ANKEM Derg 5 (No.4): 441-448 (1991)

DESIGN OF CLINICAL TRIALS OF ANTIBIOTICS: CHALLENGES IN EVALUATING NEW ANTIBIOTICS IN THE TREATMENT OF SERIOUS INFECTIONS

S. Ragnar NORRBY

Antibiyotiklerle klinik çalışmaların planlanması: Ciddi infeksiyonların tedavisinde yeni antibiyotiklerin değerlendirilmesinde karşılaşılan güçlükler.

SUMMARY

Clinical trials of antibiotics, especially when serious infections are studied, are often of a less than optimal quality. This is partly due to confounding factors such as concomitant non-infectious diseases and treatments, making the evaluation of the effects of an antibiotic *per se* difficult. Another factor of importance is often lack of knowledge with the investigators in clinical trial design and execution. This overview highlights some of the specific problems in studies of antibiotics and also some general rules for clinical trials of drugs.

INTRODUCTION

Clinical trials of antibiotics are often of a less than optimal quality. Several reasons account for that: (i) infections are often complications of other diseases and antibiotic treatment is therefore, in many cases, adjuvant to other therapeutic measures and the effects of an antibiotic *per se* becomes difficult to evaluate, (ii) efficacy end-points are in most cases dicothomous, for example, clinical cure versus clinical failure or bacteriological eradication versus bacteriological persistence, leading to demands for very large patient samples, and (iii) only rarely can a single centre accumulate enough patients to test a null-hypothesis but trials have to include several centres. While these weaknesses are difficult to eliminate, there are other deficiencies in antibiotic trials which are clearly investigator-related. Thus, Fihn and Stamm (2) demonstrated that in one of the easiest types of infection to study, uncomplicated cystitis in females, the quality of the trials was appallingly low. This could only have been due to lack of understanding and education of the principles of design and execution of clinical trails. To a large extent that is due to a misconception that clinical trials represent a second class science and can be performed on spare time.

This overview discusses the design and execution of clinical trials of antibiotics in the treatment of serious infections.

BACKGROUND INFORMATION

The clinical evaluation of new drugs is traditionally done in four phases. Phase I trials evaluate the pharmacokinetics of an antibiotic. In this phase the initial studies deal with healthy male volunteers. Later studies are performed in patients or subjects with defined organ dysfunction, for example, those with impaired renal function, as well as in children and the elderly. In phase II, clinical trials of efficacy and safety are performed in mild or moderately serious infections to establish that the antibiotics is active and that no unexpected adverse reaction occurs in high frequencies. Phase III trials are large, and preferably controlled, trials establishing efficacy and safety in the treatment of infections which the antibiotic is intended to be used in. Similar trials may be performed after registration and are then referred to as phase IV trials.

In all clinical trials, the investigator should carefully study the natural history of the ty-

Department of Infectious Diseases, University of Lund, Lund University Hospital, S-22185 Lund, Sweeden.

pe of infection to be studied. It is also essential that the investigator is well acquainted with the new drug to be evaluated as well as with the one(s) used as control. This information should be given in the trial protocol. Also in this preparatory phase of the trial, a rationale for the study should be given, that is, the purpose for why the trial to be undertaken should be clearly formulated. As discussed below, it is important to realize that this part of the protocol will and should state the bias of the investigator.

PATIENT SAMPLE

Each clinical trial tests a hypothesis in a small sample of patients. The purpose of that test is to achieve results which can be extrapolated to the population from which the sample was drawn. In other words, a well performed clinical trial should have a high external validity. This can only be achieved if the sample is representative of the population; if that is not the case, the results are only valid for the sample itself and the trial is meaningless.

Sample size

In the early phase of the planning of a clinical trial it is essential to estimate how many patients are required to achieve a meaningful study. Normally, a null-hypothesis is formulated, that is, that the difference between two treatment is not larger than specified. An example for a trial comparing two antibiotics is that the difference in efficacy between drug A and drug B is smaller than or equal to 10 %-points. This means that if the control treatment gives clinical cure in 85 % of the patients treated we will accept a cure rate with the test drug which is between 75 % and 95 % and still assume that they do not differ in efficacy. For the estimation of the sample size, the following must be defined:

(i) The efficacy in the control group. The closer this efficacy is to 50 %, the more patients will be required to test the null-hypothesis as exemplary in table 1.

(ii) The type I (alpha) error, which is the same as the significance level. This error describes the risk of falsely rejecting the null-hypothesis, that is, although no true difference exists between two treatments, the result of the study showed a (significant) difference larger than the one defined in the null-hypothesis. If the end-point is death, this error is very serious since a study rejecting the null-hypothesis can then normally not be repeated for ethical reasons. The type I error is normally, and arbitrarily, set at 0.05, that is we accept a 5 % risk that, if we reject the null-hypothesis, there is still a true difference between the treatments. In an absolute majority of trials, the type I error should be two-sided, that is, we assume that a new treatment can be both better and worse than the control.

(iii) The type II (beta) error describes the risk of falsely accepting the null-hypothesis. In antibiotic trials this is clearly the most common statistical error since most modern antibiotics are very effective. Normally, the type II error is set at 0.2, that is we accept a 20 % risk that although the null-hypothesis was accepted, a true difference between two treatments exists. Sometimes the type II error is expressed as statistical power which mathematically is $(1-\beta)$ x100. With a type II error of 0.2 the statistical power is 80 %. Compared to the type I error, the type II one is less serious since it always allows for a repetition of the study. It is commonly, even normally, overlooked. In a large number of published trials including 20-30 patients or less in each group and failing to demonstrate a significant difference, statements are made about "equality" between the two treatments studied. One should then realize that there may be upto 50 % risk that a true difference of 25 % - points exists.

(iv) The difference (delta) to be demonstrated. While the type I and II errors should be discussed with a statistician, the decision on the difference which is clinically important to find is one which must be made solely by the investigators. For practical reasons, the delta is rarely set below 10 %-points (note the difference between % and % - points). The rarer a disease, the more often it is necessary to use a high delta.

The above definitions will result in a calculation of how many patients are required for testing a null-hypothesis. Since the end-point most often is dicothomous, this figure will be high in most studies (Table 1).

Table 1. Number of patients required to test null-hypothesis that two treatments do not differ using a dichotomous end-point. The type I error (two-tailed) is assumed to be 0.05. Data from 4.

Efficacy in control group	Type II error	No. of patients per group for delta		
		5%-points	10%-points	15%-points
95%	0.1	580	190	100
	0.2	430	140	80
	0.3	340	110	60
90%	0.1	920	260	130
	0.2	700	200	100
	0.3	540	160	85
85%	0.1	>1000	330	160
	0.2	900	225	120
	0.3	720	200	100
80%	0.1	>1500	400	190
	0.2	>1000	295	150
	0.3	870	230	105

In addition to the calculation of the sample size, exclusions before and after randomization must be considered. Exclusions before randomization are those caused for administrative reasons and other reasons for eligible patients not to be considered for trial entry, for example patient refusal and fulfillment of exclusion criteria (see below). In a large study in patients with urinary tract infections, 2255 patients were screened, 1369 were excluded before randomization due to various reasons and another 252 after exclusion (Table 2) (5).

Table 2. Exclusion before and after randomization in a multicentre trial of treatment of urinary tract infections. Data from 5.

No. of patients screened	2.255
No. of patients excluded before randomization	1.369
% of screened	61
Reasons for exlusion before randomization (%)	
refusal to participate	33
administrative reasons	18
previous enrollment	15
alleged hypersensitivity to trial drug	13
no bacteriuria	9
age <18 years	6
other reasons	6
No. of randomized patients (% of screened)	886
No. of patients excluded after randomization	252
% of randomized	28
Reasons for exclusion after randomization (%)	
no significant bacteriuria	79
too short treatment time	6
no follow-up	4
other reasons	12

Inclusion and exclusion criteria

Inclusion criteria are positive ones, for example type of infection, age and sex. They are rarely controversial. Exclusion criteria, on the other hand, present major problems in many studies. The main rule is that the more exclusion criteria that are used in a trial, the less external validity will its results have. In antibiotic trials of parenteral drugs, some of the more common exclusion criteria are:

(i) Age limitations. With the exception of children, age limitations should be avoided.

In the light of the fact that serious infections are most common in elderly, it is not meaningful to set an upper age limit for a trial in patients with such infections.

(ii) Exclusion of patients with very serious infections or rapidly fatal underlying diseases. This is a common exclusion criterion. If one compares the natural history of an infection with the outcomes reported in clinical trials, there are often striking differences. For example, the minimum mortality in *Haemophilus influenzae* meningitis is 5 % but in a large series of patients randomized to either ampicillin plus chloramphenicol or to ceftriaxone, only 3/312 (1.0 %) died (1).

(iii) Exclusion of patients with various types of organ impairment are used to avoid possible accumulation or toxicity of drug due to prolonged excretion. With injectable antibiotics it is common to define a limit for the degree of renal impairment which is accepted. This is legitimate as long as the limitation does not exclude patients with physiological age-related renal impairment.

(iv) Exclusion of patients who have received previous antibiotic treatment. This is often necessary since the aetiology can otherwise not be identified and it will be impossible to separate the activity of the trial drugs from previously used treatments.

STUDY DESIGN

The ideal antibiotic trial is prospective, controlled, randomized and double-blind. For reasons discussed below, this can often not be achieved.

Controls

All antibiotic studies should be controlled. Open, uncontrolled trials can only be justified in so called explanatory studies where the purpose is to study, for example, pharmacokinetics. As soon as the purpose is to evaluate safety and/or efficacy, controls must be introduced. One important reason is that each patient sample is more or less unique, especially when serious infections are considered. Another important reason for the need of controls is investigators' bias. Many investigators claim that they are not biased. However, if an investigator does not believe that the trial drug is better or worse than existing treatment, that is, is biased for or against the test drug, it is difficult to envisage why he/she should do the trial.

There are various types of controls which may be considered. Placebo or no treatment are rarely possible to use in therapeutic antibiotic studies due to ethical limitations. In prophylactic studies, however, there are still some indications where conclusive data on protective efficacy of antibiotics is missing and placebo is then the natural control. Historical controls should not be used since evolution in the field of treatment of infectious diseases has been very rapid. Active treatment is the most common type of controls in antibiotic trials. It is then important to realize that it must be possible to verify by references the efficacy and safety of such treatment in the type of infection studied.

Randomization

Randomization is the distribution of patients to treatment groups in a way which cannot be predicted by the patients or the investigators. Notably, allocation by birth date, hospital admission number or similar methods is not randomization. Randomization is normally executed by using lists of random numbers. If the study is open, central or third party randomization should be used. This means that the investigator, examines the patient, decides that he/she is eligible for the study, obtains informed consent and then contacts a third party, identifies the patient and gets the treatment allocation. By this method, adherence to the randomization is guaranteed.

Stratification

In studies involving more than one centre, stratification is routinely done by centre.

Stratification for factors which are prognostically important should be considered if the patient sample is small (below 100/group) but is not necessary if large patients samples are recruited.

Blinding

Any study introduces considerable risk for influence of investigators' bias. If possible, studies should therefore be blinded. One can then choose a single-blind desing in which either the patient or the investigator is blinded as to treatment given. A more reliable technique is to use a double-blind design where both the patient and the investigator do not know what is given. In studies of serious infections, double-blinding is often difficult to implement; most parenteral antibiotics have characteristics when they are dissolved that cannot easily be masked. It is important to realize that if effective blinding cannot be achieved, the study should be open. An example where blinding has often been used is studies on ceftazidime. However, this antibiotic leaves a very characteristic smell on the patients treated and the investigators are therefore effectively unblinded as to which treatment was given.

In studies where blinding is not possible, the outcome of treatment and the safety of the drugs used should be evaluated by an independent person who is unaware of the treatment given (evaluator-blinded design).

Studies involving several centres

For the obvious reason that a single centre can rarely accumulate enough patients to test a null-hypothesis, antibiotic studies must often involve several centres. One can then choose between doing a true multicentre trial (MCT) or a multiple independent trial (MIT). An MCT is a fully coordinated study, that is, all investigators must meet before, during and after the trial, it must start and end simultaneously at all centres and all procedures must be the same at each centre. This is a very costly and time-consuming type of trial which should be reserved for studies which are pivotal and have high scientific value. In an MIT, the protocol is largely the same, but not necessarily identical, at all centres and the coordination is handled by the sponsor. Subprojects should be encouraged in MCTs and MITs to allow meaningful publications by the individual centres. In both the MIT and the MCT, publication of efficacy data can normally not be done by the individual centres.

TREATMENT

Dose regimens

Doses used in clinical trials are often based on in vitro activity, pharmacokinetic properties and animal toxicity of the antibiotics to be studied. Proper dose finding studies are not possible in serious infections since in such studies the lowest dose used should be significantly less effective than the other doses employed.

In clinical trials, the doses used should be consistent between patients. Dose reductions, for example in patients with reduced renal function, must be defined beforehand and carefully considered in the protocol.

Concurrent treatment

In principle, the use of other antibiotics should not be allowed unless combination therapy is studied. It is, however, often necessary to add another antibiotic due to gaps in the antibacterial spectrum. A common situation where this is the case is when cephalosporins are tested in the treatment of intraabdominal or gynaecological infections and coverage of anaerobes is required. Metronidazol can then be added.

Treatment interruptions

Often one or more doses are missed in a clinical trial, as is also the case in normal clinical practice. Such patients should in most cases be accepted for full analysis. The trial pro-

tocol must define how many doses can be missed.

Treatment time

Treatment times must be defined and should not vary considerably between patients. It is often necessary to perform large studies to find the shortest treatment time which gives maximal cure rates.

Subsequent treatment

In clinical trials on patients with serious infections, it is common to continue parenteral treatment until antibiotics are no longer needed. This is a highly artificial way of treating these infections; in clinical practice parenteral treatment is normally discontinued after a few days and the patient is switched to an oral antibiotics. This is often not possible in trials since few modern antibiotics are available both for oral and parenteral use. Treatment must then be changed to a new antibiotic, often with narrower spectrum. In the proposed new U.S. guidelines for clinical trials of antimicrobials (3), follow-up treatment with oral derivatives may be used.

Compliance

In the treatment of serious infections, patients' compliance is normally not a major problem since treatment is administered by the hospital staff and not the patient itself.

REGISTRATION OF EFFICACY

End-points for clinical efficacy

A major problem in clinical trials involving patients with serious infections is to establish objective end-points. The most commonly used ones are "cured", "improved" and "failure". Although efforts are made to define these terms, they remain subjective. It is also in most cases difficult to define "cured" and "improved" so that consistency between investigators is achieved for these classifications. Therefore, a better way of grading the outcome is to use "responded" and "failed to respond" as end-points, thereby pooling patients who were cured or who improved. Efforts have been made to define continuous or semi-continuous end-points. One type of such end-points is the scoring systems used for patients in serious conditions, for example, APACHE, APACHE II AND SAPS. It is recommended that such scoring systems are used since they serve as a means to establish balance between trial groups before start of antibiotic treatment. They may also serve as a means of relating the speed of recovery to time, although little data is available on such methods in treatment of infections. In patients with enteric infections, number of stools per day is a useful semi-continuous end-point with a better sensitivity than the dichotomous ones. Fever has been used as an end-point in some trials. However, with the frequent use of corticosteroids as well as other types of antipyretic drugs, fever may very well be masked in many patients. Laboratory surrogate markers have rarely been useful as endpoints in clinical trials of antibiotics. Of such end-points leukocytosis is of limited value as is erythrocyte sedimentation rate (ESR). C-reactive protein (CRP), if determined quantitatively may be a useful end-point infections since it decreases more rapidly than ESR.

End-points for bacteriological efficacy

Bacteriological end-points are normally included in antibiotic trials. Obvious requirements for such end-points are that the causative agent(s) was isolated pre-therapy and that adequate samples were to be taken after treatment. In patients with serious lower respiratory tract infections, the sampling technique is a major problem. Sputum bacteriology is acceptable if (i) the patient produces sputum (which is not the case in the acute phase of a pneumococcal pneumonia), (ii) the sputum is properly handled at the laboratory (washing, dissolution, microscopy to rule out oropharyngeal contaminations and proper culture) and

(iii) the patient is not tracheotomized or intubated. In patient with assisted ventilation, samples are often taken through the endotracheal tube or the tracheal cannula. Such samples cannot be taken without contaminations from the oropharyngeal flora or the skin flora, respectively. Therefore, the bacteriological diagnosis in these patients must be based on samples obtained by fiber bronchoscopy and taken by a protected brush or a broncheo-alveolar lavage. In patients with other types of serious infections, verification of the aetiology may be equally difficult. For example, in patients with deep abscesses it is not always possible to obtain pus and in patients with biliary tract infections, adequate samples require laparotomy or ERCP-technique. After treatment, lack of material for sampling must be accepted as equal to eradication of the causative pathogen.

Another major problem in the bacteriological diagnosis of serious infections is to separate true pathogens from colonizers. Often, arbitrary decisions must be taken as to classification of organisms in this respect. In MCTs and MITs, such classifications must be consistent between centres.

The registration of bacteriological findings should include susceptibility testing. MITs are often performed in several countries with various techniques for routine sensitivity testing. A recommended way to avoid this problem is to save all isolated strains for subsequent determination of MICs at one central laboratory.

REGISTRATION OF SAFETY

In seriously ill patients, it is often difficult to separate adverse reactions from reactions caused by the infection itself, underlying diseases or concomitant treatments. To avoid important adverse effects of a new antibiotic remaining undetected, registration should concentrate on adverse events rather than on known or expected side effects, that is, all unwanted or unexpected events are reported. Since this will result in registration of a large number of events not related to the drugs tested, conclusions can only be drawn from comparisons between the test drug and the control.

It is customary to ask the investigators to classify the adverse events with respect to severity and relation to the drug tested. It should be noted that such classifications are highly subjective.

ANALYSIS OF EFFICACY AND SAFETY

Decisions on what should be analysed and how the analyses should be performed should be taken before the material is divided into treatment groups. In a blinded study, all analyses must be completed before the randomization code is broken; analyses performed after that can only serve as indicators of new fields which should be studied. Thus, it is important to consider beforehand which analyses may yield results not detected in an analysis of the entire sample. An example is given in table 3.

Table 3. Bacteriological outcome in a study of ceftazidime (CAZ) versus imipenem/cilastatin (IMI) for the treatment of serious nosocomial lower respiratory tract infections. Data from Norrby et al. (to be published).

Sample	Outcome and treatment (no. of pts.)				p-value
	Cleared		Not cleared		
	CAZ	IMI	CAZ	IMI	
All patients	43	37	9	17	>0.05
Pseudomonas pneumonia	14	7	3[a]	12[b]	0.008
Staphylococcus aureus pneumonia	11	12	5	1	>0.05

(a) 1 strain emerged as ceftazidime resistant, (b) 7 strains emerged as imipenem resistant.

Intent-to-treat analysis

An intent to treat (or pragmatic) analysis is one in which all patients randomized are analysed, irrespectively of whether they fulfill criteria for evaluability. The analysis also includes patients who, after randomization did not receive the correct drug or did not receive any of the treatments to be tested. The purpose of this analysis is to evaluate the drugs in a fashion as similar as possible to normal medical practice; if a physician prescribes a drug, anything but good clinical response to that drug is failure. The intent-to-treat analysis also serves the important purpose of demonstrating investigators' bias. For example, if the death rate in a study was significantly higher in one treatment group than in the other using an analysis per protocol (see below) but several randomization errors were committed and the intent-to-treat analysis failed to demonstrate significance, it is highly likely that investigators' bias was the cause of the difference.

Analysis per protocol

The trial protocol should give the criteria for clinical and bacteriological evaluability. These should be realistic and conform with clinical practice to the highest possible extent. It is for example, not recommended to perform repeat blood cultures in a patient with septicaemia if that patient becomes afebrile, neither are wound cultures required if healing of the wound is achieved.

ETHICS

It is today well recognised and internationally accepted that clinical trials of drugs must conform to basic ethical rules as outlined in the Declaration of Helsinki and its amendments. All protocols must be approved by a research ethics committee. The patients considered for the trial or their legal guardians should receive detailed information about the purpose and conduct of the trial and should provide informed consent. The consent can be written (as is the rule in the U.S.A.) or verbal. In patients with serious infections, it is sometimes difficult to fulfill these requirements; treatment has to be instituted without delay and often no legal guardian is present and the patient is not capable of understanding the information and provide informed consent. Nevertheless, it is imperative that the ethical rules are observed; most international medical journals today refuse to publish studies unless there are clear statements on approval by a research ethics committee and on how informed consent was achieved.

CONCLUSIONS

To perform high quality clinical trials of antibiotics in patients with serious infections is difficult. However, the efforts required to be spent on a trial of high quality do not differ very much from those spent on a mediocre study. Critical issues for achieving success are to spend a long time on a detailed description of a trial and its execution in the protocol, to cooperate with others to gather a large enough patient sample in a reasonable time and to eliminate investigators' biases to the largest possible extent.

REFERENCES

1. Barson WJ, Miller MA, Brady MT, Powell DA: Prospective comparative trial of ceftriaxone vs. conventional therapy for treatment of bacterial meningitis in children, *Pediatr Infect Dis 4:* 362 (1985).
2. Fihn SD, Stamm W E: Interpretation and comparison of treatment studies for uncomplicated urinary tract infections in women, *Rev Infect Dis 7:* 468 (1985).
3. Gilbert DN, Beam TR, Kunin CM: The implications for Europe of revised FDA guidelines for clinical trials with anti-infective agents, *Eur J Clin Microbiol Infect Dis 9:* 552 (1990).
4. Huitfeldt B: Statistical aspects of clinical trials of antibiotics in acute infections, *Rev Infect Dis 8 (Suppl B):* S350 (1986).
5. The Urinary Tract Infection Study Group: Coordinated multicenter study of norfloxacin versus trimethoprim - sulfamethoxazole treatment of symptomatic urinary tract infections, *J Infect Dis 155:* 170 (1987).