PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The Bipolar Illness Onset study - research protocol for the BIO cohort study
AUTHORS	Kessing, Lars; Munkholm, Klaus; Faurholt-Jepsen, Maria; Miskowiak, Kamilla; Nielsen, Lars; Frikke-Schmidt, Ruth; Ekstrøm, Claus; Winther, Ole; Pedersen, Bente; Poulsen, Henrik; McIntyre, RS; Kapczinski, Flavio; Gattaz, Wagner; Bardram, Jakob; Frost, Mads; Mayora, Oscar; Knudsen, Gitte; Phillips, Mary; Vinberg, Maj

VERSION 1 - REVIEW

REVIEWER	Dr Rashmi Patel Department of Psychosis Studies, King's College London, UK
REVIEW RETURNED	18-Feb-2017

GENERAL COMMENTS	Thank you for kindly inviting me to review this protocol. Overall, I think this is an excellent study which has the potential to add a great deal to our understanding of the early phase of bipolar disorder and the protocol is clear and well written. The study benefits from a long period of follow-up with comprehensive data collection including clinical data from rating scales and structured questionnaires and a wide range of biological markers (such as blood, urine, retinal and MRI measures) as well as smartphone data (both user generated as well as passive data collection). A multiple biomarker approach will be used to analyse multiple potentially predictive variables and statistical analysis methods (which include GLM, integrated data analysis and penalised regression) are appropriate and will help to meaningfully analyse the large volumes of data and reduce the risk of spurious associations due to multiple comparisons. I have a few minor suggestions: (i) The authors describe the use of cross validation or splitting the sample to evaluate the accuracy of predictive algorithms but I wondered if there were any plans to evaluate these in an external cohort to consider the external validity of their findings? (ii) The authors clearly describe statistical analysis methods for the data being collected in BIO1/2/3 but I wondered if they could provide some further description on the statistical analysis methods intended for BIO4 (the prediction of illness onset by combining multimodal data from BIO1/2/3)? Would this be some form of supervised machine learning method and, if so, which method? (iii) I wondered if the authors could describe how they might address the issue of diagnostic stability? Do all patients accepted to the

	Copenhagen Affective Disorder Clinic with presumed bipolar disorder continue to have a stable diagnosis of bipolar disorder or do a certain proportion of patients turn out to have other disorders (e.g. anxiety, personality or non-affective psychotic disorders?) and if this is the case, how would this be accounted for in analysing follow-up data? (iv) What are the plans for handling missing data (e.g. due to loss to follow-up) in the statistical analyses? (v) How will the authors manage comorbid illicit substance use which can affect some people with bipolar disorder? Will individuals with comorbid illicit substance use be excluded from this study or, if they are included, how will this be accounted for in data collection and analysis? Apart from these minor issues I think this is an excellent study protocol and I wish the authors every success in their research and look forward to reading the publication of their findings.

REVIEWER	Robert M. Post Bipolar Collaborative Network, USA
REVIEW RETURNED	18-Feb-2017

GENERAL COMMENTS	Will there be a measure of the highly replicated CACNA1C gene?

REVIEWER	Elizabeth Murnane Cornell University, USA
REVIEW RETURNED	06-Mar-2017

GENERAL COMMENTS	The research is well motivated, with clearly defined goals that are highly impactful while remaining feasible. While I provided affirmative "Yes" responses to each review checklist item, I would like to list a few suggested points for the authors to explain, justify, or otherwise explicitly address in their manuscript: - Will the number of included participants with BD type 1 and BD type 2 be balanced? What of BD-NOS and Cyclothymia (i.e., how does the inclusion criteria handle these subtypes?) - Regarding the anticipated episodes and onset incidence observed during the study period (4-5 episodes; onset for 20-30% of relatives and 2% of healthy individuals), it would helpful to justify these numbers and specify references. Also, to what extent would these estimates change if the study timeframe were shorter/longer and depending on the ages of the participants? Given BD onset is the main criterion for the N=300 participants, I'm especially wondering about the likelihood that BD will appear in both a participant and an offspring within a decade of each other. - How many of the original research staff involved at the beginning of the work will remain over the multiyear study period? A brief description would be useful about how training, procedural handoffs, etc. will be managed. - It would be good to see a few additional details regarding what consent involved and how data will be managed (storage, encryption, access) given its sensitivity and the vulnerability of the population. - The frequency of assessment could be better justified — is a

	yearly/biennial followup sufficient? - "Additionally, research assistants will contact all participants regularly to identify upcoming episodes / onset of illness and to ensure continued participation in the BIO study." How regularly, more precisely? Further, I find the point about upcoming episodes/onset confusing — how will such future events be predicted? - To what extent do the authors expect to encounter missing data, and how will this be handled during statistical analyses?

REVIEWER	Gianni Faedda Child Study Center-Department of Child and Adolescent Psychiatry – NYU - Langone Medical Center, New York, NY, US
REVIEW RETURNED	11-Mar-2017

GENERAL COMMENTS	Several concerns regarding: 1) Your research hypothesis is vague and seems overambitious, making several assumptions regarding the presence of an illness' progression (page 6, first 5 lines). 2) The process of selection of putative biomarkers to be included and excluded from this study's Methods. Please provide details of what biomarkers have been studied to date, the validity and evidence of their value in diagnostic classification (affected vs. non-affected, see http://dx.doi.org/10.1371/journal.pone.0020650, DOI: 10.1111/jcpp.12520) illness' state (Depression vs Mania, see Janovsky 1994, Hasler 2006 http://dx.doi.org/10.1016/j.biopsych.2005.11.006, Valenza 2014 10.1109/JBHI.2013.2290382) and differential (i.e. Dx A vs. Dx B), illness's variables (i.e. psychosis), course (i.e. recurrence rate) and outcome (remission, relapse, recurrence) or treatment response. 3) Then please explain how you have selected and/or excluded certain variables (convenience, cost, evidence...). 5) Your definition of Onset is vague. please clearly define as first diagnosis, vs first meeting syndromal criteria, vs first psychiatric assessment, or other. 6) Your methods suggest a bias towards recruitment of patients with BP-I, or based on the presence of Mania or hypomania; this will almost certainly exclude Bipolar disorder NOS, and BP-II. 7) There is no mention of mixed states, 8) assessment for psychosis (PANSS) 9) assessment for suicide (CSSRS), or 10) mechanisms for termination of the study if required by safety concerns. 11) There is no clear definition of high risk state to onset of bipolar: is this based on past history, family history, or other variables such as HCL-32 scores? 12) Please clarify the Age range for each of the sub-studies; mention of SCID only seems to imply that only adults will be enrolled. 13) However 15 y/o controls will be recruited, how is the exclusion of a current psychiatric diagnosis going to be made in a minor? 14) There is no mention of Informed consent procedures, assent for minors in the control group, or consenting a minor becoming an adult. 15) Do you exclude subjects with HIV-AIDS? Hepatitis B or C? TB? Syphilis? Please explain consenting procedures for such tests, and privacy protection procedures as they apply to your Country. 16) In the controls, First AND Second generation relatives should be free of Bipolar disorder diagnoses.

	17) Will you exclude adopted? 18) How do you screen for presence/absence of Family history? 19) On page 10, #9 and #10: please provide operationalized definition of partial remission and full remission. Please select what type of neuropsychological assessments will you be conducting and why. 20) Please address in the limitations the issue of multiple comparisons, the lack of clinical global assessments of severity or improvement (CGI-S), the possible correlation or lack of independence between variables and mood state.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Dr Rashmi Patel

Please leave your comments for the authors below Thank you for kindly inviting me to review this protocol. Overall, I think this is an excellent study which has the potential to add a great deal to our understanding of the early phase of bipolar disorder and the protocol is clear and well written.

The study benefits from a long period of follow-up with comprehensive data collection including clinical data from rating scales and structured questionnaires and a wide range of biological markers (such as blood, urine, retinal and MRI measures) as well as smartphone data (both user generated as well as passive data collection). A multiple biomarker approach will be used to analyse multiple potentially predictive variables and statistical analysis methods (which include GLM, integrated data analysis and penalised regression) are appropriate and will help to meaningfully analyse the large volumes of data and reduce the risk of spurious associations due to multiple comparisons.

Re: Thank you for your kind comments on the paper.

I have a few minor suggestions:

(i) The authors describe the use of cross validation or splitting the sample to evaluate the accuracy of predictive algorithms but I wondered if there were any plans to evaluate these in an external cohort to consider the external validity of their findings?
Re: It is now added to page 16 that “Finally, if possible we will evaluate the biomarker(s) in external non-Danish data sets in collaboration with other international researchers”.

(ii) The authors clearly describe statistical analysis methods for the data being collected in BIO1/2/3 but I wondered if they could provide some further description on the statistical analysis methods intended for BIO4 (the prediction of illness onset by combining multimodal data from BIO1/2/3)? Would this be some form of supervised machine learning method and, if so, which method?
Re: It is now added to page 24 that “Statistical analyses: BIO4 will use a combination of penalized regression techniques, and random forests to infer the importance of the original and combined markers and to compare the similarity of prediction error from the models with and without combining the markers”.

(iii) I wondered if the authors could describe how they might address the issue of diagnostic stability? Do all patients accepted to the Copenhagen Affective Disorder Clinic with presumed bipolar disorder continue to have a stable diagnosis of bipolar disorder or do a certain proportion of patients turn out to have other disorders (e.g. anxiety, personality or non-affective psychotic disorders?) and if this is the

case, how would this be accounted for in analysing follow-up data?
Re: It is now added to the paper at page 16 that “A general principle that pertains to statistical analyses of all four sub-studies is the “intention to treat” principle. Accordingly, participants who during follow-up get a diagnosis with a higher diagnostic validity than bipolar disorder (i.e. a lower ICD-10 diagnostic number, DF00, DF10 and DF20) that may substantially influence the biomarker measures are included in the analyses until onset of symptoms from the disorder but excluded from subsequent analyses. These disorders include significant neurological disorders such as dementia, stroke, brain tumour, multiple sclerosis, Parkinson’s disease, as well as psychiatric disorders due to significant psychoactive substance use and schizophrenia“.

(iv) What are the plans for handling missing data (e.g. due to loss to follow-up) in the statistical analyses?
Re: It is now added at page 16 that “Furthermore, in all four sub-studies, the problem of missing data will alleviated by the use of mixed effect model (for longitudinal measurements) and multiple imputations using chained equations when applicable. If possible, joint modeling will be considered, depending on the missing mechanism observed”.

(v) How will the authors manage comorbid illicit substance use which can affect some people with bipolar disorder? Will individuals with comorbid illicit substance use be excluded from this study or, if they are included, how will this be accounted for in data collection and analysis?
Re: We kindly refer to the response to point (iii) above. Thus, participants who during follow-up get a psychiatric disorders due to significant psychoactive substance are included in the analyses until onset of symptoms from the disorder but excluded from subsequent analyses. Regarding less severe substance use, a covariate for comorbid substance use (mainly alcohol and cannabis in Denmark) will be included in the statistical model.

Apart from these minor issues I think this is an excellent study protocol and I wish the authors every success in their research and look forward to reading the publication of their findings.

Reviewer: 2
Reviewer Name: Robert M. Post
Institution and Country: Bipolar Collaborative Network, USA Please state any competing interests or state ‘None declared’: none

Please leave your comments for the authors below Will there be a measure of the highly replicated CACNA1C gene?
Re: It is now added to page 15 that “The BIO study sample of 600 participants is rather small for genetic analyses discriminating patients with bipolar disorder from healthy controls but some genetic analyses, including the CACNA1C gene, can be conducted in cooperation with national and international genetic network groups”.

Reviewer: 3
Reviewer Name: Elizabeth Murnane
Institution and Country: Cornell University, USA Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below The research is well motivated, with clearly defined goals that are highly impactful while remaining feasible. While I provided affirmative "Yes" responses to each review checklist item, I would like to list a few suggested points for the authors to explain, justify, or otherwise explicitly address in their manuscript:

- Will the number of included participants with BD type 1 and BD type 2 be balanced?
Re: Thank you for this comment, the answer is no, as it is not possible to balance the number due to the naturalistic design of the study. Thus all patients referred to the Copenhagen Affective Disorder Clinic as newly diagnosed/first episode bipolar patients will be assessed for participation in the BIO study.

What of BD-NOS and Cyclothymia (i.e., how does the inclusion criteria handle these subtypes?)
Re: As ICD-10 is used, BD-NOS is not an issue .Patients with a diagnosis of cyclothymia will not be included.

- Regarding the anticipated episodes and onset incidence observed during the study period (4-5 episodes; onset for 20-30% of relatives and 2% of healthy individuals), it would helpful to justify these numbers and specify references.
Re: References are now added at page 10: “Based on prior findings (2;3;19), it is estimated that patients will develop four to five affective episodes on average during the follow-up period, i.e., relapse or recurrence, defined as above. It estimated that 20-30% of the healthy first-generation relatives will develop onset of affective illness compared to 2-5% among the healthy individuals without a first-generation family history of affective disorders (24;25).

Also, to what extent would these estimates change if the study timeframe were shorter/longer and depending on the ages of the participants?
Re: It is difficult to speculate on these issues. Effects of age on the risk of recurrence is not entirely clear but the presented estimates are for patients with a first episode bipolar between the ages of 18 and 50 (2;3;19).

Given BD onset is the main criterion for the N=300 participants, I'm especially wondering about the likelihood that BD will appear in both a participant and an offspring within a decade of each other.
Re: We have indeed been speculating on this as well. As far as we are informed, no other study has been published with a similar design as the BIO study recruiting patients and their healthy first-generation relatives at the same time and with a five to ten year follow-up. Our best estimate is that 20-30% of the healthy first-generation relatives will develop onset of affective illness (unipolar or bipolar disorder).

- How many of the original research staff involved at the beginning of the work will remain over the multiyear study period? A brief description would be useful about how training, procedural handoffs, etc. will be managed.
Re: It is now added at page 11 that “Clinical assessments will be performed by by six PhD students (Mater degree as a Medical Doctor or psychologist). They will be certified in a PhD training course in the SCAN interview and will be trained in using the rating scales (HAMD-YMRS, FAST). Inter-rater sessions supervised by senior clinicians from the Copenhagen Affective Disorder Clinic and senior researchers in the BIO study will be conducted among the researchers at regular time points during the entire study period”.

- It would be good to see a few additional details regarding what consent involved and how data will be managed (storage, encryption, access) given its sensitivity and the vulnerability of the population.
Re: It has now been added at page 25 that “Data will be saved, encrypted and assessed according to the regulations from the Capital Region of Denmark. “
The BIO study has been approved by the Local Ethical Committee (H-7-2014-007) and the data agency, Capital Region of Copenhagen (RHP-2015-023). Data will be saved, encrypted and assessed according to the regulations from the Capital Region of Copenhagen. The study complies with the Declaration of Helsinki principles (Seoul, October 2008).

- The frequency of assessment could be better justified — is a yearly/biennial followup sufficient?
Re: This section at page 9 has been revised in accordance with the procedures of the BIO-study: “Patients and healthy control individuals will be face to face assessed initially and at least every year during the first four years and after this, every second year for five years.
As part of daily clinical praxis in Copenhagen Affective Disorder Clinic and as part of the BIO-2 sub-study all patients will get access to a smartphone app for electronic continuous monitoring of illness activity during a 5-year follow-up period (see sub-study BIO-2). Additionally, research assistants will contact all participants every third months to identify upcoming episodes / onset of illness and to ensure continued participation in the BIO study”.

- "Additionally, research assistants will contact all participants regularly to identify upcoming episodes / onset of illness and to ensure continued participation in the BIO study." How regularly, more precisely? Further, I find the point about upcoming episodes/onset confusing — how will such future events be predicted?
Re: It is now clarified at page 10 that “Additionally, research assistants will contact all participants every third months to identify upcoming episodes / onset of illness and to ensure continued participation in the BIO study.“
It is further added at page 9 that “At each assessment the present, clinical state (remission, manic, hypomanic, depressive, mixed episode) of all participants will be established according to ICD-10. The severity of depressive and manic symptoms (if present) will be assessed using the 17-item Hamilton Depression Rating Scale (HAMD-17) and the Young Mania Rating Scale (YMRS) with a time period of three days applied. Remission is defined as score of < 7 on the HAMD-17 and the YMRS. In this way upcoming episodes and onset of illness (for healthy individuals) will be assessed with great certainty“.

- To what extent do the authors expect to encounter missing data, and how will this be handled during statistical analyses?
Re: We believe that the frequent assessments during follow-up will add to increase adherence to the study and decrease the prevalence of missing data. It is now added at page 16 that “Furthermore, in all four sub-studies, the problem of missing data will alleviated by the use of mixed effect model (for longitudinal measurements) and multiple imputations using chained equations when applicable. If possible, joint modeling will be considered, depending on the missing mechanism observed”.

Reviewer: 4
Reviewer Name: Gianni Faedda
Institution and Country: Child Study Center-Department of Child and Adolescent Psychiatry – NYU - Langone Medical Center, New York, NY, US Please state any competing interests or state ‘None declared’: None Declared

Please leave your comments for the authors below Several concerns regarding:
1) Your research hypothesis is vague and seems overambitious, making several assumptions regarding the presence of an illness' progression (page 6, first 5 lines).

Re: The BIO-study is not hypotheses testing but rather hypotheses generating. Notably, we do not describe any hypotheses to be tested but describes the aims of the study (at page 7) as to identify composite blood- and smartphone-based markers and neurocognitive signatures during the long-term course of bipolar disorder.
We have presented evidence for a clinical progression of mood disorder in the quoted reference (reference 9), concluding that “Although the course of illness is heterogeneous there is evidence for clinical progression of unipolar and bipolar disorders”. To specify this in the BIO protocol, we have

now added a fifth aim at page 7: “To investigate whether the course of illness is progressive on average in bipolar disorder and to identify corresponding changes in biomarkers during the course of illness within BIO-1 to BIO-4”.

2) The process of selection of putative biomarkers to be included and excluded from this study's Methods. Please provide details of what biomarkers have been studied to date, the validity and evidence of their value in diagnostic classification (affected vs. non-affected, see http://dx.doi.org/10.1371/journal.pone.0020650, DOI: 10.1111/jcpp.12520) illness' state (Depression vs Mania, see Janovsky 1994, Hasler 2006 http://dx.doi.org/10.1016/j.biopsych.2005.11.006, Valenza 2014 10.1109/JBHI.2013.2290382) and differential (i.e. Dx A vs. Dx B), illness's variables (i.e. psychosis), course (i.e. recurrence rate) and outcome (remission, relapse, recurrence) or treatment response.
Re: We have responded more overall to this point at our response to point 3.
Further, thank you for these references to the use of actigrahps in bipolar disorder (and the seminal methods Hasler et al paper on endophenotypes). The BIO study includes a large number of potential biomarkers (blood-based, smartphone-based and neurocognitive signatures) and we have now included an actigraph measure that we have previously used in bipolar disorder and referred to the papers suggested by you. We have added at page 18 that “To validate smartphone based measures of physical activity and sleep, a subset of patients will wear a combined heart rate and movement sensor mounted at the thorax that has been shown to correlate with mood symptoms and affective states and differentiate between patients and controls (Actiheart; (32;33)) like other wearable actigraphs (62-64).

3) Then please explain how you have selected and/or excluded certain variables (convenience, cost, evidence...).

Re: It is now added at page11 that “Background and reasons for selection of putative biomarkers is an iterative process presented in each sub-study. Nevertheless, as the area of potential biomarkers is constantly evolving and as the sub-studies cover four different areas, the research protocol does not include systematic reviews of the literature.”

5) Your definition of Onset is vague. please clearly define as first diagnosis, vs first meeting syndromal criteria, vs first psychiatric assessment, or other.
Re: Onset is defined at page 8 and it is now specified that the diagnosis is made according to the ICD-10 (discriminating between a first manic/ hypomanic episode and bipolar disorder): “All patients referred to the Copenhagen Affective disorder Clinic as newly diagnosed/first episode patients, i.e., onset of first manic or hypomanic episode or when the ICD-10 diagnosis of bipolar disorder is made for the first time”.

6) Your methods suggest a bias towards recruitment of patients with BP-I, or based on the presence of Mania or hypomania; this will almost certainly exclude Bipolar disorder NOS, and BP-II.
Re: The BIO study will recruit patients with BP-I and BP-II according to the initial diagnostic assessment using SCID by the experienced specialists in psychiatry during the patients two year stay in the Copenhagen Affective Disorder Clinic. This will be in addition to the SCAN interview made by the researchers (as specified at page 9 top).

7) There is no mention of mixed states, Re: Mixed states are mentioned as part of the Follow-up section at page 9. We have now further specified this section.
8) assessment for psychosis (PANSS). Re: Due to the comprehensive assessments in the BIO study we do not include the PANSS

9) assessment for suicide (CSSRS), Re: Due to the comprehensive assessments in the BIO study we do not include assements for suicide or
10) mechanisms for termination of the study if required by safety concerns. Re: As this is not an intervention study and as there is no risk or side effects to the follow-up, there is no need for this.

11) There is no clear definition of high risk state to onset of bipolar: is this based on past history, family history, or other variables such as HCL-32 scores?
Re: High risk is defined according family history as described at page 23/24 of BIO-4: "….healthy (i.e., non-syndromal level) high risk population of first degree siblings and offspring to the patients with newly diagnosed/first episode mania/bipolar disorder, included in BIO-1, BIO-2 and BIO-3".

12) Please clarify the Age range for each of the sub-studies; mention of SCID only seems to imply that only adults will be enrolled.
Re: The age range is 15-70 years of age in BIO-1, BIO-2 and BIO-3 and 15 to 40 years in BIO-4 (the high-risk group). This is now specified at page 8. The SCID interview is used for adults, only.

13) However 15 y/o controls will be recruited, how is the exclusion of a current psychiatric diagnosis going to be made in a minor?
Re: This is done according to the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interview that can be used down to age 15 years.

14) There is no mention of Informed consent procedures, assent for minors in the control group, or consenting a minor becoming an adult.
Re: It is now added at page 25 that "According to these specifications, adolescents aged 15-18 years will be invited only if parents have given consent."

15) Do you exclude subjects with HIV-AIDS? Hepatitis B or C? TB? Syphilis? Please explain consenting procedures for such tests, and privacy protection procedures as they apply to your Country.
Re: Such disorders will be recorded as severe somatic disorder and excluded from the analyses as now specified at page 16 in accordance with our response to Reviewer 1, point (iii):

Re: It is now added to the paper at page 16 that "A general principle that pertains to statistical analyses of all four sub-studies is the "intention to treat" principle. Accordingly, participants who during follow-up get a diagnosis with a higher diagnostic validity than bipolar disorder (i.e. a lower ICD-10 diagnostic number, DF00, DF10 and DF20) that may substantially influence the biomarker measures are included in the analyses until onset of symptoms from the disorder but excluded from subsequent analyses. These disorders include significant neurological disorders such as dementia, stroke, brain tumour, multiple sclerosis, Parkinson's disease, as well as psychiatric disorders due to significant psychoactive substance use and schizophrenia".

16) In the controls, First AND Second generation relatives should be free of Bipolar disorder diagnoses. Re: That is correct.

17) Will you exclude adopted?
Re: It is not possible to include adopted individuals in the high risk group; however it is no problem in the patients or control groups.

18) How do you screen for presence/absence of Family history? Thank you for this question. We have now added at page 24 that "All recruited patients will be asked about the lifetime psychiatric history of first-degree relatives (their biological parents, siblings and offspring) based on the Brief Screening for

Family Psychiatric History questionnaire described by Weissman and colleagues (98).

19) On page 10, #9 and #10: please provide operationalized definition of partial remission and full remission.
Re: This is now specified at page 9/10: “At each assessment the present, clinical state (remission, manic, hypomanic, depressive, mixed episode) of all participants will be established according to ICD-10. The severity of depressive and manic symptoms (if present) will be assessed using the 17-item Hamilton Depression Rating Scale (HAMD-17) and the Young Mania Rating Scale (YMRS) with a time period of three days applied. Remission is defined as score of < 7 on the HAMD-17 and the YMRS. In this way upcoming episodes and onset of illness (for healthy individuals) will be assessed with great certainty“.

Please select what type of neuropsychological assessments will you be conducting and why.
Re: Neuropsychological tests and reasons for choosing these are described in BIO-3 under the heading “Neurocognitive testing”

20) Please address in the limitations the issue of multiple comparisons, the lack of clinical global assessments of severity or improvement (CGI-S), the possible correlation or lack of independence between variables and mood state.
Re “multiple comparisons”: This has been discussed as a limitation in the Discussion section: Fourth, due to the large number of biological and statistical tests included in the BIO study, chance findings may occur in relation to the individual biological test. However, the aim of the BIO study is to identify a composite biomarker measure related to bipolar illness, depression and mania using cross validation or alternatively split sample designs in the development and validation of the composite biomarker.

Re “lack of clinical global assessments of severity”: We use the Functional Assessment Short Test (FAST) which is specifically developed and validated to measure global function in bipolar disorder.

Re “the possible correlation or lack of independence between variables and mood state”: This is not a general limitation, but a possible association that we take into account in different ways in the different sub-type studies, e.g. in BIO-3, neurocognition is assessed only during remission.

VERSION 2 – REVIEW

REVIEWER	Dr Rashmi Patel Department of Psychosis Studies, King's College London, UK
REVIEW RETURNED	30-Mar-2017

GENERAL COMMENTS	The authors have addressed all of my queries and I now recommend their revised protocol for publication.

REVIEWER	Elizabeth Murnane Cornell University, USA
REVIEW RETURNED	17-Apr-2017

GENERAL COMMENTS	I appreciate the steps the authors have taken in this revision to address reviewers’ previous comments; the highlighting was extremely helpful. Nearly all my questions have been resolved (e.g., the likelihood of onset in relatives, references I requested be added, the frequency of followups, what episode/onset assessment entails, staff training, data management procedures, and statistically handling missing data). I still feel the research plan is highly

	worthwhile and merits acceptance, and I have specified positive "Yes" responses to all of the high level checklist items. However, I have the following specific questions I believe the authors have yet to discuss: - How does the inclusion criteria handle the subtypes BD-NOS and Cyclothymia? (Or what is the justification for excluding these types as outside the scope of this study?) - Do the authors plan to ensure balance in the number of included participants with different types of BD?

REVIEWER	Gianni Faedda Lucio Bini Mood Disorders Center, New York, NY USA
REVIEW RETURNED	14-Apr-2017

GENERAL COMMENTS	On page 27, please correct: Finally, the BIO study does not include (randomised) interventions limiting casual interpretations of ... should say causal (not casual)

VERSION 2 – AUTHOR RESPONSE

Response to Reviewer 3:
It is now specified (p. 8) that "Nearly all patients treated in the Clinic have a diagnosis of bipolar disorder type I or type II whereas patients with bipolar disorder-not otherwise specified (NOS) or patients with cyclothymia not are treated in the Clinic and consequently not included in the BIO study". It is specified (p. 9) that "There will be no attempt to balance the prevalence of bipolar subtypes in the BIO study."

Response to Reviewer 4:
The spelling error has now been corrected.