

General Points

I would like to thank both reviewers for taking the time to review my Stage 1 Registered Report and providing such helpful and constructive feedback. I feel that my manuscript is now a lot clearer having taken on board your feedback. I am very happy that I have chosen to pre-register this project using the Registered Report format, in order to gain such invaluable feedback before setting out on my review.

As a general point, I would like to highlight that the revised version of my Stage 1 Registered Report now refers to “neurodevelopmental conditions characterised by intellectual disability, epilepsy, and autism spectrum condition” rather than “autism spectrum disorder/intellectual disability”. This was my mistake in misunderstanding how to correctly define the conditions we are investigating. Additionally, we have taken the opportunity to use the term “condition” rather than “disorder” in both the case of “autism spectrum disorder/condition” and “neurodevelopmental disorders/conditions”. This was suggested by my secondary supervisor Professor Peter Kind as more appropriate language and is increasingly used in the research literature and by autistic / neurodivergent individuals.

In my responses below, I have included the initial reviewer comment in ***bold italics*** and added any additional context required in [square brackets]. I provided my response below each comment. A tracked changed document should also be attached.

Thank you both once again for your time and efforts in reviewing this manuscript to such a high standard.

Responses to Reviewer 1 (Marietta Papadatou-Pastou)

Mentioning intellectual disabilities (ID) is not needed, as this study focuses on ASD. I was actually confused while reading the abstract as to whether ID is part of the focus of this review or not. In a bit more detail, in the first sentence as well as in the phrase "summary of ASD/ID research", ID is mentioned, but in the search strategy section only ASD is mentioned. Moreover, only ASD is mentioned in the title. This is confusing to the reader as mentioned above - just by reading the title and abstract it should be clear whether ID is included in this review or not.

I apologise for the confusion caused. After speaking with my secondary supervisor, Professor Peter Kind, he clarified that the conditions that we are aiming to cover are neurodevelopmental conditions characterised by intellectual disability, epilepsy, and autism spectrum condition. This was my mistake in misunderstanding how to correctly define the conditions we are investigating. I hope that this new description improves that clarity.

The acronyms ASD and ID are not explained in the abstract.

Thank you for pointing this out, these terms have now been written in full in the abstract. Similarly, in the updated title, I have made sure that “neurodevelopmental conditions” is written in full.

Second line: Please replace "autism spectrum disorders" with "autism spectrum disorder" (singular). Please also replace "intellectual disabilities" with "intellectual disability". This is how they are mentioned in DSM-5.

Thank you for pointing this out, as the singular forms are indeed correct in the DSM-5. However, I hope you do not object that we have taken the opportunity to use the term "condition" rather than "disorder" in both the case of "autism spectrum disorder/condition" and "neurodevelopmental disorders/conditions". This was suggested by my secondary supervisor Professor Peter Kind as more appropriate language and is increasingly used in the research literature and by autistic / neurodivergent individuals.

"Living evidence summary" is a term reserved for the cases when the summary is regularly updated. In this case, the review will be just published once, with no updates after it. So, this term is not appropriate. (Please correct me if I have misunderstood something here.)

You are correct that this review will only be published once and not updated. However, the results of this study will be used to inform the development of a separate living evidence summary project, which will be regularly updated. I have removed mention of this separate project from the abstract, as it is not relevant and my word count is limited, but have added greater clarification in the final paragraph of the introduction: "Findings from this review will be used to inform the development of a living evidence summary of research using genetically-modified animals to model NDCs, a preliminary protocol for which has been preregistered on the Open Science Framework (OSF; DOI:10.17605/OSF.IO/GFTZP)."

I would suggest rewriting the abstract to take into account the PRISMA 2020 extension for Abstracts. A checklist is to be found here: https://prisma-statement.org/documents/PRISMA_2020_abstract_checklist.pdf. At the moment, a number of elements are missing from the abstract, for example inclusion and exclusion criteria as well as whether a risk of bias analysis will be conducted. I understand that there is a word limit for the abstract, but maybe the authors can fit in some of this info. They are also advised to submit the PRISMA 2020 extension for Abstracts checklist as part of the supplementary material.

Thank you for the suggestion, I have rewritten the abstract to fit the PRISMA extension for abstracts checklist as much as possible, although some sections were not applicable due to this only being a Stage 1 Registered Report. The completed checklist is included as a supplement here: <https://osf.io/bczkv> (also linked to in the revised manuscript).

The rationale behind this review is not clear. In the introduction there is some discussion about how systematic reviews identify areas of poor methodological quality or high risk of bias within a research literature. Is this the drive behind this review of systematic reviews? It is not clear to me how this will be achieved. Will the collected systematic reviews be then summarized with regards to the risk of bias they have identified? If so, shouldn't reporting risk of bias be one of the inclusion criteria? Or will the authors assess whether the reviews themselves have run such an analysis? It is mentioned that the authors will identify which tools will be used to measure risk of bias in the included studies, but not how their review will contribute towards identifying the risk of bias in the literature.

Thank you for pointing this out, the rationale is indeed unclear and I have edited the introduction to hopefully provide clarity. We will not be summarising or synthesising the results of included systematic reviews in this project, instead we are looking for an overview of what reviews are being done and how they are being done. I have added the following to the final paragraph of the introduction: "This review will not synthesise evidence from existing reviews. The rationale behind this review is that by identifying the quantity and quality of existing systematic reviews in this area, we can inform guidance on how future systematic reviews within this research field should be conducted."

Another area that is not clear is whether literature on ID is going to be reviewed (at some point it becomes clear that the answer is no, but in the beginning of the introduction the reader is made to think that it is, as ID is mentioned in the very first sentence and then a definition of ID is given in the second sentence).

Again I must apologise for this confusion caused, I hope "neurodevelopmental conditions characterised by intellectual disability, epilepsy and autism spectrum condition" provides clarity on this.

It should also be made clear (starting from the title) if this review concerns only genetically-modified animals or all animal models. Four types of animal models are used in preclinical research: (1) disease induction models, (2) xenograft animal models, (3) inbred strains, and (4) transgenic models, so it would be good to clarify if the focus of this review is only on the latter.

Thank you for pointing this out, I have now made this clear throughout by referring to "genetically modified animals as models of neurodevelopmental conditions".

An explanation (even short) of why ASD was chosen as the disease model that will be the focus of this review, should be given.

I have added the following to paragraph 2 of the introduction: "Recent evaluations of in vivo NDC research have highlighted failures to control for variability introduced by the "litter effect" as a potential source of poor reproducibility (Jiménez and Zylka, 2021)." and given reference to the fact that results from this review will help inform the development of a living evidence summary of research using genetically-modified animals to model NDCs. Hopefully this adds sufficient explanation.

The aims of this review are presented three times in the manuscript (abstract, introduction and methods), each time with a slightly different wording (minor point: the aims do not need to be repeated in the methods section). The slight change of wording is not problematic, but when the essence changes this confuses the reader. For the second aim we read "summarize the focus of these reviews" (abstract)/ "the aims of these reviews" (introduction) / "the aim of each review and what research questions does it seek to ask" (methods). These are slightly different aims. Especially with regards to "summarize the focus of these reviews" I am not sure if the authors refer to the fact that they also plan to present "a descriptive summary of data extracted from each included systematic review", as they mention in the methods section. In other words, it is not clear

to me if the main aim of this review of reviews is to map systematic reviews already conducted, if it is to identify methodological issues in primary research (as mentioned in my first comment, e.g. risk of bias in the literature), or if it (further) aims to summarize the findings in the literature, by means of presenting (or even synthesizing) the findings of the included systematic reviews.

Thank you for this comment. I have edited the wording of the aims to (a) improve clarity and (b) be consistent and have removed the repetition of the aims in the methods: "Here, we aim to conduct an umbrella review to identify the quantity, characteristics and quality of systematic reviews which synthesise research using genetically-modified animals to model NDCs." In relation to the descriptive summary of data extracted, the data refer to the bibliographic, characteristic, and reporting quality data that we have detailed in the data extraction section. I have clarified: "we will present a descriptive, tabular summary of bibliographic, characteristics and reporting quality data extracted from each included systematic review." We will not summarise the findings of included systematic reviews, which I have also clarified in the introduction: "This review will not synthesise evidence from existing reviews."

On a related note, in addition to the three aims, in the methods section, a few other aims are identified: "(i) Whether the review only included animal studies, or also included clinical or in vitro studies, (ii) Which animal models the review included, (iii) The total number of studies included in the systematic review, (iv) The total number of studies investigating relevant animal models". These are listed under "systematic review aims and research question", which is not a suitable place. These are rather characteristics of the studies. Maybe the characteristics of the systematic reviews is another aim? This goes again to the fact that the objectives of this review are not clear.

Thank you for pointing out this discrepancy. I have clarified this by making it clear that we aim to investigate the quantity, characteristics, and quality of existing reviews. I have edited the subheading from "systematic review aims and research question" to "characteristics" based on your feedback, which I feel is a much more suitable subheading.

The authors label their review an "umbrella review" (although in the methods they also refer to it as a "systematic review of systematic reviews".) 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. However, if the authors opt for an umbrella review, then the relevant literature should be summarized in the intro.

Thank you for bringing this up. I think this is a difficult one because this review doesn't fall neatly into either category of umbrella or scoping review. After a discussion with my primary supervisor Professor Malcolm Macleod, and with reference to this paper on typology of reviews (<https://doi.org/10.1111/j.1471-1842.2009.00848.x>), we decided that the review type best describing this project is "umbrella review".

The same criterion (animal models) is given both as an inclusion and as an exclusion criterion (with a slightly different wording, but it is the same criterion). Just mentioning it once suffices. The title of this section could even be modified to "inclusion and exclusion criteria", if the authors prefer it that way.

Thank you for your suggestion, I have edited the inclusion and exclusion criteria as you have advised to avoid repetition and improve clarity.

On the same note, it is not clear to me what "Other criteria - conference abstracts and preprints" means as an inclusion criterion. Do papers need to be conference abstracts and preprints to be included?

Apologies again for the confusion, I have clarified to read: "We will include systematic reviews published in peer-reviewed journals, as conference abstracts, or as preprints (where they are identified in searches)." I have also renamed "other criteria" to "publication type".

Exclusion criteria: "Publications where we cannot access the full-text". It is the first time I come across this criterion. Genetically-modified animals are rather new (I mean a few decades) so I cannot see how one will fail to access the full text, either through their library, an inter-library loan or even by purchasing the full text directly.

Thank you for pointing this out. We have removed this as an exclusion criteria as that makes most sense.

Inclusion criteria: The species of the animals should mentioned (if systematic reviews on all species will be reviewed, this should be mentioned too)

Thank you for pointing this out – I have added that any and all species will be included.

Exclusion criteria: "systematic reviews which do not search at least two bibliographic databases". Why is that an exclusion criterion? If one of the aims is to "identify the number of systematic reviews previously conducted", then shouldn't all systematic reviews be included? And then, one can perform a quality assessment of the included systematic reviews.

Thank you to both reviewers for pointing this out. On reflection I agree with you both, especially for the point that Reviewer 1 made regarding quality assessment.

It was interesting to see that the authors plan to "screen studies based on full text to avoid erroneously excluding systematic reviews which do not report their inclusion criteria in their abstract". This is fine, but -from my experience in systematic reviews and meta-analysis- a search could give back hundreds if not thousands of papers. So, the time resources needed to read each full text will not be realistic. I would suggest that the authors use a hierarchical process, by first screening titles (some papers will obviously be on a different field/irrelevant), then abstracts (or

titles and abstracts together), and then the full text. Each time, one only excludes the papers they are sure do not meet the criteria and then takes the rest of the papers to the next step to see if the relevant information is there. I understand that some of the criteria might not be identifiable in the abstracts, but this is always the case in meta-analyses. This is why we never stop at the abstracts during screening.

Thank you for highlighting this. Upon reflection with my colleagues, we agree to go ahead with our approach of screening based on full-texts as an initial stage. One of our concerns is that, from our previous experience with systematic reviews or animal studies, sometimes reviews include both clinical and animal studies but do not mention the inclusion of animal studies in the abstracts. If, to avoid including these studies, we decide to include all clinical systematic reviews this will likely involve more effort than screening all studies once based on full-text. We are confident with the feasibility of this approach.

The authors should include a PRISMA 2020 checklist for their own review (on what they plan to do) and include it as supplementary material. By doing so, they can also identify a number of elements currently missing from their registration, e.g. whether they plan on conducting a risk-of-bias analysis (I understand there is discussion on biases in the manuscript, but in the methods it is not clear if this will be conducted and how). Some other elements are missing too -please check the PRISMA 2020 checklist.

Thank you for highlighting this. I have completed and attached the PRISMA for protocols checklist (as I believed this was more appropriate for this Stage 1 Registered Report than the PRISMA 2020 checklist for completed systematic reviews) and have edited the Stage 1 Registered Report to add the missing elements required by the checklist. The completed checklist can be found here: <https://osf.io/2cmwb> (also linked in this paper). I will include a completed PRISMA 2020 checklist with Stage 2 of this Registered Report, as we will then have data to complete all the questions.

The phrase "will place no restriction on publication date or language" appears twice in the methods.

Thank you, I have deleted the additional instance of this phrase.

The definition given in the first sentence of the introduction applies to ASD only, but the way this sentence is written reads like it refers to all neurodevelopmental disorders (including ID, when a different definition for ID is given in the second sentence).

Thank you for pointing this out, you are indeed correct that the previous definition was confusing. Hopefully the text has been clarified as we now refer to neurodevelopmental conditions characterised by intellectual disability, epilepsy and autism spectrum condition.

The introduction needs to be rewritten with a focus on less repetition, better flow and better justification for the need for this review. For example, paragraphs 2 & 3 & 4 could be merged. Paragraph 3 says more or less the same thing as the beginning of paragraph 4.

Thank you for highlighting the repetition in the introduction. I have reworked the introduction to avoid this and hopefully improve clarity.

It is mentioned that the analysis code will be available online. Which analysis are the authors referring to? Reviews do not include analysis.

Apologies again for the confusion. You are correct that we will not be completing a meta-analysis. The code I am referring to is code to process the bibliographic, characteristics, and reporting quality of each included review, e.g. calculate the reporting quality score of each review and count the number of reviews which search X database. We process this data using the R programming language and share our code online so that our summary results are transparent and can be replicated by others, even if no formal statistical analysis is completed. I have reworded the sentence to clarify: "...the code used to process this data, will be shared online...".

Responses to Reviewer 2 (Richel Bilderbeek)

I would enjoy to know what happens when [the adapted PRISMA guidelines] draft is changed during this research? I assume that the draft is kept as-is during this research and it may only change when becoming a PRISMA standard

Thank you for bringing this up. I am not sure on timelines for the adapted PRISMA guidelines becoming an official extension, but I have added clarity that if an official extension is released during this research which is different to the draft published in Hunniford et al (2021), we will continue using the draft version from Hunniford et al. (2021).

[Moving the description of bias]

Thank you for your suggestion. I have edited the introduction to describe bias earlier in the text (where you have suggested) and also condensed some of the text under the suggestion of Reviewer 1.

I would enjoy to see alpha values, p values and the number of tests done, over multiple papers. E.g. if 1 paper find gene X out of 20 to be significant, where many others don't, we can expect that that paper's result was due to chance

I have added "failures to correct for false discovery rate" to the list of possible contributors to translational failure, which I hope addresses your point. The text now reads: "Several other factors may contribute to this lack of success including target choice, outcome measures insensitive to change, disease stage at which treatment is initiated, lack of construct validity, failures to correct for false discovery rates, poor methodological quality, and high risk of bias..."

Which fundamental differences [between preclinical and clinical research]? I would enjoy to at least read the most important for this study. Or give a reference

Thank you for highlighting this. I have added a reference to the Hunniford et al (2021) paper which summarises the key differences, (e.g. number of included studies, number of participants/animals per study, heterogeneity between participants/animals, etc.).

I feel the reference to Hair et al., 2022 is invalid. It is 'Elliott, Julian H., et al. "Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap." PLoS medicine 11.2 (2014): e1001603.' that describes the term 'living evidence summary' (which is cited by Hair et al.)

Thank you for your reference suggestion. I have decided to remove reference to the living evidence summary as I have edited the section introducing living evidence summaries to make it a bit less confusing. The section on living evidence summaries (last paragraph in introduction) now reads: "Findings from this review will be used to inform the development of a living evidence summary of research using genetically-modified animals to model NDCs, a preliminary protocol for which has been preregistered on the Open Science Framework (OSF; DOI:10.17605/OSF.IO/GFTZP)." I have referenced the preliminary protocol for the exact living evidence summary project that will be informed by the results of this review. The reason I had cited Hair rather than Elliot is that, while Elliot introduces and describes living evidence summaries, Hair describes the exact approach that will be used in the project I am describing. Therefore, I hope that referencing the protocol will be clearer.

Reference? [to sentence ending "as these are all genes where there is high confidence in their clinical relevance"]

Thank you for pointing this out, as the reference was misplaced (in the middle of the sentence rather than the end). I have edited the sentence to improve clarity. It now reads: "We are interested in models with alterations in any of the 102 high-confidence genes identified via large-scale exome sequencing by Satterstrom et al. (2020)".

I would enjoy to hear why [must search at least two databases] is an exclusion criterium

Thank you to both reviewers for pointing this out. On reflection I agree with you both, especially for the point that Reviewer 1 made regarding quality assessment.

I would enjoy a reference to [the statement "it is not recommended to use the general PRISMA guidelines to assess in vivo systematic review reporting quality"]

Thank you for pointing out the missing reference. I have added a reference to the Hunniford et al 2021 paper which highlights this.

***I enjoy the enthusiasm here, but I see no reason why to provide this narrative. I would enjoy to see why this narrative is provided, especially as the checklist already contain quite some 'describe X' items AND I would enjoy to see some example table here, like:
[https://github.com/richelbilderbeek/wilson_et_al_pci_rr_review]***

Thank you for your suggestion. Taking this on board, the text now reads: “we will present a descriptive, tabular summary of bibliographic, characteristics and reporting quality data extracted from each included systematic review. I will indeed use summary tables like you have suggested to present the data.