The International Society for CNS Clinical Trials and Methodology (ISCTM) welcomes this opportunity to respond to the EMA request for comment regarding the guidance document: *Guideline on clinical development of medicinal products for the treatment and prevention of bipolar disorder*

The ISCTM offers these comments for consideration based on our experience and expertise in human CNS research. The ISCTM is an independent organization focused on advancing the development of improved treatments for CNS disorders. No member of this Working Group, comprised of scientists, clinicians, trialists, statisticians and drug developers from both industry and academia, received compensation for comments provided. Comments represent individual opinions and not that of the institution, agency, or company affiliation of group members.

The ISCTM formed a group, led by Gary Sachs and Manpreet Singh, to review and provide comments on behalf of the Society. The authors of the comments provided below are (in alphabetical order):

Marc Aafjes, MSM, MSc, *Deliberate AI*

Michael Avissar, MD, PhD, *Alto Neuroscience*

Marc Cantillon, MD, *RWJ*

Dominique Demolle, PhD, *Cognivia*

Martin Dunbar, PhD, *Alkermes*

Nanco Hefting, PharmD, *H. Lundbeck A/S*

Anna Krasnow, PhD, *IQVIA*

Rasmus W. Licht, MD, *Aalborg University Hospital*

Thomas A. Macek, PharmD, PhD, *Cybin*

Atul Mahableshwarkar, MD, *Cybin*

Shishuka Malhotra, MD, *Neurobehavioral Research*

Annalisa Marotto, PharmD, *Boehringer Ingelheim International GmbH*

Tracey Petryshen, PhD, *AbbVie Inc.*

Gary Sachs, MD, *Signant Health* (Co-chair)

Navid Samad, MD, *Samad Pharma and Biotech Consulting, LLC*

Manpreet K. Singh, MD, MS, *University of California Davis* (Co-chair)

Jair Soares, MD, PhD, *McGovern Medical School at UTHealth Houston*

Stephanie Sommer, PhD, *Boehringer Ingelheim International GmbH*

Joyce Tsai, PhD, *MycoMedica Life Sciences*

Qing Wang, PhD, *Neumarker*

1. General comment:

The revised BP guidance is welcome. It is understood that for reasons of efficiency and consistency frequent reference is made to the guidance on unipolar depression. Since this guidance, however, is currently under revision, providing appropriate comments on the BD guidance is hampered because in some instances there may be reason to deviate from what would be implemented in the guidance on unipolar depression.

The guideline does not mention how many trials are needed for an indication in BD, but the topics that need to be evaluated suggest that an indication for BD cannot be achieved in one trial. For example, Would CHMP accept a clinical trial design where acute treatment effect is investigated against placebo first and then participants be re-randomized to evaluate the maintenance of efficacy? Could this be clarified in the text on maintenance of efficacy?

The guideline does not include any reference to primary or secondary prevention of bipolar disorder per se. Rather, prevention of BD episodes is discussed as recurrence prevention. As such, we suggest the following as an alternative title to more accurately reflect the main aim of the guideline: “Guideline on clinical development of medicinal products for the treatment and prevention of recurrent episodes of bipolar disorder.”

The guidelines currently lack direction on the need to identify and intervene early in individuals with prodromal or subthreshold BD symptoms, including those at high risk (such as children of affected parents). Early identification and timely intervention have the potential to delay or prevent full-blown episodes, reduce disease severity, and improve long-term outcomes. Incorporating guidance on designing trials that focus on early intervention, preventive strategies, and prodromal populations would encourage research and innovation in preventing the onset or progression of BD.

Wording alluding to the “treatment and prevention of bipolar disorder” is used throughout. This wording may be misleading since the available treatments do not prevent the incidence of bipolar disorder per se. An alternative suggested phrasing is “acute and maintenance treatments of bipolar disorder”. In addition, consider bipolar disorder an “episodic, often lifelong” disorder as an alternative to it as a “chronic” disorder.

2. Specific comment:

Definitions

Maintenance of effect: Revise for better clarity. Instead of "the whole (hypo)manic/depressive episode", we suggest "over longer-term treatment". This avoids the confusion that might arise when a participant without symptoms is labeled as still in an episode.

Relapse and Recurrence: Differentiation between relapses and recurrences is theoretically meaningful, but in clinical trials the distinction is ultimately arbitrary. For clinical trials, “Relapse” could be operationalized as an increase or “exacerbation” in symptomatology during the current (index) episode, after a clinically relevant improvement in symptomatology. The term relapse is confusing if applied within an episode. Most practitioners see it as a new episode. In ICD-10/11 and DSM-5, the arbitrary/pragmatic separation of two months without significant symptoms is used. STEP-BD used the term “Recovering” to designate periods of remission shorter than 8 weeks. A return to syndromic symptomatology after a period of “Recovering” could be considered a Relapse (to the previous episode). Those with remission sustained beyond 8 weeks would be considered “Recovered”.

In the context of trials on maintenance treatment, it is preferable to use the term “Treatment emergent episode” instead of the terms “relapses” and “recurrences”, since these latter episodes cannot be clearly separated under such conditions. Similarly, consider the terms "randomized withdrawal or continuation period" to provide greater specificity for describing that phase of a maintenance or "prevention of relapse" study design.

Remission: Has no or only few signs (e.g. no more than two symptoms of moderate severity) of a specified condition (e.g., depression) remaining, clinically significant symptoms of another mood episode (e.g., mania or hypomania) are absent and does not meet duration criteria for Recovered.

64 "…concomitant increase in symptoms of another condition…"
Suggested revision: "…concomitant increase in symptoms of another *related* condition…"

Rationale: Concomitant symptom increase has to be in a condition related to mood disorder for it to violate remission. If it is an unrelated medical condition it would still be seen as a remission.

Executive Summary:

Line number(s) of the relevant text *(e.g. 20-23)*	Comment and rationale	Proposed guidance text
70	Suggest rewording for clarity.	The main aim of the guideline is to provide up to date guidance for the development of medicinal products for the treatment and prevention of recurrent episodes of bipolar disorder.

89-91	Revise Background to clarify the concept of Bipolar as a lifetime episodic illness vs the acute episodes.	Bipolar Disorders (BD) are lifetime psychiatric disorders characterized by an irregular course in which episodes of abnormal mood states alternate with periods of full or partial remission. In Bipolar (BP) I disorder at least one of the abnormal mood states has met full criteria for mania. In Bipolar II disorder, abnormal mood states include at least one episode meeting criteria for major depression and at least one episode meeting criteria for hypomania, but no episode of mania has occurred. Over the past decades lifetime prevalence of these disorders has remained stable (BD I 0.6% and BD II 0.4%, worldwide).
95	Revise to better reflect state of scientific understanding	Large studies suggest numerous shared genetic factors may confer increased vulnerability to Bipolar Disorder, Schizophrenia, other psychotic disorders and Major Depressive Disorder. While such data may suggest considering Bipolar Disorder a bridging disease between schizophrenia and major depressive disorder, shared genetic risk factors do not mean the individual conditions are indistinguishable. The frequency of episodes varies overtime with some patients at times meeting criteria for rapid cycling (four or more episodes within a twelve month period).
108	Revise to reflect potential for preventive products	Development of new medicinal products are encouraged that should not only focus on the treatment of acute symptoms and prevention of relapse during the current i.e. index episode but explore also the potential of a medicinal product in the prevention of new episodes i.e. recurrence prevention. While prevention of a new episode is not a mandatory part of a registration package for treatment of acute episodes of BD, recurrence prevention is considered as an additional claim (section 4.4.2) and is of particular interest in regard to claims for treatment of rapid cycling.

116-125	Revise for better clarity Replace "and diagnostic accuracy is moving towards early detection and a younger patient population" with "and better diagnostic methods can facilitate early detection and identification of a younger patient population. Replace "However, these symptoms can also occur prodromally to other psychiatric disorders and are therefore not specific to bipolar disorder." with: "Thus far, however, none have been identified as not specific to bipolar disorder and these symptoms also occur prodromally with other psychiatric disorders." after "Information from both prospective studies and studies in the offspring of affected parents show that symptoms of irritability, often associated with Attention Deficit Hyperactivity Disorder (ADHD), predispose to BD and that prodromal mood symptoms are present before the onset of BD and in the offspring of patients with BD. Revise "Several behavioural, genetic, neuroimaging, electrophysiological- and (immune) biomarkers hold promises for identification of these patients at risk. However, at present the data are still too heterogeneous and not suitable for 'fingerprinting' to "While several behavioural, genetic, neuroimaging, electrophysiological- and (immune) biomarkers hold promises for identification of patients at risk for bipolar disorder, the data are not yet not suitable for 'fingerprinting'."	Although BD primarily affects adults, there is increasing evidence that the disorder often begins in adolescence, and better diagnostic methods can facilitate early detection and identification of a younger patient population. Thus far, however, none have been identified as not specific to bipolar disorder and these symptoms also occur prodromally with other psychiatric disorders. While several behavioural, genetic, neuroimaging, electrophysiological- and (immune) biomarkers hold promises for identification of patients at risk for bipolar disorder, the data are not yet not suitable for 'fingerprinting'. Studies in children and adolescents with established bipolar disorder are addressed in section 4.6.2

4. Specific considerations when developing products for the treatment and prevention of bipolar disorder episodes

Preclinical studies (suggested as a new section)

Line number(s) of the relevant text *(e.g. 20-23)*	Comment and rationale	Proposed guidance text
	The current guideline provides limited guidance on leveraging personalized medicine approaches, despite the known heterogeneity of bipolar disorder. Integrating genomic, epigenomic, and biomarker-driven strategies could help identify subgroups of patients who respond more favorably to certain treatments or reduce adverse effects. [Missing]: The guidance makes no mention of Personalized or Precision Treatment Approaches	**4.1 Preclinical studies (suggested as a new section)** The development of animal models in BD is encouraged. Bipolar disorder (BD) is characterized by episodes of abnormal mood elevation, depressive states and mixed states with intervening episodes of normal mood. Due to this heterogenic symptomatology in BD, there is a lack of adequate animal models for bipolar disorder. Specifically, most animal models mimic only mania

	It is proposed that there be a new subsection in Section 4.1.1 (Pharmacodynamics) or as a standalone subsection (e.g., Section 4.8 " Section 4.1.1 touches on biomarkers only in a very general way. - Section 4.2.1 mentions defining the estimand for different populations, which could be extended to incorporate biomarker-defined subgroups. - Section 4.2.4 and 4.2.6 discuss study populations and extrapolations but do not consider individual patient stratification based on genetic/biomarker profiles, outcomes, or how to accelerate the development of targeted therapies. Such guidance could encourage sponsors to design studies that incorporate biomarker stratification, apply pharmacogenomic testing, or tailor interventions based on individual mood cycling patterns and/or inter-episode characteristics. This would align the guideline with the evolving trend towards precision psychiatry. Personalized and Precision Approaches"):	or depression and only a few include the cyclical nature of the human condition. Guideline on clinical development of medicinal products for the treatment and prevention of bipolar disorder EMA/CHMP/406037/2024 Page 6/19 Behavioural models should help understanding the biological mechanisms that contribute to mood cycling, which is the hallmark of BD. Section 4.1.1 (Pharmacodynamics) Or Section 4.8 "Personalized and Precision Approaches Personalized and Precision Medicine Approaches: Given the heterogeneity of bipolar disorder, sponsors are encouraged to explore precision medicine strategies. This includes integrating genomic, epigenomic, and other biomarker data into study designs to identify patient subgroups with differential treatment response. For example, stratification based on genetic markers, neuroimaging findings, or validated biomarkers of treatment response may inform targeted intervention strategies. Incorporation of pharmacogenomic testing or biomarker panels into early-phase trials can guide dose selection, reduce variability in response, and enhance the interpretability of results. Additionally, considering individual-specific patterns of mood cycling and clinical trajectories when designing trials may improve the ecological validity and long-term utility of the study findings. Sponsors are encouraged to discuss such approaches during scientific advice procedures and to justify the

		selection and clinical relevance of any proposed biomarkers or genetic tests.

4.1 Clinical Pharmacology studies

4.1.1 Pharmacodynamics

Line number(s) of the relevant text *(e.g. 20-23)*	Comment and rationale	Proposed guidance text
147-148	Evidence for engagement of the drug target and/or relevant neurocircuits in the human brain should be obtained through receptor occupancy PET and/or through functional modalities such as fMRI, EEG, and cognitive testing. Data may be useful for dose finding and establishing proof of mechanism in early drug development. Pharmacodynamic data support engagement of relevant neurocircuits that provide proof of drug mechanism, and evidence for changes in functional activity relevant to a patient population act as proof of concept for clinical efficacy Given the guidance text in lines 147-148 commenting on inadequacy of animal models in BD, the requirement for preclinical supportive evidence of claims on cognition should be removed.	The requirement for preclinical supportive evidence of claims on cognition should be removed.
149	"Behavioral models…" It is not clear to the reader what is meant here exactly. Are we talking about animal models? Since previous paragraph states there are no models of the cyclical change in mood.	Please clarify what models these are or delete.
155-156	It is unclear why the reference goes to 4.4.3 "Improvement in cognitive function" and that cognition is gender dependent? Furthermore, so far there is no robust evidence that PD is gender dependent in this disease, apparent differences between genders may be attributed to PK.	References to findings from preclinical studies can be moved to the preclinical study section or deleted.
158-159	We are unaware of any pathways associated with rapid onset in humans.	The sentence needs clarification.

4.1.2. Pharmacokinetics

160-169	This section on PK/PD does not cover any considerations related specifically to BD. It just seems generic. Is it necessary?	Consider removing this section.

4.1.3 Interaction studies

174	Rather than being prescriptive in the compounds with which DDI studies should be conducted, the selection of appropriate compounds for these should depend on the mechanism of action (MoA) of the medicinal product under study.	Interaction with relevant CNS active medicinal products should be investigated. If the purpose is to see side effects, this will become ethically problematic. Conducting studies requiring administration of alcohol to patients might not get EC approval. With respect to other CNS active drugs it is too unspecific.

	4.1.4 Biomarkers Add new section before Section 4.2 and add the recommended text.	The use of biomarkers may be useful to delineate patient heterogeneity in symptoms and treatment response, and to support engagement of relevant neurocircuits. It is encouraged to use functional biomarkers such as continuous digital metrics, fMRI, EEG, cognitive testing, and neuroendocrine markers or neuropeptides. Selection of functional biomarkers should consider feasibility in later phase studies, and biomarkers should be implemented in later phases if shown to increase effect size of primary and/or key secondary endpoints. Rapid advances in -omics technologies may enable identification of reliable omics biomarkers that help delineate patient biotypes or predict treatment response.

4.2. Assessment of therapeutic efficacy

178	Please clarify that treatment effects do not need to be shown in both acute treatment and prevention phases. Please add: A difference in response/remission rate does not necessarily need to meet traditional levels of statistically significance in each individual clinical trial.	Results should be discussed in terms of both clinical relevance and statistical significance, and the effect should be shown to be insensitive to the analysis used. When an effect is quantified in terms of change from baseline to end of treatment using a validated measurement tool, this effect has to be addressed also as rates of responders and remitters. The statistical significance and the clinical relevance of the effect together form the basis for the benefit/risk assessment. (Usually, a binary outcome analysis, such as responder or remitter analysis, has less statistical power than a continuous outcome of change from baseline on a rating scale. Is it expected that responder/remitter analyses provide statistically significant results in individual trials and that clinical trials are powered for these analyses? That would require larger sample sizes and would make clinical development in BD less attractive.)

4.2.1 Target of estimation in bipolar disorder

205	“death by suicide" is the preferred term now. Address adjunctive treatment dose changes in the document.	death due to committed suicide might require incorporation into the estimand Adjunctive treatment dose changes are common and should be addressed in the document

4.2.2. Placebo effect and strategies to address high placebo response

236	Some specific variables related to bipolar disorder appear to be more frequently associated with the placebo effect, for example low symptomatic	ADD TEXT: In this context, some variables associated with high placebo response may be

	severity, higher fluctuation in episodes with mixed features, patients with first episode, and rapid cycling with spontaneous remission. Appropriate population selection, careful screening process and appropriate training of investigators for example in terms of scoring may help to reduce high placebo response during the study. While Meta-analysis of RCTs in acute mania suggest manic patients to be less responsive to placebo as compared to depression, some acute mania trials report substantial placebo response. As emphasized in this section, several prognostic variables have been associated with a high placebo response, ultimately negatively influencing the estimation of treatment effect. However, no guidance is provided on how to mitigate this effect, but the EMA guidance on the adjustment for baseline covariates in clinical trials (EMA/CHMP/295050/2013), clearly explains how to mitigate variables associated with treatment response. Therefore, it would be helpful to see this mentioned in the current guidance.	considered as baseline prognostic variables for adjusted analyses (EMA/CHMP/295050/2013).
241	Clarify rationale: Screen failing participants with high response (>20%) to a short placebo lead-in is a common practice. Please clarify why is use of a placebo lead-in period problematic. In clinical practice we want to maximize response to all aspects of treatment, while in clinical trials we want to know what added benefit the pharmacological effect brings. Reducing placebo effects helps to better define that effect and avoid flooring effects due to exaggerated non-specific (placebo) response.	Use of a placebo run-in period (single- or double-blind) is considered problematic with regard to the generalizability of the results to the population treated in clinical practice, since patients included in the studies may not correspond to the target population. This concern also applies to study designs implementing a second randomisation step of placebo non-responders and also potentially placebo responders.
246	Clarify rationale to accept P2 but not p3 since both are considered for approval.	Enrichment strategies that identify and segregate placebo responders from the primary analysis could be provided in the case of phase II studies and not for confirmatory phase III studies.

4.2.3. Investigation of relapse and recurrence

264	Please add references for these time frames.	The duration of the maintenance phase is usually set at about 12 weeks for mania and 3 to 6 months for bipolar depression, to correspond with the average duration of an episode of either mania or depression.

266	Clarify what is meant by these sentences and what the guidance is. The individual duration of any mood episode cannot be predicted as historic episode duration is not predictive for future ones, so how to take episode duration into account? The guideline states below and in section 4.3.1 that maintenance of efficacy in the current episode should be investigated, but this sentence suggests that this is not needed for mania. Is that correctly understood?	In any individual, however, the duration of an episode varies considerably and may be more (or less). As this might affect the interpretation of the results, the 6 months cut-off point is not used for regulatory purposes. Instead, the guideline focuses on showing effect during the current episode and/or prevention of the next episode.
276	Please clarify acceptability of claim preventing episodes of one polarity Is a recurrence another episode of either polarity after the index episode?	Recurrence should be prespecified as either a manic or a depressive episode that fulfils current DSM-5 criteria and a certain degree of severity on a validated rating scale.
279	Please clarify if this simply encourages accession of a sample at higher risk or is the intent to include a proportion of rapid cycling patients in the sample.	Patients with a history of higher frequency of manic or depressive episodes should be included in the recurrence prevention investigation and the recent recurrence rate should be taken into account when planning the duration and power of the study.

4.2.4. Study population

288	Revise to reflect current terminology. Note: SCID-5 is the updated version of the structured interview for DSM-5 Diagnosis. Also DSM-5 does not use the multiaxial system	SCID-5 (Structured Clinical Interview for DSM 5 Diagnosis)
292-294	Examples would be useful here of internal and external validity and how they may interact.	"In contrast to proof-of-concept…criteria should be clearly justified"
299	Please clarify/justify/strike: Given the time to diagnosis of bipolar disorder (10 years), this does not seem to be a feasible and the rationale not justified. This seems to be a carryover from schizophrenia and may not apply or be needed. Instead of “disease history less than 5 years” consider “less than 5 years since first diagnosis of Bipolar Disorder”.	It is recommended to include at least 20% of patients with a disease history of less than 5 years.

4.2.5. Comedication

305	Medication implies drug treatment, but supplementary non-drug therapies are also addressed in this section	Consider changing header to Concomitant Therapies
311	Standardised psychotherapy, psychoeducation, support or counselling may be offered as supplementary treatment, but their use should be prospectively defined in the protocol and documented in the study report, including a discussion on the effects on treatment.	Please clarify if this means conducting analyses on subgroup(s) with and without additional therapies?

4.2.6. Extrapolations

319-320	MDD and BD are mentioned as two different ”entities”. Given that MDD is only MDD until change to BD, it may be more appropriate to describe MDD and BD as two different “categories”.	Describe MDD and BD as two different “categories”.
322-324	Please add comment to clarify: Is extrapolation of results from bipolar I depression to bipolar II depression permissible?	A major depressive episode as seen in BD II also occurs in the framework of major depressive 318 disorder. Based on current evidence, and due to the different disease characteristics (or different nature) of the two disease entities (BD and MDD), extrapolation of short term and maintenance of efficacy in adults from bipolar depression to unipolar depression does not seem possible.
322	It’s not clear why “A switch from depression to mania should always be considered in BD, but not in MDD.””	A switch from depression to mania should always be considered an undesirable outcome in BD.

4.3. Methodological features

4.3.1. Study design

336-337	Confirmatory short-term studies should be double-blind, randomised, parallel group, placebo-controlled studies. ..... A two-arm non-inferiority study is not an option as the sole basis for demonstrating efficacy as a reliable non-inferiority margin is difficult to determine.	Would a superiority trial with a comparator other than placebo qualify as evidence for label claim (e.g. with psychedelics)?

338	A two-arm non-inferiority study is not an option as the sole basis for demonstrating efficacy	Revise for clarity: A two-arm non-inferiority study is not an option as the sole basis for demonstrating ACUTE efficacy
339-340	E.g. in some studies, lithium, which is believed to be the gold standard for the treatment of bipolar disorder, failed to separate from placebo.	Replace Gold standard ref with: E.g. in some studies, lithium, which is a commonly used treatment of bipolar disorder, failed to separate from placebo.
343-345	Suggest that the agency expand the range of acceptable washout periods beyond a few days Note: A short duration of washout may not be adequate to uncover worsening due to washout and may result in varying amount of worsening in symptomatology in the recruited population. This may impact efficacy evaluation and hence suggest that the agency expand the range of acceptable washout periods beyond a few days	Regarding the wash out of prior medication, tapering in a single blind placebo run-in period of sufficient duration aiming at elimination of prior treatment without substantial worsening of symptoms should be applied. While A few days MAY BE appropriate for run-in, it is acceptable to expand the range of wash out to reflect half-life of relevant compounds or concerns about rebound effects.
342-348	Consider deleting or revising to reflect agency thinking.	Following screening, qualitative and quantitative baseline assessments should be conducted in a short run-in period. Regarding the wash out of prior medication, tapering in a single blind placebo run-in period of sufficient duration aiming at elimination of prior treatment without substantial worsening of symptoms should be applied. Typically, a few days will be appropriate for run-in. Placebo responders should not be excluded from randomisation. In some instances, screening, baseline assessments, randomisation and start of study medication may be performed in a single day, especially if patients are severely ill.

4.3.1.1. Short-term studies

4.3.1.1.1. Acute manic episodes

350	Instead of using the term “acute manic episodes”, we suggest stating “acute treatment of manic episodes” since all manic episodes can be considered acute like all depressive episodes	acute treatment of manic episodes
359	Please list which specific scales are being referred to, e.g., by listing in brackets after 'a usual rating scale'	50% improvement of a patient on a usual rating scale is accepted as a clinically relevant response.
382	To avoid confusion with 338: Please clarify under which circumstance non-inferiority margin is challenging to determine	The choice of comparator should be justified and non-inferiority to the comparator in maintenance of effect should be demonstrated. Non inferiority margin should be pre-defined and justified in the study protocol.

4.3.1.1.2. Major depressive episode in the framework of bipolar disorder

364	This statement could effectively exclude patients previously treated with Li from any monotherapy trial. Was this intended? Furthermore, a substantial period of time is too unspecific. Would it be acceptable to stratify the randomization for subjects who had discontinued lithium over the month prior to randomization? What other treatments need be washed out (Benzodiazepine)?	For a monotherapy claim, patients need to be off lithium or other mood stabilisers at baseline for a substantial period of time because of possible rebound phenomena. While the potential rebound effect need be considered, prolonging the time off mood stabilizers might increase the risk of switch during the study.
366	Consider striking. Relevance is limited because it is contrary to real world practice. What would the alternative claim be? Monotherapy of MDE in lithium non-responders?	~~Moreover, they should not be lithium resistant (depression under lithium therapy) as this has consequences for the claim requested.~~
364-366	Consider defining lithium/mood stabilizer non-response in terms of dose/serum levels and duration of treatment.	

4.3.1.2. Long-term studies

372-373	Consider specifying that the preferable outcome measure is time to any episode (new).	Due to the character of the disorder, longer studies are necessary to demonstrate that the acute effect

	Also comment on the acceptability of time to intervention for mood symptoms or time to any intervention.	is maintained during an episode (relapse prevention). Episode as part of this outcome does not need to be an episode meeting full diagnostic criteria. It might, for example, just be indicated by the necessity of adding new medication. It should be specified that breaking down treatment emergent episodes into mania and depression could be possible, but only as a secondary outcome, since interpretation of results based on this breaking down is not straightforward. If, for example, a given drug compared to placebo is highly preventive for only one pole, the risk of occurrence of the opposite pole may be overestimated.

4.3.1.2.2. Major depressive episode in the framework of bipolar disorder

385-388	Carefully cross-check with the currently unpublished unipolar depression guidance. As the current unipolar depression guidance is not issued, it may not apply.	

4.3.2. Efficacy endpoints

389-395	This will require larger trials. Could one trial be replaced by a single relapse prevention trial? Or, if pursuing the route of a non-inferiority trial, can one combine 2 acute trials in one maintenance trial? For the assessment of response, specific developed rating instruments are necessary Clarify: Does the agency agree that currently used depression rating scales such as MADRS and HAM-D are adequate to measure response?	Non-inferiority margin should be pre-defined and justified in the study protocol (in the mania maintenance of effect scenario). Efficacy should be assessed by adequate, validated rating scales. The choice of instruments should be 390 justified from the test quality criteria (reliability, validity and sensitivity to change). For the assessment 391 of response, specific developed rating instruments are necessary.

4.3.2.1. Manic episodes

402-403	The CGI-BP Overall score may be a misleading score for interpretation purposes. It may be best to use the individual CGI-BP scores for Mood Elevation and Depression instead or make the CGI-BP-Overall an exploratory endpoint.	Furthermore, a global scale, e.g. the CGI – BP, can be used as further endpoint. Placement of an endpoint within a testing hierarchy should be driven by the scientific question to be answered and aligned in scientific advice procedures.
404	Revise for clarity: "Other tools like actigraphs and electronic devices AS EXPLORATORY ENDPOINTS are currently applied to provide additional information on wellbeing of patients." This is Interesting, and welcome but requires clarification. How should one use Life Charting prospectively in the light of maintenance of efficacy and prevention of recurrence? Should life charts include rating scale results, be clinician administered/completed or be purely PROs? Please clarify further. Could results from actigraphs and electronic devices (please describe which type of information is intended, as eCRF data are also collected via electronic devices) be used for claims in the SmPC?	Other tools like actigraphs and electronic devices as exploratory endpoints are currently applied to provide additional information on wellbeing of patients. The use of life charting or other measures may also be used as a secondary endpoint. Other tools like actigraphs and electronic devices are currently applied to provide additional information on wellbeing of patients.

4.3.3. Statistical considerations

428-429	Consider Adding TEXT after Line 428: As mentioned in Section 4.2.2, prognostic variables should be considered to improve the estimation of the treatment effect, and therefore it is worth adding guidance in the statistical section.	As discussed in section 4.2.2, some variables related to bipolar disorder might be associated with high placebo response, such as low baseline symptom severity or rapid cycling. As baseline prognostic covariates, they should be considered for adjustment (EMA/CHMP/295050/2013). By accounting for key prognostic factors, adjusted analyses can enhance the precision of efficacy evaluations while preserving the generalizability of trial results.

431	Recommend removing the parenthesis and provide an example of an approved product here. Otherwise, it may be implied it is recommended to try to use a co-primary endpont which statistically is harder to achieve than a dual or single primary endpoint measure using a sequential trial design Control for multiplicity should be a topic for scientific advice	approach to properly control for multiplicity should be pre-specified (usually co-primary assessment)
After Line 431	[Missing] Please add Guidance on Digital Health Measures.	Both multimodal data integration and digital health technologies represent modern approaches to capturing more granular, ecologically valid, and patient-centric data. Continuous passive sensing devices, smartphone applications, and digital biomarkers, combined with advanced analytical tools like AI and machine learning, may enhance the detection of subtle changes in mood states, predict relapse more accurately, and improve patient engagement.

4.4. Specific Claims

4.4.2. Recurrence prevention

448	Consider deleting the term “either” as the target would be recurrence of a mood episode, irrespective of its polarity.	Duration and Sample size sufficient to address recurrence of mood episodes, irrespective of its polarity (including both mania (hypomania) and depression).
450	Recurrence of any mood episode is the preferred endpoint. Requesting clarification whether this was the intended outcome here. Generally, the polarity of recurrences largely reflects the polarity of the index episode at trial entry (e.g. mania to mania, MDE to MDE). Powering for an effect to the opposite polarity would require a much larger sample. In addition, when recurrence to the opposite pole occurs, the treatment intervention needed is likely to be prohibited by protocol. And consequently, the patient is dropped from the blinded study/data truncated or at least compromised for assessment for the recurrence of an episode of the initial pole. This may make the trial infeasible.	The number of patients should be sufficient to address both recurrences of depression and of mania.

4.4.3. Improvement in cognitive function

466	Suggest deleting. Since these two sentences are confusing and even contradictory. Since scientific advice should be sought for this claim, a discussion of the appropriateness of the patient population to be included in studies supporting this claim can occur most meaningful there.	~~As in schizophrenia, a relatively younger patient population might be more appropriate for testing effect on cognition to avoid confounding by age. However, in this case, extrapolation to older patient population will need to be justified.~~
469	To contextualize results on cognitive performance, information on functional capacity should be captured.	The effect of treatment on cognitive performance should be demonstrated on an appropriate instrument with an appropriate estimand addressing the research question at hand. Scientific advice for choice of instrument and estimand approach is recommended.
470-471	Revise: 471: ~~changes in performance in a neuropsychological or cognitive test is not sufficient.~~ The current text could be interpreted that all items within a cognitive test battery must change.	Whatever tool is used, mere reduction on specific items of a larger test battery is not acceptable However, it is unlikely that one therapeutic will improve all cognitive domains. Development of treatments should be in the context of cognitive domains (e.g. social cognition, learning/memory).

4.5. Bipolar disorder with specifiers

477-478	If indication in sub-population is sought, can the program entirely consist of participants with the specifier?	If a claim for a sub-population as defined by a specifier is pursued, a dedicated study with specific inclusion criteria and adequate endpoints is required.

4.5.1. Rapid cycling

487	This may need to be qualified and/or clarified. If a reduction in cycle number is considered as an efficacy endpoint, isn't that the same preventing relapses?	Evidence-based data are too limited to provide further advice, but as rapid cycling is defined as four or more acute mood episodes within the past 12 months, it might be considered to define efficacy by a clinically relevant reduction of cycles. Next to the acute efficacy, the potential of a

		medicinal product to prevent relapsing is important in rapid cycling bipolar disorder, necessitating long-term studies.
486-487	The statement may need to be qualified. Reducing the number of episodes/rate of cycling as a consequence of producing persistent manic or depressive episodes should not count as a good outcome.	…define efficacy by a clinically relevant reduction of cycles and/or prolongation of euthymia.

487-488	Please provide guidance as given in the other sections on the duration on long term specification: 3-6 months. Next to the acute efficacy, the potential of a medicinal product Without guidance or reference to section 4.3.1.2 it is unclear what is an expected duration for the studies . Please provide guidance as given in the other sections on the duration on long term specification: 3-6 months.	

4.5.2. Mixed features

490	Request for clarification and elaboration of this section. Currently, the guidelines do not offer concrete strategies for selecting appropriate endpoints, managing rapid symptom fluctuation, or determining optimal timing of assessments in patients with mixed states. Without this guidance, sponsors may struggle to design trials that accurately capture the complexity of mixed features. Providing detailed methodological recommendations and examples of suitable trial designs would strengthen the guideline's utility and help ensure that treatments claiming efficacy in mixed states are rigorously and consistently evaluated.	Mixed features present significant heterogeneity and variability in symptomatology over short periods of time, making traditional trial designs and endpoints potentially inadequate. Additional Considerations for Mixed Features: In clinical trials targeting patients with mixed features, sponsors might consider specialized designs and measurement strategies to address the variability and instability of symptoms. This may include:

	Offer concrete strategies for selecting appropriate endpoints, managing rapid symptom fluctuation, or determining optimal timing of assessments in patients with mixed states. Episodes featuring symptoms of either depression in mania or mania in depression are acknowledged a particular state (eg Depression with mixed features) would it be permissible or would it also be necessary to study Mania/hypomania with mixed features? Requesting clarification whether the agency is asking for separate studies in each of these groups; would a study specifying	- Employing adaptive trial designs or stratified randomization based on mixed-feature symptom counts, severity and symptom stability at baseline. - Utilizing composite endpoints that capture both poles of the disorder and incorporate time-to-event analyses to handle rapid fluctuations. - Increasing the frequency of assessments (e.g., weekly or even more frequent) to better characterize the temporal dynamics of mixed features. - Implementing flexible visit schedules or digital tools to capture symptom changes between scheduled visits. Sponsors are encouraged to seek scientific advice to determine the most appropriate methodologies and endpoints for evaluating treatments in patients with mixed features. Episodes featuring symptoms of either depression in mania or mania in depression are acknowledged. Requesting clarification whether the agency is asking for separate studies in each of these groups; would a study specifying a particular state (eg Depression with mixed features) be permissible or would it also be necessary to study Mania/hypomania with mixed features?

4.6.2. Children and adolescents

518- ?	The rationale to restrict studies to >13 years should be justified. The EMA has approved Lithium down to the age of 12, so the rationale to restrict studies to >13 years should be justified.	There is inconclusive evidence for the existence of BD in childhood. Studies proposed including subjects <13 years of age require clear rationale and definition.

4.6.3. Sex issues / differences

542-543	Suggest to revise based on available evidence.	Up to now there is no robust evidence on sex differences in response to mood stabilizers or to antidepressants.
545	Request the agency to discuss how results of such analyses could be included in the label to inform clinicians.	Hence, during the development of medicines for BD, predefined analyses of gender specific groups are welcomed.

4.7. Safety evaluation

549	Add missing comma after AE. Please clarify what is meant by "frailty." Is it medical frailty? How is it defined?	Characterized in relation to duration of treatment, dosage, duration of the AE, age, frailty, and other.
550-551	Consider alternatively to add IF USED, This is only relevant if these are used in trial. Context for comment: Requesting clarification on whether these scales are required. In general, quite a number of scales are included in clinical trials in BD, contributing to patient burden and non-specific effects. Will data from scales like UKU be included in the SmPC, additionally to the ADR table(s)? If not, can one do without and submit analyses of AEs of special interest (AESIs) based on MedDra standardized search terms?	Adverse event scales, if used, should be standardised for use in studies with psychotropic drugs.
556	Suggest alternatively to include a clarifying example: way for comparisons of event rates, SUCH AS PATIENT-YEARS OF EXPOSURE	way for comparisons of event rates
558-560	Suggest replace by mode of action (MoA). Dependent on the MoA it will not make sense to focus on non-MoA relevant AEs, thus it gives clarity to what the EMA is expecting.	Particular attention should be paid to mode of action (MoA) ~~anti-dopaminergic, anti-cholinergic or cholinergic, anti-histaminergic, serotoninergic and adrenergic, and to glutamatergic or anti-GABAergic~~

4.7.1. Specific adverse events to be monitored

4.7.1.1. Psychiatric adverse events

565-567	Clarification on second part is requested, better to compare to Placebo, if applicable.	In order to explore the risk of an adverse effect on the… Clarification on second part is requested, better to compare to Placebo, if applicable. Potential for misunderstandings: the risk of an AE does not depend on deterioration measured by efficacy, as in both patient population (investigational drug and placebo) AEs can occur
568-569	~~As part of the adverse event data, undesirable psychiatric effects including depression and anxiety should be measured using validated rating scales.~~ The scales and AE reporting are disconnected and not clear what this is referring. It also implies that lack of efficacy should possibly be reported as an adverse event or that validated rating scales conducted in such instances.	In addition to collecting adverse event data, undesirable psychiatric effects, such as, depression and anxiety may be assessed using validated rating scales.

4.7.1.2. Adverse effects on cognitive functioning

571-572	As written it implied the MADRS or YMRS for monitoring cognition. Requesting clarification whether this is what was meant or did the agency mean something like Bond-Lader VAS?	A detrimental effect on cognition should be monitored using validated rating scales, which may be identical to those used to support an efficacy claim. Effects on cognition, reaction time, driving and severity of sedation should also be studied. Or alternatively: A detrimental effect on cognition should be monitored ~~using validated rating scales, which may be identical to those used to support an efficacy claim.~~
573-574	Suggest to strike: ~~using validated rating scales, which may be identical to those used to support an efficacy claim~~. In the adolescent population specific issues such as memory, learning, school performance, etc. should be studied in relation to both the safety and efficacy perspective. Requesting clarity.	In the adolescent population specific issues such as memory, learning, school performance, etc. should be studied in relation to both the safety and efficacy perspective.

4.7.1.3. Overdose

577-578	~~Depending on the mechanism of action risks and effects of overdose should be studied particularly with regard to serotonin-syndrome, QT-prolongation and delirium.~~ Serotonin syndrome, delirium and QT prolongation may be unrelated to overdose. Whether or not they need to be characterized should be dependent on the medicinal product's MoA.	Depending on the mechanism of action risks and effects of overdose should be studied; if relevant, the medicinal product's effect may need to be characterized with regard to serotonin-syndrome, QT-prolongation and delirium.

4.7.1.4. Suicide

584-585	~~narrative summaries of suicidal patient statements or behaviours should be provided.~~ Suicidality is acknowledged as a major risk in this patient population. Yet, generating patient narratives may be most meaningful for those instances that reach a certain level of severity, such as type 4 or 5 on the C-SSRS.	narrative summaries of suicidal patient statements or behaviours reaching a certain threshold should be provided.

4.7.1.6. Haematological adverse events

591	While standard lab should be collected in all clinical trials, detailed analyses of haematologcial parameters may specifically warranted only for medicinal products with MoAs that may exert an effect on these.	Suggested revision: Special attention should be paid to incidence of neutropenia, agranulocytosis and aplastic anaemia, if relevant based on the MoA of the medicinal product.

4.7.1.7. Endocrinological adverse events and sexual dysfunction

593-594	~~Special attention should be paid to effects on sexual functioning, libido, galactorrhoea, and gynaecomastia. Investigation of neuro-endocrinological parameters relating to prolactin is necessary.~~ Safety monitoring should be based on the MoA.	Special attention should be paid to effects on sexual functioning, libido, galactorrhoea, and gynaecomastia, if relevant. Investigation of neuro-endocrinological parameters relating to prolactin may be necessary.

594	Rationale: assumed this was reference to antipsychotics but if not warranted by MOA - it should not be a requirement	Investigation of neuro-endocrinological parameters relating to prolactin is necessary, IF WARRANTED BY TOXICOLOGICAL FINDINGS OR KNOWN MECHANISM OF ACTION SUCH AS DOPAMINE RECEPTOR ANTAGONISTS

4.7.1.9. Extrapyramidal symptoms (EPS)

612	~~Tardive dyskinesia occurs late in treatment and is reported for both atypical and typical antipsychotics.~~ Clarification needed. Suggest revision for clarification.	Tardive dyskinesia HAS BEEN REPORTED FOR BOTH ATYPICAL AND TYPICAL ANTIPSYCHOTICS AND MAY OCCUR AT ANY TIME DURING TREATMENT AND SPECIAL ATTENTION SHOULD BE MADE TO MONITOR PATIENTS FOR TARDIVE DYSKINESIA, WHEN APPROPRIATE. Or alternatively, since safety monitoring should be based on the MoA: Tardive dyskinesia occurs late in treatment and may be reported for both atypical and typical antipsychotics, if relevant.

4.7.1.10. Serotonin syndrome/ Neuroleptic malignant syndrome

614-618	Suggest to delete this paragraph, which is a summary of current knowledge and does not provide any guidance. Alternatively, suggested revision: Add to end of this section - NMS AND SEROTONIN SYNDROME: Subjects SHOULD BE MONITORED FOR SEROTONIN SYNDROME OR NMS, AS APPROPRIATE AND WHEN WARRANTED Clarification to include these are AEs of interest and patients should be monitored for these specific adverse events when warranted.	Serotonin syndrome (SS) can be caused by excessive serotonergic agonism in central and peripheral nervous system serotonergic receptors and has been described for many antidepressants. The clinical symptoms include neuromuscular hyperactivity, autonomic hyperactivity and altered mental status. Neuroleptic malignant syndrome (NMS) consists of similar clinical symptoms and has been reported for all antipsychotics.

4.7.1.11. Rebound/ withdrawal phenomena/ dependence

621-625	Since the exact approach to capturing potential withdrawal effects may largely depend on the exact setting (adjunct vs monotherapy, type of population) as well as the PK of the medicinal product, being overly prescriptive in this guidance does not seem to capture various scenarios that may need to be considered.	In some of the short-term and long-term clinical studies, treatment should be stopped abruptly, and patients should be followed for a suitable duration, in other studies careful tapering off might be more appropriate depending on the mechanism of action of the medicinal product. Occurrence of rebound and/or withdrawal phenomena should be scored at the appropriate time. The approach to characterizing these effects should be topics of a scientific advice procedure.
626-627	Suggest to delete.	Animal studies will be needed to investigate the possibility of dependence in new classes of medicinal products or when there is an indication that dependence may occur. These requirements are laid out in ICH M3(R2) and the CHMP guidance on nonclinical studies for the characterisation of dependence potential (EMEA/CHMP/SWP/94227/2004)
628	Suggest to delete or encourage seeking scientific advice. The current statement does not provide any clarity on what may need to be considered by the applicant. Since further studies are expected to be largely data-driven, a more appropriate guidance text may suggest seeking scientific advice.	Depending on the results of these studies further studies in humans may be needed.

4.7.1.14. Children and adolescence

641	Validated questionnaires/ scales/ tests should be used for the assessment of adverse events Need clarification and to specify/distinguish this relates to this subset of Aes and not all AES	Validated questionnaires/ scales/ tests should be used for the assessment of THESE TYPES OF adverse events

642-643	Suggest to strike: ~~but appropriate protocols should be available when the use in children is applied for.~~ The current text suggests the need for a PASS. Since its design will be part of discussion during the MAA review, it may be more realistic to develop its protocol during this process.	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.

5. References

264-265	Suggestion to add references to support claims for epidemiology of maintenance phase time frames	