

Author response to reviews of

Benefits and risks associated with children's and adolescents' interactions with electronic screens: An umbrella review

Taren Sanders on behalf of co-authors
submitted to *Nature Human Behaviour*

[RC] Reviewer comment

Manuscript text

Dear Dr Charlotte Payne,

Thank you for considering our manuscript for publication at *Nature Human Behaviour*. We appreciate the feedback that you and the reviewers have provided. In the following itemised list we respond to each comment point-by-point.

1. Editor Comments

RC: COMMENT 1

Reviewer 3 points out that your last search date is May 2020. We agree that this means the review is not as timely as it could be and ask that you update your search to September 2022, and redo your analyses with all newly included studies.

AR: We have updated the search to September 2022. We found a substantial amount of new literature had been published since our initial search, with an additional 108 meta-analyses included in the review. While many of these were updates to existing reviews, we also included 110 new exposure/outcome combinations (or existing combinations for new age groups) that were not found in the original search. We think that this increase in research demonstrates the level of interest in this area.

We'd also like to thank the editor and reviewers for their patience while we screened and extracted data from these additional reviews.

RC: COMMENT 2

Reviewer 2's concerns about the appropriateness of your technical approach (notably: subjective characteristics of review selection, your statistical approach, and your choice to harmonize metrics of effect sizes) are significant and we ask that you prioritise addressing these issues when redoing your analyses.

AR: Please see our response to Reviewer 2, in particular comment 29. The reviewer raised some important concerns, especially regarding odds ratios which we have decided should not have been included. For the other issues, we have provided stronger justification for our decisions.

RC: COMMENT 3

Both reviewers 2 and 3 are concerned about the lack of information on risk of bias assessment and population characteristics. Please ensure this information is clearly available and add data on both to Table 1.

AR: We apologise that the quality assessment information was not included in the original submission. We have added this information to Table 1, along with the information about the sample characteristics requested by the reviewers. We note that this increases the size of the table, and we have removed it from the text of the manuscript and instead included it as a separate file.

RC: COMMENT 4

Reviewer 1 highlights the inappropriateness of causal language given that the majority of included studies speak to association, not causation. Please remove all inappropriate use of causal language accordingly.

AR: Please see response to Reviewer 1, Comment 4. We agree that the use of causal language from an evidence-base that is mostly cross-sectional would be misleading. We did not intend to make causal statements, and have reviewed the manuscript for any remaining instances.

2. Reviewer 1

RC: COMMENT 1

As the authors note, “screen time” as a metric is highly criticized and problematic. This begs the question, though, of whether a wide-scale review of this sort, which retains a focus on screen time (although it also acknowledges other more nuanced metrics of exposure), is further perpetuating the issues seen by the broader field.

AR: We agree with the reviewer that the term ‘screen time’ is problematic. Indeed, as our results show, the influence of screen time on children is a less a function of a physical screen, and more a function of the activity that the screen provides. But, we feel that a review such as this is needed in order to emphasise the diverse influence that screens can have. If the results are not collated into a single place, it is difficult to see how the field will begin to progress away from ‘screen time’ as a concept. We have included this in the *Implications for Future Research* section:

Our results highlight the need for the field to more carefully consider if the term ‘screen time’ remains appropriate for providing advice to parents. Instead, our results suggest that more nuanced and detailed descriptions of the behaviours to be modified may be required. Rather than suggesting parents limit ‘screen time’, for example, it may be better to suggest that parents promote interactive educational experiences but limit exposure to advertising.

Pg. 20

RC: COMMENT 2

The authors note that “no review has synthesized the evidence available across a broad range of outcome domains, such as physical health, education, physical and cognitive development, behaviours, and wellbeing.” The authors end up concluding that there are different magnitude and directions of associations for different outcomes and exposures (and, indeed, the mechanisms of these are highly different depending on the predictor-outcome combination), which makes me wonder whether casting such a wide net is helping the reader understand more or less about the research to date on technology use and child/adolescent outcomes? I would want to hear more justification in the manuscript itself on why they assert it is helpful to try and paint them with one broad brush?

AR: Our goal with this review was to provide an overview of the field of screen time research. A summary of an entire field will, out of necessity, be more broad than a typical systematic review. We believe that there is significant value in providing a single overview to summarise the state of the field. We have added some additional information to the introduction on this point:

No review has yet examined the evidence available across a broad range of outcome domains, such as physical health, education, physical and cognitive development, behaviour, and well-being. By summarising and synthesising all evidence in one overview, we provide a reference point for the field and allow for easier comparison of risks and benefits for the same behaviour.

Pg. 6

RC: COMMENT 3

Page 10 notes that the study “examined the combinations of exposure and outcomes and removed any effects that appeared more than once, keeping the effect with the largest total sample size.” Did this take into consideration whether some meta-analyses of the same exposure-outcome pairs may have included mostly different studies (i.e., because they focused on different date ranges or some other inclusion criteria)?

AR: We intended to include the review which captured the most information on the exposure/outcome combination for each effect size. The assumption was that the largest review would be the most recent, and that the newer reviews would contain the studies located in previous reviews. Manual inspection of the reviews which did not contribute any effects suggests that this was successful - older reviews tended to only be included if they focused on the general population while a newer review focused on a sub-population.

RC: COMMENT 4

Causal language- The present study sought to “provide a holistic perspective on the influence of screens on children’s lives across a broad range of outcomes,” but I am not convinced that the studies included in the meta-analyses are designed in ways that allow us to understand causal IMPACT at all. Many have criticized this literature for being largely self-reported (with flaws inherent to this method, which are reviewed in the present manuscript’s discussion), observational, and cross-sectional, with very few rigorous experimental or well-controlled longitudinal designs that would enable researchers or the public to draw robust conclusions about the direction of associations. I fear that this comprehensive review, though rigorous, does not adequately attend to the fact that we do not usually know whether screens are causing helps or harms for youth, if different types of youth behaviors may be causing them to engage more with screens (e.g., sedentary youth may seek out more screen time; depressed youth may turn more to social media as an avoidance strategy), or if indeed there are third variables driving observed associations. Although the review here does a great job of summarizing the direction and strengths of associations across studies, I fear that it does not attend to these important issues in terms of how one ought to interpret them.

AR: Yes, we agree that the body of evidence does not lend itself to strong causal conclusions. We note that we deliberately chose to avoid using strong causal language as much as possible, instead indicating that the evidence suggests associations with outcomes. We have reread the manuscript and removed any remaining causal language that we found. We have also added the following to the *Discussion* section:

Our high-level approach also means that we could not engage with the specific mechanisms behind each association, and as such, we cannot comment on the evidence for causality. Instead, readers who wish to more deeply understand one specific relationship are directed to the cited review for that effect, where the authors could engage more deeply with the mechanisms.

Pg. 21

Note that we do use causal language when describing hypothesised mechanisms, and when describing guidelines. We feel that this is sensible — guidelines are based on causal assumptions, and while our evidence does not always support them, we are referring to them in the context that they are made (i.e., to instruct parents).

RC: COMMENT 5

The manuscript notes that the nature of the data allowed for few age-based conclusions- I fear that this is a substantial limitation that deserves more discussion. It seems difficult to characterize all children’s

media use from age 0-18 in the same way, especially across such a wide range of developmental outcomes. A more nuanced treatment of developmental considerations would have been appreciated.

AR: One of the benefits of the other reviewers' and editor's suggestion to update the search is that some of this limitation has been addressed in more recent literature. Newer papers were more likely to provide moderation by age group, or new reviews targeting specific age groups were now available. Unfortunately, many of these still failed to meet our criteria for statistical certainty. Inspecting the results in the supplementary material still reveal fewer differences by age group than may be expected. In some cases this is because there is still only an effect for a single age group, as many reviews target behaviours or outcomes which are more common in one age group (e.g., sexting in adolescents, or developmental outcome in young children). For the effects where there were multiple age groups, differences tended to be small (e.g., body composition).

3. Reviewer 2

RC: COMMENT 1

I appreciate the opportunity to review this manuscript focusing on the impact of screen use on child well-being. This umbrella review organized findings from meta-analyses by child outcomes and screen types, converted pooled data into Pearson correlations as a unified estimate to facilitate cross-study comparisons, and reached a conclusion that general screen use and social media were associated with harmful impacts on child learning, literacy, body composition, and depression, and the potential risks could be mitigated by the content and context of screen use. This conclusion is somewhat similar to that of another umbrella review conducted by Stiglic and Viner in 2019 (PMID: 30606703). However, the current one made advances on including studies examining educational screen time and social media and harmonizing data from meta-analyses.

AR: We are grateful to the reviewer for seeing the value in our review.

RC: COMMENT 2

Abstract:

Page 3 Line 38: In my opinion, “examined” is a better than “synthesized”. Because the data from meta-analyses were not pooled.

Page 3 Line 39: I suggest replacing “common metric” with “Pearson’s r”.

Page 3 line 41-42: “extracted 165 unique exposure/outcome combinations from 66 studies.”

AR: We have made all of these changes.

RC: COMMENT 3

Summary:

Page 4 line 59: the lack of comprehensive evidence hampers efforts to make an informed decision.

AR: We have made this change.

RC: COMMENT 4

Introduction:

Page 5 line 84-86: please further justify how a review study would make the limited evidence base less lacking or generate new findings.

AR: The quote from the *Lancet* editorial is that the understanding is lacking, rather than the evidence. Our understanding is lacking partly because benefits of screen time (as seen in some educational literature) and reviewed separately from harms. Our review aims to address this. See additional context in the *Introduction*:

No review has yet examined the evidence available across a broad range of outcome domains, such as physical health, education, physical and cognitive development, behaviour, and well-being. By summarising and synthesising all evidence in one overview, we provide a reference point for the field and allow for easier comparison of risks and benefits for the same behaviour.

RC: COMMENT 5

Page 5 line 87-90: I suggest improving the logic flow and citing more recent studies (2016 or newer) to highlight the potentially new additions of the current review.

AR: We have revised the introduction with new papers, with a particular focus on other umbrella reviews.

RC: COMMENT 6

Page 5 line 91-94: Please provide a more in-depth explanation on the advantages of a review study to include a broad spectrum of health outcomes.

AR: One of the key strengths of our review is that it summarises what is available in the literature, rather than trying to *a-priori* predict what areas should be investigated. That is, our review provides a comprehensive summary of what meta-analytic evidence exists. We have added some additional text to this effect in the *Introduction*:

This review synthesises evidence on any outcome of electronic media exposure. We deliberately do not pre-specify outcomes, in order to get a comprehensive list of areas where there is meta-analytical evidence. Adopting this broad approach allowed us to provide a holistic perspective on the influence of screens on children's lives. By synthesising across life domains (e.g., school and home), this review provides evidence to inform guidelines and advice for parents, teachers, pediatricians and other professionals in order to maximise human functioning.

Pg. 6-7

RC: COMMENT 7

Page 6 line 96-104: these guidelines provide age-staged recommendations and mainly target on recreational screen time. New evidence to support or reject these recommendations are important breakthroughs. So how does the current review contribute to the refinement of these guidelines? It could be proposed here to strengthen the justification of this study, and also need to be addressed in the discussion section.

AR: We feel that we have amply addressed the contribution of this review to the field. Guidelines should reflect both the most recent evidence, and the most comprehensive evidence. Many of the previous reviews used to justify guidelines focused on narrow ranges of outcomes (e.g., the impact on adiposity). We think this is a mistake - you cannot make an informed decision about an action if you only examine the risks, you must also examine potential benefits.

RC: COMMENT 8

Page 6 line 104-107: this is not a direct contrast. Longer screen time is needed to pose mental health risks, but short screen time can lead to other health issues. While guidelines are built on a comprehensive examination of health outcomes. Thus, I consider here as a weak argument/transition. The organization of the introduction section need to be improved.

AR: We had difficulty understanding the reviewers exact concern here. We are happy to revise this further if the reviewer does not feel that the other changes made to the introduction improve the set of arguments.

RC: COMMENT 9

Page 6 line 117-119: Stiglic and Viner’s umbrella review summarized evidence across a broad range of outcome domains. Please further justify the current study in addition to theirs.

AR: We have included information on this in our summary of previous umbrella reviews.

While there have been other overviews of reviews on screen time, these have tended to focus on a single domain (e.g., health²¹), focus on a particular exposure (e.g., social media^{22,23}) or provide only a narrative summary of the literature.²⁴ No review has yet examined the evidence available across a broad range of outcome domains, such as physical health, education, physical and cognitive development, behaviour, and well-being. By summarising and synthesising all evidence in one overview, we provide a reference point for the field and allow for easier comparison of risks and benefits for the same behaviour.

Pg. 6

RC: COMMENT 10

Page 7 line 122: please define “any plausible outcomes” to make the scope more scientifically sound and robust. In addition, the result section included findings of behavioral outcomes (physical activity and sleep duration) which I was not considered as academic, health, and social-psychological outcomes. Thus, it is better to be clear about the types of outcomes at the beginning of the manuscript.

AR: We did not *a-priori* choose which outcomes would be included. Instead, we chose to summarise the literature on for any meta-analysis where screen exposure was the independent variable. We have removed the word “plausible”, as we agree that makes the sentence unclear and the revised phrasing is more aligned with the eligibility criteria. Our intention in the introduction is not to preempt the results, which means we do not outline the outcomes found in the introduction.

RC: COMMENT 11

Page 7 line 123: the authors stated the application of “a broad approach”, while highlighted the importance of nuanced findings and guidelines. The introduction section would be improved by balancing these two important aspects.

AR: We agree that “broad approach” does not correctly convey our meaning. We intended to indicate that in order to make nuanced recommendations, you must have all of the information. Our approach allowed for a broad set of exposures and outcomes in order to capture all of this information, which is what we intended to convey. We think our changes elsewhere (see [comment 2](#) to Reviewer 1, and [comment 6](#) above) address this.

RC: COMMENT 12

Methods:

Page 8 line 146-148: the effects of screen use on youth with a clinical condition are likely be different from those of the general population. If this review included any, the population characteristics should be specified in the result section or in the figures.

AR: Yes this is a good suggestion. We have added a column to indicate if the review was conducted on a general population or not in all of the forest plots. This information is also available in Table 1.

RC: COMMENT 13

Page 8 line 152: please define outcomes.

AR: Per [comment 10](#) above, we did not pre-specify outcomes and feel it would be inappropriate to revise our inclusion criteria.

RC: [COMMENT 14](#)

Page 10 line 181-187: these two sections could be combined.

AR: We agree and have made this change.

RC: [COMMENT 15](#)

Page 10 line 190: Table 1 does not contain RoB info.

AR: Our apologies for this oversight. We initially intended to include the quality assessment in Table 1, but felt it made the table too large and moved it to a separate table. It seems we inadvertently did not upload the separate file when submitting the review. We have now included the quality assessment in Table 1.

RC: [COMMENT 16](#)

Page 10 line 195-196: the description of data conversion is not transparent. Pearson's r assumes linear relationship. But associations between screen use and well-being are often non-monotonic (Twenge 2020, PMID: 32303719). In addition, meta-analyses reporting odd ratios defined exposure outcomes in various ways. For example, Zhang et. al. 2016 compared the obesity risk by the lowest and highest TV viewing time, whereas Poorolajal et. al. 2020 compared watching TV > 1-2h/d with an unclear group. Can the harmonization of OR into R guarantee that the two values are comparable? Therefore, further justification is needed.

AR: We agree that the relationship between some outcomes and screen time may not always be linear. However, our previous research (e.g., 10.1186/s12966-019-0881-7) looking at a range of health and education outcomes in a high-quality dataset found little evidence for curvilinear relationships. We therefore felt that the linear model was appropriate for the majority of the outcomes in this review. Additionally, almost none of the reviews we included reported information that would allow us to calculate non-linear effects. We agree that this is an important limitation though, and have added information on this to the *Discussion* section:

We converted all effect sizes to a common metric (Pearson's r) to allow for comparisons of magnitude, but acknowledge that this assumes a linear relationship. Some previous research suggests that associations are typically linear.¹⁸ However, others have identified instances where non-linear relationships exist, especially for very high levels of screen time.^{17,57,58} Additionally, our conversion may not always adequately account for differences in study design or measures of exposures and outcomes. Care is needed, therefore, when interpreting the effect sizes.

Pg. 21

Regarding odds ratios, we initially included these based on several papers that suggested that a conversion to r was possible (e.g., 10.1037/0003-066X.62.3.254). On further consideration, we agree with the reviewer that it is difficult to ensure that this conversion produces comparable estimates when the categories used across studies is inconsistent. We have therefore removed effect sizes calculated from odds ratios from our results. In our efforts to remain completely transparent, these data remain available in the supplementary material.

RC: [COMMENT 17](#)

Page 10 Line 201-206: The inclusion/exclusion criteria are here and there (page 10 line 197-198, page 11 line 221-222, etc.). Please provide specific and full criteria of effect size inclusion/exclusion in one

section.

Page 11 line 211 and 223-226: same issue as above.

AR: We agree this could be confusing to the reader. The previous sections outlined the inclusion criteria for including a *review*, whereas these sections focus on reasons for excluding an *effect size* from the review. We did not require all effect sizes within a review to be eligible or usable for the review to still be included.

Two researchers independently extracted all quantitative meta-analytic effect sizes, including moderation results. We excluded effect sizes which were reported as relative risk ratios or odds ratios, as meta-analyses did not contain sufficient information to meaningfully convert to a correlation. We also excluded effect size estimates when the authors did not provide a sample size. Where possible, we also extracted effect sizes from the primary studies included in each meta-analysis.

To facilitate comparisons, we converted effect sizes to Pearson's r using established formulae.^{27,28} Effect sizes on the original metric are provided in Supplementary File 2.

Pg. 10

RC: COMMENT 18

Page 11 line 222: Egger's test is not suitable for evaluating publication bias of ORs.

AR: Per comment 16, we decided not to include ORs in the meta-analyses.

RC: COMMENT 19

Page 11 line 229-231: An effect with statistical credibility but showing $P < .05$ indicate no association of interest.

AR: We have made this change.

RC: COMMENT 20

Page 12 line 232-233: "include systematic reviews without meta-analyses" conflicts "alongside the main meta-analytic findings".

AR: We are not clear on what the reviewer means here. Per the quoted sentence, we intended to include all systematic reviews (not just meta-analyses), but the volume of meta-analyses was such that we felt this was unneeded. This was a change from our protocol, and thus reported as such. We are happy to revise this section if the reviewer has a suggestion.

RC: COMMENT 21

Page 12 line 235: Harmonizing the metric of effect size does not guarantee the comparability of the strength of associations. Please fully justify the current practice with a thorough analysis of the pros and cons.

AR:

RC: COMMENT 22

Page 12 line 238-243: A significant P was not a criterion to consider studies as credible in other reviews. Not sure why it is a reason for altering the original plan.

AR: It's possible the reviewer has misunderstood. In our original plan, we wrote that $P < 0.001$ would be *one of* the criteria for indicating strong evidence (alongside other metrics, such as sample size and heterogeneity). We realised that this would restrict our 'strong' evidence to only statistically significant effects. What would mean that we could not include evidence for no effect as strong evidence, which is clearly a mistake. Our deviation from the protocol (removing the requirement for a P-value threshold) was intended to rectify this.

RC: COMMENT 23

Results

Page 12 line 247-248: please add a description that 66 studies contributed unique effect sizes to the current review.

AR: This information is available here (note that the updated search increased the number of reviews):

After removing reviews with duplicate exposure/outcome combinations, our process yielded 255 unique effect/outcome combinations (retaining multiple effects for different age groups or study designs) contributed from 103 reviews.

Pg. 13

RC: COMMENT 24

Figure 1: 1) what's the difference between references and studies in this figure? For those seven entries categorized as "non-studies", what are they? 2) "other(explain)" is not self-explanatory. 3) what's the difference between non-target samples and wrong population? In addition, the total number in this box is 1848. 4) "larger study available": I'm not confident about using sample size as a sole criterium of selecting studies.

AR:

1. Yes, we agree that this is confusing. The use of both 'references' and 'studies' comes from the screening software we used (Covidence; <https://www.covidence.org/>). It is intended to capture instances where there are multiple papers ('references') reporting on the same study. In our case, there were seven instances where a review was conducted but the results of the review were reported in multiple papers. While we merged these in Covidence for completeness, in all seven cases all of the information that we needed could be retrieved from just one of the papers (which is the one we provide the citation to). To make this clearer for readers, we've removed this confusing additional extra information.
2. Yes, we agree. We've collapsed all of the "wrong exposure" codings together.
3. We originally used these when screening full-texts for those where the inclusion criteria targeted the wrong population (e.g., the exclusion criteria was >18 years; coded as "wrong population") vs those where the inclusion criteria included youth, but none of the meta-analytic results could be used because they included adults. We think for most readers this is an unimportant distinction, and have collapsed these into one category. We also appreciate the attention to detail. The updated PRISMA diagram has the correct value in this box.
4. We used sample size as the heuristic for selecting which effect to keep in instances where the same exposure and outcome combination was present, and where the effects represented the same population with the same study designs. To create a non-subjective method for choosing which effect size to report we chose to use the sample size, under the assumption that the larger study size would provide the most

information. Inspection of the effects that were excluded reveal that this decision is not that impactful: we would have ended up with almost exactly the same set of results had we instead chosen to use the number of included studies (k) or the most recent review by publication year.

After extracting data, we examined the combinations of exposure and outcomes and removed any effects that appeared multiple times (i.e., in multiple meta-analyses, or with multiple sub-groups in the same meta-analysis), keeping the effect with the largest total sample size. In instances where effect sizes from the same combination of exposure and outcome were drawn from different age-groups (e.g., children vs adolescents), or were drawn using different study designs (e.g., cross-sectional vs longitudinal) we retained both estimates in our dataset.

Pg. 10

RC: COMMENT 25

Page 12 line 249-256: 1) I suggest adding a summary table (counts and percentages) of the basic characteristics of the included references/studies which including year range of publication, types of study designs, sample age categories, outcome types, and exposure types, quality assessment scores, etc. 2) The current description is quite confusing to me, for example, what does “cases” mean? How did you calculate up 197 cases? What appeared twice or more times, the combination of exposure/outcome by age group? 3) table 1: I suggest adding data of the original effect sizes.

AR:

1. We have added a table to summarise the effects (including some information on the reviews these effects were drawn from) as Supplementary File 9.
2. We agree that this section was not entirely clear, and we should not have introduced the term ‘cases’. We have revised the language in this section:

Of the 274 unique exposure/outcome combinations, 242 occurred in only one review, with 23 appearing twice, and 9 appearing three or more times. Full characteristics of the included studies are provided in Table 1.

Pg. 13

3. While we agree that this information might be useful for some readers, we found that this made the table difficult to read. Most reviews presented multiple effect sizes, and once the information to describe each effect was added (e.g., the specific outcome, exposure, age group, and moderator), the table became overwhelming. All of the effect sizes (in their original, converted, and recalculated forms) are available in the supplementary material.

RC: COMMENT 26

Page 26 line 260-268, Information on quality assessment was provided in a supplementary file, not table 1. Please revise.

AR: As noted in [comment 15](#), this has been corrected.

RC: [COMMENT 27](#)

Page 27 line 291: Please provide pre-determined criteria of the magnitude of effects with supporting references.

AR: We have included this information in the *Methods* section:

Throughout the results section we interpret the size of the effects using Funder and Ozer's guidelines:²⁹ very small ($0.05 < r \leq 0.1$), small ($0.1 < r \leq 0.2$), medium ($0.2 < r \leq 0.3$), large ($0.3 < r \leq 0.4$), and very large ($r \geq 0.4$). These are similar to other interpretations based on empirical data.³⁰

Pg. 10

RC: [COMMENT 28](#)

Page 28 line 299-304: 119-31-30-43 = 15, not 17.

AR: Thank you for the attention to detail. The reported numbers are now correct (note that they are different due to the updated search).

RC: [COMMENT 29](#)

Three methodological issues need to be clarified or justified. 1. The selection of reviews to generate "unique" comparisons appeared to be subjective/unclear. 2. The statistical approach extensively reduced the evidence pool. For example, 46 unique effects on education outcomes was reduced to 12 and 119 unique effects on health outcomes was reduced to 17. 3. Harmonizing the metrics of effect sizes into r may be invalid and placing the converted effect sizes in forest plot is somewhat deceiving. Because across-study comparisons are not guaranteed due to variations in study design, setups of exposure and outcome variables.

AR:

1. We agree that the our reporting of the method could have been clearer, and have revised the text in this section to more clearly explain our method. We disagree with the reviewer that the method was "subjective". If we had multiple effects for the same exposure/outcome combination, and which had the same sample age restrictions and the same study design restrictions, we chose the one with the largest sample size. We can certainly see how others might use a different method for eliminating duplicate effects (and we state this in the discussion section), but it is difficult to see how using the sample size could be considered "subjective".

After extracting data, we examined the combinations of exposure and outcomes and removed any effects that appeared multiple times (i.e., in multiple meta-analyses, or with multiple sub-groups in the same meta-analysis), keeping the effect with the largest total sample size. In instances where effect sizes from the same combination of exposure and outcome were drawn from different age-groups (e.g., children vs adolescents), or were drawn using different study designs (e.g., cross-sectional vs longitudinal) we retained both estimates in our dataset.

2. We were also surprised by the volume of effects which did not meet the criteria for statistical credibility. We were surprised, for example, that almost a quarter of all effects did not provide study-level data (as might be presented in a forest plot), which we thought would be a low bar to clear. Almost 40% of those that did have study-level data were derived from samples of less than 1,000 participants (a threshold that has been used by others; 10.1038/s41467-021-24861-8). And, it seems unwise to place too much emphasis on results where there is an indication of publication bias (i.e., statistically significant Egger's test). Given that we pre-specified this criteria, we think that it would be inappropriate to lower the standard because the majority of reviews did not reach it. Instead, we think it reflects the need for the field to raise its standards, especially when it comes to reporting (i.e., making data available, reducing file-drawer problem).

After extracting data, we examined the combinations of exposure and outcomes and removed any effects that appeared multiple times (i.e., in multiple meta-analyses, or with multiple sub-groups in the same meta-analysis), keeping the effect with the largest total sample size. In instances where effect sizes from the same combination of exposure and outcome were drawn from different age-groups (e.g., children vs adolescents), or were drawn using different study designs (e.g., cross-sectional vs longitudinal) we retained both estimates in our dataset.

3. We understand the reviewer's concerns regarding the conversions of the effect sizes (although we disagree with the characterisation of our results as "somewhat deceiving"). We chose Pearson's r as we wanted to focus not on the precise size of the association, but instead on the *direction* and the *magnitude*. We think it is worth considering the counterfactual to our approach: had we not converted effect sizes to a common metric readers would likely still make comparisons, but these would be subjective. We agree that readers should interpret the effect sizes carefully, and have included text to this effect in the *Discussion*:

We converted all effect sizes to a common metric (Pearson's r) to allow for comparisons of magnitude, but acknowledge that this assumes a linear relationship. Some previous research suggests that associations are typically linear.¹⁸ However, others have identified instances where non-linear relationships exist, especially for very high levels of screen time.^{17,57,58} Additionally, our conversion may not always adequately account for differences in study design or measures of exposures and outcomes. Care is needed, therefore, when interpreting the effect sizes.

RC: COMMENT 30

Discussion

Page 31 line 362-368: I think these are important nuances. If the analytical plan was subjective altered in order to generate more consistent associations. Then, the practice can introduce selection bias. I

suggest adding qualitative syntheses of those studies removed by not meeting the “statistical credibility” criteria.

AR: We are unclear what the reviewer is referring to here, although we think many of the concerns were addressed in [comment 29](#). We feel that a qualitative synthesis of more than 200 effect sizes is outside the scope of this paper.

RC: COMMENT 31

Page 31 line 380-383: both guidelines mentioned recreational sedentary screen time. It is not in all instances.

AR: Even within recreational screen time, our findings do not support guidelines which focus just on minimisation. Video games, for example, are a recreational form of screen time which appear to have some benefits for educational and cognitive outcomes.

RC: COMMENT 32

Page 32 line 389-391: I assume that the data of primary studies was retrieved by the team whenever possible. But “failed to provide study-level data” make me wonder did not those reviews provide reference information of the included studies?

AR: There were two ways that the data from the primary studies could be included in our review:

- The data were provided in the paper (i.e., in forest plots, tables, or supplementary files); or
- The authors provided the data when contacted.

We contacted authors who did not provide primary study data in their published article. Where authors did not provide data in a format that could be re-analysed, we used the published results of their original meta-analysis.

Pg. 11

We did not retrieve data from the primary studies when the data were not provided in the paper or by contacting the author. Our results include more than 3,000 primary studies, and even after accounting for those that provided data in the paper, it would not be feasible to extract all of the remaining data.

RC: COMMENT 33

Page 32 line 395: from my point of view, the evidence strength generated from data from ten loosely designed cross-sectional studies is weaker than 5 rigorously designed intervention studies or longitudinal studies. I wonder how the quality and quantity of studies were balanced in the current study.

AR: We agree that study design should have been a larger factor when choosing which effects to use. We have rectified this by including study design as a factor in our method of choosing effects. That is, if there was the same exposure and outcome for the same age group, but one review used cross-sectional evidence and another used longitudinal, we retained both in the dataset. However, we found that there were not a lot of reviews which examined the same exposure and outcome, but differed on choice of study design. When we included study design as part of our method of choosing effects, it only increased the number of effects by 5 (from 241 to 246).

We agree that readers should include study design in their interpretation of the results, and have therefore included this information in all forest plots.

[RC 3.1.] Reviewer 3

RC: COMMENT 34

OVERALL IMPRESSION

The article by Sanders et al is an umbrella review of the literature on benefits and risks associated with screen interaction in children and adolescents. This is a noteworthy and hotly debated topic, with significant social and political implications. Therefore, the current study can certainly inform and impact policy and society in general.

Overall, the methodology is sound and follows current guidelines on how to carry out systematic reviews and meta-analysis. Moreover, an umbrella review is an apt choice to summarize the wide and heterogeneous literature on the topic. It is the first of its kind, and therefore represents a significant evidence-based advance in summarizing a complex topic. All in all, there are not any unexpected or highly controversial results but the final document is a good quality common-sense summary of the literature that should probably be read by anybody with extreme positions in the topic.

The umbrella review was registered in PROSPERO and the (rather small) deviations from the protocol are adequately reasoned by the authors. Indeed, I would like to praise authors for centering their efforts in interpreting mainly “credible” effects that are either significant or non-significant.

AR: Thank you for the kind comments. We are glad that you saw the value in our review.

RC: COMMENT 35

MAJOR COMMENTS

My biggest issue with the article as it stands is in regards to the date of the last search. This was carried out in May 2020, so two years and a half have already passed, so the review is quite outdated. Moreover, these two years and half with the Covid pandemic have likely changed how many children interact with technologies and screens. I suspect that many primary articles have been published on the issue at hand, and, quite likely some meta-analyses too. I would recommend the authors to update their search and results, even acknowledging that that means quite a lot of work. That being said, this update might yield some meta-analyses that include primary articles carried out during the pandemic, when the positive and negative outcomes in relation to screen-time might be correlated with factors that differ greatly from previous years, and this should probably be mentioned in the discussion.

AR: We agree that our search was out of date. We reran the search in September 2022, which revealed a substantial amount of new literature.

We conducted an initial search on August 17, 2018 and refreshed the search on September 27, 2022.

Pg. 9

Our updated search did reveal some meta-analyses on screen use during or post-pandemic, but these tended to be small and it would be difficult to draw too many conclusions from these reviews. The historic nature of reviews is an issue we acknowledge in our limitations:

In addition, reviews provide only historical evidence which may not keep up with the changing ways children can engage with screens. While our synthesis of the existing evidence provides information about how screens might have influenced children in the past, it is difficult to know if these findings will

translate to new forms of technology in the future.

Pg. 22

RC: COMMENT 36

Another main issue that I have with the article is that I feel that a better description of supplementary materials is needed. I have not been able to find a consistent description of the materials that can be found in them. Moreover, in the case of the CSV documents a variable dictionary would be most welcome. Somewhat relatedly, in line 190 (and 260) it says “see table 1”, so I was hoping to see some summary on the risk of bias evaluation in the table. I have not been able to find it there. In the Prisma checklist item 18 it says that risk of bias assessments can be found in OSF, but no mention of OSF is provided in the main text. After going through all the supplementary materials, I have been able to find the ratings in a file. My reading from all this is that the article has gone through several iterations and some systematizing and cleaning should be carried out to present in a more clear way the supplementary materials (an index of contents with a short description, table and figure captions, variable dictionary...)

AR:

Our apologies for this oversight. Per our response to Reviewer 2 ([comment 15](#)), we initially intended to include the quality assessment in Table 1. We removed it because it made the table very large, but did not include the new file with the separate table. We have now included the quality assessment in Table 1.

We agree that the supplementary materials are difficult to navigate. We have added a codebook (as Supplementary File 3) as suggested.

RC: COMMENT 37

MINOR COMMENTS

The selection of the “the National Health, Lung and Blood Institute’s Quality Assessment of Systematic Reviews and Meta-Analyses” as opposed to the more standardized AMSTAR-2 is slightly surprising (even if it does not likely have any real consequences in the final product).

AR: Our rationale for choosing the NHLBI was that we felt it could be applied equally to all fields than AMSTAR-2. Reporting standards differ by discipline, and several of the items in AMSTAR-2 are typically only reported in medical fields. While these should maybe be reported in all fields, we felt that the NHLBI was a more appropriate standard for our review.

RC: COMMENT 38

The paragraph on age within the eligibility criteria is unclear to me: “(i.e., the highest individual study from the meta-analysis had a mean age was < 18 years).”: is it a mean age below eighteen (hence, potentially including some adults), or is it the oldest participant being under 18 (more clear-cut, but on the other hand limiting the number of outcomes)?

AR: It is the former, which we have now clarified here:

We included meta-analyses containing studies that combined data from adults and youth if meta-analytic effect size estimates specific to participants aged 18 years or less could be extracted (i.e., the highest

mean age for any individual study included in the meta-analysis was < 18 years). A meta-analysis was still included if the age range exceed 18 years, provided that the mean age was less than 18. We excluded meta-analyses that only contained evidence gathered from adults (age >18 years).

Pg. 8

We chose this as most reviews provided mean ages for individual studies but not age ranges, which meant we would not have been able to consistently apply the criteria.

RC: COMMENT 39

“In this review we adopted a population-level perspective, meaning that we examined electronic media exposure that occurs during typical daily living activities (e.g., home, school-based electronic media exposure).”: I do not think this paragraph is fully clear.

AR: We agree that this was not clear, and have revised it:

In this review we focused on electronic media exposure that would be considered typical for children and youth. That is, exposure that may occur in the home setting, or during schooling. Consistent with this approach, we excluded technology-based treatments for clinical conditions.

Pg. 8

RC: COMMENT 40

“Outcomes: We included all reported outcomes.”: it might be worth adding “on benefits and risks”

AR: We have updated this text.

Outcomes: We included all reported outcomes on benefits and risks.

Pg. 8

RC: COMMENT 41

Lines 200-204: it is unclear to me whether this refers to effects found within the same meta-analysis, across meta-analysis on the same outcome, or both. Some additional clarification would be welcome.

AR: We agree this could have been clearer. We have revised the text to indicate that it was both:

After extracting data, we examined the combinations of exposure and outcomes and removed any effects that appeared multiple times (i.e., in multiple meta-analyses, or with multiple sub-groups in the same meta-analysis), keeping the effect with the largest total sample size. In instances where effect sizes from the same combination of exposure and outcome were drawn from different age-groups (e.g., children vs adolescents), or were drawn using different study designs (e.g., cross-sectional vs longitudinal) we retained both estimates in our dataset.

Pg. 10

RC: COMMENT 42

While PRISMA is named along the review, it is not specifically indicated that it was followed. I would recommend that authors explicitly say it.

AR: We have noted this under the *Methods* section.

RC: COMMENT 43

Education outcomes: text indicates 12 effects from 8 reviews, but figure 2 only shows 11 effects from 7 reviews. A similar issue happens with Helath related outcomes: text indicates 17 effects, but I only see 15 in figure 3. What is going on?

AR: Thank you for this attention to detail. This is the same issue raised by Reviewer 2 ([comment 28](#)), and we have corrected these numbers.

RC: COMMENT 44

“For example, a meta-analysis of the effect of television watching on learning among adolescents diagnosed with depression would be included.” This seems a strange choice, as it mixes very different populations. At the very least, the type of population should be indicated in table 1, and the impact of this discussed in the discussion.

AR: Yes, we agree that this information is important. We have included information about the sample in Table 1, and have indicated it on all forest plots. We also note that our criteria for choosing effect sizes in the case of duplicates (i.e., taking the one with the largest N) meant that we would generally take the effect size for the ‘typical’ population. More concretely: in the example cited we would typically choose the effect size of television watching on learning for the general population, rather than for those who are depressed, because the sample size would be larger for the former.