents when children are the subjects, have been discussed many times elsewhere. But one aspect not often mentioned is the consequence of recruiting subjects in blocks, *e.g.*, classrooms in schools. This can mean that all children in a classroom are to be mobilized for a dental examination or a session of supervised brushing. There are administrative advantages for the teacher in not having to hold back certain children. Thus, there is a tendency to assume that all children for whom parental consent is obtained are suitable subjects. But some may be inappropriate, as, for example, children under orthodontic treatment, those with an appreciable number of banded teeth. Exclusion of such children could fall under the rubric of self-selection considered by Dr. Heifetz, and thus be a potential source of bias. Because there seems to be some evidence that the proportion of children under orthodontic care is on the increase, attention to this aspect of planning a clinical trial assumes increased importance. In particular, if possible, identification of subjects under orthodontic treatment should be made *before* allocation to treatment. The same suggestion applies, of course, to any other condition that might lead to subject ineligibility. Be it noted that once parental consent has been given, there seems to be an ethical responsibility to include the subject in all activities of the trial, even though the subject's data may not be used in the analyses of the trial results.

(2) *Random allocation to treatment.* – This item is sometimes dismissed in a trial report by merely noting that "subjects were randomly assigned to treatment". Ethical reporting requires that the entire procedure for random assignment be described in detail, so that others may judge if the randomization protocol appears to be satisfactory. Sources of random number sets used should be given, and when these are computer-generated, even a note on the random number generator that is used may be relevant. In addition, tests of the data to evaluate differences among treatment groups with respect to baseline parameters should be made and reported upon in detail.

(3) *Blindness to treatment.* – We are all aware that "double-blindness" is an important *desideratum* of a clinical trial, but it is not automatic. A report on a trial must detail the ways in which blindness of both patient and observer are assured, as well as aspects of the trial and the protocol that might compromise blindness. With children in a dental clinical trial, it may not be easy to maintain the blindness to treatment. For example, when two dentifrices under study are two different colors, or a gel is compared with a non-gel, children will have these kinds of information, will share them with others, and can inadvertently "blab" such to those who should not know. Thus, ethical considerations require that such possibilities be reported in detail, as well as the ways in which the examining dentist takes precautions against the subject "blowing the cover".

(4) *Treatment complications.* – These possibilities, often of great importance in medical trials, are less likely to become problems in dental trials. Nevertheless, ethical considerations, both to the subjects and to the research community, require careful attention to the potential for complications and side-effects, together with detailed reporting on anything of more than trivial importance.

(5) *Loss to follow-up.* – Most clinical trials experience attrition over the period of the trial, for a variety of reasons and to a variety of extents. Again, the ethics of reporting to the research community require careful recording of all reasons for drop-out, insofar as they can be determined, with a summary reporting of such in the trial report. Dr. Heifetz has pointed out that compliance failure as a reason for drop-out is a potential source of bias, and there may be other reasons as well. Any such should be reported carefully.

(6) *Statistical analyses.* — It is incumbent upon the reporting investigator to report on any statistical methodology used that is beyond the usual means, standard deviations, simple analysis of variance, etc. With so much analytical work now being done by computer, using statistical packages, the name of the package used and the procedures within the package should be identified, and justified if necessary. While an unlikely source of trouble, it is still conceivable that such information may be relevant to judging the credibility of the trial results.

(7) *Power considerations.* — This is a most important item, but I do not feel that I need try to add to the excellent discussion of this topic by Dr. Heifetz.

In summary, ethical considerations require that all relevant information regarding the conduct and outcome of a clinical trial be reported. The research community ultimately judges the credibility of the results of a trial, and we all have strong ethical responsibilities to make sure that everything relevant to the making of this judgment is available in black and white. To this end, the items covered above constitute a useful check-list for both the planning and reporting stages of a clinical trial.

REFERENCE

DerSIMONIAN, R.; CHARETTE, L.J.; McPEEK, B.; and MOSTELLER, F.: Reporting on Methods in Clinical Trials, *N Engl J Med* 306: 1332-1337, 1982.

General Discussion

PETERSON: I would like to comment on Dr. Stamm's allusion to Lou Ripa finding that 80% of the DMFS was in the F component. Dr. Bagramian mentioned that we are finding dentist interference with our diagnosis. Lou Ripa finds 80%. In my last study, I found about 60% of the DMFS was in the filled component. A number of years ago, Dr. Art Braddock expressed this feeling, that quadrant dentistry, lower caries rates, and hungry dentists would ultimately make it difficult, if not impossible, to obtain valid results from these studies. I have maintained for years that the reason we have obtained lesser results from preventive programs on pit and fissure surfaces than on distal surfaces is that significant numbers of pits and fissures are and always have been filled prophylactically by dentists. My point is that dentists are now filling a higher proportion of prophylactic pit and fissure fillings than in the past, and that because they are hungry, because there are more dentists, because the caries rate has gone down, this is going to affect our results more and more.

RIPA: I would also like to comment that the tone of this Conference is certainly different from the tone of the Conference in 1968 where we were discussing more technical questions. For want of a better term, perhaps we are discussing more philosophical questions now.

Early in his paper, John Stamm said, "Randomized clinical trials are the gold standard in population-based clinical research". Later he said, "Caries clinical trials represent a full application of the true scientific experiment". These are rather strong statements for John. I was beginning to wonder about them until still later I think he provides some balance and moderation when he discusses the pragmatism associated with clinical trials. The classically controlled clinical trials that we run usually do not test the test agent the way that it is ultimately meant to be used. An example is fluoride dentifrices. We will study a population for two or three years, possibly under supervision, with the subjects specifically selected on the basis of age and the number of permanent teeth that they have available. In real life, fluoride-containing dentifrices are to be used in the primary dentition. They are generally not supervised, so that daily brushing frequency varies, and they are theoretically meant to be used for the lifetime of the individual, so these real-life conditions are dramatically different from the test situations that we seek to set up. In general, the controlled clinical trial may underestimate the value of the product or the agent under study.

GLASS: For one who attended the last Conference, it is only natural to think of what's new and different. What progress have we made? Certainly we have new and faster computers, and we are able to improve substantially on our record systems. We have gone to optical scan systems from marking cards, etc. But I was a little disappointed this morning in that I heard nothing new about the measurement of dental caries. In the final analysis, that's what it is all about. It is the "shoe leather" of the field of epidemiology that counts. Caries prevalence has decreased (in the western industrialized countries), and as it has decreased, the co-efficient of variation has increased substantially. All of these factors combine to make it more difficult to obtain a statistically significant treatment effect if one exists. It seems to me that we in the United States have been extremely provincial in not following through with some of the alleged improvements in diagnostic criteria that our Scandinavian compatriots have developed — mainly white spots. I have been lucky enough to spend two or three winters in these cold areas and have learned their application through Professor Scheinin. I have used them in two studies. I find that this is a method of increasing the sensitivity of diagnosis, and I think we should give more consideration to it.

Some people seem to confuse descriptive epidemiology with analytic epidemiology, and if one is considering secular changes, one in fact must look back at the prevalence of diseases and compare them with those in the future. In a clinical trial or in intervention studies, the use of historical controls is quite different. An intervention study or clinical trial is best defined as a cohort study in which groups differ only with respect to that variable under study. Time is an extremely important epidemiological variable, and if the controls are not concurrent, one does not have a good clinical trial. Today, such a trial is being described as a demonstration study.

HOROWITZ: I think it's a very good idea to understand the caries process better, to see what happens to white spot lesions and whether they can be remineralized. But in our efforts to gain greater sensitivity, and because of our fear of not being able to run studies because of low incre-

ments of true cavitation, and without going to huge numbers of subjects, I don't think we should be measuring things that do not have clinical importance. Because we know that the white spot lesion can be remineralized or reversed, we must ask ourselves, does it truly represent decay or disease? To count them (the lesions) as true disease initially in a study may be a misrepresentation in the real sense. I am concerned about changing criteria to the point that they may become meaningless in the real world.

GLASS: We must keep in mind, though, that the present DMF index is essentially an outgrowth of the one developed by Klein and Palmer. This was a major step forward, but I think we are terribly provincial in not adopting, working with, or evaluating a scheme that was developed elsewhere. These white spots do become carious. I worked in at least two studies where they were carious. I have done some work where white spots which appear hard can be detected radiographically as well. These need not confound the study, because they can be categorized separately in this day of modern recordkeeping and high-speed computers.

MARTHALER: I was thinking of a particular question: Are we doing studies for populations of children which at age 12, 13, or 14 have three or fewer DMFT? What is the significance of our studies for the population at large when the caries rate is so low?

RIPA: Certainly caries activity in the US, with which I am familiar, is going down, but there are pockets of children in the United States, at least, with high caries activity. On Long Island, New York, where we conduct all of our experiments in a radius of only 20 miles, there are wide swings in the caries prevalence rates in the different groups that I examine.

GLASS: We found a similar thing, too. Epidemiologists will consider those variables of time, people, and place. For example, in a very heavily fluoridated area in New England, we found tooth decay which was absolutely reminiscent of what we have observed 25 or 30 years ago. We asked a few questions and found that they were immigrants from Portugal and had been brought into this specific area because one of the state senators is of that extraction and has made it a political issue to help his compatriots enter the country.

BOHANNAN: There have been two references thus far to demonstration programs. I have to take exception to Dr. Glass' comment that the demonstration programs are not research. I think we have to consider carefully the design of demonstration programs. Dr. Bagramian this morning, I believe, referred to the sealant demonstration program in New Mexico. I am not aware of the fact that the people in New Mexico even considered that as a demonstration program. The definition of the demonstration program may not be as well-defined as that of a clinical trial, but I don't think that the blanket statement that demonstration programs are not research can go unchallenged in a group such as this. We may, however, of necessity, require a greater emphasis on the scientific design of demonstration programs.

CASH: It might be good to take a long-range view of the interpretation of clinical significance. Research has slowed down considerably in terms of its ability to discover more effective agents. We are sort of moving slowly in steps. I think that it might be impractical to consider making huge clinically significant jumps in a given study. Over the long range, a slowly building effectiveness could add up to a clinically significant result.

YANKELL: I'd like to address myself to clinical significance *vs.* statistical significance. We also talked this morning about the decrease in the level of caries incidence. I would like to pose this question upon the question that was just asked: I wonder what the odds are today that, admitting that there is a decrease in the caries level, a new caries-prevention substance could be found which would further decrease the caries incidence, and what the chances are of convincing public or private sources of supporting such efforts, which may result in clinically significant effects.

ZIMMERMAN: How many of you think that the odds are still good to decrease the incidence for clinical significance further? Let's have a show of hands.

HEFFERREN: It looks like a majority.

LU: I would like to answer this question with a trace of optimism. Twenty years ago people asked us, can you go to the moon? They say you are crazy. Today it is commonplace. Right now we seem to have hit a plateau with anticaries agents. There will be new ones. Better ones are yet to come.

Just what is clinical significance? Clinical significance and statistical significance are not necessarily mutually exclusive. I would like to propose that statistical significance is the minimum requirement for clinical significance. If it is statistical and not clinical, there is no way it can be clinically significant. Clinically significant from whose point of view? Yours, mine, the dentist, the public health official, or some politician?

ROSS: Dr. Lu pretty well expressed the thoughts I wanted to present, but perhaps they should be expanded. We must decide whether caries has declined enough so that we are willing to let it remain at this state, or do we wish to have an even further decline. If we wish an even further decline, then anything which results in a greater decline is clinically significant. We may have to alter our thinking to consider how much we can expect a new agent to be able to demonstrate.

SCHEININ: I have two questions of Dr. Stamm: The first one relates to randomized and non-randomized clinical trials: Let's assume that there is a randomized study on the effect of sucrose restriction; then one group can have sugar, with sugar restriction in the other group. The outcome will presumably be that there will be some improvement in caries incidence due to the restricted intake. If there would be a third non-randomized group containing all the subjects willing to reduce sucrose intake, there would be a tremendous effect showing the maximum effects one could achieve. In the second randomized group, this could be considered as a public health measure actually within the present standards of what the dental profession can do. Still, the value would be that if attitudes were changed, then a lot could be achieved. I refer to a study published in 1981 by McDonald and co-workers in the *British Dental Journal* about the effect of such measurements.

STAMM: I have two comments: One, whatever you measure for the group that acts as volunteers could be subject to criticism because of selection bias. People who have a desire to do something for dental health, probably not by sugar intake restriction but by some other means, select themselves into a control group. I would guess that they might have lower caries increments than you would find in the sugar restriction group that was used in the random application. The second comment I would offer is that, by and large, in terms of efficiency, you always have to balance the benefit of multiple control groups against using all the subjects allocated to just two groups. With the two groups, you are going to get a more efficient test than you would get with the multiple comparison

groups. Out of necessity, three groups will each have smaller sample sizes than would two groups.

SCHEININ: Thank you. That would mean that one could benefit from something that Dr. Horowitz called compliance bias. The other question was about blind analysis. When actually looking for caries, there could be several factors influencing the "blindness" – for example, systemic fluorides could produce milk fluorosis or there could be staining, as in using stannous fluoride. The type of plaque found in heavy sugar users and even the race of the subjects can also influence the "blindness" of the trial.

BURCHELL: I just wanted to make a point on the judgments which are necessary in determining clinically important differences. I think that judgment is necessary with clinical significance. The differences we see in the short-term exploratory clinical trial are underestimated compared with what we are going to get in the community or pragmatic trial. Therefore, we must take this into account in making a judgment of whether that difference is clinically significant. The other factor we need to take into account is the fact that, in using means for the clinical difference, we aren't getting the full picture. We really need to express our clinical benefit in terms of the frequency response against factors such as caries susceptibility.

HOLLOWAY: Since good clinicians now no longer take radiographs routinely, but do so only when there is clinical indication for each individual child, is it still ethical for us to take routine radiographs in our clinical trials?

RIPA: A determination when to use X-rays is to ask the question: What are they going to show you? Certainly, with proximal caries prevalence low, and with a high F-to-DMFS ratio in a population, there is little need to take radiographs. However, with advances in the use of tooth-colored filling materials in both the anterior and posterior teeth, we may have to rely on X-rays in order to identify fillings.

GLASS: The big problem with radiographs is that many general practitioners are still taking them at regular six-month intervals for all patients, regardless. As a concerned dentist, I would first of all focus my attention on that. Let's think in terms of a clinical trial that's going to last for three years in an area where there is a moderate amount of tooth decay, and we do expect to have the radiographs contribute something to the overall increment. I am specifically concerned with increments on proximal surfaces. I would, in certain circumstances, consider taking radiographs at the beginning of the study and making certain they are available for treatment for the youngsters involved, and at the end of the study, three years later, repeating the radiographs and requesting the dentists in the interim not to take radiographs except for emergencies. In these circumstances – *i.e.*, good communication with the parents, with the public health authorities, and with the dentist (which we are supposed to maintain in well-planned studies anyway) – I think there is no question about ethics in using dental radiographs.