The European Agency for the Evaluation of Medicinal Products
Evaluation of Medicines for Human Use

London, 25 April 2002
CPMP/EWP/518/97, Rev. 1

COMMITTEE FOR PROPRIETARY MEDICINAL PRODUCTS (CPMP)

NOTE FOR GUIDANCE ON CLINICAL INVESTIGATION OF MEDICINAL PRODUCTS IN THE TREATMENT OF DEPRESSION

DISCUSSION IN THE EFFICACY WORKING PARTY	July 1997
TRANSMISSION TO CPMP	March 2001
RELEASE FOR CONSULTATION	March 2001
DEADLINE FOR COMMENTS	September 2001
DISCUSSION IN THE EFFICACY WORKING PARTY	November 2001
TRANSMISSION TO CPMP	April 2002
ADOPTION BY CPMP	April 2002
DATE FOR COMING INTO OPERATION	October 2002

Note:

This revised Note for guidance will replace the previous Note for guidance, adopted in 1988

7 Westferry Circus, Canary Wharf, London, E14 4HB, UK
Tel. (44-20) 74 18 84 00 Fax (44-20) 74 18 8613
E-mail: mail@emea.eu.int http://www.emea.eu.int

NOTE FOR GUIDANCE ON CLINICAL INVESTIGATION OF MEDICINAL PRODUCTS IN THE TREATMENT OF DEPRESSION

This note for guidance should be read in conjunction with Part 3 of the Annex of Directive 75/318/EEC, and 83/570/EEC and current and future guidelines, especially those on:

The extent of population exposure to assess clinical safety of drugs intended for long-term treatment in non-life threatening conditions

Biostatistical methodology in clinical trials in applications for marketing authorisations (ICH E9)

Choice of control group (ICH E10)

Clinical investigations of medicinal products in geriatrics (ICH E7)

Clinical investigations of medicinal products in children (ICH E11)

Clinical investigations in bipolar disorders (CPMP/EWP/567/98)

This note for guidance is intended to assist applicants in the application of these Directives with respect to specific problems presented by clinical investigations of medicinal products intended to be used for treating depression.

1. INTRODUCTION

A medicinal product is to be considered an antidepressant only when it is shown to be effective in treatment of major depressive episodes. Major depressive episodes are depressive syndromes, which are distinguishable from the symptom alone.

Depressive disorders are classified in various classification systems, e.g., DSM IV and ICD 10. According to DSM IV, depressive disorders can be divided into Major Depressive Disorder, Dysthemic Disorder and Depressive Disorder Not Otherwise Specified.

This guideline will only deal with Major Depressive Disorder. Major Depressive Disorder is characterised by one or more Major Depressive Episodes.

Major Depressive Disorder is a common disorder, with a lifetime prevalence of about 15 %, perhaps as high as 25 % in women. Major Depression is not a benign disorder and risk of suicide is considerable. It tends to be chronic and both relapse and recurrence are seen frequently.

Treatment of Major depression focuses on the treatment of acute symptoms and maintenance of the effect during the index episode. Additionally purpose of treatment may be the prevention of new episodes, or recurrence prevention.

A Major Depressive Episode may also be seen in the framework of a Bipolar Disorder. In that case it is sometimes referred to as ‘bipolar depression’

2. SPECIFIC CONSIDERATIONS WHEN DEVELOPING PRODUCTS FOR THE TREATMENT OF DEPRESSION

In developing medicinal products for the treatment of depression specific problems can be encountered. These include:

2.1 Use of placebo

Clinical studies should provide unambiguous evidence of the antidepressant activity and of the effective dose or dose range. In depression comparisons between a test medicinal product and reference substances are difficult to interpret since there is a high and variable placebo response in depression and the absence of a significant statistical difference does not necessarily indicate a therapeutic equivalence. Actually in about one-third to two-third of the trials, in which an active control is used as a third arm, the effect of the active control could not be distinguished from that of placebo. As the effect rate in a specific trial is thus uncertain, a non-inferiority margin can not be determined and a non-inferiority trial is not an option, as the sole basis for demonstrating efficacy

Therefore, from a scientific point of view, randomised double blind comparisons versus placebo are needed, to permit adequate evaluation of efficacy, though showing superiority over an active comparator could be an acceptable alternative. Comparison to a placebo is also of value for distinguishing disease manifestations from adverse reactions of the medicinal product.

Ethically, however, the use of a placebo is a controversial issue, especially when performing studies during acute episodes and/or in out-patients, but it would be detrimental to public health and ethically unacceptable to grant a license to a medicinal product to be used in depression without providing unambiguous evidence of its activity.

Precautions to minimise the impact of the study should be taken however, e.g., by limiting the duration of the study - generally a duration of about 6 weeks should be sufficient and a longer duration should be justified – and by using a fail-safe provision whereby a serious deterioration of the patients condition will allow withdrawal from the trial and standard therapy to be give under open conditions.

Three-arm trials including both a placebo and an active control are recommended.

2.2 Relapse and Recurrence

Relapse is reappearance of the symptoms of the index episode, whereas recurrence is the appearance of symptoms in a new episode. Rebound and withdrawal are seen when withdrawing the medication

Relapse is defined as an increase in symptomatology soon after medication is stopped. It usually indicates that treatment duration was too short. The symptoms are considered to be part of the index episode.

Recurrence is defined as a re-emergence of depressive symptoms after a time without or nearly without symptoms and without medication. It is seen as the start of a new episode.

Relapse should be distinguished from rebound and withdrawal phenomena, which are due to dependence on and/or discontinuation of the medicinal product. Rebound is defined as an increase of symptoms immediately after treatment is stopped, whereas withdrawal is the development of symptoms different from the original ones. One way to deal with it might be by not taking into account the first month after stopping the medication as a secondary analysis.

Depression covers a heterogeneous group of patients and there is a large variance in natural course. In the literature a distinction is made between treatment in the acute phase, the continuation phase and if required the maintenance phase. The purpose of the latter is to prevent new episodes, whereas the continuation phase is meant to prevent deterioration during the index episode. The duration of the continuation phase is usually set at about 6 months, to correspond with the average duration of an episode of depression. In any individual however it should be noted that the duration of an episode varies considerably and maybe more (or less) than 6 months. As this might affect the interpretation of the results, the

6 months cut-off point is not used for regulatory purposes. But instead, the guideline focuses on showing effect during the index episode and/or prevention of the next episode.

For licensing it should be shown that a short-term effect can be maintained during the episode. For this a randomised withdrawal study, also called a relapse prevention study, is probably the best design. In this design, responders to treatment of sufficient duration, with the test product, are (re-)randomised to test product or placebo. In the first period, the test product is usually given open, uncontrolled. The duration of either treatment phase is hugely variable in the literature. It will depend among others on the type of patients included and on the time of inclusion. The optimal duration is not known at the moment, but a duration of e.g., 8 to 12 weeks for the first period appears acceptable, whereas the period after (re-)randomisation usually has a duration of up to 6 months. For such study, the protocol must include specific measures to prevent complication of the disease (especially risk of suicide), like close monitoring and the possibility to use rescue medication or to switch deteriorating patients to appropriate treatment.

A placebo-controlled extension study is less advisable, as there is a risk, that the results will be ambiguous, due, for example, to a differential drop-out in the first few weeks, the fact that placebo responders may continue to be responders, the patients groups are not comparable anymore or the possibility that after a certain time the time-effect curves may become parallel.

Prevention of the next episode(s) or recurrence prevention is not an obligatory part of a registration package. When a claim is made, specific studies are needed. In non-manic depressive patients, definitive comparisons of the test substance should be performed versus a placebo. For prevention in bipolar patients, the relevant guideline should be consulted.

For a given patient, the duration of treatment depends on the rate of his/her recurrences. Patients with a history of several depressive episodes should be included and the recent recurrence rate should be taken into account when powering the study.

As a general principle, the duration of such studies is at least one year, although 2 years or more may be necessary. This will depend on the recurrence rate in the population.

2.3 Extrapolations

As is indicated in the introduction, patients included in the trials will be diagnosed as having Major Depression using accepted diagnostic criteria, e.g., DSM IV. However, depressive symptoms are also seen in other psychiatric disorders or other types of depression. Specific studies should be provided for each indication claimed in this field, even if the antidepressant activity has been correctly documented in major depressive episodes.

The issue of mixed depression/anxiety requires a specific approach. The issue is twofold: anxiety symptoms may be a part of depression or a due to a co-morbid disorder like GAD. In the first situation the anxiety symptoms are seen as secondarily to depression and therefore they will clear with the improvement of the depression. The effect is therefore a part of the antidepressant effect and no claims are expected; in the second situation patients have to be diagnosed correctly and it has to be shown that the product improves anxiety independent of the effect on depression

Additional studies would be required to support a claim in psychotic depression or other types of depression.

Major depression may be further classified as mild, moderate and severe. Clinical trials will usually recruit patients, who are moderately ill, as it is difficult to demonstrate an effect in mildly ill patients. Demonstration of an acceptable benefit/risk in moderately ill patients will be considered sufficient for a registration package to get a license for “Episodes of Major

Depression". Additional studies, especially in a group of severely ill patients, when well defined, may lead to additional claims. The definition should be justified

As mentioned in the introduction a Major Depressive episode may also occur in the framework of a Bipolar Disorder. In general the development of a product in this patient group will be the same as for unipolar depression. However, there are some specific issues, like switching rates, which are addressed in the guideline on Bipolar Disorder.

2.4 Therapy resistant patients

Though a number of antidepressant agents of different pharmacological classes are accepted and used widely in the treatment of Major Depression, their efficacy is insufficient in about 30% of the patients. There are various treatment algorithms for depressed patients, but usually a patient is considered therapy resistant when consecutive treatment with two products of different classes, used for a sufficient length of time at an adequate dose, fail to induce an acceptable effect. What happens then depends on the algorithm used.

At the moment no specific treatment is approved for these patientsThough specific guidance can not be given at present, the following should be considered: therapy resistance has to be well defined and it should be shown prospectively, that the patients are indeed therapy resistant. Non-compliance for instance should not be the reason for insufficient effect. The duration of the trials should be sufficiently long to show an effect and to show that the effect is maintained. The endpoints should be relevant to this patient group, e.g., remission may be more important than a mean change on a scale. The pivotal trials should include placebo and/or an acceptable comparator. The choice of the comparator(s) has to be justified. Consideration should be given to the product that was insufficiently effective and other available treatment modalities.

2.5 Elderly

Depression in the elderly is not uncommon, but certainly not all elderly patients with depressive symptoms will have Major Depression. In ICH E7 it is indicated that the efficacy and safety for the elder population can be derived from the total database, unless there are specific reasons not to do this.

Recently studies have been conducted in the elderly, that could not distinguish between test product and placebo, even though the design of the studies and the dose of the test product were as expected and efficacy of the product had already been shown in adults

Moreover extrapolation of the adult dose may be difficult due to pharmacokinetic properties of the product and/or to a different sensitivity in the elderly for the pharmacodynamics of the product.

Therefore not only efficacy, but defining a safe dose (range) in these patients is a main concern. Usually this should be addressed before licensing.

In principle two approaches are possible. One is an analysis of the whole database, whereas the other would be to conduct specific trials in a specified patient population. The optimal design would be a placebo-controlled dose response study.

The first approach may be accepted as pivotal information for agents of known pharmacological classes, provided that sufficient elderly patients are included to allow a prospective subgroup analysis. As both efficacy and the optimal dose should be addressed, this may be difficult. Specific studies will be more informative and are preferred.

For new products with a new mechanism of action specific trials are needed.

In both situations pharmacokinetic studies may support the choice of the dose and should be conducted.

2.6 Children

Research of effective pharmacological treatment in depression in children and adolescents has increased over recent years, but experience is still limited.

Trials of a product in this area should differentiate between children and adolescents and specific trials for both age categories are needed, as presentations of symptoms and natural course may be different. Rating scales should be specific for and validated in the age groups. Development of the product solely in adolescents could also be considered, but the results would not be generalisable to younger children. Difference in impact of adverse effects, seen in adults, in children and adolescents should be considered. In line with the paediatric guideline (ICH E11), trials may be conducted after a licence for adults is obtained

Moreover, in line with the relevant guideline, effects on cognition, learning, development, growth and endocrine functions should be addressed,: cognition and learning should be studied pre-licensing using recognised tests, validated for the age and patient group. Also the direct effect on endocrine functions in adolecents should be studied before licensing. Long-term effects on learning, development, growth and sexual function may be studied post-marketing, but appropriate protocols should be available when the use in children is applied for.

3. ASSESSMENT OF EFFICACY CRITERIA

Results should be discussed in terms of both clinical relevance and statistical significance. When a statistically significant effect is found and it has been shown that the effect is robust and insensitive to the analysis used, this effect has to be addressed in clinical terms, depending on the purpose of the trial. It should be noticed that the relevance of the effect is a basis for the benefit/risk assessment. The sample size of the studies should take this into account (see 5.2) Multiple studies will ameliorate the problem of the unreliability of studies in Major Depression.

3.1 Short-term trials

Improvement should be documented as the difference between baseline and post-treatment score in symptomatology, but should also be expressed as the proportion of responders. In Major Depression a 50% improvement on the usual rating scales is accepted as a clinically relevant response. Other definitions of responder may be used, e.g., remission in mildly depressed patients, but these have to be justified in the trial protocol

Remission is defined as a condition where no or only few signs of illness remain; the cut-off on a validated rating-scale has to be defined in the protocol and should be justified

3.2 Long-term trials

In randomised withdrawal trials, efficacy usually is expressed as number of patients worsening (relapsing) and/or time to this event. Both efficacy criteria are of interest and should be submitted. The choice of one of them as primary and the relevance in clinical terms will depend on the type of patients included and the purpose of the trial and have to be justified in the protocol. The analysis should carefully consider the possible biases arising from drop-outs and the statistical methods of dealing with them.

Worsening or relapse has to be defined in the protocol and should be clinically relevant. Usually a clinically relevant increase in symptoms, scored on a validated rating scale at one or more visits is used.

Also in the case of prevention of recurrence, recurrence has to be defined in the protocol. Usually recurrence will include reappearance of clinically relevant depressive symptoms, scored on a validated scale.

4. METHODS TO ASSESS EFFICACY

Efficacy will be assessed by rating scales. The choice of rating scales should be justified from the test quality criteria (reliability, validity) and the sensitivity for change should be known. For the assessment of improvement specifically developed rating instruments are necessary.

Acceptable scales for use as primary endpoint to determine symptomatic improvement include the Hamilton Rating Scale of Depression, preferably the 17 item scale, and the Montgomery Asberg Depression Rating Scale. The protocol should indicate which scale is used as primary variable.

In addition global assessment (e.g., item 2 of the Clinical Global assessment scale) may be used as a secondary endpoint.

5. DESIGN FEATURES

5.1 Study population

The disorder should be classified according to an internationally acknowledged classification system, preferably DSM IV, using the diagnostic criteria herein. ICD 10 may also be used if additional information on the actual type of depression is provided. The same classification system should be used for the whole development of the medicinal product. A rating scale alone is insufficient and is not equivalent to a diagnosis.

Further descriptive parameters, like severity of the episode, as well as a detailed history, e.g., duration of the depression and of the index episode, number of episodes per time interval, previous treatment outcome, should be documented.

In addition cut-off scores, based on an appropriate scale may be used as inclusion criteria.

It is highly desirable that the study population is homogenous with respect to the indication for the dose finding and pivotal studies (see also section 2.3)

Though some of the earlier studies may be done in hospitalised patients, the majority of the database will be in out-patients.

When prevention of recurrence is studied, patients should be included, who fulfil the criteria of major depression (recurrent) in sustained remission. To distinguish between relapse and recurrence, patients should be stable and off medication for a predefined period in case of Depressive disorder. For prevention of recurrence in patients with Bipolar disorder, the relevant guideline should be followed.

5.2 Study design

In principle, to assess the effect of medicinal products parallel, double blind, randomised placebo controlled trials are necessary (see also section 2.1). In addition, comparison with a standard product in an adequate dose is generally needed. The dose and the comparator should be justified.

Generally a placebo wash-out period to exclude placebo responders is not useful and may impair generalisation of the results. When patients are already treated with an antidepressant agent, a wash-out period may be necessary. Any reason to exclude placebo-responders should be discussed.

The sample size should be justified, using clinical (responders) and biostatistical criteria. Statistical analysis should include various analyses, among others intention-to-treat and per protocol. The handling of dropouts and missing data should be prospectively planned in the trial protocol. The risk of under- or overestimation of the effect should be addressed. See further the biostatistical guideline.

Investigators should be properly trained in evaluating the patient. Inter-rater reliability scores (kappa) should be documented for each investigator in advance and if necessary during the study, both with regard to the diagnosis and to rating scales used for efficacy and/or safety, where relevant.

Prior and concomitant medication has to be documented in detail. Relevant medication has to be washed out. If appropriate, rescue medication should be provided.

If anxiolytic or hypnotic medication can not be avoided in the beginning of treatment, stratification may be useful and the effect on the treatment effect should be analysed.

Drop-outs and reasons for drop-out have to be documented in detail. Moreover information concerning time to drop out should be provided. In addition information of patients screened but not included in the study should be documented.

If necessary, standardised psychotherapy, psycho-education, support or counselling may be given as supplementary treatment, though it may enhance the placebo effect, but it should be prospectively defined in the protocol. It should be documented in detail and its effect on treatment effect should be analysed. Potential centre effects should be evaluated carefully.

6. STRATEGY

6.1 Pharmacodynamics

A variety of tests can be performed, but there is no specific model in humans for depression. Studies in cognition, reaction time or on sleep architecture may be informative concerning the side effect pattern of the product.

6.2 Pharmacokinetics

The usual pharmacokinetic studies should be performed. (see guideline on pharmacokinetic studies in man). Especially in dose response studies individual plasma levels may be studied.

6.3 Interaction studies

In general the guideline on Interaction should be followed to investigate possible pharmacokinetic and pharmacodynamic interactions. Concerning the latter, interactions with alcohol and other CNS active medicinal products should be investigated.

6.4 Short-term trials

Controlled, parallel fixed dose studies, using at least 3 dosages to establish are needed to establish as far as possible the lower end of the clinical effective dose range as well as the optimal dose. Generally it is useful to add a placebo arm as well as an active comparator.

The dossier should also include parallel group studies against placebo and/or active comparator. Three-arm or multi-arm studies are preferred, as the trial will be internally validated therefore addressing the problem of assay sensitivity. The aim of the study may be superiority over placebo or active comparator or demonstration of at least a similar balance between benefit and risk of the test product in comparison with an acknowledged standard antidepressant agent.

The duration of these trials usually is around 6 weeks. (see section 2.1)

6.5 Long-term trials

Due to the character of the disorder, longer double blind trials are necessary to demonstrate that the effect is maintained during an episode. Design and duration of these trials are discussed in section 2.2.

Studies demonstrating prevention of a new episode are not required for licensing, though of major interest. (see section 2.2)

The usefulness of including more than one dose of the test product to investigate the optimal dose for long-term treatment should be considered.

7. SAFETY ASPECTS

Identified adverse events should be characterised in relation to the duration of treatment the dose and/or plasma level, the recovery time, age and other relevant variables. Clinical observations should be supplemented by appropriate laboratory tests and cardiological recordings. Specific rating scales are preferred where relevant.

All adverse events occurring during the course of clinical trials should be fully documented with separate analysis of adverse drug reactions, drop-outs and patients who died while on therapy. Specific attention should be paid to the possibility of suicides and where relevant Serotonergic syndrome.

Any information available concerning clinical features and therapeutic measures in accidental overdose or deliberate self-poisoning should be provided.

Special efforts should be made to assess potential adverse effects that are characteristic of the class of products being investigated depending on the action on various receptor sites, like serotonergic, dopaminergic, or noradrenergic receptors or anticholinergic or antihistaminergic effects.

7.1 Rebound /Withdrawal phenomena / Dependence

When pharmacological treatment is stopped, rebound and/or withdrawal phenomena may occur. Trials should be designed in such a way, that these phenomena can be studied. In some of the short-term and long-term clinical trials, treatment should be stopped abruptly and patients should be followed for a suitable duration. Occurrence of rebound and/or withdrawal phenomena should be scored at the appropriate time.

Animal studies will be needed to investigate the possibility of dependence in new classes of compounds or when there is a indication that dependence may occur,

Depending on the results of these studies further studies in humans may be needed.

7.2 Psychiatric adverse reactions

Depending on the class and the interactions with various receptor it should be needed to study effects on cognition, reaction time and/or driving and the extent of sedation. Specific claims always have to be based on specific studies.

7.3 Haematological adverse reactions

Special attention should be paid to incidence of neutropenia. Agranulocytosis and aplastic anaemia.

7.4 Cardiovascular adverse reactions

Cardiovascular events like hypertension, orthostatic hypotension and effects on the heart, e.g. QT-interval should be investigated.

7.5 Sexual dysfunction

Special attention should be paid to the effect on sexual function and libido.

7.6 Long-term safety

The total clinical experience should generally include data on a large and representative group of patients in line with the guideline on population exposure.