PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Eye-tracking in adult depression: protocol for a systematic review and meta-analysis
AUTHORS	Noyes, Blake; Biorac, Aleks; Vazquez, Gustavo; Khalid-Khan, Sarosh; Munoz, Douglas; Booij, Linda

VERSION 1 – REVIEW

REVIEWER	Sprenger, Andreas University of Lübeck, Neurology
REVIEW RETURNED	09-Nov-2022

GENERAL COMMENTS	The authors present a protocol to perform a systematic review and a meta analysis on eye movement behavior and adult depression. Eye movements have been shown to be sensitive parameters for dissociating healthy participants and patients in several studies including major depressing disorder (e.g. Takahashi et al. Frontiers in Psychiatry, 2021). The authors present a straightforward strategy to perform the review with experts from several neuroscience disciplines including eye movements (DPM). Some comments - The authors should have an eye on the method of saccadic peak velocity calculation in the selected papers. There are numerous articles using fixed target amplitudes and averaging the peak velocity. A calculation like this underestimates the real peak velocity because patients may perform smaller saccades (lower gain) that have lower but normal peak velocity (Leigh & Zee, 2015). Saccade peak velocity should be calculated by using a linear regression (for saccades < 10°) or 2nd order/polynomial fits for larger saccades - I highly recommend to use the expertise of DPM in the decision process of selecting articles. - Instead of using Excel sheets the authors might think about using a review managing software (e.g. RevMan, Cochrane Training, https://training.cochrane.org/online-learning/core-software/revman) which is free of charge.

REVIEWER	Reuter, Benedikt MSB Medical School Berlin GmbH
REVIEW RETURNED	06-Jan-2023

GENERAL COMMENTS	Review BMJ open 2022 069256 Article: Noyes et al., Eye tracking in adult depression: protocol for a systematic review and metanalysis

	The manuscript presents the protocol for a systematic review and meta-analysis on eye-tracking research in adult depression. The topic is of high scientific interest. The ms is clearly written and specifies the aims, methods and procedures of the planned study in a very comprehensible way. Nevertheless, there are several issues that should be clarified before publication. The only major issue refers to the conceptual framework of the planned review. It seems to be defined by a purely technical issue ("eye tracking methodology"). Although this is legitimate, the authors should elaborate the rationale of choosing a technical issue rather than a theoretical framework to define the scope of their review. If theoretical issues actually play a role, these should be outlined. Eye tracking measures can be used to investigate features of the oculomotor response system, but they may also be used, for example, to analyze attentional and cognitive processing of different content (e.g. emotional). Minor issues: Abstract 1) p. 3: The term "psychiatric" refers to a medical discipline rather than to a well-defined set of disorders. I suggest to use the term "mental disorders" instead (applies to several passages in the manuscript) Background 2) General: Although the background is well presented, the current version does not acknowledge the heterogeneity of depression. Is there any pervious evidence of how different features of depression (e.g., episodic vs. persistent, severity, acute vs. remitted, clinical vs. subthreshold) influences eye tracking measures? 3) p. 5, l. 15: You may also refer to the ICD-10/11 as the second major and international classification system 4) p.5, l. 26: "is" instead of "are"? 5) p. 5, l.49: The term "eye tracking task" is a bit misleading as tracking is not the task but a method of assessing the dependent variable. You may wish to speak of "eye movement tasks" or "eye tracking measures" 6) p. 6, l. 8: Eye tracking methods are also used to identify impairments of cognitive or motor processes in order to improve psychological models of symptom developments (e.g., attentional processes). As both your search strategy and your eligibility criteria do not seem to exclude such studies, you may mention that the potential of eye tracking methods goes beyond identifying "neural markers, which may lead to early identification in at-risk individuals." 7) p. 6, l. 12: As the terms task and paradigm are often used in a nearly synonymous way, you may write "tasks or paradigms" 8) p. 6, 2nd section: As eye-tracking appears to be a superior method in assessing attentional biases – a prominent area of

	research in depression – you may explicitly include (overt) attention shifting (e.g., Sanchez et al, J Abnorm Psychol, 2013 122(2):303-13. doi: 10.1037/a0031529) 9) p. 7, l. 15: “and” before “increased error rate…” 10) p. 7, l. 38: Why “also”? 11) p. 8, l. 38: You might mention attention shift tasks here (see issue 7) 12) p. 8, l. 42: Please specify accuracy. The term can refer to both qualitative (e.g., correct response, error) and quantitative issues (e.g., degree of visual angle, location of saccade landing point) 13) p. 9, top of page: How does a specific parameter qualify as a biomarker? Can you specify criteria? Method 14) p. 9, l.26: What means “clinically rated or diagnosed”? Would it be a sufficient to have a symptom rating by a clinician without any specified procedure of classificatory diagnostics? 15) p.9, types of participants: Although the inclusion of individuals with a primary diagnosis of major depressive disorder actually implies that patients remitted from MDD will be considered, it might be helpful to explicitly state whether you are planning to this. 16) p.9, l. 45: Do you mean studies that used mood induction procedures to induce depression-like states in healthy participants or does the exclusion also refer to studies that used mood induction to vary mood states in patients with diagnoses from the depression spectrum (e.g., patients remitted from MDD)? 17) p. 9, l. 49: As some studies choose to include control participants without current mental disorders but do not exclude or even assess a history of all mental disorders, this criterion might lead to the exclusion of actually relevant studies. It might be better to include studies with less selective control groups but to consider the type of control group in the analyses. Regarding the latter it might even be relevant whether standardized clinical interviews are applied to exclude previous depressive episodes in control groups (see von Koch et al., 2023, Behav Res Ther, 160:104231, for a discussion of this issue) 18) p. 11, l. 49: You may reconsider the inclusion of conference papers, as quality control is typically lower compared to journal articles and many conference papers might be related to journal articles. 19) p. 11, l.51: Should it be “will not be limited”? Discussion 20) p.13, effect measures: Please specify, whether you are planning to analyze data separated by groups (e.g., major depressive disorder, dysthymia, etc.). Which groups will be built? I also suggest to separate different subgroups of major depressive disorder, especially acute and remitted groups. But you may also consider to distinguish between severe depressive episodes and

	moderate to mild depressive episodes. 21) p.15, first two sentences: regarding the term “psychiatric”, please see comment #1

VERSION 1 – AUTHOR RESPONSE

REVIEWER 1
1. The authors should have an eye on the method of saccadic peak velocity calculation in the selected papers. There are numerous articles using fixed target amplitudes and averaging the peak velocity. A calculation like this underestimates the real peak velocity because patients may perform smaller saccades (lower gain) that have lower but normal peak velocity (Leigh & Zee, 2015). Saccade peak velocity should be calculated by using a linear regression (for saccades < 10°) or 2nd order/polynomial fits for larger saccades.
Thank you for bringing this to our attention. We are well aware of the main sequence relationships and will carefully scrutinize any velocity comparisons that are not explicitly amplitude matched. This has been added on page 8. We will keep an eye out for this when reading the literature.

2. I highly recommend to use the expertise of DPM in the decision process of selecting articles.
We will increase the role of DPM in the decision process. DPM and LB will now oversee the article selection as a whole, in addition to handling inter-rater discrepancies. DPM will be providing input on the eye tracking (e.g., evaluation of methodology), and LB will provide input on the clinical aspects of the studies. This has been added on page 10.

3. Instead of using Excel sheets the authors might think about using a review managing software (e.g. RevMan, Cochrane Training, which is free of charge.
Thank you for this suggestion. We will use RevMan for our review managing and have identified this on page 10.

REVIEWER 2 – Major Issue
1. The only major issue refers to the conceptual framework of the planned review. It seems to be defined by a purely technical issue (“eye tracking methodology”). Although this is legitimate, the authors should elaborate the rationale of choosing a technical issue rather than a theoretical framework to define the scope of their review. If theoretical issues actually play a role, these should be outlined. Eye tracking measures can be used to investigate features of the oculomotor response system, but they may also be used, for example, to analyze attentional and cognitive processing of different content (e.g. emotional).
Thank you for your concern. The rationale of choosing a technical issue was that this would allow for easier comparability between studies. Eye-tracking is a tool to interpret systems neuroscience, and each eye movement task targets specific parts of the brain, responsible for attention and cognitive processing as you mentioned, as well as oculomotor control. Prior reviews, such as Carvalho et al (2015), organized their results based on eye movement tasks.
When introducing these eye movement tasks in the systematic and meta-analytic review, we will describe the brain areas and networks necessary to execute the tasks properly. After reviewing the results of how individuals with depression perform on each of these tasks, we will discuss what this means for areas such as processing of emotional information, attentional biases, and cognitive control.
To address the rationale of this framework in the text, we have edited the paragraph on uses of eye-tracking on page 1.

REVIEWER 2 – Minor Issues
1) p. 3: The term “psychiatric” refers to a medical discipline rather than to a well-defined set of disorders. I suggest to use the term “mental disorders” instead (applies to several passages in the

manuscript)
Thank you for the suggestion. We have adapted the use of “mental disorders” throughout the manuscript.

2) General: Although the background is well presented, the current version does not acknowledge the heterogeneity of depression. Is there any pervious evidence of how different features of depression (e.g., episodic vs. persistent, severity, acute vs. remitted, clinical vs. subthreshold) influences eye tracking measures?
Thank you for this important thought. There is some previous evidence on how different features of depression impact eye tracking measures. We have addressed this on page 4 with the following passage: “Although much of the research in this area includes the use of one patient group in comparison to controls, some studies have looked at performance on eye-tracking tasks in relation to differential features of depression. For example, non-melancholic depressed patients have been found to perform more similarly to control participants than melancholic depressed patients for most saccade tasks.[17] Sears and colleagues[18] found that while currently dysphoric participants exhibited initial attention orientation biases for both depression-related and positive images, previously depressed participants only exhibited a bias for depression-related images. This pattern of attention to happy faces was also identified by Isaac and colleagues[19], where currently depressed individuals looked at happy faces less than healthy controls, but there was no difference between individuals with remitted depression and healthy controls in looking time. Severity of depressive scores has also correlated with more abnormal performance on their free-viewing task[20]. Different symptom groupings of depression, patient history, and severity of depression have all been found to impact eye-tracking measures, and thus need to be considered when synthesizing results.”

3) p. 5, l. 15: You may also refer to the ICD-10/11 as the second major and international classification system
Thank you for this suggestion. We had added mention of the ICD-11 on page 1.

4) p.5, l. 26: “is” instead of “are”?
Thank you for pointing this out. We have addressed this and rearranged this sentence on page 1.

5) p. 5, l.49: The term “eye tracking task” is a bit misleading as tracking is not the task but a method of assessing the dependent variable. You may wish to speak of “eye movement tasks” or “eye tracking measures”
Thank you for the suggestion. We have changed this to “eye tracking measures” on page 6.

6) p. 6, l. 8: Eye tracking methods are also used to identify impairments of cognitive or motor processes in order to improve psychological models of symptom developments (e.g., attentional processes). As both your search strategy and your eligibility criteria do not seem to exclude such studies, you may mention that the potential of eye tracking methods goes beyond identifying “neural markers, which may lead to early identification in at-risk individuals.”
Thank you for the suggestion. We have edited the paragraph regarding the usage of eye-tracking to incorporate the above implications, on page 1.

7) p. 6, l. 12: As the terms task and paradigm are often used in a nearly synonymous way, you may write “tasks or paradigms”
Thank you for the suggestion. We have changed this to “tasks or paradigms” on page 2.

8) p. 6, 2nd section: As eye-tracking appears to be a superior method in assessing attentional biases – a prominent area of research in depression – you may explicitly include (overt) attention shifting (e.g., Sanchez et al, J Abnorm Psychol, 2013 122(2):303-13. doi: 10.1037/a0031529)
Thank you for this suggestion, we will include such studies. This addition has been addressed in the abstract and on pages 2, 5, 7, 8, and 10.

9) p. 7, l. 15: “and” before “increased error rate…”
Thank you for noticing this error. We have addressed it on page 3.

10) p. 7, l. 38: Why “also”?
We have changed the wording here, see page 3.

11) p. 8, l. 38: You might mention attention shift tasks here (see issue 7)
Thank you for the suggestion. This has been addressed on page 5.

12) p. 8, l. 42: Please specify accuracy. The term can refer to both qualitative (e.g., correct response, error) and quantitative issues (e.g., degree of visual angle, location of saccade landing point)
Thank you for this suggestion. We will consider both qualitative and quantitative accuracy and have added this clarification in the manuscript on pages 5 and 7.

13) p. 9, top of page: How does a specific parameter qualify as a biomarker? Can you specify criteria?
Thank you for that critical question. Our laboratory uses the National Institute of Health’s definition of biomarkers, which states “biological markers (biomarkers) are characteristics that can be objectively measured and used as an indicator of normal biological processes, disease processes, or pharmacologic responses to a therapy”. However, we acknowledge that the research into eye movement differences is relatively new, and we are far from identifying a consistent biomarker of illness. Thus, we have toned down the usage of the word biomarker throughout the manuscript.

14) p. 9, l.26: What means “clinically rated or diagnosed”? Would it be a sufficient to have a symptom rating by a clinician without any specified procedure of classificatory diagnostics?
Thank you for clarifying. We will accept individuals with depression that are identified with a formal diagnosis through standardized instrument, or by clinical judgement. We will not accept groups based solely on self-reported scores with no clinical oversight. We have updated this on page 6.

15) p.9, types of participants: Although the inclusion of individuals with a primary diagnosis of major depressive disorder actually implies that patients remitted from MDD will be considered, it might be helpful to explicitly state whether you are planning to this.
Thank you for the suggestion. We plan to consider such studies, and this has been clarified on page 6.

16) p.9, l. 45: Do you mean studies that used mood induction procedures to induce depression-like states in healthy participants or does the exclusion also refer to studies that used mood induction to vary mood states in patients with diagnoses from the depression spectrum (e.g., patients remitted from MDD)?
In regard to mood induction, we meant we would exclude studies that use mood induction procedures in otherwise healthy controls. This has been clarified on page 6.

17) p. 9, l. 49: As some studies choose to include control participants without current mental disorders but do not exclude or even assess a history of all mental disorders, this criterion might lead to the exclusion of actually relevant studies. It might be better to include studies with less selective control groups but to consider the type of control group in the analyses. Regarding the latter it might even be relevant whether standardized clinical interviews are applied to exclude previous depressive episodes in control groups (see von Koch et al., 2023, Behav Res Ther, 160:104231, for a discussion of this issue)
Thank you for this suggestion. Though it would be ideal for studies to screen for history of mental disorders, we will include all studies with control participants even if they are not screened for history. That is because we expect many studies do not screen their controls for psychiatric history.

Furthermore, not much is known whether psychiatric history affect our planned outcome measure. This change has been addressed on page 6 with the following passage: “Studies must also have a control participant group with a mean age between 18-60 years. Though screened controls (no history of mental disorders or psychiatric medication usage) would be preferred, we will also include studies that do not screen their controls and take this into consideration upon analysis and discussion of results.”

18) p. 11, l. 49: You may reconsider the inclusion of conference papers, as quality control is typically lower compared to journal articles and many conference papers might be related to journal articles.
Thank you for this important consideration. We have reevaluated and decided not to include conference papers in this review. This change is reflected on page 9.

19) p. 11, l.51: Should it be “will not be limited”?
Yes, thank you for noticing this error. This change has been made on page 9.

20) p.13, effect measures: Please specify, whether you are planning to analyze data separated by groups (e.g., major depressive disorder, dysthymia, etc.). Which groups will be built? I also suggest to separate different subgroups of major depressive disorder, especially acute and remitted groups. But you may also consider to distinguish between severe depressive episodes and moderate to mild depressive episodes.
Pending the final study sample, we will analyze dated based on subgroups if three or more studies are available. The proposed subgroups include differential diagnoses (e.g., major depressive disorder versus dysthymia), severity of episode (mild, moderate, severe) as well as distinguishing between current and remitted depression. This has been stated on page 11.

21) p.15, first two sentences: regarding the term “psychiatric”, please see comment #1
Thank you, this has been addressed.

VERSION 2 – REVIEW

REVIEWER	Reuter, Benedikt MSB Medical School Berlin GmbH
REVIEW RETURNED	18-Apr-2023

GENERAL COMMENTS	Thank for considering my comments in the manuscript. All concerns have been adressed adequately. I am looking forward to see the reults.