Report on the Fourth Consensus Conference on the Methodology of Clinical Trials with Anxiolytic Drugs

J. Angst, M. Ansseau, P. Bech, R. R. Engel, H. Hippius, M. Lader, Y. Lecrubier, O. Lingjaerde, M. Paes de Sousa, A. Rossi, E. Rüther, B. Saletu, C. N. Stefanis, K. D. Stoll, H. G. M. Westenberg, B. Woggon

1. Introduction

1.1. Organizational Structure

The Consensus Conference on the Methodology of Clinical Trials with Anxiolytic Drugs was organized by an Organizing Committee consisting of members of the Association for Methodology and Documentation in Psychiatry (AMDP), the Collegium Internationale Psychiatriae Scalarum (CIPS), and the European College of Neuropsychopharmacology (ECNP), together with staff of the Psychiatric University Hospitals of Munich and Zurich. Clinical Experts from 14 European Countries served as National Advisers to the Organizing Committee and suggested participants. The theme of the conference was divided into 10 major topics and 10 speakers were asked to review current controversial issues. Each session was chaired by a clinical expert and Rapporteurs recorded the content of the papers, kept the minutes of the discussion, and reported both back to the Consensus Committee.

The Consensus Committee consisted of the Organizing Committee and the National Advisers. After the main part of the program the Consensus Committee formulated a first draft of the consensus statements. After some editing, the president of the CC presented these statements to the plenum. The presentation was followed by a plenary discussion with all 110 participants of the conference. The results of this discussion were incorporated into a draft of the final paper, which again was sent to all participants of the conference for further review.

1.2. Previous Conferences

Previous Consensus Conferences on the Methodology of Clinical Trials were held in 1988 in Zurich on Antidepressant Drugs (*Angst* et al., 1989), in 1989 in Munich on Nootropic Drugs (*Amaducci* et al., 1990) and in 1990 in Zurich on Neuroleptic Drugs (*Angst* et al., 1991).

1.3. Scope of the Conference

For the present conference, anxiolytic drugs were defined as *agents that reduce anxiety in a variety of clinical conditions,* namely anxiety disorders, panic disorder, depressive disorders, obsessive compulsive disorder, and various others. From a chemical point of view, *benzodiazepines* play the major role, but other groups are also used in this indication, namely MAO-inhibitors (in panic disorder), tricyclic antidepressants (in panic and obsessive compulsive disorder), selective 5-HT reuptake blockers (in panic and obsessive-compulsive disorder), and the new "Nonbenzodiazepine Tranquilizers" that consist of various chemical groups by themselves.

One of the more general issues that had to be decided upon before the conference was whether it was possible to construct *general guidelines* that would be valid for all anxiolytic drugs in all clinical conditions with symptoms of anxiety. If not, specific suggestions would be needed for the different disorders. The decision to ask four different speakers to summarize the methodological conditions in four different clinical conditions reflects some scepticism about the practicality of general guidelines. The design and the particular assessment instruments to be used in a study may indeed largely depend on the disorder studied and may be specific to a given condition. Nevertheless some methodological issues arise in *all* trials with anxiolytic drugs, whatever the disorder.

1.4. Previous Guidelines

Several guidelines for anxiolytic drug trials have been published in the past. Most were issued before the publication of the DSM-III-R (APA, 1987) and thus use a different terminology to classify the patients. Of great influence have been the so-called *Wittenborn* guidelines (*Wittenborn,* 1977) and the FDA guidelines from 1978 (FDA, 1978). The WHP Regional Office for Europe has published guidelines for anxiolytic drug trials (WHO, 1983) and more recently the NIMH held a Consensus Conference on the Treatment of Panic Disorders, where some of the methodological problems in drug evaluation were discussed (NIMH, 1991).

2. Studies in Healthy Human Volunteers

In contrast to antidepressants and antipsychotics, different aspects of anxiolytic drug actions can be faithfully assessed in animal as well as in human models. The importance of preclinical and human pharmacology studies in determination of therapeutic and safety profile of anxiolytic drugs is therefore emphasised.

Studies in healthy human volunteers may play an important role in assessing tolerability, pharmacokinetics and -dynamics, as well as the *potential* psychoactive effects of a substance *before it has been given to a single patient.* Typical pharmacodynamic questions in studies with human volunteers include the following: has the substance any effect on the CNS?

Pharmacopsychiat. 26 (1993) 1–5

Received: 10. 8. 1992
Accepted: 18. 8. 1992

Has it specific effects that resemble previously studied drugs? What is the minimal centrally active dose? Does the central effect change with increasing dose? What is the equipotent dose to well-known drugs? What is the time course of its central effects? What are the differences between various galenic formulations? Does the substance impair memory or psychomotor performance?

Sensitive measures are available for the following functions: *sleep efficiency,* measured with objective sleep laboratory procedures and subjective measures; *vigilance,* assessed by EEG, self-assessment or objective tests; *memory functions,* assessed by a range of memory tests; *visuomotor skills,* also assessed by objective tests; and *subjective feelings,* usually assessed by self-report or by observer enquiry. Various other laboratory procedures and imaging techniques may also be of value, sometimes combined with experimental test procedures.

The *liability of abuse* of an anxiolytic may also be assessed in healthy volunteers. For instance, drug-seeking behaviour can be measured by drug preferences or by subjective report.

3. Patients

3.1. Inclusion Criteria

The clinical diagnosis will usually be made according to structured diagnostic manuals like the *DSM-III-R or ICD-10.* In addition to the appropriate diagnosis most studies call for a *minimal degree of severity* of the disorder, usually assessed with a rating scale specifically constructed and validated for that syndrome. The inclusion criteria very often also specify a *minimal duration* of the disorder, and sometimes this criterion is integral to the diagnosis. However, diagnostic definitions may change: in the DSM-III the minimal duration for a diagnosis of Generalized Anxiety Disorder was one month, in DSM-III-R it is six months. Thus it may be wise to include certain time specifications within the inclusion criteria independently of the diagnostic criteria.

There is much *overlap* between the disorders treated by anxiolytic drugs. As a consequence there are few "pure" cases. Studies that use only such cases may be representative of only a minority of the patients that will be treated with the drug in clinical practice. Although pure cases will be studied initially, it is highly desirable to include a broader spectrum of a disorder in later trials. If necessary the other relevant factors can be controlled by analyses of covariance or by factorial designs.

Caution is needed when using rating scales for stratification or exclusion because most rating scales for anxiety have items that overlap extensively with those for depression. Consequently, a patient scoring high in anxiety may also score high in depression. If a study on an anxiety disorder includes only patients with a low depression score the *most severe cases* of anxiety may actually be *excluded* because they score high on the depression scale as well. Therefore, the *relative expression* of both measures needs to be taken into account.

3.2. Description of the Setting

Most studies with anxiolytic drugs are conducted on *outpatients.* As a consequence, the setting of the study may vary considerably between studies, and the recruitment procedures can be very different. The published studies with panic disorder patients show very different placebo response rates, indicating a strong influence of factors unrelated to the study medication. These may concern the definition of the patients, but may also result from differences in the parameters of the trial context. The authors of a study should make reasonable efforts to describe carefully their setting and to standardize treatment and milieu during the course of the trial.

3.3. Value of Biological Markers

Biological measures have not been widely used in the clinical assessment of anxiety because of limited relevance and logistical restraints. However, in special areas of application (for instance in EEG sleep laboratory studies) very useful information can accrue that is not otherwise available.

4. Design of Trials

4.1. Type of Studies

Clinical studies with new drugs should follow a particular order: first come studies that establish a *clinical hypothesis of activity* (pharmacological evidence from animal experiments., incidental experiences within other studies, open trials); second are studies delineating a *dose-response-relationship;* and thirdly *double-blind, randomized, parallel-group studies against placebo or (better) against both placebo and a reference drug.* From a current point of view, the dose-response studies are very important. In the last few years some large randomized pivotal studies with fixed dosages have been conducted using dose ranges that have only later (in clinical practice) proved to be unnecessarily high.

The safety of new drugs is usually assessed together with their efficacy. Special safety studies are, however, necessary if a drug is suspected to show *rebound effects* (i. e., *after drug discontinuation,* the symptoms return with greater severity than before the treatment) or to induce physical dependency. In these cases, possible rebound effects and/or physical dependency should be measured after a controlled, double-blind discontinuation of the drug, following adequate treatment. With benzodiazepines, drug discontinuation is done gradually. The influence of *duration of treatment* on rebound and/or dependency can be studied by switching patients to placebo at different times after the start of the experimental treatment.

4.1.1. Short-Term Studies

Short term trials last for four to eight weeks. Beyond this, the trial is medium term.

The short-term efficacy study is the commonest with anxiolytic drugs and is usually the first type performed with a new drug. Since many eligible patients have usually received some kind of previous drug therapy, trials have to incorporate an initial placebo wash-out phase, whose duration depends on the prior drugs. In studies comparing nonbenzodiazepine drugs to benzodiazepines, a very long wash-out

phase is needed for prior benzodiazepine users. This has been noted in studies that have stratified patients according to their previous use of benzodiazepine drugs. It is not that important for studies comparing two active benzodiazepine compounds. It is also recommended that studies end with a single-blind placebo run-out phase to assess the effects of drug withdrawal. It may also be interesting to have a follow-up measurement several weeks after the experimental therapy to monitor the further course of the disorder and the effects of further (nonexperimental) treatment.

4.1.2. Long-Term Studies

Long-term studies have a duration of more than six months. Many of the disorders treated by anxiolytic drugs are chronic conditions and therefore anxiolytic efficacy under long-term conditions needs to be demonstrated. It is even more important to assess rebound or withdrawal after long-term than short-term use by placebo substitution with or without drug-dosage tapering.

4.1.3. Maintenance Studies

Maintenance studies require the patient to reach a certain criterion of improvement on a drug and then the patient is allocated randomly to drug or placebo. They test whether continuous drug use can reduce the risk of relapse.

4.2. Control Groups

A placebo controlled parallel-group trial is judged by many experts to be the best design to test the efficacy and safety of an anxiolytic drug. Most regulatory authorities insist on such trials. Where a standard treatment exists, it should be incorporated into a three-arm design.

4.3. Dosage

As described above, the appropriate dosage should be established by means of dose-range studies. If a drug is believed to be effective in more than one disorder, dose-range studies should be performed separately in each proposed indication, since effective dosages may differ considerably. To give an example: in the past few years the dosages needed for patients with major depressive disorder have been used without changes for large confirmative trials in patients with obsessive compulsive disorder and panic disorder. *After* those trials were done, the dosages were criticized as excessive for these disorders.

4.4. Sample Size

The sample size is governed by the primary purpose of the study. In the past, many studies have used samples that were too small to give interpretable results. In recent years, some studies seem to use sample sizes that are unnecessarily large and this is also a waste of resources. Therefore only reasonable assumptions (often accessible by means of metaanalysis of older trials) should be put into the power calculations that predict the adequate sample size of a trial.

4.5. Concomitant Drug and Psychotherapy

In most trials with anxiolytics, concomitant medication is not necessary and should be avoided. Exceptions are, however, possible. With a nonsedative anxiolytic for instance, a hypnotic may be needed in addition to the anxiolytic. As in other drug trials, any concomitant medication or self-medication including alcohol, caffeine, and nicotine should be documented.

In some disorders, especially in obsessive-compulsive and panic disorders, a certain level of behavioral, psychoanalytic, or supportive treatment is common. The more effective such a treatment, the more remote the chance of finding a significant difference in efficacy between drug and placebo. If such a treatment is initiated – for whatever reasons – its time course must not coincide with the study drug schedule.

4.6. Multicenter Trials

Trials with anxiolytics will often result in large multicenter trials because this is the only way to assure sufficient numbers of patients within a reasonable time. Multicenter trials pose problems regarding the homogeneity of centers. The number of patients treated in each center must be large enough for detection of any drug efficacy by center interaction. Thorough, case-related training of the clinicians and careful monitoring throughout the trial are important to maintain standardisation of all procedures.

4.7. Disorder-Related Design Consideration

4.7.1. Trials in Panic Disorder

Trials assessing the effectiveness of a drug in the treatment of panic disorder should not be shorter than eight weeks.

4.7.2. Trials in Generalized Anxiety Disorder

In generalized anxiety disorder the duration of a short-term trial may be between four and six weeks.

DSM-III-R defines GAD with a time criterion of at least six months. Many people regard this criterion as needlessly long and most patients will have been treated with several drugs in this time. On the other hand, GAD has a variable course at its beginning and placebo response may be more probable within the first few months. Therefore it may be wise to exclude patients with a very short history.

4.7.3. Depression Studies

New drugs may have a potential for both anti-anxiety and antidepressant activities. If so, the drug should be tested in both disorders according to the established rules relevant to each. Therefore, the duration of a trial with depressed patients should last at least four weeks, preferably six weeks. In trials assessing the antidepressant activity of known anxiolytic drugs, stratification of the sample according to the severity of depression yields important information on differential effectiveness in mild versus severe cases.

4.7.4. OCD Studies

The duration of a short-term trial in patients with OCD should not be shorter than eight weeks. Although effects have been seen much earlier, improvement increases linearly with treatment duration.

5. Assessment

5.1. Compliance Assessment

Compliance should be assessed carefully. Otherwise, compliance may be so bad that the results of a study are limited or even useless. Drug tablet counting, keeping a diary, and returning unused medication are simple behavioral approaches towards improving compliance. Urine or blood analysis may be helpful and may also detect whether *unauthorised* medication is being taken.

5.2. Assessment of Clinical Efficacy

5.2.1. Global Clinical Assessment

A global clinical rating has been the most sensitive measure in many studies. With regard to the sensitivity and statistical power of a study, a clinical global measure may be the best main outcome variable. Its drawback is ignorance as to *which* features of the clinical picture influence the clinician's rating and thus the *exact* effect of the drug remains undefined (= lacking *specificity* of the measure).

Possible alternatives to the use of a global measure are the combination of two or more measures (for instance number *and* severity of panic attacks). If an explicit combination of measures is interpreted *only as a combination,* the alpha risk of the study is not inflated. If more than one measure is interpreted, precautions must be taken against increasing the alpha risk.

5.2.2. Rating Scales

An observer-rating of anxiety (either global or narrowed towards specific manifestations of anxiety) remains the standard outcome measure, especially if made by a clinically trained person.

As a conseqence, many rating scales exist in this field as well as, for example, in depression. Despite some criticisms, it is probably wise to include one of the popular scales as a benchmark, against which newer scales are calibrated.

Self-rating scales are fairly useful in anxiolytic drug trials, because many of the symptoms of anxiety are only accessible through subjective experience. Again, there exist more global scales and more specific ones. In some conditions, as for instance OCD, it is accepted practice to construct a short "customised" scale according to the main symptoms complained of by each patient ("Prime Symptom Scale").

The many scales assessing "anxiety" are not necessarily interchangeable: scales that are useful in assessing anxiety in GAD may fail totally to estimate the global anxiety of a patient with panic attacks.

5.2.3. Behavioral Measures

Behavioral measures are often used in trials with obsessive compulsive disorder. In this condition the exact assessment of a specific behavior can constitute a very useful target measure. In other conditions, behavioral measures are usually less reliable and less sensitive than the more global rating scales that are essentially integrations of many specific behaviors over a given period of time.

5.3. Biological Measures

As stated earlier, biological measures may be of great value, but only in a *specific context,* e.g., in a sleep laboratory evaluation.

5.4. Assessment Interval

During the first four to six weeks, assessments should be made weekly. Later on, assessments may be done every two weeks or even monthly; this depends very much on the aim and the total duration of the study.

5.5. Assessment of Safety

5.5.1. General Adverse Drug Reactions

As a general rule the use of checklists for symptoms is recommended. If, for whatever reason, this practice is not followed, and if spontaneously volunteered descriptions of putative adverse drug reactions are recorded instead, then the procedure used has to be described in full detail. In anxiety disorders, many symptoms of the illness mimic side-effects of a drug. Therefore the first assessment of possible "adverse events" must be made at baseline before any drug treatment.

5.5.2. Specific Adverse Drug Reactions

Many anxiolytic drugs have the adverse effects of *symptom rebound* after treatment discontinuation and of *physical dependency.* Both side-effects have to be addressed in safety studies using specific rating instruments within experimental drug discontinuation designs (see 4.1 above). Withdrawal effects and return to old anxiety levels can be discriminated, not by the symptoms themselves, but by the *different time course after* drug discontinuation. Withdrawal symptoms are usually time-limited, though protracted abstinence syndromes are sometimes seen. It is not yet clear whether the dependency risk correlates with the time of exposure, as most data about this relationship do not stem from independent observations. If this is a potentially important aspect of a new drug then the respective designs have to be used (see the design section above). Multiple regression analyses have shown that personality factors may also play an important role.

Many anxiolytic drugs also have adverse effects on cognitive and psychomotor functions. These side-effects need addressing in separate experimental studies using sophisticated but appropriate batteries of psychological tests. Nevertheless, simpler tests can be employed in routine efficacy studies in outpatients and even in primary care settings.

6. Statistical Analysis

The statistical treatment of anxiolytic drug trials should folllow the usual procedures. The problem of differential drop-out rates is particularly problematic. Drop-out rates have sometimes been high in anxiolytic drug trials and have been unequal in the treatment groups. In such situations, an analysis of the completers only is insufficient and may be misleading. Therefore an *Intent-to-treat-analysis* is a necessary part of the statistical evaluation of anxiolytic drug trials.

References

Amaducci, L., J. Angst, P. Bech, O. Benkert, J. Bruinvels, R. R. Engel, C. G. Gottfries, H. Hippius, R. Levy, O. Lingjaerde, J. J. López-Ibor Jr., J. M. Orgogozo, C. Pull, B. Saletu, K. D. Stoll, B. Woggon: Consensus conference on the methodology of clinical trials of "nootropics", Munich, June 1989: Report of the consensus committee. Pharmacopsychiatry 23 (1990) 171–175

American Psychiatric Association: DSM-III-R: Diagnostic and Statistical Manual of Mental Disorders. APA, Washington DC (1987)

Angst, J., P. Bech, D. Bobon, R. R. Engel, H. Hippius, G. J. Janzen, Y. Lecrubier, O. Lingjaerde, H.-J. Möller, St. Montgomery, M. Paes de Sousa, A. Rossi, B. Saletu, G. Sedvall, C. Stefanis, K.-D. Stoll, B. Woggon: Report on the third consensus conference on the methodology of clinical trials with antipsychotic drugs. Pharmacopsychiatry 24 (1991) 149–152

Angst, J., P. Bech, P. Boyer, J. Bruinvels, R. Engel, H. Helmchen, H. Hippius, O. Lingjaerde, G. Racagni, B. Saletu, G. Sedvall, J. T. Silverstone, C. N. Stefanis, K. Stoll, B. Woggon: Consensus conference on the methodology of clinical trials of antidepressants, Zurich, March 1988: Report of the consensus committee. Pharmacopsychiatry 22 (1989) 3–7

Food and Drug Administration: Guidelines for the clinical evaluation of psychotropic drugs. Antidepressant and anti-anxiety drugs. Psychopharmacology Bulletin 14 (1978) 45–63

Wittenborn, J. R. (Ed.): Guidelines for clinical trials of psychotropic drugs. Pharmacopsychiatry 10 (1977) 205–231

World Health Organization: International Guidelines for Anxiolytic Drug Trials. WHO Regional Office for Europa, Copenhagen (1983)

National Institutes of Health: Consensus Development Conference on the Treatment of Panic Disorder, September 25–27, 1991. NIH, Bethesda, MD (1991)

Prof. Dr. R. R. Engel
Psychiatrische Universitätsklinik
Nußbaumstr. 7
D-8000 München 2
Germany

Appendix

Active Participants

J. Angst	Switzerland	CC		OC	M	
M. Ansseau	Belgium	CC	A			R
P. Bech	Denmark	CC	A			
D. Bobon	Belgium			OC		
J. A. den Boer	Netherlands		S			
M. Bourin	France		S			
R. Buller	Germany		S			
R. R. Engel (**)	Germany	CC		OC		R
J. D. Guelfi	France		S			
I. Hindmarch	United Kingdom		S			
H. Hippius (*)	Germany	CC		OC		
G. Laakmann	Germany		S			
M. Lader	United Kingdom	CC	A		M	
Y. Lecrubier	France		A		M	
O. Lingjaerde	Norway	CC	A			
J. J. López-Ibor, Jr.	Spain		A			
D. B. Montgomery	United Kongdom		S			
S. A. Montgomery	United Kingdom		S			
M. Paes de Sousa	Portugal	CC	A			
H. Pétursson	Iceland		S			
C. Pull	Luxemburg		A			
K. Rickels	USA		S			
A. Rossi	Italy	CC	A			
E. Rüther	Germany	CC	A			R
B. Saletu	Austria	CC	A			R
G. Sedvall	Sweden		A		M	
C. N. Stefanis	Greece	CC	A		M	
K.-D. Stoll	Germany	CC		OC		
H. G. M. Westenberg	Netherlands	CC	A			R
B. Woggon	Switzerland	CC		OC	M	

CC	Consensus Committee
A	National Advisor
S	Speaker
OC	Organizing Committee
M	Moderator
R	Rapporteur

Four Advisors, *D. Bobon, J. J. López-Ibor, Jr., C. Pull,* and *G. Sedvall,* and one speaker, *J. A. den Boer,* were unfortunately unable to participate in the conference. Prof. *den Boers'* paper was read by *H. G. M. Westenberg.*

(*) President of the Consensus Committee
(**) Local Organizer