Dear Prof Chris Chambers,

We are pleased to resubmit our stage 2 Registered Report manuscript titled 'Genetically-modified animals as models of neurodevelopmental conditions: a review of systematic review reporting quality' following peer review comments.

We would like to thank both our recommender and reviewer for providing detailed and helpful comments, including providing greater clarity on what is and is not allowed in the Registered Report format. We're eager to get this right, so thank you for all your help.

Please find our responses to each of the review comments below. Review comments are in **bold** and author responses are in normal text.

Responses to Recommender Comments

One of the reviewers from Stage 1 kindly returned to evaluate your Stage 2 manuscript; the other was unavailable, but I have decided that we can proceed with an interim Stage 2 decision based on this reviewer's comments and my own reading.

Overall I think the Stage 2 manuscript is promising, with the caveat that there needs to a lot more clarity over the scope and reasons for the numerous text deviations between the approved Stage 1 manuscript and corresponding components within the Stage 2 manuscript. As the reviewer notes (and I picked up myself), there are several points where deviations are made but not explained, and some where they are made but not flagged (e.g. the tracked-changes version of the Stage 2 manuscript didn't flag some text changes in the Introduction, such as the statement about conducting an umbrella review.

In revising please revert all changes to the text of the Introduction unless they are necessary to correct an error of fact, to clarify an ambiguity, or to appropriately alter the tense. Beyond these conditions, no new literature, rationale or justification (or redescription of unchanged procedures/rationale) should be made. Elsewhere, other tracked changes are not fully explained, such as the alterations to the Data Synthesis section, leaving the reader (and the reviewer in this case) unclear as to how much is purely descriptive change vs. how much is actual changes to the methods. As a result, the description of deviations from protocol appears to be incomplete.

I want to be clear that there is no accusation of impropriety here; the requirements of RRs are somewhat different (and more strict) than with regular papers, so some confusion is understandable particularly for authors who are

new to RRs. But please attend carefully to this general point in revising, along with the reviewer's many additional points.

Thank you for taking the time to read and review the manuscript and for making these points. This is the first/corresponding author's first Registered Report and the process is a big change from more traditional publishing approaches.

We apologise that some deviations from Stage 1 were made but not flagged. All changes were made using tracked changes but some changes may have erroneously been accepted.

Additionally, (see responses to review 1) we have ensured in our resubmission that all deviations from stage 1 are clearly outlined and explained in the "deviations from the preregistration" section. That said, we have gone back over the changes we have made and reverted changes which were unnecessary and not permitted by the Registered Report format (see responses to reviewer 1).

Responses to Reviewer 1 Comments

I would first like to congratulate the authors for completing their review in a timely manner (esp. since their searches were based on full text; this is a good practice by the way).

When I started reading the Stage 2 manuscript, I noticed a lot of differences between Stage1 and Stage 2 (obviously I am not referring to the results and discussion section here, but the earlier sections), starting with the title, which was changed from "Genetically-modified animals as models of neurodevelopmental conditions: an umbrella review" to "Review evaluating the reporting quality of systematic reviews 1 describing animal studies of neurodevelopmental conditions: a Registered Report". This is not just a slight change of wording. For instance, the new title does not describe "genetically-modified" animals nor an umbrella review.

In hindsight, we agree that the dramatic changing of the title wording is necessary and the new wording leaves out details like genetically modified animals. This change was made based on wording from the in-principal acceptance of stage 1 which described the work as "Evaluating the quality of systematic reviews in preclinical animal studies of neurodevelopmental conditions".

In our resubmission, we have reverted back to the original title, however we have changed "umbrella review" to "review of systematic review reporting quality". The reason for dropping umbrella review is described in the response to the below comment. Additionally, we felt that it was important to clarify in the title that this is

not a review of animal models per say, but a review of systematic reviews on this topic, including their reporting quality.

Another example of a difference is that the aim is now worded differently in the introduction. The Stage 1 aim was: "Here, we aim to conduct an umbrella review to identify...." and the Stage 2 is: "Here, we aim to identify....". I still wonder why this work is not termed an umbrella review in Stage 2 (a review technique with specific characteristics), but more importantly why it was not an umbrella review as accepted in Stage 1.

In our stage 1 submission, our choice to describe our review approach as an "umbrella review" was questioned by reviewer 1 in the following comment in revision 1: "However, the aim of an umbrella review is to determine what is known on a topic, what remains unknown, and to make recommendations for what requires further research. This is very different to giving a descriptive summary of data from each included systematic review -an umbrella review is about synthesizing these findings (and reaching overarching conclusions), not presenting summaries side by side. So, if the authors only want to focus on methodological issues (e.g., identify risk of bias) or to just present the number and different characteristics of published systematic reviews, this sounds more like a "scoping review of systematic reviews". Scoping reviews aim to map a research area and not to synthesize findings."

Although we argued in our stage 1 response to reviewer comments that our approach could be described as an umbrella review, in hindsight we have decided that using the term, especially in the title, may mislead readers as to the purpose of our study. Given our work related to clear and transparent reporting, changing our terminology from "umbrella review" to "review of systematic reviews" seem most appropriate.

Also, the "full text retrieval" section does not appear in Stage 1. There are also a lot of minor changes, new sentences, and up-to-date references in the intro and methods section. The most important change is that from "Data extraction" onwards, all titles have been modified (the title "Data extraction" has been removed altogether - I suggest it is reinstated) and there are subtle changes in the text too.

Thank you for highlighting this. These changes were intended to improve the readability and flow of the work, but we understand now that such drastic changes are not permitted. Most of these changes have now been reverted. The changes to the introduction are detailed in the response to the final comment.

Regarding the methods section, we have made the following changes:

- Bibliographic search
  - The date the search was conducted has been added.
  - Databases has been changed to data sources as this is a more accurate description of these resources.
  - Autism spectrum condition has been changed to autism.
  - We have added that the full search terms are also available in the stage 1 Registered Report
  - Sentence on PROSPERO methods have been moved to its own section as these were not fully detailed in stage 1.
  - Initials EW are added to duplicate removal methods to show which author conducted that part of the methods.
  - The reference to Hair et al., 2021 has been updated to 2023 (changed from preprint to citing publication).
- Full text retrieval (now part of screening)
  - We had moved the text describing full text retrieval from the screening section to its own section as we believed this improved the flow and clarity of the work. However, we have reverted this change.
  - We planned to use in house code to retrieve full texts but in practice for this dataset found EndNote's full text retrieval function more convenient. We have explained this change in deviations section.
  - The number of interlibrary loans we required to access all full texts has been added.
  - We planned to use Google Translate to translate any non-English publications, however we did not retrieve any in our search. We have explained this in text.
- Screening
  - We had combined the screening and annotation sections into one as they use the same methods (two independent reviewers grade each study using the SyRF platform and disagreements are reconciled by a third reviewer) which we thought reduced duplication of text. We have reverted this change and screening and annotation are as they were.
  - However, we added details on how SyRF displays records in a random order and reviewers are unaware of other reviewers decisions. As we are concerned with reporting quality in our work, we thought it appropriate to mention our own use of randomisation and blinding.
  - We also added the research resource identifier (RRID) for SyRF.
  - We added the initials of the two reviewers and reconciler.
  - Initials of reviewers and reconcilers have been added.

- o Wilson et al has been updated as a reference (preprint version to published version).
- Inclusion and exclusion criteria
  - o Angelman syndrome has been added (discussed in deviations to the preregistration).
  - o Autism spectrum condition is now autism.
- Data extraction
  - o This has been moved back to its own section as in stage 1.
  - o Headings has been reverted to remove the "annotation of" prefix.
  - o In bibliographic data, "included study" has been changed to "included review" to make it clear that the publications we are reviewing are systematic reviews.
  - o In reporting quality, we have added that the checklist is also available in the stage 1 registered report.
- PROSPERO search
  - o In our stage 1 registered report, we stated only that we would contact authors of ongoing reviews registered on PROSPERO to check the status of those reviews. This is not a complete methodology as it did not detail how we would identify reviews or what data we would present, so these details have been added to the methods and described in the deviations to the preregistration.
  - o Due to time constraints we were unable to contact authors individually. However, we did search for published versions of PROSPERO registrations. This is described in the deviations section.

The data synthesis paragraph has been changed from "We will not conduct a meta-analysis. However, we will present a descriptive, tabular summary of the bibliographic, characteristics, and reporting quality data extracted from each included systematic review. We will score each included systematic review using the PRISMA-Pre checklist and provide a summary table detailing which items of the checklist each review met. Additionally, the checklist will allow us to assess which tools are currently being used to conduct systematic reviews in this field, including the tools used to screen studies (Checklist Item 6a), extract numerical data (Checklist Item 17a), and measure study quality or risk of bias (Checklist Item 19)."

to "The purpose of this work is not to conduct a meta-analysis based on the findings of included systematic reviews, but rather to identify the quality and reporting quality of these reviews. All data annotated are presented as descriptive summaries, and no formal statistical analysis was undertaken."

Here, there is an important change, namely that the "tools used to screen studies, extract numerical data, and measure study quality or risk of bias" were not assessed, as promised in Stage 1.

Apologies, the exclusion of these results is fully on our part. They have now been added to the results where tools were reported in included reviews.

I guess whether such changes are allowed is something to be discussed with the recommender. However, it is my understanding from my experience in publishing Stage 2 RRs that only the verb tense is allowed to change from future tense (Stage 1) to past tense (Stage 2). These changes created more workload for me, as I had to compare the two submissions instead of just evaluating the results and discussion section (maybe the authors could have annotated their manuscript so that these changes were easily spotted?). But most importantly, this practice goes against the rationale of Registered Reports, whereby in principle acceptance is given to a study at Stage 1 and at Stage 2 what was agreed earlier is what should be done and reported. Of course, understandably things don't always go as planned, therefore it is very important for transparency purposes that Stage 2 includes a section entitled: "Deviations from Preregistration" (as is also suggested by PRISMA).

Thank you for highlighting the challenges in reviewing stage 2 of this work. We can only apologise for unknowingly including changes that aren't permitted in a Registered Report. We have reverted most of the changes and any remaining changes are included in the deviation from the preregistration section.

Such a section exists in the manuscript, but the title is "Protocol registration". Actually, "Protocol registration" and "Deviations from Preregistration" should be two different sections. Having said that, this section only reports two changes (the inclusion of studies on the UBE3A gene and the search strategy for PROSPERO) and none of the changes I listed above. Therefore, this section should be more detailed.

Thank you, we agree that registration and deviation should be two sections. The deviation section now details all changes made.

Below are some more specific comments:

-Results - systematic search results: "The same systematic review was identified as a conference abstract (Zhang et al., 2022) and a peer-reviewed journal article (Zhang et al., 2021). As our evaluation is primarily concerned with reporting quality these publications, we assessed the conference abstract and journal article as separate publications." (Of note, there is a "the" missing in the second sentence). I am not sure I follow this rationale - researchers and policymakers, who use systematic reviews to make decisions, would read the published paper and not the conference abstract.

Thank you for raising this point. We agree with your comment. However, as we said that we would include and assess conference abstracts in our stage 1 methods, this was the methods we stuck with. In hindsight, the PRISMA-Pre reporting quality checklist is not designed for conference abstracts, which are severely limited in their word count and hence their reporting. We have raised this limitation in the discussion.

-In the "Characteristics of included reviews" I would also like to hear which neurodevelopmental conditions were targeted by the reviews.

We agree this is a good addition to the review. We have added the following: "FMR1, MECP2, and UBE3A alterations are strongly associated with Fragile X syndrome, Rett syndrome, and Angelman syndrome, respectively. Additionally, PTEN, SHANK3, ADNP, and NRXN1 are associated with Cowden syndrome, Phelan-McDermid syndrome, Helsmoortel-Van der Aa syndrome, and Pitt-Hopkins-like syndrome 2, respectively (SFARI Gene, 2023). CACNA2DE, CHD8, and TBR1 are associated with autism and intellectual disability generally but not with any named syndromes.".

-Figures 2 and 3 are not needed, as they duplicate information already presented in the text of the manuscript (only a few words in parenthesis are missing, which can be added). Actually, this also concerns Figures 6 and 7.

While we agree that the data in figures 2, 3, 6 and 7 are also reported in the text, we would argue that visual presentations of findings may be easier to interpret than text and we don't see an issue with having information presented in multiple ways (visual and text).

-line 348: "neurodevelopmental conditions or autism." Autism is also a neurodevelopmental condition (best use the term "autism spectrum disorder" throughout)

I believe this comment refers to the line. "Many of the protocols did not specify specific models of interest, instead often including "all" animal models of neurodevelopmental conditions or autism." Apologies for the ambiguity, I was not implying that autism is not a neurodevelopmental condition, I was referring to inclusion criteria either referring to all neurodevelopmental condition models or all autism models.

I have edited the sentence to read: "Many of the protocols did not specify specific models of interest, instead often stating that they will include "all" animal models of neurodevelopmental conditions or "all" animal models of autism.

In relation to the second point, my word choice was guided by best practices given in "Avoiding Ableist Language: Suggestions for Autism Researchers" (Bottema-Beutel et al., 2021; https://doi.org/10.1089/aut.2020.0014) which recommends autism as opposed to ASD.

-The flow diagram is fine, but I would suggest that the authors use the PRISMA template (here: http://prisma-statement.org/prismastatement/flowdiagram.aspx?AspxAutoDetectCookieSupport=1)

The PRISMA flow diagram is designed for systematic reviews of RCTs. We adapted the PRISMA flow diagram to fit our review approach. For instance, PRISMA flow diagrams display the reasons for exclusion for each publication. It was not possible for us to record reasons for exclusion of each of the 1,325 unique publications retrieved in our search, so we had to adapt the information displayed in our flowchart.

Minor comments/typos

Since the authors already made some changes in Stage 1, I will take the liberty to make a couple of comments on that part too:

1. First sentence of the abstract: "Using genetically-modified animals to model neurodevelopmental conditions (NDCs) helps better our understanding of their underlying biology". It is not clear what "their" is referring to, maybe some rewriting will make this clearer.

Thank you for highlighting this. Their should be referring to the neurodevelopmental conditions. We have edited the text to read: "Using genetically-modified animals to

model neurodevelopmental conditions (NDCs) helps better our understanding of their underlying biology underlying these conditions."

2. Second sentence of the abstract: "In vivo research has unique characteristics not shared with clinical research, meaning approaches to systematic review must be adapted to this context." To "systematically review" what? Again, maybe some rewriting will make this clearer.

Thank you for highlighting this ambiguity. Here, we intended systematic review as a noun rather than a verb. We have edited the text to read: "In vivo research has unique characteristics not shared with clinical research, meaning approaches to systematic review methods must be adapted to this context."

3. Abstract, methods, line 3: "is" - should it be "was"?

We agree with this tense change, thank you for highlighting.

4. Abstract, methods, line 4: the first comma needs to be removed I think.

The extra comma has been deleted.

5. There are a lot of one-sentence paragraphs. The APA publication manual asks that paragraphs contain more than one sentence. However, maybe this is something that is different from discipline to discipline.

We do not plan to submit to any journals that require APA formatting. We argue that where we have used single sentence paragraphs, it is to break apart text and increase the readability of the work.

6. Line 354: "all which are ongoing." An "of" is missing here.

Thank you, we have now corrected this.

7. There are other minor grammatical issues in the manuscript too, but I guess they will be spotted during type-editing.

We have checked for additional typos and grammar issues and have corrected these.

Overall, if the recommender thinks that the changes made to the first part (corresponding to Stage 1) can be accepted, then I would suggest publication following the minor revisions I list above.

We have reverted all changes to the introduction that do not correct errors, clarify ambiguity, or correct tense. Where we added extra content to the introduction in the previous submission, we have moved this to the discussion where it is more appropriate.

The changes that we have kept are:

- Introduction paragraph 1:
    - "Characterised by" has been changed to "presenting with"
- Introduction paragraph 4:
    - "we aim to conduct an umbrella review to identify" has been changed to "we aim to evaluate" to remove reference to the term umbrella review which we are no longer using and because the word evaluate is more accurate than identify to describe our aim (e.g. evaluating rather than identifying reporting quality).
    - Numbers (1) and (2) have been added to improve the readability of the sentence describing our genes of interest.
    - Models of Angelman syndrome have been added (the reasoning for which is given in the deviations from the preregistration section) and the sentence has been restructured to account for that insertion.
    - Tense changes where appropriate
- Introduction paragraph 5:
    - "identifying the quality and reporting quality" has been changed to "identifying the quality of and assessing the reporting quality" for the same reason as the first point in paragraph 4.
    - "preliminary protocol for which has been preregistered" in relation to future work informed by this project has been changed to "preliminary protocol which is available" as our living evidence summary protocol is not an official preregistration but rather a living document in an OSF project.