PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Clinical, neurocognitive, and demographic factors associated with functional impairment in the Australian Brain and Mind Youth Cohort Study (2008-2016)
AUTHORS	Lee, Rico; Hermens, Daniel; Naismith, Sharon; Kaur, Manreena; Guastella, Adam; Glozier, Nick; Scott, Jan; Scott, Elizabeth; Hickie, Ian

VERSION 1 – REVIEW

REVIEWER	Sarah Kertz Southern Illinois University Carbondale, USA
REVIEW RETURNED	02-Jun-2018

GENERAL COMMENTS	Overall this is a well written paper that has the potential to make a nice contribution to the literature. The strengths include the large sample size, the inclusion of broad, and dimensional transdiagnostic measures of psychopathology, the inclusion of neurocognitive factors, and a focus on functional outcomes. There is a lot to value about this manuscript. However, there are a few areas that would benefit from additional attention. I have outlined them in detail below, but broadly I would suggest focusing on (1) giving more attention to the developmental range of the sample (and clarifying language around “youth” given the sample includes people in their 30s) and (2) including greater detail in the results and method section. I hope the authors find these comments helpful. Introduction 1. Overall I find the rationale compelling and timely. The focus on transdiagnostic symptoms and functional outcomes is important and consistent with the direction of the field. That said, a more cohesive theory or more rationale for the associations between symptoms, functioning, and substance use would be helpful, especially in terms of the causal direction of the effects. Although this is cross-sectional data, pulling on the directionality of effects from theory might help place the findings in a broader context. 2. While the authors explain in a concise way their rationale for the transdiagnostic and dimensional framework, the issue of development is completely overlooked. This is problematic given the significant age range covered in the study. Most studies do not consider people in mid to late 20s and 30s “youth” and so this could be misleading. Additional attention to the influence of cognitive and emotional development on relevant outcomes is warranted, as is additional justification for collapsing across what I would consider multiple different developmental phases. Similarly,

	additional rationale and justification for splitting the groups by age (and at the cut-off points selected) is needed. 3. The authors use the term “early stage mental disorders” but it’s not clear what this means. Onset before age 36 is not typically considered “early.” Please clarify. Method 4. Additional details on study staff would be helpful. How many of each kind of team member were there (clinical psychologist, clinical neuropsychologist, research psychologist (not clear what this means) and what kind of training did they receive? 5. Please explain how diagnoses were made, who made them, and what training and credentials the interviewers had. Were there any estimates of reliability? 6. Please provide estimates of the measure psychometric properties in this sample (internal consistency). 7. Please provide additional detail on the measure (if space allows), including at a minimum the scale of the items and how scores are interpreted. 8. Please provide means, SDs, (and correlations if space allows) between study variables. 9. How much data for each variable were missing? Please describe. 10. What was the distribution of the data? Please provide additional justification for using the ML estimator if data were non-normal, as this is typically not acceptable and an ADF or robust adjusted estimator may be more appropriate. 11. I found the description of the moderator analyses (testing for model invariance across groups) a little lacking intechnical detail and relying heavily on the built in function of the statistical software. Please revise this section to better reflect the analytic approach rather than the AMOS procedure. Results 12. The term “psychotics” on page 11 should be changed. 13. Please include additional details about the “developmental or behavior disorders.” 14. It’s not clear what “best fitting factor structure” for the neuropsyc measures means. Please first provide information on the model fit for each of the latent variables, then move on to testing the full model. 15. The description of the core clinical dimension model is also unclear. The authors refer to “only three” but do not explain relative to what other measures? What does the “only” mean? 16. On page 12 the authors note that a two factor model emerged as the best fitting model but additional details need to be provided.

	Explain to the reader how this conclusion was made and include all relevant fit statistics and tests. 17. It's not clear to me why the authors first test single predictor models and then the full model? Please explain. 18. There was no discussion or hypothesis about gender effects, diagnosis, or medication use in the introduction to warrant the moderator analysis. Please revise. Discussion 19. The authors write on page 14 "The findings are relevant as they demonstrate that whilst neurocognitive impairment may undermine functioning in those with psychotic disorders, they are not specific to such cases." Please discuss a little more as I'm not sure I understand the meaning. 20. The authors refer to a lack of research in older adults on page 15 but given the explicit focus on this paper on younger populations, this seems out of place. Other notes: 21. Please consult the APA style manual for notation in the results. For example, n and SD should be italicized. Similarly, the alphabetical notation used to describe the models is not consistent with the headings outlined by the style manual. Statistics should be offset with commas, not parentheses.

REVIEWER	Alexandre Dumais Department of psychiatry and addictology, University of Montreal, Canada
REVIEW RETURNED	04-Jun-2018

GENERAL COMMENTS	This is an interesting study that focuses on young people aged 12 to 36, particularly adolescents and young adults. The purpose of the study was to determine the association between clinical, neurocognitive, and alcohol and drug use and the functioning of individuals. However, the article has several problems: in the introduction at the level of the conceptual framework and the literature review as well as methodological problems that make the conclusions not well supported. More specifically, the introduction brings us to a dimensional approach. We are told towards the end of the introduction that a transdiagnostic approach allows us to specify what is shared and unique according to the diagnoses but we do not specify what is unique to each diagnosis and in particular on the dependent variable of this study, the functioning. It would have been interesting if a more in-depth discussion had been made on the differential functioning according to the different disorders. It would also have been interesting to have a discussion on personality disorders that have a significant functional impact when present, but unfortunately this is missing from the introduction. In addition, it would have been interesting to discuss the functional aspects according to the type of consumption, that is to say, to be a substance user versus someone with a disorder and to specify the functioning according to the substance.

	In the method section, I found several problems: The measure of social functioning is quite broad and does not allow to see which part of the functioning is more deficient in an individual. As far as I understand, we only refer to the whole functioning of a person. The BPRS is an interesting tool for measuring psychiatric symptoms in general and in particular in cases of schizophrenia but does not allow to specify well symptoms such as anxiety and depression. Psychosis and mania are also evaluated in a rough way in this evaluation. The measurement of alcohol and drug use is essentially a screening measure. It is not in my opinion possible to be able to determine who has a consumption problem, which could have a significant impact on the functioning. The analysis was made in an undifferentiated manner according to the diagnoses which are multiple as shown in Table I. It would have been interesting to have an analysis according to the diagnoses to specify what is unique according to the diagnosis. This would have been interesting because we can observe that the participants who have a diagnosis of psychosis have a lower overall functioning. In addition, it can be observed that the symptomatology of the patients were relatively low during the evaluation, which may bias the final model. We therefore find patients who are not very symptomatic and evaluated in a coarse manner on the majority of the measurements, which favors the neurocognitive disorders which have been evaluated in a more exhaustive way and which are usually disorders that persist over time. The analysis does not therefore make it possible to separate the effect of symptomatology, alcohol and drugs consumption, personality elements, as well as neurocognitive aspects of the functioning of the individual.

REVIEWER	Antonio Vita Professor of Psychiatry, University of Brescia, Italy
REVIEW RETURNED	04-Jun-2018

GENERAL COMMENTS	In this paper, 1003 outpatients were recruited, aged between 12 and 36 years with baseline diagnoses of affective, psychotic, developmental or behavioural disorders. Social and occupational functioning was used to index level of functional impairment. Structural equation modelling was employed to examine associations between neurocognition, core clinical symptoms, alcohol and substance use, sleep and circadian changes, and clinician- and researcher-rated functional impairment. Independent of diagnosis, neurocognitive impairments, and depressive, anxiety and/or negative symptoms, were significantly associated with functioning. The authors conclude that, in a clinically representative sample of youth, the key determinants of functioning may not be disorder specific. This is an interesting, well presented, study, on a relevant issue, of both clinical and heuristic value. The Authors may wish to consider the following issues: a. Some of the assessment tools used in the study had been especially developed or used for specific diagnoses, and may not

	capture the variety and heterogeneity of functioning dimensions, symptomatology or cognitive characteristics of such diverse conditions as those included in the analysis. It may be well that the associations found with functioning, independent from diagnosis, coukd have been somehow affected by the assessment instruments used. This possibility should be discussed b. The rationale of including the treating clinician's evaluations besides those of independent raters is not well clarified. Also, the concordance between raters' and clinicians' evaluations should be provided c. Could an analysis of the moderating role of type and dose of medication be performed?

VERSION 1 – AUTHOR RESPONSE

R1: A more cohesive theory or more rationale for the associations between symptoms, functioning, and substance use would be helpful, especially in terms of the causal direction of the effects. Although this is cross-sectional data, pulling on the directionality of effects from theory might help place the findings in a broader context.

Response to R1: We have added a sentence in the final paragraph of the Introduction highlighting a systematic review we had previously conducted, which identified the influence of these specific, key domains to functioning. We have also added in the Discussion more detailed discussion of the issue of causality, drawing upon past literature.

R2: While the authors explain in a concise way their rationale for the transdiagnostic and dimensional framework, the issue of development is completely overlooked. This is problematic given the significant age range covered in the study. Most studies do not consider people in mid to late 20s and 30s "youth" and so this could be misleading. Additional attention to the influence of cognitive and emotional development on relevant outcomes is warranted, as is additional justification for collapsing across what I would consider multiple different developmental phases. Similarly, additional rationale and justification for splitting the groups by age (and at the cut-off points selected) is needed.

Response to R2: In line with the generally accepted definition of what youth constitutes as well as past studies, we have conducted an additional sensitivity analysis constricting our modelling to those aged 15 to 25 years of age and added these findings the Results section. Briefly, negligible changes to the findings occurred when we age-restricted our sample.
Previously, in our data analysis section of Methods we stated that our moderators were chosen to be dichotomous to maintain statistical power within sub-groups for this statistical procedure. We have added an additional clarifying statement to ensure it is clear why we have split the groups by age for the moderator analysis: "For instance, the median-split on age was performed to determine whether the model held for both younger and older individuals whilst maintaining statistical power." The moderator analysis together with our sensitivity analysis provides strong support that our findings are robust and hold across our age range. We have also added an additional sentence in limitations in the Discussion to acknowledge the potential influence of cognitive and emotional development on outcomes (which is beyond the main aims of the current paper, but we acknowledge its importance for further research).

R3: The authors use the term "early stage mental disorders" but it's not clear what this means. Onset before age 36 is not typically considered "early." Please clarify.

Response to R3: Given that a small percentage of individuals are in the older end of the age range and may not be necessarily in the earliest stages of mental illness, we have removed the modifier "early-stage" from the first line of the Discussion. In the Introduction, 'early stage' was used to summarise findings from clinical cohorts either; 1/ where the clinical syndrome had not met diagnostic threshold; or 2/ in first-episode psychosis or first-episode depression. Given that these were referenced to specific papers of these cohorts, we argue it would have interfered with the flow of arguments in the first paragraph of the introduction to further clarify further what was meant by early stage.

R4: Additional details on study staff would be helpful. How many of each kind of team member were there (clinical psychologist, clinical neuropsychologist, research psychologist (not clear what this means) and what kind of training did they receive?

Response to R4: The clinical cohort study was conducted over 8 years, with a number of clinical psychologists and neuropsychologists conducting the assessments or supervising research psychologists to conduct these assessments over these 8 years. We have clarified the term 'research psychologist' to avoid any confusion (i.e., graduate-level academic psychologist).

R5: Please explain how diagnoses were made, who made them, and what training and credentials the interviewers had. Were there any estimates of reliability?

Response to R5: As outlined in Methods, senior, treating consultant psychiatrists (>20 years of clinical experience) made the diagnoses according to DSM-IV, with any disagreements resolved at a consensus meeting with the whole treating team. The latter has been added in Methods for clarification. We concede these were not SCID-confirmed diagnoses as we have no specific reliability data of our diagnoses. However, we argue that our approach is transdiagnostic in nature and there is less of an emphasis on strict DSM diagnoses. We have accordingly added this point in limitations paragraph to reflect this, as well as clarification in methods that all clinicians were board-certified.

R6: Please provide estimates of the measure psychometric properties in this sample (internal consistency).

Response to R6: All our measures (structured scales and neuropsychological tests) are well-validated in the literature with good to excellent psychometric properties (these are all cited in the measures section).

R7: Please provide additional detail on the measure (if space allows), including at a minimum the scale of the items and how scores are interpreted.

Response to R7: We have added details on how to interpret the scores for each of the primary measures (directionality). We have referenced the appropriate papers for further details on the structure of the scale, including the scale of the items and argue that all measures used are established scales with strong psychometric properties, and are well known and not obscure or bespoke measures.

R8: Please provide means, SDs, (and correlations if space allows) between study variables.

Response to R8: All means and SDs for symptom domains, sleep, alcohol and substance use and social and occupational functioning were provided in Table 1 by diagnostic groups. We have added an additional table (Table 2) with means and SDs of the neuropsychological measures (we did not add this in Table 1 to avoid clutter and assist with ease of interpretation).

R9: How much data for each variable were missing? Please describe.

Response to R9: Diagnostic, demographic and functioning data were available for all participants. In total, 9.1% of data were missing for functioning, 18.8% for neurocognition, 12.7% for clinical symptoms and disturbed sleep, and 17.7% were missing for alcohol and substance use. Nonetheless, each analysis (univariate and multivariate) had >80% of cases with complete data. All this is now outlined in Methods.

R10: What was the distribution of the data? Please provide additional justification for using the ML estimator if data were non-normal, as this is typically not acceptable and an ADF or robust adjusted estimator may be more appropriate.

Response to R10: All endogenous variables passed inspection of the Q-Q plot test. Please see analysis section in methods for more details (i.e. formal statistics were not used as these are overly sensitive in large samples like with the present study).

R11: I found the description of the moderator analyses (testing for model invariance across groups) a little lacking in technical detail and relying heavily on the built-in function of the statistical software. Please revise this section to better reflect the analytic approach rather than the AMOS procedure.

Response to R11: We have modified our description as per the reviewer's suggestion.

R12: The term "psychotics" on page 11 should be changed.

Response to R12: This was our oversight – it was a typo and was meant to read "psychotic [disorder]". We have now corrected this error.

R13: Please include additional details about the "developmental or behavior disorders."

Response to R13: We refer you to the specific diagnoses in the footnote of Table 1 for details on the specific developmental and behavioural disorders described in our study.

R14: It's not clear what "best fitting factor structure" for the neuropsyc measures means. Please first provide information on the model fit for each of the latent variables, then move on to testing the full model.

Response to R14: We have provided clarification on this matter and provided the full model fit statistics for the tested factor structures.

R15: The description of the core clinical dimension model is also unclear. The authors refer to "only three" but do not explain relative to what other measures? What does the "only" mean?

Response to R15: We have clarified in the results that 'only' meant depression and anxiety, negative, and positive symptoms, whereas mania and disorientation constituted the other clinical dimensions

R16: On page 12 the authors note that a two-factor model emerged as the best fitting model but additional details need to be provided. Explain to the reader how this conclusion was made and include all relevant fit statistics and tests.

Response to R16: We have accordingly added details on how we reached the conclusion of a best-fitting two-factor model in results.

R17: It's not clear to me why the authors first test single predictor models and then the full model? Please explain.

Response to R17: We have further clarified this issue in methods (italics is added text). "…we used SEM to test the structural model (i.e., the relationship between predictors and social and occupational functioning) at both the single-predictor and the overall levels in order to explore potential predictors and delineate unique contributions. This was done in a two-step process – first, by testing individual predictors and then by testing the combined predictors - to quantify the amount of overlapping and unique explanatory power."

R18: There was no discussion or hypothesis about gender effects, diagnosis, or medication use in the introduction to warrant the moderator analysis. Please revise.

Response to R18: We have aimed a secondary aim to determine the influence of clinical and demographic factors given the heterogeneity of a youth mental health sample. Given the aim was secondary, we argue that it is exploratory and specific hypotheses were not essential.

R19: The authors write on page 14 "The findings are relevant as they demonstrate that whilst neurocognitive impairment may undermine functioning in those with psychotic disorders, they are not specific to such cases." Please discuss a little more as I'm not sure I understand the meaning.

Response to R19: As per suggested, we have added a sentence to further clarify our intended meaning.

R20: The authors refer to a lack of research in older adults on page 15 but given the explicit focus on this paper on younger populations, this seems out of place.

Response to R20: As per request, this has been omitted.

R21: Please consult the APA style manual for notation in the results. For example, n and SD should be italicized. Similarly, the alphabetical notation used to describe the models is not consistent with the headings outlined by the style manual. Statistics should be offset with commas, not parentheses.

Response to R21: We have followed traditional formatting requirements of medical journals and not that BMJ Open do not use APA formatting.

In response to the comments by Reviewer 2, please see replies to each point as follows:

R22: We are told towards the end of the introduction that a transdiagnostic approach allows us to specify what is shared and unique according to the diagnoses but we do not specify what is unique to each diagnosis and in particular on the dependent variable of this study, the functioning. It would have been interesting if a more in-depth discussion had been made on the differential functioning according to the different disorders. It would also have been interesting to have a discussion on personality disorders that have a significant functional impact when present, but unfortunately this is missing from the introduction. In addition, it would have been interesting to discuss the functional aspects according to the type of consumption, that is to say, to be a substance user versus someone with a disorder and to specify the functioning according to the substance.

Response to R22: We have added more in-depth discussion of the diagnostic findings and how this informs what is unique and different to disorders in the neurocognition paragraph of discussion (given neurocognition was the main domain implicated in the diagnosis moderator analyses). We argue that

personality disorders were beyond the scope of the current paper as we focused on Axis I disorders (according to DSM-IV-TR). However, we do acknowledge its importance and have accordingly added this to limitations. Further, we examined the impact of substance use in our SEM analyses and argue that examining the functional implications of whether an individual is categorically a substance user would be a blunt approach to analysis of our data (whereas we examined substance use on a dimension/continuum).

R23: The measure of social functioning is quite broad and does not allow to see which part of the functioning is more deficient in an individual. As far as I understand, we only refer to the whole functioning of a person.

Response to R23: We agree that there are different approaches to indexing different domains of functioning (which we did not seek to do). However, the current inter-rater reliability (ICC = 0.7) of SOFAS scores between two trained raters supports the reliability of our SOFAS score.

R24: The BPRS is an interesting tool for measuring psychiatric symptoms in general and in particular in cases of schizophrenia but does not allow to specify well symptoms such as anxiety and depression. Psychosis and mania are also evaluated in a rough way in this evaluation. The measurement of alcohol and drug use is essentially a screening measure. It is not in my opinion possible to be able to determine who has a consumption problem, which could have a significant impact on the functioning.

Response to R24: We had originally included in our discussion the limitations of our measures. We have further elaborated on these in the limitations section. Of note, this is the first time a well-powered attempt has been made at disentangling this broad range of factors that may contribute to functional impairment in a very large sample of young patients (whereas the vast majority of past studies have focussed narrowly on one or two domains that may be linked to functioning, and looked exclusively at chronic schizophrenia).

R25: The analysis was made in an undifferentiated manner according to the diagnoses which are multiple as shown in Table I. It would have been interesting to have an analysis according to the diagnoses to specify what is unique according to the diagnosis. This would have been interesting because we can observe that the participants who have a diagnosis of psychosis have a lower overall functioning. In addition, it can be observed that the symptomatology of the patients were relatively low during the evaluation, which may bias the final model.

Response to R25: We have added in our manuscript a detailed discussion about the influence of diagnosis. Further, we argue that the range in symptomatology in our sample reflects those typically seen in outpatient youth mental health settings (our population is clearly outlined in methods to ensure this is clear to the reader), and those with more severe symptomatologies (e.g., inpatients) are beyond the scope of our paper.

R26: We therefore find patients who are not very symptomatic and evaluated in a coarse manner on the majority of the measurements, which favors the neurocognitive disorders which have been evaluated in a more exhaustive way and which are usually disorders that persist over time. The analysis does not therefore make it possible to separate the effect of symptomatology, alcohol and drugs consumption, personality elements, as well as neurocognitive aspects of the functioning of the individual.

Response to R26: We have added a sentence in limitations to reflect the issue that future studies need to use more detailed measures of these dimensions to clarify more definitively determine whether the current effect size for neurocognition holds.

In response to the comments by Reviewer 3, please see replies to each point as follows:

R27: Some of the assessment tools used in the study had been especially developed or used for specific diagnoses, and may not capture the variety and heterogeneity of functioning dimensions, symptomatology or cognitive characteristics of such diverse conditions as those included in the analysis. It may be well that the associations found with functioning, independent from diagnosis, could have been somehow affected by the assessment instruments used. This possibility should be discussed.

Response to R27: The limitations around the measures used are now discussed in limitations.

R28: The rationale of including the treating clinician's evaluations besides those of independent raters is not well clarified. Also, the concordance between raters' and clinicians' evaluations should be provided.

Response to R28: The inter-rater reliability was good (ICC = 0.70; in manuscript). Multiple ratings were conducted to ensure reliability of this single score, which was highlighted by our original justification: "This composite score was derived to obtain a more reliable estimate of real-world functioning."

R29: Could an analysis of the moderating role of type and dose of medication be performed?

Response to R29: The sheer breadth and diversity of medications prescribed, doses, and polypharmacy meant that a quantitative analysis of how specific types of medications influenced the current findings was not possible. We have added this briefly as a limitation.

VERSION 2 – REVIEW

REVIEWER	Sarah Kertz Southern Illinois University
REVIEW RETURNED	

GENERAL COMMENTS	Were data missing at random or was there a systematic pattern to the missingness? Did subjects with and without missing data differ on important variables? It would be helpful to know if there were other factors that might be involved. I still have a few questions about normality of the data. The authors note that the endogenous variable was normally distributed based on the Q-Q plot, but what about the other study variables?

VERSION 2 – AUTHOR RESPONSE

R1: Were data missing at random or was there a systematic pattern to the missingness? Did subjects with and without missing data differ on important variables? It would be helpful to know if there were other factors that might be involved.

Response to R1: We have conducted a supplementary analysis and found that out of all the variables we have on those who are missing, those with missing data were more likely to be younger and have an anxiety disorder diagnosis (so not MAR), albeit the effect sizes were small. This has been added to methods, as well as in limitations in the discussion.

R2: I still have a few questions about normality of the data. The authors note that the endogenous variable was normally distributed based on the Q-Q plot, but what about the other study variables?

Response to R2: Based on visual inspection of the frequency histograms and assessment of the Q-Q plot, the predictor/exogenous variables that departed from normality were positive symptoms, negative symptoms, mania, disorientation, trail making test-a, and trail making test-b, which were all observed to have a slight positive skew (no others were skewed). We note that prior studies have found that MLE methods (as currently conducted) are robust in cases where variables depart from normality (such as in the current case) in terms of both the overall model fit and parameters estimates when N > 600 (Muthén and Kaplan, 1985). We concede that there are other approaches to conducting SEM in such cases (as the reviewer had rightly suggested), however, procedures such as ADF SEM require no missing data (as conducted in AMOS) and would affect the generalisability of findings as well as statistical power in the current study. We understand the reviewer's concern in this regard and have additionally added this as a limitation in the discussion. Further, the normality findings stated here as also incorporated into the manuscript.