
Consensus Meeting

ECNP Consensus Meeting March 2000 Nice

Guidelines for Investigating Efficacy in Bipolar Disorder

Chairpersons: Stuart Montgomery, Barbara van Zwieten-Boot

Committee: J. Angst, C.L. Bowden, J.R. Calabrese, R. Chengappa, G. Goodwin, Y. Lecrubier, R. Licht, W.A. Nolen, G. Sachs, A. Saint Raymond, J. Storosum, P. Suppes, J.M. van Ree

Rapporteur, D.B. Montgomery

The guidelines on investigating the efficacy of treatments for bipolar disorder presented here are the product of a consensus meeting of experts on the disorder held under the auspices of the European College of Neuropsychopharmacology in Nice in March 2000. The guidelines will need to be interpreted and possibly modified in the light of new data. The guidelines are developed principally for the investigation of pharmacological treatments but the same principles apply to nonpharmacological treatments.

1. Introduction

Bipolar disorder was first defined as an illness by Falret in 1851 and 1854. His longitudinal observations led him to propose the condition “folie circulaire” defined by manic and melancholic episodes separated by asymptomatic periods. Bipolar disorder is a recurrent condition where depressive episodes and episodes of abnormally and persistently elevated, expansive, or irritable mood occur in the same patient, sometimes at the same time. It is a serious condition that has a disruptive effect not only on the lives of those with the disorder but also those near to them.

Bipolar disorder is a common disorder with reported lifetime prevalence rates in Bipolar I and Bipolar II between 0.4% and 3.7%, as reported in a recent review by Angst (1998). In Bipolar I disorder as defined in DSMIV (American Psychiatric Association, 1994), characterised by the presence of at least a single manic episode, with or without depressive episodes, the distribution is approxi-

mately equal between the sexes; in Bipolar II disorder, where, besides major depression hypomanic episodes but never full manic episodes occur, the incidence is higher in women.

The lifetime occurrence of manic or hypomanic episodes is the defining feature of the disorder and it is therefore distinct from unipolar depression where the recurrence of illness is always a further episode of depression. If a single episode of mania or hypomania supervenes in a history of recurrent unipolar depression that is not attributable to other factors such as antidepressant use the diagnosis automatically changes from unipolar depression to bipolar disorder.

Compared with unipolar depression bipolar episodes and periods of remission tend to be shorter, residual states and chronicity more frequent, and high levels of comorbidity with other disorders are frequent. The cycling tends to become more frequent over the first three to four episodes and some individuals develop rapid cycling, defined by at least four episodes per year, a condition that seems to be more resistant to conventional prophylaxis.

Bipolar disorder and unipolar major depression are both disabling conditions. However, bipolar disorder is thought to be the more disruptive because of its earlier onset, with a peak in the late teenage years, and more frequent cycles. The rapid and unpredictable shifts of mood are particularly damaging in a social or occupational context and this is reflected in a divorce rate three times higher than the normal population. The poor judgement associated with mania or hypomania leads to reckless and impulsive acts which may have lasting detrimental consequences for the individual, their family, and friends (Montgomery and Cassano, 1996). The disorder is associated with a high suicide rate (Dilsaver et al., 1994) and has one of the highest rates of associated substance abuse of the major psychiatric disorders.

Bipolar disorder is a complex condition involving different targets for treatment, which require separate investigation. Investigation of the efficacy of treatments has to address the treatment of manic or hypomanic

episodes, depressive episodes, and mixed states, and also the long term prophylaxis, or mood stabilisation, of bipolar disorder.

2. Trial design

2.1. Diagnosis

In DSMIV, bipolar disorder includes Bipolar I, which requires the lifetime presence of at least one manic episode, and Bipolar II, which is characterised by major depressive episodes and at least one hypomanic episode but no full manic episodes. A mixed episode needs to meet criteria for both manic episode and major depressive episode nearly every day for at least one week. These guidelines will not focus on cyclothymic disorder, which is characterised by a two year period (or in the case of children a one year period) of frequent hypomanic symptoms and depressive symptoms that do not meet the criteria of major depressive episode, or bipolar disorder not otherwise specified.

The DSMIV definitions for manic episode, and major depressive episode are clear and have been used in sufficient studies for investigators to be confident in their appropriateness to identify these conditions; there is less consensus regarding the use and study of patients with mixed episodes.

In most respects the ICD10 criteria for Bipolar Affective Disorder (World Health Organisation, 1992) are in agreement with DSMIV though both type and severity of presentation define the categories. The ICD10 classification excludes single episodes of mania or hypomania, which are categorised separately until there is a recurrence of some kind, and also do not have a separate categorisation for Bipolar II except under the rubric Other Bipolar Affective Disorders. Like the DSM system the ICD10 research criteria are operationally defined though they have not yet been used in sufficient studies to validate their use for efficacy studies.

The separation of Bipolar I and Bipolar II in the DSMIV system, which mostly depends on determining a difference between a manic and a hypomanic episode, may not always be easy. Distinguishing clinically between mania and hypomania is often relatively difficult and some of the suggested criteria for establishing boundaries, for example the need for hospitalisation, are affected by variations in practice and availability of hospital beds both between and within different countries. The same number of symptoms determines the diagnosis of a manic or hypomanic episode, the distinction depending on a judgement of the duration and degree of impairment of occupational or social functioning. In hypomania the change in behaviour must be "obvious to others" and in mania the impairment must be marked and last a week, or cause hospitalisation. The separation between mania and hypomania, which is a close

judgement in some cases, may be even more difficult in retrospect with the passage of time.

The diagnosis of bipolar disorder is dependent on the judgement of the lifetime presence of at least a single manic or hypomanic episode. This judgement is often made retrospectively and carries the risk of distortions of memory. Therefore the diagnosis needs to be confirmed by another source such as a family member or the history in the medical records.

Features that can help to differentiate bipolar disorder from unipolar depression include a frequent positive family history of bipolar disorder and also an early age of onset, though rarely before the age of 12, with a peak in the late teenage years and early adulthood. Nevertheless when assessing early episodes of recurrent depression it is not possible to know which individuals will continue to suffer from unipolar depression or will subsequently develop bipolar disorder.

A small proportion of patients with manic episodes have psychotic symptoms. The presence of these symptoms can lead to confusion in the diagnosis between schizophrenia and bipolar disorder. It is important to exclude schizophrenia. It is possible that a treatment might be effective in ameliorating the psychotic symptoms without necessarily exerting an effect on the manic symptoms. If psychotic symptoms are present they should be documented and appropriate scales used to separately identify changes in symptomatology.

2.2. Severity

In general the DSMIV diagnostic criteria have a requirement for significant distress or impairment. For Bipolar I disorder this reflects the manic episodes. However for Bipolar II the criteria for a hypomanic episode require only an observable change in behaviour without marked impairment in function while the distress or impairment is related to the major depressive episodes. The studies carried out in bipolar patients with less severe manic symptoms have not always been able to establish efficacy so that a minimum severity criterion may be considered for those entering studies in acute mania. For similar reasons mild depression is often excluded from studies of major depression by the use of a minimum severity entry criterion and this would clearly be needed in studies in bipolar depression.

The aim for investigations of efficacy would be to identify a patient population suffering from moderate to high moderate severity of the disorder, but allowing for generalisation to a wider population. In studies of acute mania a minimum score at entry to the study of 18 and 20 on the Young Mania Rating Scale (YMRS) (Young et al., 1978) have both been used to select a group of patients in whom it was possible to demonstrate a significant difference between an active drug and placebo. A minimum score of 16 on the Manic Rating Scale items of the Schedule for Affective Disorders and Schizophrenia –

Change Version (SADS-C) (Endicott and Spitzer, 1978) has also been used successfully. On the Bech-Rafaelsen Mania Scale (BRMAS) (Bech et al., 1978) a score typically higher than 15 has been used. Minimum entry scores for studies in major depression have been well established in many studies. A minimum score of 22 is generally accepted for the Montgomery & Asberg Depression Rating Scale (MADRS) (Montgomery and Asberg, 1979) and 18 to 20 for the Hamilton Depression Rating Scale (17 item) (HAMD) (Hamilton, 1967). These minimum severity entry criteria would serve equally for studies in bipolar depression.

2.3. Comparator group

The most scientifically rigorous demonstration of efficacy for monotherapy is derived from positive results in well-conducted, randomised double-blind placebo-controlled studies in separate acute and long-term treatment investigations. Despite the practical difficulties it has proved possible to demonstrate efficacy in short term acute phase treatment of mania in placebo controlled studies. The same applies to the treatment of depression where well-defined methodologies have been developed. The efficacy of long term prophylaxis of bipolar disorder has also been established in placebo-controlled trials but none in the last 25 years and none using currently acceptable methodology.

The use of an active comparator in placebo-controlled trials is optimal in order to validate the study population tested and to allow an assessment of clinical relevance. The most frequently used active comparator has been lithium, although there is only one placebo-controlled study in acute mania with lithium as comparator that employed current methodology, with random assignment to parallel arms (Bowden et al., 1994).

When evaluating the acute antimanic efficacy of an antipsychotic drug the choice of active comparator is more complicated. Lithium has been proposed as a reference comparator for new antipsychotic drugs in acute mania but it would seem rational to use a comparator of similar mechanism, such as haloperidol, although this drug has not been evaluated in acute mania studies employing current methodology despite being widely used as an antimanic agent and licensed as such in some countries. The ideal solution would be to use both type of comparator in the trial programme.

The demonstration of superior efficacy of an agent against a comparator standard treatment can also be used to establish efficacy. This requires careful safeguards to ensure that a fair comparison is made. The population studied must be appropriate and the choice of comparator and the dose used has to be justified. Failure on previous treatments, which treatments were unsuccessful, and the degree of treatment resistance, should also be documented.

The finding of similar efficacy to standard treatments in

studies that did not include a placebo control is not, however, regarded as providing sufficiently rigorous evidence to establish efficacy since the underlying placebo response rate can be substantial and varies across and within studies.

2.4. Assessment scales

An important consideration in assessing treatments for bipolar disorder is the possibility of provocation of a swing to the other pole during treatment of an acute episode. Studies investigating the efficacy of treatment in bipolar disorder therefore need to include scales to assess the severity of symptoms of both mania and depression regardless of whether the population under investigation is primarily suffering from acute mania, a mixed episode, or acute bipolar depression. In studies of long-term treatment both types of scales should also be included as primary measures.

The pivotal scales for either pole needs to be defined in advance and preferably the same scales should be used in studies through the full development programme so that analysis of the data in larger numbers is possible.

The choice of scales to assess severity should be based on documented sensitivity to treatment change and ability to detect drug placebo differences. The scales need to be valid and internationally recognised.

The scales considered appropriate for rating acute mania are mostly observer rated. Self rating scales are not thought to be useful for studies of mania or hypomania because the insight of the individual is frequently compromised and they fail to acknowledge their symptoms.

The studies that have been able to establish efficacy in the acute treatment of mania have tended to use the Young Mania Rating Scale (YMRS), the SADS-C mania subscale, and the Bech-Rafaelsen Mania Scale (BRMAS). These scales are clearly internationally recognised, valid and reliable.

The severity of depression during treatment is usually measured on either the Montgomery Asberg Depression Rating Scale or the Hamilton Depression Rating Scale, both of which have been shown to establish antidepressant efficacy in pivotal placebo-controlled studies. There is evidence that the MADRS may be more sensitive than the HAMD in measuring the change in symptoms in bipolar depression (Calabrese personal communication). The same scales are appropriate to rate deterioration and the re-emergence of depression in studies of long term treatment.

It is helpful in both mania and depression to include additionally a Clinician Global Impression (CGI) rating of severity and change to register the global state as distinct from changes measured on the disorder specific scales. For bipolar disorder a specific version of the CGI has been developed and validated, the CGI-BP (Spearing et al., 1997). A measure of change in the functional aspects of the disorder, such as the Sheehan Disability Scale

(Sheehan et al., 1996), may be a useful secondary measure. Mood charts such as the Life Chart Method may provide valuable extra information as a clinician aid but are not recommended as an efficacy measure.

2.5. Clinical relevance

A significant difference on the pivotal scales needs to be judged in relation to measures of clinical relevance. The most commonly used criterion of response is the 50% reduction on the pivotal efficacy scale whether in acute mania or depression, which is regarded as clinically relevant. A statistically significant difference on a clinically relevant outcome measure has been accepted as a measure of a clinically relevant response. Similarly a significant difference on a global measure such as the CGI has also been used as a clinically relevant measure. However, for long term studies and studies to assess prophylactic efficacy other tactics, such as expected maxima on specific scales, may be needed.

Comparison of the rates of withdrawals from treatment, or overall survival, in placebo-controlled studies because of lack of efficacy also provides an assessment of clinical relevance. In a placebo controlled study both clinicians and their patients have the right to withdraw from the study because of unsatisfactory clinical response. These withdrawals, defined beforehand, contribute to the judgement of both efficacy and the clinical relevance of the observed drug placebo difference. The most sensitive method for capturing these data would be survival analysis with withdrawals for efficacy reasons representing the event.

Significant differences on functional scales used as secondary measures may also be helpful in making a judgement that the significant difference on the pivotal scales is also clinically relevant.

3. Acute treatment and relapse prevention

For convenience the efficacy of acute treatments in bipolar disorder need to be considered separately for mania, or the less severe state hypomania, and bipolar depression.

In Europe the demonstration of long-term efficacy is a necessary part of establishing the overall efficacy of treatment in all chronic disorders where treatment is likely to continue in the long term and this is supported by regulatory advice. Thus treatments for bipolar disorder are required to demonstrate both acute and long-term efficacy before being licensed.

The term long-term treatment covers a period of continuation treatment for relapse prevention following response of the acute episode to establish that efficacy is maintained; it also covers the concept of prophylaxis or prevention of the recurrence of new episodes. The compli-

cation in bipolar disorder is that symptoms or episodes of either pole may arise and these need to be documented.

Different groups have used some terms relating to treatment and response to denote diverse aspects. These guidelines adopt the definitions used by the CPMP Note for Guidance on clinical investigation of medicinal products for the treatment of Bipolar Disorders.

Response: clinically relevant improvement (defined in advance)

Responder: patient with clinically relevant predefined improvement

Maintenance of effect: the effect of treatment, seen in the short-term is maintained during the whole episode.

Relapse: an increase in symptomatology immediately or almost immediately after medication is stopped. It usually indicates the treatment duration was too short.

Recurrence: a re-emergence of symptoms (new episode) after a time with no or minimal symptoms.

Mood stabiliser: a compound that has been shown to be efficacious in recurrence prevention of manic and depressive symptoms. A mood stabiliser may not necessarily be a good agent for improving manic or depressive symptoms although it would be an advantage if it were.

3.1. Acute mania or hypomania

The separation between Bipolar I and Bipolar II is not sufficiently clear cut to mandate the separate development and testing of treatments for the two subcategories. The evidence supporting the efficacy of treatments for mania indicates that these treatments are also effective in hypomania. There is no evidence to suggest that effective treatments in the more severe mania or the manic symptoms in mixed episodes are not also effective for the treatment of lesser states. It therefore seems reasonable to establish the efficacy of an agent in mania and to extrapolate from the results to assume efficacy in hypomania. However, the reverse cannot be assumed. A treatment shown to be effective for hypomania will not necessarily be effective in treating manic episodes.

The serious nature of mania puts certain constraints on trial design for efficacy studies. The recommended design is a double-blind randomised parallel group comparison with placebo but the rate of early discontinuations on placebo is expected to be high. Studies therefore need to be of short duration and analysis should take the drop outs due to lack of efficacy into account. One commonly used method is to use the intention to treat (ITT) last observation carried forward (LOCF) method of analysis. Studies with a duration of three weeks have been able to demonstrate reliably a significant advantage of drug compared with placebo despite an uncomfortably high number of dropouts.

Three weeks is considered a sufficient length of time in studies to demonstrate a significant drug placebo difference in acute mania. However the definition of a patient as a

responder at three weeks may be premature, as it would miss some cases. A further period of treatment of the episode of mania with the agent would seem to be desirable to allow full response or remission. This may not occur until 8 to 12 weeks or in some cases up to 16 weeks. There are few data to support exactly how long the period of acute treatment should be before a reliable judgement of remission may be made. However, it seems reasonable to suppose that a judgement could be made at 8 or preferably 12 weeks after the start of acute treatment.

Some treatments are currently licensed for the treatment of acute mania for which there is no evidence of their usefulness in the general treatment of bipolar disorder. There is some evidence, not from placebo-controlled studies, to suggest that, for example, haloperidol may alleviate the manic symptoms but increase the depressive symptoms (Kukopulos et al., 1980). New treatments for acute mania (pure or mixed) should demonstrate that they do not precipitate or exacerbate depressive symptoms, nor cause switching to depression. While it is possible to envisage new treatments for mania that are effective only for the acute episode, the risk of subsequent depressive and manic symptoms in these cases needs to be quantified in controlled studies.

3.2. Continued response in acute mania

An effective agent for acute mania should also establish that the effect is maintained and that it is effective in preventing, or minimizing the likelihood of, the early return of symptoms of mania. The three-week duration is now established as the accepted period to demonstrate efficacy in placebo-controlled studies. A further period of continuation treatment to cover the natural course of the episode and ascertain sustained response is necessary. Three weeks is considered too early to serve as the point at which long-term treatment studies to investigate efficacy in preventing the recurrence of new episodes should commence. It is not entirely clear what the duration of this continuation or stabilisation should be but after 12 weeks it would appear that stabilisation of the response has largely been achieved. It is useful to focus not only on percentage responders but also on the percentage of patients in remission.

The preferred design to establish that an effective agent in the treatment of the acute episode maintains efficacy through the episode is the placebo-controlled discontinuation of treatment in responders to acute treatment with the agent and comparison of the subsequent relapse rates. The usual methodology is to compare the cumulative relapse rate on drug and placebo using survival analysis. The definitions for response to acute treatment and for relapse should be predefined. One proposal for acute mania is to continue the treatment openly and to include in long-term prophylactic treatment studies only the responders at three weeks whose response is sustained to 8 or 12 weeks. This

period between 3 and 12 weeks is too short to test efficacy in relapse prevention separately from the three-week acute treatment period.

A study comparing the efficacy of a new antimanic agent with an established comparator over the acute 3 week period and the sustained period to 8, or to 12 weeks will provide reassurance that the level of protocol-defined response and protocol-defined remission is similar or at least not different. In the absence of a placebo control formal testing of efficacy over this period is not possible.

3.3. Acute treatment of bipolar depression

Current evidence and clinical experience suggests that effective treatments for unipolar major depression are also effective in treating episodes of bipolar depression (Cohn et al., 1989). However, the reverse is not necessarily true and it is possible that a treatment demonstrated to be effective in bipolar depression may not be effective in unipolar depression. For example lithium is thought to be effective in the treatment of bipolar depression, although the studies are very limited, but it is not regarded as effective in unipolar depression.

In bipolar depression, there is an associated risk of provoking a switch from the depression into mania. Some switches to hypomania, mania, and mixed states in bipolar disorder are expected to occur naturally on placebo. In a seven-week study of patients with Bipolar I depression (excluding rapid cyclers) a switch rate of 5.4% has been reported (Calabrese et al., 1999). It is thought that antidepressants may vary in their propensity to accelerate switching to mania: the TCAs have the highest reported switch rates up to 50% and SSRIs and bupropion the least, with rates below 5%. Effective treatments in unipolar depression therefore need to be tested for their potential to increase swings to mania. It is important to demonstrate that a treatment is effective in monotherapy and that it is associated with a relatively low risk of switches to mania in the treatment of bipolar depression. Switches during combination treatment are difficult to interpret. Finally care should be taken to distinguish between switches to mania or hypomania and simple improvement in depression.

The evaluation of a drug intended for use in the acute treatment of bipolar depression needs to be made in parallel group, randomised double blind comparisons against placebo and three way studies including the agent, placebo and an active comparator are preferred. The choice of active comparator needs to be justified. Some consider that lithium should be an active comparator but opinion on its antidepressant efficacy is divided and others hold the view that the appropriate comparator would be an antidepressant associated with a low rate of switches to mania or hypomania. The ideal solution would be to include lithium in one study and an active antidepressant comparator in another.

The study period for investigations of treatment for acute bipolar depression should be in line with the studies in unipolar depression, which are generally of 6 weeks' or sometimes 8 weeks' duration.

3.4. Continued response in bipolar depression

Treatments that have been shown to be effective in the acute treatment of unipolar depression are required before being granted a licence in the EU to demonstrate that the effect is maintained in long-term treatment for 6 months. The preferred evidence comes from the classic relapse prevention study where responders, or preferably remitters, to acute treatment are randomised to substitute placebo, with appropriate tapering of the active drug, or to continue with the same treatment. During the six month period the rate of relapses are compared using survival analytic techniques.

Evidence of efficacy in relapse prevention is normally considered sufficient demonstration of long term efficacy to license treatments for unipolar major depression and a similar demonstration of long term efficacy will be sufficient for bipolar depression. There is some suggestion that the continuation phase may be shorter in bipolar disorder than in unipolar depression and therefore the evidence of efficacy in this phase may be derived from placebo-controlled studies lasting between 3 and 6 months.

Responders to acute treatment with the agent at 6 to 8 weeks who meet predefined remission criteria, eg a score of 12 or less on the MADRS, should be randomised to continue on the agent or placebo for a period of 3 to 6 months. Relapse criteria need to be predefined, eg a score of 18 or more on the MADRS or 16 or more on the HAMD have been used successfully, as has the withdrawal from the placebo-controlled study for reasons of clinical deterioration.

The definitions used in the study for remission and relapse of depressive symptoms need to be defined in advance. Criteria for deterioration reflected in the appearance of symptoms of mania or hypomania also need to be predefined and the studies in bipolar depression need to include the relevant assessment scales for mania as well as for depression.

Some evidence of efficacy in continued treatment may be adduced from extension data from double blind placebo-controlled studies in acute treatment where the responders are continued under double blind conditions with the same treatment. The rationale for this design is the demonstration that the relapse rates in placebo responders continued on placebo treatment is very similar to the relapse rates seen in responders to antidepressant treatment subsequently treated with placebo (Montgomery et al., 1993; Anton et al., 1994). The design has the advantage of being unaffected by potential discontinuation effects but the patients continuing treatment may not be the same in the various groups.

A certain number of breakthrough episodes of mania can be expected during a six month treatment period in bipolar disorder patients even when treated with established mood stabilisers. The investigation of the potential switch rate to mania or hypomania of treatments for bipolar depression should therefore ideally be carried out in controlled comparisons with established mood stabilisers in monotherapy in studies of sufficient power in order to establish a low relative rate of switches to mania. In the absence of a placebo control formal testing of efficacy over this period would not be possible.

3.5. Treatment of mixed states

Where an individual meets full criteria for both an episode of mania and an episode of major depression at the same time the appropriate treatment will need to be effective in resolving the symptoms of both simultaneously. There are as yet no placebo-controlled studies that have evaluated efficacy in a prospectively defined cohort of patients presenting in the mixed phase of bipolar disorder. As well as showing efficacy a treatment for mixed states will need to show that it will not exacerbate the symptoms at either pole. Thus a treatment that provokes depression or manic switches is unlikely to be helpful in the acute treatment of mixed states.

There is some evidence that those suffering from mania accompanied by depressive symptoms that do not reach full depressive syndrome, referred to as dysphoric mania, have a different response to treatment. For example in the study of Bowden et al. (1994) carried out in this population, valproate was effective and lithium not. This suggests that those with mania who have some depressive symptoms should be investigated separately in subgroup analyses to identify possible pharmacological differences. The number of such depressive symptoms is not clear cut but analyses of the group showing two or three core symptoms have been undertaken.

4. Mood stabilization and recurrence prevention in mania and bipolar depression

Establishing efficacy in preventing new episodes of mania or depression requires prophylactic studies in stabilised patients. Establishing long-term efficacy in bipolar disorder is conceptually more complicated than, for example, unipolar depression. Patients who have a return of symptoms after being treated successfully for an acute episode of mania (or hypomania), or of depression may suffer a recurrence showing either the same type of symptoms, or symptoms of the opposite pole, or both.

The purpose of mood stabilisers is to prevent the appearance of symptoms or episodes of either pole. While it is possible but unlikely for bipolar disorder to manifest itself in recurrent manic episodes alone, a mood stabiliser

for long-term treatment should ideally have evidence of long-term efficacy in preventing both depressive and manic (or hypomanic) episodes (Calabrese and Rappaport, 1999). An alternative suggestion is that a mood stabiliser is a medication that possesses efficacy in one phase of the illness without causing a negative effect on other phases of the illness (Bowden, 1998).

The studies to establish long-term prophylactic efficacy should preferably be placebo-controlled since this provides the most unequivocal evidence of efficacy. The return of both depressive symptoms and manic symptoms needs to be measured in order to judge whether the agent was able to protect against the return of symptoms of either pole.

The favoured design is placebo-controlled discontinuation of treatment in patients who have responded to treatment of the acute episode, defined preferably by an absolute score on a severity rating scale, whose response has been stabilised. The number of patients showing symptoms of a new episode of either depression or mania, or hypomania to a predefined criterion is compared. The preferred methodology is to compare the cumulative relapse rate on drug and placebo using survival analysis.

It would be clinically inappropriate to use a criterion for recurrence that required the development of the full syndrome since most clinicians and patients would withdraw from a placebo-controlled study before this was reached in order to intervene with treatment of the impending episode. There are insufficient data to support more precise recommendations on where the level of deterioration should be set for mania, but deterioration to a score of 12 on the YMRS in a remitted population has been suggested. In the studies of the long term treatment of major depression the withdrawal of the patient from the study due to concerns about lack of efficacy is often the most sensitive definition of deterioration. Other criteria that have proved useful are a score of 18 or more on the MADRS or 16 or more on the HAMD. The use of these criteria for deterioration of depression is recommended for studies in bipolar disorder.

To establish unequivocal efficacy evidence is needed from long-term treatment studies that the proposed treatment is able to prevent the emergence of depressive and manic episodes compared with placebo. The size and length of such studies depend on the potential recurrence rates of the population studied. Power calculations suggest that longer studies are more likely to test efficacy and depending on the morbidity of the population a difference between drug and placebo is unlikely to be seen in less than a year.

Placebo-controlled trials of this length are difficult from a practical point of view. The alternative to a placebo-controlled study, which also produces good evidence of efficacy, is the demonstration of significant superiority of a new agent relative to an established mood stabiliser. Evidence of significant superiority of efficacy compared to an established mood stabiliser could be accepted as being

at least as strong as a significant superiority to placebo. However, based on current knowledge this is an unrealistic goal. Establishing equivalent efficacy to a comparator is methodologically difficult and is unlikely to be persuasive.

5. Combination treatment

In establishing the efficacy of a new agent the demonstration of the response of monotherapy compared to placebo takes precedence. Where monotherapy has already been established as effective in well-conducted, placebo-controlled studies it may be useful to examine whether the combination of the new agent with an established mood stabiliser in placebo-controlled add-on designs might provide additional therapeutic advantages. For example, where breakthrough occurs in spite of prophylactic treatment with a mood stabiliser a design using a combination with another agent may be appropriate. However, the demonstration of an advantage of the combination treatment in a fair comparison with the established mood stabiliser cannot be taken as evidence that the new agent itself is effective as monotherapy in long-term treatment.

The combination study will need to be carefully controlled for potential pharmacokinetic and pharmacodynamic interactions and to address the increased concerns about safety.

6. Discontinuation effects

There is no evidence from the clinical field that established treatments for bipolar disorder have any dependence producing properties. Individuals with mania or hypomania or those with bipolar depression apparently experience little hesitation in discontinuing treatment and indeed compliance with medication is often a major management problem.

There is however some evidence that discontinuation effects that occur when lithium is withdrawn may compromise the usefulness of this treatment. There have been a number of reports of an increased recurrence of mania over the expected rate associated with discontinuing long-term treatment with lithium (Suppes et al., 1991). This phenomenon is thought to be particularly apparent where discontinuation is abrupt which seems to advance or provoke mania in some individuals (Faedda et al., 1993). Discontinuation of lithium in studies needs to be carried out with a gradual taper.

The design of efficacy studies in bipolar disorder needs to take account of these discontinuation effects that make lithium a problematic comparator mood stabiliser. They are a particular problem because many patients considered for clinical trials in bipolar disorder will have received prior treatment with lithium and their inclusion in studies may complicate interpretation of the results. Abrupt dis-

continuation effects with an increased risk of manic episodes have not been reported with other mood stabilisers but for new products documentation of this is of importance.

Abrupt discontinuation of certain antidepressants is associated with discontinuation symptoms, which are maximal in the first few days to a week and reduce thereafter. Discontinuation studies carried out over a period of one to two weeks may help to determine the potential for discontinuation effects with agents used in bipolar disorder.

7. Rapid cycling

Rapid cycling is the term applied to those with bipolar disorder who have at least four defined episodes of mania or depression a year. They are considered by many to be a subgroup of bipolar disorder that needs to be studied separately. Efficacy in the treatment of bipolar rapid cycling cannot be assumed from efficacy demonstrated in bipolar disorder generally.

Determining the diagnosis is sometimes difficult partly because the history obtained is often unreliable. In many cases the rapid cycling is seen to be transient when followed up prospectively. This may be because of the variability of the cycling process but could also be due to stopping a provoking agent. Rapid cycling may be provoked by the use of substances of abuse, by tricyclic, and possibly some other antidepressants, and spontaneous remission may occur after the withdrawal of a provoking agent. For this reason the optimistic reports of efficacy from open treatment studies are regarded as unsubstantiated anecdotes which may at best be hypothesis generating.

Superiority compared with placebo provides the best evidence of efficacy and this strategy can be defended since one of the management options is the withdrawal of all treatment. A washout period of one to three months is recommended to help to confirm the diagnosis. In this lengthy washout period attempts can be made to distinguish between intrinsic rapid cycling, which may be more resistant to treatment, and the rapid cycling that may respond to the withdrawal of extrinsic provoking factors.

Survival analysis, which takes both the number of episodes and the time to the onset of the next episode compared with placebo, provides the most sensitive assessment of efficacy. The study will need to predefine the relapse criteria for both depressive and manic episodes. There is evidence that the withdrawal from a placebo-controlled study for efficacy reasons is a sensitive measure of efficacy in rapid cyclers particularly where the relapse criteria used are stringent. It has been suggested that response to an effective treatment might be seen not in the prevention of episodes but in a change in the course of the subsequent episodes in duration or severity for example.

This would however be difficult to determine since it would require a lengthy study with no alternate treatment provided during the episode and this would be impractical.

In rapid cycling bipolar disorder the subtype, whether bipolar I or bipolar II, may help to differentiate the response to treatment, particularly in long term treatment. It might therefore be useful to identify the subtype in advance for a subanalysis.

8. Children and adolescents

Bipolar disorder is a condition with an early onset and is seen in both adolescents and young adults. There are data indicating that the earlier the onset the more severe the course of the disorder. For this reason treatment of bipolar disorder in the young is regarded as a high priority. The DSM diagnostic criteria for bipolar disorder have been difficult to apply in children below the age of twelve and the disorder has been considered to be rarely observed in this age group although there are conflicting data (Geller, 1997; Kowatch, 1997). It would therefore be extremely difficult to comment on relative efficacy of treatments in children.

There is no evidence to suggest that the treatments used in adults are any less effective in adolescents but there is a paucity of controlled studies. Studies in acute mania might be possible in adolescents although the practical problems are seen to be challenging. Long-term treatment studies are more difficult because of the length of time often needed to establish the diagnosis in bipolar disorder and that, combined with the length of the study, may take the individual out of the adolescent period. Data to address safety concerns, especially potential risks related to the age group, with a particular treatment may be collected from open treatment studies.

9. Elderly

Bipolar disorder has a lifetime recurrent course and by the time those with the condition become elderly the cycle length has tended to stabilise. The strategies for treatment have often become more complex as resistance to monotherapy may increase with time and a variety of combinations of mood stabilisers may be in place. There is evidence, for example, that the efficacy of lithium declines with the passage of time in many individuals. Even in patients who showed a good response initially there may be a loss of efficacy during prolonged treatment. Maj et al. (1988) reported that only 44% of those who responded for two years remained responders after five years. This reduction in efficacy with time has been reported with carbamazepine and to a lesser extent with valproate.

The elderly who have suffered the disorder for many years appear to be a largely resistant subgroup and

consequently are a difficult group in which to investigate efficacy. The widespread use of polypharmacy raises the issue of safety and pharmacokinetic interactions which are more likely in the elderly. Safety issues will therefore need to be thoroughly investigated in the elderly.

10. Conclusion

Establishing the efficacy of treatments in bipolar disorder is necessarily complicated because of the differing manifestations of the disorder.

Efficacy may be established separately in the treatment of acute mania, bipolar depression and in the long-term mood stabilisation of the disorder.

Efficacy in the acute treatment of mania can be established at least in two positive placebo-controlled studies over a three week period but will need to be followed by a demonstration of efficacy that is similar, or at least not significantly inferior, to an established antimanic agent up to 12 weeks. A treatment that is effective in mania may be assumed to be effective in hypomania but not vice versa.

Efficacy in bipolar depression will need to be established in placebo and reference controlled studies using the same methodology established for unipolar depression over a period of 6 to 8 weeks. The efficacy in acute depression needs to be supported by placebo-controlled evidence of efficacy in relapse prevention over a 3-6 month period following response to acute treatment.

Effective treatments of bipolar depression will need to demonstrate a low potential for provoking manic switches, comparable with either placebo or an established mood stabiliser. The choice of comparator is between standard antidepressant (with a low switch rate), needed to establish a relevant antidepressant effect size and to generate data on switch rate, and a mood stabiliser such as lithium to compare relative switch rates. Two positive placebo-controlled studies in bipolar depression or one in bipolar and one in unipolar depression should be sufficient to establish efficacy in bipolar depression. Treatments shown to be effective in bipolar depression may not be assumed to be effective in unipolar depression.

Mood stabilisers are long-term treatments which reduce the chances of the return of either manic or depressive episodes or symptoms and are widely seen to be the key to maintaining a normal life. Efficacy can be established by a comparison of the number of episodes or near episodes occurring on the treatment compared to placebo over a prolonged period of a year or more in individuals who have both responded to acute treatment and been stabilised. The validity of the design and sensitivity of the population studied as well as the clinically relevant treatment effects in this population might well be established by comparing the efficacy of a standard mood stabiliser such as lithium with placebo in the same study. Both mania and depression rating scales will need to be

used in such a design with predefined criteria for recurrence.

The licensing of new treatments for bipolar disorder will also depend on a risk benefit assessment. This consensus document has explored in some detail the various methods of establishing efficacy. The risks of the treatment will also need to be carefully documented particularly in relation to the known risks of existing treatments for bipolar disorder.

11. BIPOLAR DISORDERS – DSMIV

11.1. Bipolar I Disorder

- (i) Single Manic Episode
- (ii) Most recent episode Hypomanic plus one past Manic or Mixed Episode plus significant distress or impairment
- (iii) Most recent episode Manic plus one past Major Depressive, Manic, or Mixed Episode
- (iv) Most recent episode Mixed plus one past Major Depressive, Manic or Mixed Episode
- (v) Most recent episode Major Depressive plus one past Manic or Mixed Episode
- (vi) Most recent episode Unspecified plus one past Manic or Mixed Episode plus significant distress or impairment

11.2. Bipolar II Disorder

- Present or past Major Depressive Episode plus
- Present or past Hypomanic Episode
- No Manic or Mixed Episodes
- Plus significant distress or impairment

12. MAJOR DEPRESSIVE EPISODE

Five or more of the following symptoms of which one must be:

- Depressed mood or loss of interest or pleasure
- Weight loss or gain or increased or decreased appetite
- Insomnia or hypersomnia
- Psychomotor agitation or retardation
- Fatigue or loss of energy
- Feelings of worthlessness or inappropriate guilt
- Loss of concentration
- Thoughts of death, suicidal ideation, or attempt

Symptoms have had a duration of two weeks, cause distress or represent a change from previous functioning.

12.1. Manic Episode

Distinct period of abnormally and persistently elevated,

expansive, or irritable mood. If only irritable mood four, otherwise three or more of the following symptoms:

Inflated self-esteem or grandiosity
 Decreased need for sleep
 Increased talkativeness or pressure to keep talking
 Flight of ideas or feeling of thoughts racing
 Distractibility
 Increase in goal-directed activity or psychomotor agitation
 Excessive involvement in pleasurable activities with potential for painful consequence

Mood disturbance severe enough to cause marked impairment, or to necessitate hospitalisation, or there are psychotic features.

Duration one week. No duration criterion if hospitalisation required.

12.2. Hypomanic Episode

As for Manic Episode with the following exceptions:

Change in functioning is observable by others but is not severe enough to require hospitalisation
 No psychotic features
 Duration 4 days

12.3. Mixed Episode

Meets criteria for both Manic Episode and Major Depressive Episode nearly every day for at least one week.

References

- American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders DSM-IV. Washington: APA, 1994.
- Angst, J., 1998. The emerging epidemiology of hypomania and bipolar II disorder. *Journal of Affective Disorders* 50, 143–151.
- Anton, S.F., Robinson, D.S., Roberts, D.L., Kensler, T.T., English, P.A., Archibald, D.G., 1994. Long-term treatment of depression with nefazodone. *Psychopharm. Bull* 30, 165–169.
- Bech, P., Rafaelsen, O.J., Kramp, P., Bolwig, T.G., 1978. The mania rating scale: scale construction and inter-observer agreement. *Neuropharmacol.* 17, 430–431.
- Bowden, C.L., 1998. New concepts in mood stabilization. *Neuropsychopharmacol.* 19, 194–199.
- Bowden, C.L., Brugger, A.M., Swann, A.C., Calabrese, J.R. et al., 1994. Efficacy of divalproex vs. lithium and placebo in the treatment of mania. The Depakote Mania Study Group. *JAMA* 271, 918–924.
- Calabrese, J.R., Rapport, D.J., 1999. Mood stabilizers and the evolution of maintenance study designs in bipolar I disorder. *J. Clin. Psychiatry* 60, 5–13.
- Calabrese, J.R., Bowden, C.L., Sachs, G. et al., 1999. A double-blind, placebo-controlled study of lamotrigine monotherapy in outpatients with bipolar I depression. *J. Clin. Psychiatry* 60, 79–88.
- Cohn, J.B., Collins, G., Ashbrook, E., Wernicke, J.F., 1989. A comparison of fluoxetine, imipramine and placebo in patients with bipolar depressive disorder. *Int. Clin. Psychopharmacol.* 4, 313–322.
- Dilsaver, S.C., Chen, Y., Swann, A.C., Shoaib, A.M., Krajewski, K.J., 1994. Suicidality in patients with pure and depressive mania. *Am. J. Psychiatry.* 151, 1312–1315.
- Endicott, J., Spitzer, R.L., 1978. A diagnostic interview: the schedule for affective disorders and schizophrenia (SADS). *Arch. Gen. Psychiatry* 35, 837.
- Faedda, G.L., Tondo, L., Baldessarini, R.J., Suppes, T., Tohen, M., 1993. Out-come after rapid vs. gradual discontinuation of lithium treatment in bipolar disorders. *Arch. Gen. Psychiatry* 50, 448–455.
- Falret, J.P., 1851. Marche de la folie. *Gazette Hopitaux* 24, 18–19.
- Falret, J.P., 1854. De la folie circulaire ou forme de maladie mentale caractérisée par l'alternative régulière de la manie et de la mélancholie. *Bull. Acad. Med. (Paris)* 19, 382.
- Geller, B., 1997. Child and adolescent bipolar disorder: a review of the past 10 years. *Journal of American Academy of Child and Adolescent Psychiatry* 36, 1642.
- Hamilton, M., 1967. Development of rating scale for primary depressive illness. *Br. J. Soc. Clin. Psychol.* 6, 278–296.
- Kowatch, R.A., 1997. Mania in young children. *Journal of American Academy of Child and Adolescent Psychiatry* 37, 1005.
- Kukopulos, A., Reginaldi, D., Laddomada, P., Floris, G., Serra, G., Tondo, L., 1980. Course of the manic-depressive cycle and changes caused by treatment. *Pharmakopsychiatrie-Neuropsychopharmacol* 13, 156–167.
- Maj, M., Priozzi, R., Kemali, D., 1988. Long term outcome of lithium prolix in patients initially classified as complete responders. *Neuropharmacol.* 98, 535–538.
- Montgomery S.A., Cassano G.B., Management of bipolar disorder. London: Martin Dunitz, 1996.
- Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. *Br. J. Psychiatry* 134, 382–389.
- Montgomery, S.A., Rasmussen, J.G.C., Tanghøj, P., 1993. A 24 week study of 20 mg citalopram, 40 mg citalopram and placebo in the prevention of relapse of major depression. *Int. Clin. Psychopharmacol.* 8, 181–188.
- Sheehan, D.V., Harnett-Sheehan, K., Raj, B.A., 1996. The measurement of disability. *Int. Clin. Psychopharmacol.* 11, 89–95.
- Spearing, M.K., Post, R.M., Leverich, G.S., Brandt, D., Nolen, W.A., 1997. Modification of the Clinical Global Impressions (CGI) scale for use in bipolar illness (BP): the CGI-BP. *Psychiatr. Res.* 73, 159–171.
- Suppes, T., Baldessarini, R.J., Faedda, G.L., Tohen, M., 1991. Risk of recurrence following discontinuation of lithium treatment in bipolar disorder. *Arch. Gen. Psychiatry* 48, 1082–1088.
- World Health Organisation: International classification of diseases 10th revision (ICD-10). Geneva: World Health Organisation, 1992.
- Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A., 1978. A rating scale for mania: reliability, validity and sensitivity. *Br. J. Psychiatry* 133, 429–435.