Clinical trials in cardiovascular medicine: are we looking for statistical significance or clinical relevance?

R Willenheimer, B Dahlöf, A Gordon

During the last two decades several large clinical trials have documented beneficial effects of different pharmacological interventions on morbidity and mortality in various cardiovascular conditions. Based on the results of these trials physicians in general feel convinced that there is a considerable reduction of mortality and morbidity as a result of treatment with: (a) β receptor blockers in patients with acute myocardial infarction as well as in the postinfarction period; (b) acetylsalicylic acid (ASA; aspirin) and thrombolytic drugs in patients with acute myocardial infarction; (c) angiotensin converting enzyme (ACE) inhibitors in patients with left ventricular systolic dysfunction irrespective of symptoms of heart failure and, as recently shown in the HOPE study, in patients with high cardiovascular risk but without heart failure or obvious left ventricular systolic dysfunction; (d) β receptor blockers and spironolactone in addition to an ACE inhibitor in patients with chronic heart failure; (e) diuretics, β receptor blockers, calcium channel blockers, and ACE inhibitors in hypertension; and (f) statins in secondary prevention.

However, trials not showing benefit of the intervention have conveyed important messages also. Thus, based on trials showing lack of benefit or even harm from the intervention, most cardiologists do not use class I (and perhaps class III) antiarrhythmic drugs to treat supraventricular or ventricular arrhythmias, as well as (most) calcium channel blockers and positive inotropic drugs in the treatment of heart failure. Furthermore, the role of digitalis in the treatment of heart failure has been clarified since digitalis has been shown not to affect mortality in these patients. The message about lack of effect or harm is not clear, however. As regards antiarrhythmic drugs some studies show benefit of amiodarone, and benefit from calcium channel blocking agents has been indicated in heart failure patients with preserved left ventricular systolic function. Furthermore, preliminary mortality data on calcium sensitisers, such as levosimendan, suggest that certain positive inotropic drugs may improve prognosis.

Have the cardiovascular clinical trials been well designed and conclusive?

Are the interpretations and conclusions of the mortality trials correct? We believe that they are not entirely correct. As clinicians we tend to extrapolate results from clinical trials to the population we meet and treat in our daily work. The ultimate purpose of clinical trials is, indeed, that we should be able to do just that. This has, however, quite frequently not been the case so far.

THE AGE PROBLEM
Old patients are usually not included in cardiovascular mortality trials, probably because of competing comorbidities. This has obvious relevance when mortality is an end point. For example, most of the trials examining the effects on mortality of ACE inhibitor intervention in left ventricular systolic dysfunction and heart failure have included a patient population considerably younger than the typical heart failure patient. Consequently, although some ACE inhibitor intervention trials in cardiac failure after myocardial infarction have included a population more relevant to the clinical situation in terms of age, we know very little about the effects of ACE inhibitors in the elderly. Since elderly heart failure patients constitute the main target population for ACE inhibitor treatment, this patient category should obviously have been more extensively represented in the ACE inhibitor trials. The same is true for the mortality trials with β receptor blockers, spironolactone, and digitalis in heart failure.

THE SEX PROBLEM
In general, cardiovascular mortality trials have included few women. Yet women constitute half of the target population in clinical practice. Indeed, it is difficult to draw conclusions regarding a treatment effect in women from virtually all of the cardiovascular mortality trials conducted to date. A likely partial reason for the relative paucity of data on females is the interrelationship with the problem of age; the onset of cardiovascular disease is normally later in women. In future trials the design will certainly have to be improved in order to include a more representative sample of women.
women. However, since it is important to include patients consecutively it is possible that many men will have to be unnecessarily included in order to include a relevant number of women in a trial. A possible solution to this problem is to conduct separate trials in women, or to include consecutively only every second or third man, but all the women.

THE PROBLEM OF EXTENSIVE EXCLUSION
Most cardiovascular mortality trials have presented a long list of exclusion criteria. Besides excluding many women and elderly patients, these criteria have led to the exclusion of an additional, substantial part of the patient population with the condition relevant to the intervention. One example is the very common exclusion criterion “other serious diseases which may affect survival during the study period”. Another reason for exclusion is when comorbidities dictate the choice of drug to be used, which often is the case in hypertension trials. In clinical practice we rarely exclude patients from a certain treatment because they meet exclusion criteria applied in the trials providing the knowledge about the very same treatment effect. This is obviously not evidence based medicine since such patient categories have not been investigated and we, thus, cannot know if the results shown for the intervention are applicable to these patients. It seems reasonable to use fewer exclusion criteria in future trials, in order to reproduce the clinical reality.

THE PROBLEM OF SELECTING LOW RISK PATIENTS
The inclusion and exclusion criteria employed in trials have quite frequently caused a low risk study population to be selected. Examples are the CAST trial,14 15 and to some extent the SWORD trial.16 These trials examined the effects of class I14 15 and class III16 antiarrhythmic drugs in survivors of a myocardial infarction. In these trials the mortality among placebo treated patients was lower than intended. The intervention was shown to cause harm and these studies are largely responsible for the negative view on antiarrhythmic treatment among clinicians in general, including use in high risk patients (perhaps with the exception of amiodarone). However, it may not be correct to extrapolate these trial results to the entire population of patients with arrhythmias, since we know little about the effects of these drugs in a high risk population. If these drugs have the potential to cause harm as well as benefit, it will obviously not be possible to demonstrate any beneficial effects without examining a high risk population. The DIAMOND study,17 which examined the effect of the class III antiarrhythmic agent dofetilid, included a study population at considerably higher risk compared to those in the CAST and SWORD trials. Furthermore, the study design aimed to lower the risk of proarrhythmic effects and permitted the exclusion of patients with early proarrhythmic reactions. The study showed a neutral effect on mortality and a significant reduction of heart failure hospitalisations.18 It is likely that the result would have been different if a low risk population, as in CAST and SWORD, had been included. Thus, it is vital not to extrapolate results of clinical trials to patient populations other than those actually studied.

Another reason for a lower than intended mortality in a trial is because power calculations are based on old data, and do not take into account the constantly improving prognosis in many cardiovascular conditions.

THE PROBLEM OF CHANGING CONCOMITANT TREATMENT
The evidence that β blocker treatment has a beneficial effect on mortality and morbidity in acute and postmyocardial infarction patients originates mainly from studies performed in the prethrombolytic era. Thus, we cannot be confident that β blocker treatment causes the same benefit in patients who have received thrombolytic treatment. It is possible that the effect is less pronounced since the thrombolytic treatment has already reduced the risk substantially in these patients. It is, however, also possible that the effect is greater since, as a result of the thrombolytic treatment, some patients who would have died without thrombolysis will survive to become high risk patients.

THE STUDY DURATION PROBLEM
We usually extrapolate a treatment effect documented during a limited time of follow up to a longer period of time, perhaps life long. How long should β blocker treatment postmyocardial infarction be given? We tend to consider this a life long treatment. However, with the exception of a six year open follow up of the Norwegian timolol study,19 no trial showing benefit from β blockade in postinfarction patients has had a study duration longer than three years. Thus, there is little evidence of a life long beneficial effect of this treatment in these patients. Furthermore, the importance of safety is amplified when long term effects are considered.

THE PROBLEM OF TESTING THE HYPOTHESIS ADEQUATELY
It is of course extremely important to design a trial so that the hypotheses can be adequately tested. An example of a mega-trial with a design that did not allow for conclusive testing of a hypothesis is the ISIS-4 trial.20 The interpretation of the study result was that intravenous magnesium given to patients with acute myocardial infarction no beneficial effect. This result was in contrast to a prior, smaller study,21 showing that magnesium significantly reduced mortality in patients with acute myocardial infarction. Since ISIS-4 was a much larger study clinicians ceased (or never started) using magnesium for this indication. It has, however, been argued that magnesium can only be of benefit if given before thrombolysis,21 22 which was not the case in ISIS-4.23 Thus, contrary to the general belief, a beneficial effect of magnesium in the acute myocardial infarction setting was not precluded by ISIS-4. A new study with a proper design will investigate this hypothesis.22
THE PROBLEM OF IMPLEMENTING TRIAL RESULTS INTO CLINICAL PRACTICE

It is well known that it usually takes a long time before the results of clinical trials change clinical practice. In some cases it seems almost impossible to implement the trial results despite seemingly well documented beneficial effects of an intervention. One example is the use of ACE inhibitors in heart failure, which has been reported to be only around 30% in most western societies despite documented beneficial effects on mortality, morbidity, and physical capacity in several trials. Furthermore, if prescribed at all, in clinical practice ACE inhibitors are frequently used in lower doses than in clinical trials. This is alarming since the effects on mortality and morbidity have been proven for “target doses” only. Indeed, it has been shown that the benefit of a lower dose of an ACE inhibitor is less pronounced compared to a higher dose.

A factor strongly contributing to the low prescription of ACE inhibitors to heart failure patients is, according to a British investigation, a fear among physicians of possible side effects, especially hypotension and renal failure. How can this be when virtually all major studies on ACE inhibitors in heart failure have shown these agents to be safe with a low incidence of these side effects? A possible explanation is that the patients physicians usually deal with are different from those included in the trials. The most obvious difference is that patients encountered in clinical practice are usually substantially older than those in the trials. Most often these patients also have serious concomitant diseases, which is not the case for patients in the clinical trials. Consequently, clinicians do not feel convinced that the results of the clinical trials are applicable to their own heart failure patients, and might fear that the side effect profile is substantially more unfavourable compared to what has been shown in the trials. This could of lead to a reluctance to prescribe ACE inhibitors in elderly patients with multiple diseases. In order to avoid problems of this nature in the future we believe that it is of the utmost importance to perform trials involving patients such as those generally found with the disease in question. Perhaps the recently published HOPE study may help to increase the use of ACE inhibitors, since it showed clear benefit of ACE inhibition in patients we commonly meet in our everyday clinical practice.

In this context it is of course also important that the study results are convincing in terms of absolute differences in mortality and morbidity. One may question whether the result of the large mortality trials examining the effects of ACE inhibitors in heart failure really are that impressive, especially in view of the fear of side effects, obviously shared by many physicians. The SOLVD treatment trial, for example, shows that the chance of surviving for the next 3.5 years without an ACE inhibitor is 60.3%, whereas it is 64.8% with an ACE inhibitor. This difference might not be convincing enough for the clinician, especially if he or she fears that the result might be even less pronounced in the elderly patient with recently diagnosed heart failure in whom he/she is about to start treatment. A clear result from an elderly patient cohort would obviously be of much help in this situation.

There are numerous examples of clinical trials in cardiovascular disease which show impressive relative reductions of mortality and morbidity but which, owing to a relatively low risk in the population examined or to a short study duration, do not show much increase in the chance of survival or event free survival in absolute terms. It is of course the latter figure which is important to the physician and patient when making a decision about treatment. Consequently, it is imperative to express the study results in absolute terms and to design the study in order to get the best possible results in absolute terms. The importance of an adequate study duration is discussed further below.

A consistency of results in different trials of the same agents or the same class of agents should help to convince the clinicians and facilitate the implementation of a certain treatment. Thus, it is of importance to verify the results of any clinical trial in other, independent trials.

Should clinical trials be stopped early because of a significantly beneficial effect of the new intervention?

In the best interests of patients in the intervention group it is obvious that studies showing a negative effect of an intervention must be stopped as soon as the negative effect is evident and has been reported to the steering committee by the data safety monitoring committee. Does this necessarily mean that clinical trials should be stopped early because the new intervention is having a significantly beneficial effect? Such early termination should of course always be based on clearly defined prespecified stopping rules. However, one might argue against this practice. The reason for early termination of trials showing benefit is that the patients in the control group must be given the opportunity to receive the new, better treatment a soon as possible. However, the overall purpose of a clinical trial is that, if proven superior, the new treatment should be given as soon as possible to all patients like those in the trial. The concern about the control group may be in conflict with the best interest of the general population with the condition studied. Some examples of such situations are presented below.

THE PROBLEM OF EARLY TERMINATION

An example is where the number of deaths differing between two study groups may be quite small although there is a highly significant beneficial effect of the intervention. In the US carvedilol trials 1094 heart failure patients were randomised to double blind treatment with placebo (n = 398) or carvedilol (n = 696) in addition to standard treatment, including an ACE inhibitor. After a mean follow up of just over six months the study was prematurely terminated because of a highly significant
(p < 0.001) 65% relative mortality reduction in the carvedilol group. The number of deaths was 22 (3.2%) in the carvedilol group versus 31 (7.8%) in the placebo group—a difference of nine deaths. Had the groups been equally large the difference would have been 25 deaths (given a similar mortality rate). The clinician might question whether a result like that should change the therapeutic approach in heart failure. Indeed, many clinicians were not convinced by that result. Consequently, most heart failure patients did not receive this new treatment. Thus, the result of the early termination of the study was that, while the relatively small group of patients in the study control group (perhaps) received the new treatment, the vast majority of the heart failure patients in the world did not. If the new treatment is truly effective this is certainly an example of how the concern of the study patients can be a disadvantage to the general population with the condition studied. Had the study duration been longer the effect would have been more evident, provided that the mortality curves had been as in fig 1A. In that case it is likely that clinicians in general would have been more convinced of a beneficial effect. Consequently, the vast majority of heart failure patients would have been given the benefit of the new treatment, whereas the control patients of the study would have had to wait a little longer to receive the new treatment. It is surely worth discussing which approach is the best from an ethical point of view.

THE PROBLEM OF A RELEVANT FOLLOW UP PERIOD
An early termination of a study may lead to a different mean time of follow up than the one intended. Strictly scientifically the hypothesis may not have been properly examined in such a case. The shorter follow up may also be biologically questionable. How valid are six month mortality results in a heart failure population with an expected mean survival time of around six years? Do we know if the mortality curves would have been of type A, B, or C (fig 1), if the study had gone on for a longer time? It is obviously important to determine a biologically valid period of follow up and design the study accordingly.

Trial acronyms
AIRE: Acute Infarction Ramipril Efficacy
ATLAS: Assessment of Treatment with Lisinopril And Survival
CAST: Cardiac Arrhythmia Suppression Trial
CIBIS: Cardiac Insufficiency Bisoprolol Study
CONSENSUS: Cooperative North Scandinavian Enalapril Survival Study
DIAMOND: Danish Investigations of Arrhythmia and Mortality On Dofetilide
DIG: Digitalis Investigation Group
HOPE: Heart Outcomes Prevention Evaluation
ISIS: International Study of Infarct Survival
LIMIT: Leicester Intravenous Magnesium Intervention Trial
MERIT -HF: Metoprolol CR/XL Randomized Intervention Trial in Heart Failure
RALES: Randomized Aldactone Evaluation Study
SAVED: Survival And Ventricular Enlargement
SOLVD: Studies of Left Ventricular Dysfunction
SWORD: Survival With Oral d-Sotalol
TRACE: Trandolapril Cardiac Evaluation

THE PROBLEM OF THE IMPACT OF “PROFESSIONAL STUDY PATIENTS”
In many studies of chronic conditions a considerable part of the study population comprise “professional study patients”. These are survivors who are able to participate in many clinical trials simply because they survive long enough. They may also be exceptionally compliant to treatment. Consequently, they constitute a selected group of patients who are not necessarily representative of the general population with the chronic condition studied. The inclusion of these patients in clinical trials always makes it questionable to extrapolate results from clinical trials to the general population with the condition. These “professional study patients” are often included in the beginning of the study period, because they are
available, whereas “ordinary” patients usually are included later on. The “professional study patients” will therefore contribute the most to the total time of follow up. Given that the inclusion period is relatively short compared to the total time of follow up, the relative contribution of these patients to the follow up time is greater the shorter the study duration. Consequently, if a trial is stopped early it may be more questionable to extrapolate the results because of the inclusion of “professional study patients” than if the study had gone on for a longer time.

**THE PROBLEM OF THE CONSIDERATION FOR THE CONTROL PATIENTS**

One may object to the argument that a trial showing benefit of a new intervention should be stopped early in the interest of the study control patients. If the benefit of the intervention is mostly caused by an early effect the control patients may not have much benefit when given the new treatment later on. This may be the case in postmyocardial infarction trials. Furthermore, if the study duration is longer the early treatment effect will appear more clearly. This information is important to obtain since it will teach us to start treatment early, which will be of benefit to future patients in clinical practice. This is another reason for not stopping trials earlier than intended because of a beneficial effect of the intervention.

**Conclusions**

Clinical trials constitute the backbone of evidence based medicine. Clinical trials in various cardiovascular conditions have indeed supplied us with extremely valuable information during the last 20 years. Without these trials prognosis in hypertension, myocardial infarction, and heart failure most likely would have been worse than it is today. Yet the design of most trials has not been optimal, and knowing what we know today we probably would have designed most of them differently. The design of clinical trials in cardiovascular medicine has improved during recent years. In order to secure further improvements it is important that we continuously discuss the pitfalls and problems outlined in this article.