Editorials

Evidence-based therapeutic strategies. There is the need to bridge the gap between simplified megatrials and individual prescriptions

Luigi Tavazzi

Cardiology Department, IRCCS Policlinico San Matteo, Pavia, Italy

(Ital Heart J 2001; 2 (2): 73-77)

Received November 28, 2000; accepted December 14, 2000.

Address:

Prof. Luigi Tavazzi

Dipartimento di Cardiologia IRCCS Policlinico San Matteo Piazzale Golgi, 2 27100 Pavia E-mail: l.tavazzi@smatteo.pv.it

After 20 years of clinical research based on randomized, controlled trials which revolutionized research methodology, changed the attitudes of regulatory bodies, and directed the behavior of clinicians, a debate has arisen within the international medico-scientific community centered predominantly on the diffusion of guidelines and their application. The subject of debate is how to proceed with clinical research.

In the 1980s, the ISIS-1[1] and GISSI 1[2] trials opened the era of the megatrials which Robert Califf[3] in an editorial published in *Circulation* in December 1998, commenting on the paper concerning 10 years of follow-up of the GISSI 1 trial, described in this way: "the broad scale clinical research collaboration initiated by the ISIS and GISSI groups has changed the fate of cardiovascular medicine. Before the demonstration by these two collaborative groups that large numbers of patients could be entered into randomized trials of acutely ill patients, assessment of the impact of therapies on true clinical outcomes was considered to be impossible. With very little investigator payment and strict adherence to the principle of focus on simplicity, these groups opened the field to the dramatic benefit of evidence-based medicine", and also "Before the modern era we were in an age of therapeutic inefficacy ... Demonstration of benefit was straightforward compared with placebo or conventional care. The GISSI and ISIS studies represent the best of this era", and finally "the spirit of the GISSI organization must be replicated to provide a mechanism that can deliver the evidence for the practice of evidence-based medicine". In fact, the era of trials was born from the conviction, held by rather few, that therapies were too empirical and vague and that it was essential and possible to reach certitudes concerning the efficacy and safety of therapeutic regimens. The trials have been the practical instrument of this new approach to therapy and, at the same time, a mechanism for amplifying the concept of evidence-based medicine, which has become the soul and routine procedure of research in medical practice in the last years of the past century. The conceptual milestone of this process was the article by Hampton[4] "The end of clinical freedom" published in the *British Medical Journal* in 1983. The practical consequences are the guidelines which have multiplied over recent years.

Is this the direction in which to continue? Again, Robert Califf[5] says yes, and indeed, in another recent editorial he suggests simplifying and extending the process yet further. It is important to realize that as far as new treatments are concerned, development strategies are directed not so much by researchers as by regulatory bodies which define the quality and quantity of proof required for a new drug to be approved for therapeutic use. Nowadays, regulatory bodies and the scientific community believe that nothing other than randomized trials can provide sufficient weight to make results above proof. In reality, megatrials are not necessary to reach certain results. The proof of the efficacy of vaccines or antibiotics, of

vitamin B12 and iron in anemias, of cardiac stimulators in high-grade atrioventricular blocks, and of diuretics in acute heart failure did not require sophisticated trials for these treatments to become swiftly and justly accepted into clinical practice.

In general, a megatrial is necessary in the following situations[6]:

1. when the benefit of the treatment is small (and difficult to distinguish from the spontaneous natural variability of the disease evolution);
2. when the patients treated are heterogeneous and only a small, unidentified group of them draw benefit. Attilio Maseri[7] likes to give anemia as an example of this type of situation. If we did not know about vitamin B12 deficiency and were not, therefore, able to identify patients with this type of disorder, but nevertheless carried out a large trial testing vitamin B12 supplementation in all anemic patients (and if by chance we included a high enough number of vitamin B12 deficient subjects) we would discover that, on average, vitamin B12 gives a small benefit and it would therefore become a recommended treatment for all anemic patients. In all likelihood, this is exactly the behavior that we use day in day out for various classes of drugs. Incidentally, the irrationality of the behavior is obvious but, in the absence of specific knowledge, the therapeutic approach remains sensible; if the treatment was not recommended the few or many potential beneficiaries would not, in fact, benefit at all;
3. when it is difficult to measure the benefits because the symptoms affected by the treatment are mild or erratic or because the long-term incidence of events is very low.

In the situations described above there is no adequate alternative to large trials.

Nevertheless, if we estimate that large numbers are needed, it is because we expect the benefit to be modest. The large numbers serve, in this case, to give statistical certitudes. But, what if the small benefit is clinically significant? Individually, only the patient and his or her doctor can establish this. Generally speaking, a small benefit becomes relevant if it concerns a very common disease, a disease which carries a high risk, or high cost, or when there are no side effects to the treatment such that the risk/benefit ratio remains favorable even if the benefit is small (as long as the cost/benefit ratio is also acceptable).

However, as the years have passed and the experience of large trials has become consolidated, the strategy has begun to reveal its limitations. I shall mention some:

1. the large, sometimes enormous number of participants necessitates multinational enrollment with considerable logistic and organizational problems;
2. the costs are so high and the risks so great (it should not be forgotten that most trials give neutral or negative results) that only few drugs can be tested in this way, the few that have the most commercial potential and that only large companies can bear the costs (and this underlies the many company mergers underway);
3. for ethical reasons already recommended drugs cannot be withheld from patients – this means that new drugs can only be tested "on the top" of the recommended treatment, and thus each new treatment tends to add to, but does not replace, the preceding ones. It is clear that this process cannot continue indefinitely;
4. in the context of a class of drugs there is the problem of how to manage the analogs (with some problems also in defining the class). Must each new ACE-inhibitor or thrombolytic agent follow the same sequence of clinical studies, including a megatrial on mortality and morbidity, as the first ones? For a thrombolytic agent, for example, the additional advantage cannot be greater efficacy but the possibility of being given as a single bolus dose, which would allow therapy to be started before hospital admission, or a lower cost for equivalent efficacy. This leads to the equivalence trials. Is this strategy appropriate or feasible in the long term?

Are there alternatives to the megatrials?

In order to increase the feasibility and decrease the costs of trials while maintaining their clinical significance, some alternatives have been suggested and tried.

One is the use of *combined endpoints*. In general these include several relevant clinical events, for example "death, reinfarction and stroke". Increasing the number of endpoints decreases the population sample size required to reach statistically significant results. There are a series of methodological problems which I shall not consider in detail here, such as the equivalence of clinical significance between events (and few are equivalent to death), competition between the endpoints (who dies early has a lower probability of having a reinfarction) and others[8]. A well-designed trial considerably reduces the possible biases associated with the choice of endpoints and I, therefore, believe that the practice of combined endpoint trials is valid and feasible if the aim and design of the study are appropriate.

The discussion about *surrogate endpoints* is different. The history of clinical research is littered with errors concerning surrogate endpoints. The principle is that of evaluating the efficacy of a treatment through an effect induced on a variable which is considered to express the central pathophysiological mechanism of the evolution of the disease and a substitute prognostic indicator of clinical events. The weakness of the strategy lies in this discretional evaluation, which strongly depends on the correctness of the conceptual model of the disease. In general, surrogate endpoints are excellent indicators for pathophysiological studies and for verifying the activity (but not necessarily the efficacy) of drugs but cannot replace strong clinical endpoints when the aim of a trial is to determine the clinical efficacy or safety of a treatment.

A crucial problem concerning both methodological and clinical aspects is the analysis of *subgroups*. Since

the universe of patients enrolled in a trial is usually so vast and heterogeneous, it is common practice to analyze not only the results in the whole population but also the results in subgroups, distinguished according to age, gender, and parameters characterizing the disease under study. Since 1998 the Food and Drug Administration has required demographic analyses of subgroups divided according to gender, race and age before approving new drugs[9]. Only rarely is the size of the population sample calculated taking into account the subgroup analyses and consequently the power of the study is frequently inadequate for this type of analysis. For example, a recent review of 67 randomized cardiology trials, including more than 1000 patients, showed that more than half of these trials reported analyses of five or more subgroups without having carried out formal statistical tests to verify interactions[10]. There are numerous methodological defects inherent in subgroup analyses[8], but the fundamental weakness is that they often do not correspond to precise pathophysiological requisites and thus cannot be useful to the essential aim which should generate them: identification of those patients for whom the tested drug would be really effective. Here, the vitamin B12 or iron supplementation in anemia example shows its limitations. In this example, at the end of the trial we would be able to recognize responders, although we would not know why those particular patients responded. The problem is much more complicated when the effect is less identifiable, for example because we are evaluating only the reduction in an endpoint such as mortality, but we are unable to distinguish between those who survive because of the treatment and the rest of the subjects who would have survived anyway. Similarly, by treating everyone with lipid-lowering drugs, we were unable to distinguish those who will not have clinical events because they are being treated from those who will not have clinical events because that is the natural evolution of their disease. In this case a well-structured, culturally in-depth study design should be able to provide answers to precise questions, particularly by analysis of subgroups.

Observational and outcome studies

In the "pre-trial" era, observational studies were the rule. Besides epidemiological purposes, they were undertaken with the same aims as the randomized trials: to verify the efficacy and safety of a treatment. In the 1970s systematic comparative reviews of observational outcome studies and randomized trials highlighted the poor reliability of the results of the former. In particular, it was noted that observational studies tended to emphasize the positive results of treatments. Chalmers et al.[11], in a review of studies of that period, noted that 56% of the non-randomized studies found favorable results compared to 30% of randomized studies on the same treatments. The same group, in another review, reported that the percentages were even 79% and 20% respectively[12]. There are numerous examples, some very recent, of observational studies which generated hypotheses of treatment efficacy not confirmed by randomized studies[13-15]. Nevertheless, two recent systematic comparative analyses of observational and randomized studies have re-opened the debate on the weakness of observational studies in evaluating treatments[16,17]. Most of the studies carried out in the 1980s and 1990s led to conclusions at least qualitatively similar to those obtained in the randomized trials. One important reason is probably methodology. The design of observational studies has gradually been honed, and some errors of setup are no longer made; the effect on the quality of the results has been positive.

Nevertheless, observational studies should not be considered as alternative to randomized trials, but rather as complementary. Firstly, they can represent a way of trying to understand the pathophysiological context before testing drugs in poorly understood diseases. They can help to explore or generate pathogenetic hypotheses which could form the rationale for therapeutic studies. Secondly, once an effective and safe drug has been identified, the problem of transferring its use into clinical practice remains. The patients enrolled in trials are "always" selected, and the population treated, when a therapy is approved for prescription, is "always" different from that originally studied. Then, there is the problem of whether the safety and efficacy of treatment recorded in a trial are maintained when the therapeutic strategy is applied in the "real world". Furthermore, the practical interpretation of what has been demonstrated in a trial, in terms of indications, contraindications, doses and the individual doctor's perception concerning the efficacy and risks of a given therapy have a strong effect on its real use. For example, the dose of beta-blockers used in the treatment of heart failure must be slowly up-titrated. Once the treatment has begun, and this usually occurs in hospital, the up-titration must be continued at home under the care of the family doctor until the recommended dose or the maximum dose tolerated by the patient is reached. In fact, as shown by the observational (not yet published) BRING-UP study, the up-titration is frequently not completed and the doses maintained at inadequate levels or at least levels other than those tested in the randomized trials. All these factors make it necessary to have studies that monitor the process of implementing therapies, providing information useful for guiding this implementation and generating hypotheses for correcting it. Indeed, this is the central point. It is not so much having a neutral "photograph" of what is happening in a given clinical situation that matters as much as being able to use the observational information to guide the implementation in such a way as to maximize the benefits and minimize the risks, thus creating the premises for eliminating inappropriate overuse or underuse of a given therapeutic opportunity. Outcome studies, constituting a new phase 4 research model, to

be carried out with rigorous methodology, should be viewed as an attempt to respond to the central question implicit in the introduction of any new class of drugs: are the epidemiology and natural history of a given clinical problem really modified, in terms of both public health and economics, by the introduction of the new treatment? It is in this therapeutic perspective, besides the interpretative setting mentioned above, that observational studies acquire the same scientific and social dignity awarded to randomized trials of which they are the natural, complementary evolution[18].

And the future?

I believe there are two key issues in this perspective.

The first concerns the multiplication of therapies. Will it be endless? It is clear that it is not possible, even simply from an economic point of view, to administer preventive therapies indiscriminately to a multitude of unsusceptible subjects. Nor can therapies, particularly if aggressive or expensive, be added to subjects who will gain no incremental benefit from them. As Maseri[6,7] has been saying for some time, the central aim of future clinical research will be to identify the "responders" to specific treatments and to measure the incremental benefits of associated therapies in them. This need becomes particularly clear when proposing invasive, expensive treatments. An example for all: ventricular electrical stimulation in heart failure[19]. In order to get unambiguous results, trials must be designed appropriately and the opportunities offered by biotechnology, e.g. pharmacogenetics, exploited to the full. There will be two important limitations. One is the degree of understanding of the subject under study. We look for what we know and if we do not ask intelligent questions we will probably get stupid answers. Another limitation is the availability of resources. It will be very difficult to find economic support from companies if the aim is to circumscribe the number of potential users of a drug. It is essential that institutional bodies which distribute funds support not only basic research but also make funds available, as the National Institute of Health is doing, for orphan drugs or trials with aims that are not considered financially rewarding by drug companies, and above all, in accordance with a recent European Community directive (May 2000), that more funds be allocated to research.

The second issue concerns a critical reappraisal of the development pathway of a drug. I cannot discuss this in detail here. It is worrying that in the last 10 years, of the more than ten classes of drugs tested in heart failure, two thirds have failed to pass the hurdle of the morbidity-mortality trial because of evidence of an excess of events in the treated population compared to the placebo group. This highlights the inadequacy of both our knowledge and of the methodological approach preceding mortality trials. The cost of developing a drug has increased by one order of magnitude over the last 30 years, from about 50 million dollars in 1970 to the present day 500 million dollars. Since the bottle-neck in developing a drug is the mortality-morbidity trial, companies tend to minimize the studies preceding it in order to limit costs and economic risk, relying directly on the large trial. This custom, besides exposing patients to avoidable risks, reduces pathophysiological research to such a degree as to prevent understanding of the mechanism of the efficacy of the drugs and thus, research and identification of responders. This is the opposite of desirable.

In the light of this, a greater role in directing clinical research must be played by medico-scientific societies, intended as ethically-based professional and social groups which recognize the development of knowledge at the service of patients as an aim.

Collaboration with the commercial world is not only a necessity but also an opportunity, provided that each and everyone carries out his own role with clarity and coherence. If not, the spectre of conflicts of interest, which is raising its head in the public debate on scientific research[20-22], will risk sullying the world of medicine, further undermining the trust on which the relationship between patients and physicians is based and devitalizing the efforts of many good-willed doctors to maintain the efficient and untainted health care that we want.

References

1. ISIS-1 (First International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous atenolol among 16 027 cases of suspected acute myocardial infarction: ISIS-1. Lancet 1986; 2: 57-66.
2. Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). Long-term effects of intravenous thrombolysis in acute myocardial infarction: final report of the GISSI study. Lancet 1987; 2: 871-4.
3. Califf RM. Ten years of benefit from a one-hour intervention. Circulation 1998; 98: 2649-51.
4. Hampton JR. The end of clinical freedom. BMJ 1983; 287: 1237-8.
5. Califf RM. Dual randomization in cardiovascular trials. (editorial) Am Heart J 2000; 140: 1.
6. Maseri A, Cianflone D, Paceri V, Crea F. The risk and cost-effective individual patient management: the challenge of a new generation of clinical trials. Cardiovasc Drugs Ther 1997; 10: 751-8.
7. Maseri A. Evidence-based medicine: progress but not a final solution. Clin Cardiol 1998; 21: 463-4.
8. Boissel JP, Leizorovicz A. Disegno e conduzione di uno studio clinico. La sperimentazione clinica in cardiologia 1999. Roma: Il Pensiero Scientifico Editore, 1999: 3-38.
9. Food and Drug Administration. Final rule on investigational new drug applications and new drug applications (GENERIC). 21 CFR, parts 312 and 314, 1998.
10. Parker AB, Naylor CD. Subgroups, treatment effects, and baseline risks: some lessons from major cardiovascular trials. Am Heart J 2000; 139: 952-61.
11. Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. N Engl J Med 1983; 309: 1358-61.

12. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. Am J Med 1982; 72: 233-40.
13. Hulley S, Grady D, Bush T, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. JAMA 1998; 280: 605-13.
14. GISSI-Prevenzione Investigators. Dietary supplementation with n-3 polyunsaturated fatty acids and vitamin E after myocardial infarction: results of the GISSI-Prevenzione trial. Lancet 1999; 354: 447-55.
15. Omenn GS, Goodman GE, Thornquist MD, et al. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. N Engl J Med 1996; 334: 1150-5.
16. Concato J, Shav N, Howitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000; 342: 1887-92.
17. Benson K, Hartz A. A comparison of observational studies and randomized, controlled trials. N Engl J Med 2000; 342: 1878-86.
18. Maggioni A, Tavazzi L. Introducing new treatments in clinical practice: the Italian approach to β blockers in heart failure. Heart 1999; 81: 453-4.
19. Tavazzi L. Ventricular pacing: a promising new therapeutic strategy in heart failure. For whom? Eur Heart J 2000; 21: 1211-4.
20. Krimsky S. Conflict of interest and cost-effectiveness analysis. JAMA 1999; 282: 1474-5.
21. Safeguarding participants in clinical trials. (editorial) Lancet 2000; 355: 2177.
22. Shalala D. Protecting research subjects - what must be done. N Engl J Med 2000; 343: 808-10.