

Recreational Screen Use and Internalizing Problems from Preadolescence to Young Adulthood in a Longitudinal Twin Cohort with Co-Twin Comparison

Corresponding Author: Dr Tong Gong

Version 0:

Decision Letter:

21st October 2025

Dear Dr. Gong,

Thank you once again for your manuscript, entitled "Recreational Screen Use and Internalizing Problems from Preadolescence to Young Adulthood: A population-based Cohort and Co-Twin Control Study", and for your patience during the peer review process.

Your Article has now been evaluated by 3 referees. You will see from their comments copied below that, although they find your work of potential interest, they have raised quite substantial concerns. In light of these comments, we cannot accept the manuscript for publication, but would be interested in considering a revised version if you are willing and able to fully address reviewer and editorial concerns.

We hope you will find the referees' comments useful as you decide how to proceed. If you wish to submit a substantially revised manuscript, please bear in mind that we will be reluctant to approach the referees again in the absence of major revisions. We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

To guide the scope of the revisions, the editors discuss the referee reports in detail within the team, including with the chief editor, with a view to (1) identifying key priorities that should be addressed in revision and (2) overruling referee requests that are deemed beyond the scope of the current study. We hope that you will find the prioritised set of referee points to be useful when revising your study. Please do not hesitate to get in touch if you would like to discuss these issues further.

In particular, please address the following (as well as all other reviewer comments):

- (1) Causal claims should be reserved for only the specific results of the co-twin control analyses that hold after correction for multiple testing, and such interpretations must acknowledge the limitations of the co-twin control method, which cannot rule out alternative explanations for observed associations.
- (2) Ensure that all interpretations and conclusions accurately reflect the pattern of results. Non-statistically significant results in the twin comparisons should be described as such, and interpreted as showing no evidence of an effect (rather than consistent with cross-sectional results).
- (3) Reviewer 2 mentions the need to correct for multiple comparisons, and we ask that you include such corrections. Please note that in cases where both corrected and uncorrected results are presented, all interpretations of the results must be based on corrected (rather than uncorrected) tests.
- (4) Carry out the additional analyses recommended by Reviewer 1, including the use of additional control variables and the direct calculation of effect attenuation.
- (5) Ensure that your underlying conceptual model and the limitations of your dataset are entirely clear, as highlighted by Reviewer 3.
- (6) We appreciate Reviewer 3's suggestion to reduce this paper to presenting only the co-twin comparisons. While we agree that the paper would benefit from being more streamlined, we ask that you do not remove the correlational results entirely. However, we agree that it may increase the readability and reach of the work if you move some of these to the SI, to ensure that your main text is focused on the most informative and novel results.

Finally, your revised manuscript must comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its consideration. To assist you in this process, I have attached a

checklist that lists all of our requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

If you wish to submit a suitably revised manuscript, we would hope to receive it within 4 months. I would be grateful if you could contact us as soon as possible if you foresee difficulties with meeting this target resubmission date.

With your revision, please:

- Include a "Response to the editors and reviewers" document detailing, point-by-point, how you addressed each editor and referee comment. If no action was taken to address a point, you must provide a compelling argument. When formatting this document, please respond to each reviewer comment individually, including the full text of the reviewer comment verbatim followed by your response to the individual point. This response will be used by the editors to evaluate your revision and sent back to the reviewers along with the revised manuscript.
- Highlight all changes made to your manuscript or provide us with a version that tracks changes.

Please use the link below to submit your revised manuscript and related files:

Link Redacted

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

Thank you for the opportunity to review your work. Please do not hesitate to contact me if you have any questions or would like to discuss the required revisions further.

Sincerely,

[Redacted]

[Redacted]

Nature Human Behaviour

Reviewer expertise:

Reviewer #1: child mental health, twin studies

Reviewer #2: mental health, twin studies

Reviewer #3: psychiatric epidemiology, screen use and mental health

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

It was great to read this very interesting and well-conducted study examining the relationship between screen time and internalising problems. The study used a large longitudinal sample of Swedish twins and conducted multiple analyses, including co-twin control methods to strengthen causal inference. The paper was very well written, methodologically sound, and the figure in the introduction providing the study overview was excellent. Below I provide some suggestions for improvement.

Abstract:

1. It would be helpful to state an effect size or range of effect sizes, if possible.

Results

Section "Cross-sectional associations between recreational screen use and internalizing problems":

2. For the sentence beginning "at age 9" it would be helpful to state that this is in the full sample/phenotypic analyses, to clarify to the reader. It would also be helpful to briefly describe the effect sizes in the text (e.g., "standardised effect sizes ranging from 0.05-0.14"). Similarly, for the sentence beginning "at ages 15 and 18" it would also be beneficial to state that this is in phenotypic analyses.

3. For the paragraph on discordant twin comparisons at ages 15 and 18, the text seems misleading, as it says "most associations remained significant within discordant twin pairs, where shared genetic confounding and familial environment factors were implicitly accounted for". However, in Figure 3D, the results were non-significant for 3 out of 4 screen time exposures at age 15. I therefore suggest rewording this to specify that the associations remained significant at age 18, while at age 15, most associations became non-significant in MZ twins.

4. So that readers can understand the extent of genetic and environmental confounding, it would be helpful to calculate a %

reduction in effect size from the phenotypic to the MZ co-twin control analyses.

Discussion

5. Related to the above point, I suggest rephrasing the sentence “Cross-sectional associations were strongest at ages 15 and 18 and more pronounced among girls, even after adjusting for prior symptoms and shared genetic and familial environmental factors” given that the twin-findings were largely non-significant at age 15.

6. Similarly, the sentence “the observed associations in our cohort largely persisted within twin pairs cross sectionally in mid-to-late adolescence” should be rephrased to specify this is mainly in late adolescence, not mid adolescence.

7. The sentence “These findings suggest that increased screen use may independently contribute to internalizing problems beyond shared familial factors” seems a bit of an overinterpretation, as many associations were fully attenuated in the co-twin control analyses. E.g., in longitudinal co-twin analyses, only 3 of 8 analyses were statistically significant (Figure 4). I would recommend rephrasing this sentence to reflect a more cautious and nuanced interpretation of the findings.

8. Sentence “Together, these findings, along with previous literature⁵², emphasize the specific need of focusing on clinically vulnerable subgroups and developing targeted interventions (for example, coping strategies for negative feedback or social comparison) to prevent severe outcomes, including the rising incidence of self-harm among youth^{14,53}”. As self-harm wasn't examined here, it doesn't seem justified as an implication of this study.

9. Regarding the paragraph on effect sizes being too small to warrant policy change, I would suggest two changes here. First, it doesn't seem appropriate to discuss the effect sizes of the analysis on clinically ascertained outcomes, as these results don't represent causal estimates. Instead, I would recommend comparing the effect sizes from the co-twin control analysis (i.e., the most likely causal estimates) to effect sizes of other developmental risk factors from other quasi-experimental work. For example, meta-analyses of quasi-experimental studies have estimated the causal effects of bullying victimisation (Cohen's $d=0.27$; Schoeler et al., 2018), childhood maltreatment (Cohen's $d=0.24$; Baldwin et al., 2023) and parental anxiety ($r=0.13$; Ahmadzadeh et al., 2021) on internalising problems. This would provide a helpful comparison, a bit like in the Orben et al (2019) paper on screen time (although these effect sizes were unadjusted).

10. Second, it is noted “Assuming a similar degree of effect attenuation due to shared familial factors (as observed in symptom-based co-twin analyses), these results suggest that high screen engagement during adolescence have clinically meaningful implication”. This is a fair point, but it could be made more convincing and evidence-based if you (i) calculated the degree of effect attenuation in symptom co-twin control analyses (i.e., on average across outcomes), and (ii) applied it to the clinically ascertained outcomes, to estimate the likely effect on clinically ascertained outcomes if those estimates accounted for familial confounding.

11. As a final point, it is stated in the limitations section “Fourth, residual confounding from non-shared factors may have influenced the findings. For example, physical activity, peer relationship, and social environments during adolescence could displace screen use and may vary within twin pairs.” However, is it not possible to control for these factors in CATSS data? To my knowledge, there are measures of physical activity, peer bullying, and social environments (e.g., family, school, peers) available.

Reviewer #2 (Remarks to the Author):

Thank you for inviting me to review this manuscript. I have expertise in the methods utilised and the main outcome of interest (adolescent mental health).

The manuscript aims to investigate the relationship between recreational screen use and anxiety and depression. The topic is of importance as anxiety and depression are increasing in young people and screen use, in particular social media, has been associated with these outcomes. As the authors note, much, but not all, previous research has been correlational in nature and there a cause – effect relationship has yet to be established. In the current analysis, the authors utilise CATSS a twin registry based in Sweden. They also use linked medical records to explore this relationship. They use the discordant twin approach and longitudinal design to try and get closer to understanding the causal relationship.

The three main RQs are well defined and they use suitable statistical methods to answer the questions. There are a lot of analyses but overall, I think the authors do well to presenting these clearly and summarising the key points. My assessment of this manuscript is very positive.

I have some comments below that the authors can consider if invited to re-submit.

Abstract:

“Associations between longer screen time and elevated internalizing symptoms during adolescence persisted in co-twin comparisons, supporting a potential causal link”

Is this too strong? This is related to a general point I make in the result section. Not all associations were significant at the level of MZ discordant pairs. I wonder if the key finding is that there are likely to be familial influences on this relationship (given the attenuating associations) however there is also some evidence for causal relationship. I know that sometimes a simpler message is good but want to make sure that you are happy that this is exactly what your results are implying.

Methods:

Can you give a bit more detail about the discordant twin analysis and give a reference for the approach taken? There are a 2 main ways that this can be done (to my knowledge) and be good to be explicit.

Results:

Did you or have you considered correcting for multiple testing?

P10 "most" I think maybe need to acknowledge that the twin differences were a bit more mixed

I understand that this could be personal preference, but I do not find the Sankey diagrams particularly helpful for interpreting the pattern of the trajectories. This might be the best way to represent this data but perhaps it could be a bit clearer?

Discussion:

P18"... Consistent with these concerns, we observed attenuated associations in discordant DZ and MZ twin comparisons, supporting a role of shared genetic and environmental liability. Nonetheless, the observed associations in our cohort largely persisted within twin pairs cross..."

I think this does represent what you found but wonder if this message is aligned with how you have framed it in abstract and some sections of the results (see earlier comments).

I liked the inclusion of the cox analysis and linkage with medical records. I do think though you can emphasise the strengths of symptom based approaches. Something doesn't have to be clinically relevant to have impact on behaviour/wellbeing. Also acknowledging that not everyone with clinically relevant symptoms will approach the health service for treatment.

I thought the discussion was well considered.

Reviewer #3 (Remarks to the Author):

The present study is a comprehensive, longitudinal and well designed study to investigate associations between screen time and adolescent to adult mental health using a range of methods including cross-sectional, longitudinal, twin comparisons, and time to event analyses. The sample size is large, and the sample is well characterized. This study present a wide range of results using a novel study design and substantially adds to the literature. The design of the study is rigorous. I have several comments for the reviewers to consider.

While comprehensive, there are also hundreds of models and results for a reader to wade through. In addition to six figures and table, each of which often includes many screen time measures and several outcomes, there are 28 supplemental tables and 13 supplemental figures. I suggest that the authors pare down this paper to the most novel results - the twin comparisons. For example, cross-sectional associations are not very novel and the results are largely similar to any large study. In my opinion, the paper would have a bigger impact if the authors select the most informative and novel results rather than presenting every possible comparison in the data.

This is especially the case because time spent on digital media is not a very sensitive measure of online exposure. Indeed in my opinion the biggest limitation of the study, one that cannot be overcome with additional analysis, is that the consensus of the field is now that 'time spent' is very limited in terms of the information it provides about risks associated with digital media use. Without information on what the adolescents are doing online, whether they are distressed about their use, or any other more relevant factor, the present study does not add substantially to the literature except that it includes twins which provides more rigorous control for shared genetic and environmental factors, which is a major advantage of the design. Authors divide their measures into 'interactive' and 'passive', but it's not always the case that gaming for example is interactive, or that video watching is passive (e.g. if users are interacting with others on YouTube while watching). These limitations should be taken into account.

Some additional minor comments are about framing and interpretation. The authors make the case in the introduction that twin comparisons are necessary for this topic to control for shared genetic and environmental factors, but do not make the case that there is substantial evidence for genetic and shared environmental confounding or what these sources of confounding may be. Greater attention to the underlying conceptual model that motivates the twin comparisons would benefit the argument. Aligned with this comment, the authors interpret higher screen time correlations among MZ compared with DZ twins as evidence of genetic contributions. It's not clear how genetic factors would influence the amount of screen time someone engages in, it may be a more plausible scenario that there are violations of the equal environments assumption when it comes to activities that MZ and DZ twins engage in on a regular basis. The authors could address this point and consider whether there is sufficient literature on the genetic of screen time to make this case.

All in all, this is a very well done study and adds to the literature. I think the paper would be stronger with greater attention to paring down the core set of novel results, and giving additional attention to the plausibility of various arguments.

Version 1:

Decision Letter:

20th January 2026

Dear Dr. Gong,

Thank you for submitting your revised manuscript, "Recreational Screen Use and Internalizing Problems from Preadolescence to Young Adulthood in a population-based Twin Cohort with Co-Twin Comparison". After careful consideration and discussion with my colleagues, I am sorry to have to tell you that we do not feel that the referees' and editors' previous comments have been sufficiently addressed to justify sending this revision back for peer review.

This unusual course of action is taken occasionally to avoid unproductive rounds of review that result in reviewer fatigue and damage the chances of the manuscript obtaining a fair and objective evaluation. Such situations are not in an author's best interest so we try to avoid them when it seems prudent to do so.

In order to consider this manuscript further we would request that you please do your best to fully address all of the comments of the reviewers and editors. In particular, please ensure that you address the following:

1) Following a request from Reviewer 2 and the editors, please ensure that all conclusions are based on results corrected for multiple comparisons. Please note that it is journal policy that in cases where both corrected and uncorrected results are presented, interpretations must be based on corrected tests.

2) Ensure that your manuscript is entirely clear about which analyses have a potential causal interpretation and which do not. For example, the Discussion section (around line 403) currently describes most aspects of your study as providing "triangulation of causal evidence", but these methods are entirely correlational. Although your co-twin comparison analyses may provide evidence consistent with a causal interpretation (as highlighted by the reviewers), the other analyses do not. This distinction must be reflected throughout the text, including in the statement of research questions and discussion.

3) Please ensure that your presentation and discussion of results makes the full pattern of evidence clear and avoids selectively presenting results that were statistically significant. As an example of what we mean, pg. 7 refers to Supplementary Table 17 and presents results for 18-year-olds, stating "within-pair estimates showed only slight additional attenuation within MZ pairs. For example, heavy use of interactive screen activities during weekends was associated with standardized β (95% CI) on age-18 CESD scores of 0.31 (0.24–0.38) in the full sample, 0.27 (0.15–0.39) within DZ pairs, and 0.26 (0.07–0.54) within MZ pairs". While it is true that more than 6 hours of interactive activities on weekends was significantly associated with CESD scores at age 18 for within MZ pairs, this was just one of several variables considered (i.e. there were a variety of combinations of activity type and intensity evaluated). Of the 12 estimates of CESD scores for 18-year-old within MZ twins, only 4 appear to be statistically significant. The discussion of these results should be revised to reflect the mixed pattern of evidence and to avoid mentioning only the subset of results that were statistically significant. Please carefully revise throughout the manuscript with this point in mind.

Should you be able to adequately respond to these and the reviewers' other concerns, we would be happy to look at a revised manuscript.

We shall hope to receive your revised version as soon as possible. If you anticipate a delay of more than six weeks, however, please let us know. We will be happy to consider your revision so long as nothing similar has been accepted for publication at Nature Human Behaviour or published elsewhere. Should your manuscript be substantially delayed without notifying us in advance and your article is eventually published, the received date may be that of the revised, not the original, version.

If you are not interested in submitting a suitably revised manuscript in the future please let me know immediately so we can close your file. If you have any questions, please contact me.

Please use the link below to submit a suitably revised manuscript and updated response to referees when they are ready:

Link Redacted

Sincerely,

[Redacted Signature]

[Redacted Name]

Nature Human Behaviour

Version 2:

Decision Letter:

Our ref: NATHUMBEHAV-25083638B

8th June 2026

Dear Tong,

Thank you for submitting your revised manuscript "Recreational Screen Use and Internalizing Problems from Preadolescence to Young Adulthood in a population-based Twin Cohort with Co-Twin Comparison" (NATHUMBEHAV-25083638B), and for your

patience while waiting for a decision.

Your revision has now been seen by the original referees, and their comments are below. As you can see, the reviewers find that the paper has improved in revision. We will therefore be happy in principle to publish it in Nature Human Behaviour, pending revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements within approximately three weeks. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Please do not hesitate to contact me if you have any questions.

Sincerely,

[REDACTED]

[REDACTED]

Nature Human Behaviour

Reviewer #1 (Remarks to the Author):

The authors have done an excellent job of addressing my comments.

Reviewer #2 (Remarks to the Author):

I have reviewed the response to reviewers and believe that the authors have adequately addressed the comments I have made.

I also read the revised manuscript. The re-draft is much clearer and the analyses presented are more streamlined which helps with readability. I think the authors have modified language to reflect the mixed picture regarding causality.

Overall, I think the manuscript is an important addition to the literature on screen use and mental health.

Reviewer #3 (Remarks to the Author):

The authors have addressed concerns and comments rigorously and reasonably and I have no additional concerns with the manuscript. It is a rigorous and well done study that contributes novel data and interpretation to the field.

Version 3:

Decision Letter:

Dear Tong,

I am happy to let you know that your Article "Recreational Screen Use and Internalizing Problems from Preadolescence to Young Adulthood in a Longitudinal Twin Cohort with Co-Twin Comparison", has now been accepted for publication in Nature Human Behaviour.

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to <https://www.springernature.com/gp/open-science/plan-s-compliance> Plan S principles or the <https://www.springernature.com/gp/open-science/us-federal-agency-compliance> NIH public access policy) then you should select the gold OA route, and we will direct you to the compliant route where possible. Because authors warrant under our subscription licensing terms that they haven't committed to licensing any version of their article under a licence inconsistent with the terms of our agreement – including the applicable embargo period – publication under the subscription model isn't suitable for authors whose funders require no embargo.

Once your manuscript is typeset and you have completed the appropriate grant of rights, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see <http://www.nature.com/nathumbehav/info/gta>). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our web site).

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

An online order form for reprints of your paper is available at <https://www.nature.com/reprints/author-reprints.html>. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Human Behaviour as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Human Behaviour logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

In approximately 10 business days you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

We look forward to publishing your paper.

With best regards,



Nature Human Behaviour

P.S. Click on the following link if you would like to recommend Nature Human Behaviour to your librarian
<http://www.nature.com/subscriptions/recommend.html#forms>

** Visit the Springer Nature Editorial and Publishing website at http://editorial-jobs.springernature.com?utm_source=ejp_NHumB_email&utm_medium=ejp_NHumB_email&utm_campaign=ejp_NHumB for more information about our career opportunities. If you have any questions please click [here](mailto:editorial.publishing.jobs@springernature.com).

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source. The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Response to Editors and Reviews

– NATHUMBEHAV-25083638 “Recreational Screen Use and Internalizing Problems from Preadolescence to Young Adulthood in a population-based Twin Cohort with Co-Twin Comparison”

We thank the editor and reviewers for evaluating our manuscript and providing valuable suggestions to improve the quality of our work. In this revision, we have responded to the comments point-by-point and highlighted the corresponding changes in the manuscript (STR_screentime_depanx_Main_20260107_track.pdf). We have included a clean manuscript (STR_screentime_depanx_Main_20260107_clean.pdf).

Editor Comments:

In particular, please address the following (as well as all other reviewer comments):

(1) Causal claims should be reserved for only the specific results of the co-twin control analyses that hold after correction for multiple testing, and such interpretations must acknowledge the limitations of the co-twin control method, which cannot rule out alternative explanations for observed associations.

We appreciate the Editor’s comment and agree that causal interpretations should be more tightly aligned with the specific results of the co-twin control analyses while acknowledging the limitation of co-twin control method. Specifically, we have made the following changes in this revision:

1. To strengthen causal interpretation, we added co-twin control design for the time-to-event analysis on clinically ascertained outcomes (see R1.9).
2. We applied multiple test correction using the Benjamini–Yekutieli procedure¹ (see R2.3).
3. We revised interpretations in the Abstract, Results, and Discussion.

(2) Ensure that all interpretations and conclusions accurately reflect the pattern of results. Non-statistically significant results in the twin comparisons should be described as such, and interpreted as showing no evidence of an effect (rather than consistent with cross-sectional results).

We have revised text throughout the manuscript to cautiously interpret the non-statistically significant results.

(3) Reviewer 2 mentions the need to correct for multiple comparisons, and we ask that you include such corrections. Please note that in cases where both corrected and uncorrected results are presented, all interpretations of the results must be based on corrected (rather than uncorrected) tests.

As per Reviewer 2’s suggestion, we applied multiple testing corrections across all analyses (see R2.3).

(4) Carry out the additional analyses recommended by Reviewer 1, including the use of additional control variables and the direct calculation of effect attenuation.

In response to these valuable suggestions, we conducted new sensitivity analyses in which we controlled for measures of physical activity and peer bullying (R1.11) and presented the effect attenuation comparing effects estimated from all twins versus those from co-twin control analyses (R1.4, R1.10).

(5) Ensure that your underlying conceptual model and the limitations of your dataset are entirely clear, as highlighted by Reviewer 3.

We have added the underlying conceptual model and clearly stated the limitations now (see R3.2, R3.3).

(6) We appreciate Reviewer 3’s suggestion to reduce this paper to presenting only the co-twin comparisons. While we agree that the paper would benefit from being more streamlined, we ask that you do not remove the correlational results entirely. However, we agree that it may increase the readability and reach of the work if you move some of these to the SI, to ensure that your main text is focused on the most informative and novel results.

We have now moved the figure describing twin correlations of screen time and internalizing problems (previously Figure 2) into supplementary information (now Supplementary Figure 3), leaving with one figure/table per RQ in the main text.

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author): *It was great to read this very interesting and well-conducted study examining the relationship between screen time and internalising problems. The study used a large longitudinal sample of Swedish twins and conducted multiple analyses, including co-twin control methods to strengthen causal inference. The paper was very well written, methodologically sound, and the figure in the introduction providing the study overview was excellent. Below I provide some suggestions for improvement.*

We thank the reviewer for the considerable time and effort devoted to reviewing our manuscript and for the insightful and constructive feedback provided.

Abstract:

R1.1. It would be helpful to state an effect size or range of effect sizes, if possible.

Thank you. The sentences have been revised accordingly and now read as “Cross-sectional associations between longer screen time and elevated internalizing symptoms (β ranging 0.09-0.38) were attenuated in co-twin comparisons but largely persisted in late adolescence, suggesting that shared familial factors may not fully explain these associations”. (page 2 lines 37-42)

Results Section “Cross-sectional associations between recreational screen use and internalizing problems”:
R1.2. For the sentence beginning “at age 9” it would be helpful to state that this is in the full sample/phenotypic analyses, to clarify to the reader. It would also be helpful to briefly describe the effect sizes in the text (e.g., “standardised effect sizes ranging from 0.05-0.14”). Similarly, for the sentence beginning “at ages 15 and 18” it would also be beneficial to state that this is in phenotypic analyses.

Thanks for the suggestion. These two sentences have been revised as follows:

“At age 9, we observed associations with depressive and anxiety symptoms across different types of screen use, when comparing the most frequent use (daily) to restrictive use (≤ 2 times/week) in the full sample (standardized β ranging from 0 to 0.12, **Supplementary Table 17**).” (page 12 lines 310-312) and

“At ages 15 and 18, the association analyses between screen time and internalizing problems in all twins demonstrated a dose-response pattern, with longer screen time per day corresponding to higher symptom scores (standardized β ranging from 0.21 to 0.38 when comparing heavy use [>6 hours/day] to restrictive use [≤ 2 hours/day], **Supplementary Table 17**).” (page 12 lines 317-321)

R1.3. For the paragraph on discordant twin comparisons at ages 15 and 18, the text seems misleading, as it says “most associations remained significant within discordant twin pairs, where shared genetic confounding and familial environment factors were implicitly accounted for”. However, in Figure 3D, the results were non-significant for 3 out of 4 screen time exposures at age 15. I therefore suggest rewording this to specify that the associations remained significant at age 18, while at age 15, most associations became non-significant in MZ twins.

We revised the paragraph on page 7 lines 152-160 to reflect cautious interpretation of these results: “In discordant twin comparisons at ages 15 and 18 (**Figure 2D-F**), we observed attenuation of effect estimates within DZ pairs and a more pronounced attenuation within MZ pairs, particularly at age 15. At this age, within-MZ pair associations were attenuated to non-significance (effect attenuated by 32-95%; **Supplementary Table 17**), suggesting substantial confounding by shared genetic and familial environmental factors. By contrast, at age 18, within-pair estimates showed only slight additional attenuation within MZ pairs. For example, heavy use of interactive screen activities during weekends was associated with standardized β (95% CI) on age-18 CESD scores of 0.31 (0.24–0.38) in the full sample, 0.27 (0.15–0.39) within DZ pairs, and 0.26 (0.07–0.54) within MZ pairs (**Figure 2D-F; Supplementary Table 17**).

R1.4. So that readers can understand the extent of genetic and environmental confounding, it would be helpful to calculate a % reduction in effect size from the phenotypic to the MZ co-twin control analyses.

Thank you for the great suggestion. We have now added a column with percentage of effect size change for cross-sectional and cross-age associations, as well as for time-to-develop clinical condition analyses (see **Supplementary Tables 17, 19, 22, and Table 1**). For example, at age 15, the effect of heavy use [>6 hours/day] on depressive/anxiety symptoms was attenuated by 32-95% from all twins to the MZ co-twin comparison; while the effect attenuation was more modest at age 18 (e.g., heavy use effect by interactive screen activities was attenuated by 9-61%).

Discussion

R1.5. Related to the above point, I suggest rephrasing the sentence “Cross-sectional associations were strongest at ages 15 and 18 and more pronounced among girls, even after adjusting for prior symptoms and shared genetic and familial environmental factors” given that the twin-findings were largely non-significant at age 15.

We have rephrased this sentence, which now reads “Cross-sectional associations between longer screen time and internalizing symptoms were strongest at ages 15 and 18 and more pronounced among girls. However, after adjusting for prior symptoms and shared genetic and familial environmental factors, these associations were attenuated but remained largely positive, although no longer statistically significant at age 15 (RQ1).” (page 24 lines 554-559)

R1.6. Similarly, the sentence “the observed associations in our cohort largely persisted within twin pairs cross sectionally in mid-to-late adolescence” should be rephrased to specify this is mainly in late adolescence, not mid adolescence.

We have rephrased this sentence accordingly, which reads “...the observed associations in our cohort largely persisted within twin pairs cross-sectionally in late adolescence”. (page 24-25 lines 594-596)

R1.7. The sentence “These findings suggest that increased screen use may independently contribute to internalizing problems beyond shared familial factors” seems a bit of an overinterpretation, as many associations were fully attenuated in the co-twin control analyses. E.g., in longitudinal co-twin analyses, only 3 of 8 analyses were statistically significant (Figure 4). I would recommend rephrasing this sentence to reflect a more cautious and nuanced interpretation of the findings.

Thank you for the recommendation. We have rephrased this sentence accordingly, which reads

“During these developmental periods, the observed pattern of effect attenuations in co-twin comparisons was most consistent with the hypothesis of a causal effect after taking familial confounding into account². Alternatively, the residual effect could still be explained by residual confounding from non-shared factors, chance findings, measurement errors, or carry-over effects (e.g., heavy screen user influencing his or her co-twin’s screen behavior or internalizing problem)³.” (page 25 lines 598-608)

R1.8. Sentence “Together, these findings, along with previous literature⁵², emphasize the specific need of focusing on clinically vulnerable subgroups and developing targeted interventions (for example, coping strategies for negative feedback or social comparison) to prevent severe outcomes, including the rising incidence of self-harm among youth^{14,53}”. As self-harm wasn’t examined here, it doesn’t seem justified as an implication of this study.

Thanks for noticing this. We have now removed the self-harm part of the sentence, which now reads “... to prevent worsening internalizing symptoms and related adverse outcomes”.

R1.9. Regarding the paragraph on effect sizes being too small to warrant policy change, I would suggest two changes here. First, it doesn’t seem appropriate to discuss the effect sizes of the analysis on clinically

ascertained outcomes, as these results don't represent causal estimates. Instead, I would recommend comparing the effect sizes from the co-twin control analysis (i.e., the most likely causal estimates) to effect sizes of other developmental risk factors from other quasi-experimental work. For example, meta-analyses of quasi-experimental studies have estimated the causal effects of bullying victimisation (Cohen's $d=0.27$; Schoeler et al., 2018), childhood maltreatment (Cohen's $d=0.24$; Baldwin et al., 2023) and parental anxiety ($r=0.13$; Ahmadzadeh et al., 2021) on internalising problems. This would provide a helpful comparison, a bit like in the Orben et al (2019) paper on screen time (although these effect sizes were unadjusted).

Thanks for the great suggestion on effect size comparisons and causal interpretations.

To strengthen causal interpretation of the results from survival analysis of clinically ascertained outcomes, we have extended the follow-up period and performed co-twin control analyses on these clinical outcomes (results were added to **Table 1** and **Supplementary Table 22**). Similar to the results in cross-sectional and cross-age associations, the estimates from co-twin control comparison on clinically ascertained outcomes were attenuated compared to the estimates from the full sample but the risk remained elevated. Take interactive screen use at age 15 for example, the hazard ratios for clinically ascertained depressive or anxiety disorders were only partially attenuated, with 40-62% reduction in log(HR) in the co-twin comparisons (e.g., for weekday heavy use: aHR 3.37, 95% CI 1.98-5.73 in all twins compared to aHR 1.65, 1.07–2.56 in within-twin pairs); whereas the results for passive screen use were mostly null.

Regarding the comparison of the effect sizes with other known risk factors, we have z-standardized our internalizing problem measures before fitting all models (mean 0, SD 1) and the current fixed-effect regression coefficients β are standardized effect sizes, similar as the within-pair Cohen's d . For direct comparison with Orben et al (2019)⁴, we need to assume linear dose-response relationship: the standardized β on internalizing symptoms were indeed small (ranging from 0.03 to 0.05) per additional hour of interactive screen use at age 18 (Supplementary Table 20). These estimates were broadly in line with those reported by Orben et al⁴, where effect sizes ranging (-0.03 to -0.07) on psychological well-being by electronic device/social media use. Accordingly, we have substantially revised the text in the paragraph, which reads as:

"...Consistent with previous literature⁵, we found relatively small effect sizes associated with per-hour increase, while the impact on internalizing symptoms was most pronounced among those in the heavy use group and among adolescent girls. For example, the co-twin controlled estimates corresponded to approximately 0.03–0.05 SD higher internalizing symptoms per additional hour of interactive screen use at age 18, which is similar to the magnitude reported in Orben et al. across three large cohorts from the US and the UK. These per-hour effects are much smaller than those reported for established risk factors for internalizing problems (e.g., bullying victimization⁶ and childhood maltreatment⁷ with 0.2–0.3 SD range). However, symptom elevations do not necessarily translate into clinical conditions ..." (page 26 lines 652-659)

R1.10. Second, it is noted "Assuming a similar degree of effect attenuation due to shared familial factors (as observed in symptom-based co-twin analyses), these results suggest that high screen engagement during adolescence have clinically meaningful implication". This is a fair point, but it could be made more convincing and evidence-based if you (i) calculated the degree of effect attenuation in symptom co-twin control analyses (i.e., on average across outcomes), and (ii) applied it to the clinically ascertained outcomes, to estimate the likely effect on clinically ascertained outcomes if those estimates accounted for familial confounding.

Thank you. In connection to your prior points, we have added the relative percentage of effect size reduction in the co-twin controlled survival analyses on clinically ascertained outcomes directly (**Table 1**).

R1.11. As a final point, it is stated in the limitations section "Fourth, residual confounding from non-shared factors may have influenced the findings. For example, physical activity, peer relationship, and social environments during adolescence could displace screen use and may vary within twin pairs." However, is it not possible to control for these factors in CATSS data? To my knowledge, there are measures of physical activity, peer bullying, and social environments (e.g., family, school, peers) available.

The reviewer was right that crude measures of physical activity (at 9, 15, and 18 yr), peer bullying (at 15 yr), and aspects of the school environment (at 9 yr) are available in our cohort (distributions added in **Supplementary Tables 1-15**). We, however, chose not to include these measures in our primary analyses, due to conceptual reasons and incomplete data (detailed below). The new sensitivity analyses controlling for these measures at the prior wave (as more plausible confounders and with better data coverage than the current wave) led to only modest reductions in the estimates and did not materially change our conclusions.

There are two main reasons why we did not include these variables as time-specific covariates in our primary models.

1. Conceptual/potential causal role of physical activity, peer bullying and social environment

Physical activity, peer relationships and school/family environment are plausibly related to both screen use and internalizing problems. However, their precise roles are not straightforward. Prior work on sedentary behavior and physical activity has highlighted that physical activity can potentially act as a confounder, an intermediary (mediator), or an effect modifier, and that adjusting for physical activity without a clearly specified causal role can lead to biased estimates of the exposure of interest (screen use) rather than reducing confounding⁸ [see **Table R1** below with directed acyclic graph illustrations]. In particular, if physical activity and screen use are just separate activities within the same 24-hour activity scale, and physical activity is neither a confounder nor an effect modifier, then we should not adjust for physical activity when estimating the effect of screen use. Similar considerations apply to peer bullying and the school environment, which may themselves be partially shaped by earlier internalizing problems or screen use.

However, we consider it might be more appropriate to adjust for these variables at a prior time point, when they are more plausibly confounders of subsequent screen use and later internalizing problems. For example, in our cross-sectional analyses, we consider physical activity and bullying victimization at age 15 as a potential confounder of the association between screen use and internalizing problems at age 18. We have clarified this rationale in the Methods and Limitations sections.

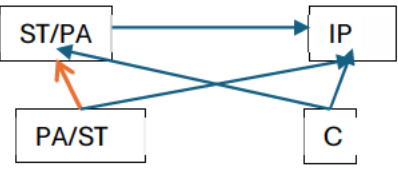
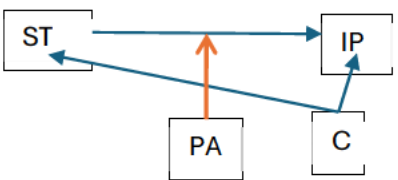
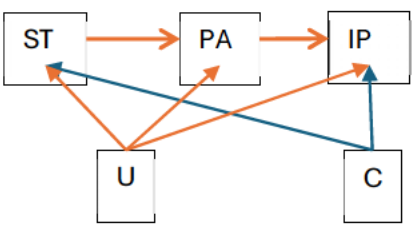
2. Missing data on physical activity and social environment and potential risk of selection bias

For the variables recommended by the reviewer, there are substantial percentages with missing data once we require information at multiple time points (compared to parental sociodemographic factors). When physical activity, peer bullying and school environment are all included as time-varying covariates across ages, >50% of participants have missing data on at least one of these measures (see **Supplementary Tables 1-15**). Adjusting for them in the cross-age or screen use trajectory analyses would either (a) substantially reduce the sample size and increase the risk of selection/collider bias, or (b) require extensive multiple imputation under strong assumptions about the missingness mechanism and the causal structure. Given these concerns, we judged it more transparent to keep the primary models as they are and to acknowledge residual confounding from such non-shared factors as a limitation.

In response to the reviewer's and editors' helpful suggestion, we have conducted sensitivity analyses in which we additionally adjusted for physical activity and peer bullying at the immediately preceding wave, where these data were available to the majority of the participants and conceptually plausible as confounders (i.e. cross-sectional associations at 18 and cross-age associations at 18 and 24 while controlling for physical activity and peer bullying at 15). These adjustments had only a modest impact on the estimates and did not materially change our conclusions. We describe these analyses in the Methods and present the results in **Supplementary Tables 26-27**.

Overall, we have revised the methods and limitations, to a) explicitly acknowledge that these measures are available in CATSS with descriptive statistics, b) clarify why we did not include them as covariates in main analyses (see page 31 lines 880-886), and (c) added the new sensitivity analyses (see page 22, lines 536-543, **Supplementary Tables 26-27**) when applicable to consider physical activity and peer bullying at prior time points.

Table R1. Causal assumptions underlying physical activity (PA), screen time (ST) and internalizing problems (IP) with selected literature.

How PA were considered	Simplified DAG	Results from literature	Why can it be problematic?
Confounder (C) or two exposure variables: mutually adjusted in the model		Both ST and PA were significantly associated with internalizing problems. ⁹⁻¹¹	Alternative causal pathways are very likely too, as illustrated below.
Effect modifier: stratified, or interaction model		Effect of phone time at all PA levels but not consistent with other ST ⁹ . Some combinations of lower ST and higher PA levels seem to have positive effect on mental wellbeing in adolescents. ¹¹	Biologically plausible, but it assumes PA does not cause ST, or vice versa, which is questionable.
Mediator: adjusted when looking at direct effect of ST.		PA partially mediated the associations ($\beta=0.048/0.12$) ^{12,13}	Plausible with displacement theory, but as illustrated in the DAG, conditioning on PA may open alternative paths from ST $\leftarrow$ U (unmeasured confounders) $\rightarrow$ PA $\rightarrow$ IP and introduce further bias.

Reviewer #2 (Remarks to the Author):

Thank you for inviting me to review this manuscript. I have expertise in the methods utilised and the main outcome of interest (adolescent mental health).

The manuscript aims to investigate the relationship between recreational screen use and anxiety and depression. The topic is of importance as anxiety and depression are increasing in young people and screen use, in particular social media, has been associated with these outcomes. As the authors note, much, but not all, previous research has been correlational in nature and there a cause – effect relationship has yet to be established. In the current analysis, the authors utilise CATSS a twin registry based in Sweden. They also use linked medical records to explore this relationship. They use the discordant twin approach and longitudinal design to try and get closer to understanding the causal relationship.

The three main RQs are well defined and they use suitable statistical methods to answer the questions. There are a lot of analyses but overall, I think the authors do well to presenting these clearly and summarising the key points. My assessment of this manuscript is very positive.

I have some comments below that the authors can consider if invited to re-submit.

We thank the reviewer for the considerable time and effort devoted to reviewing our manuscript and for the insightful and constructive feedback provided.

Abstract:

R2.1 “Associations between longer screen time and elevated internalizing symptoms during adolescence persisted in co-twin comparisons, supporting a potential causal link” Is this too strong? This is related to a general point I make in the result section. Not all associations were significant at the level of MZ discordant pairs. I

wonder if the key finding is that there are likely to be familial influences on this relationship (given the attenuating associations) however there is also some evidence for causal relationship. I know that sometimes a simpler message is good but want to make sure that you are happy that this is exactly what your results are implying.

Thank you for the thoughtful comment. Carefully considering all reviewers' comments, we have now revised the sentence to be more specific:

"Cross-sectional associations between longer screen time and elevated internalizing symptoms (β ranging 0.09-0.38) were attenuated in co-twin comparisons but largely persisted in late adolescence, suggesting that shared familial factors may not fully explain these associations". (see page 2 lines 37-42)

Methods:

R2.2 Can you give a bit more detail about the discordant twin analysis and give a reference for the approach taken? There are a 2 main ways that this can be done (to my knowledge) and be good to be explicit.

Thank you for the suggestion for further clarifications. We now clarified in the manuscript that "We estimated within-pair associations in discordant MZ and/or DZ twins using fixed-effect linear regression and stratified Cox regression models, conditional on the family strata. Models were further adjusted for sex (if opposite-sex twin pair), and childhood adverse experiences or prior depression/anxiety symptoms. As described by Sjölander & Zetterqvist³, this exposure-outcome contrast measurement can be interpreted as a potential causal effect within families, provided the effect is (approximately) constant across families." (see page 32 lines 906-912)

Results:

R2.3 Did you or have you considered correcting for multiple testing?

We thank the reviewer for raising the important issue of multiple testing. We are aware that the volume of our analyses can increase the risk of false positive findings. However, the analyses in this manuscript were conducted and interpreted primarily within an estimation framework, emphasizing effect sizes and confidence intervals rather than dichotomous hypothesis testing. Additionally, our primary goal is to triangulate evidence across multiple analytic approaches to strengthen causal inference (instead of interpreting the significance of single test).

Our rationale is that applying multiple testing correction in this context is not straightforward, as it requires defining a clear family of hypothesis tests. Our analyses involve multiple correlated outcomes and exposures, as well as different levels of adjustments and sensitivity analyses, making the number of tests and the appropriate level of correction non-trivial and potentially arbitrary.

In response to the reviewer's suggestion, we now report all p-values, applied the Benjamini–Yekutieli procedure¹ for multiple test correction, which controls the FDR under arbitrary dependence (with global significance threshold $p < 0.00138$, described in the Methods), presented alongside effect estimates and confidence intervals. We have also clarified in the Results and Discussion which findings remain robust after correction.

We respectfully disagree that interpretation should be limited solely to estimates passing multiple testing correction, as statistical significance thresholds do not necessarily reflect scientific and/or clinical relevance. We believe this revised presentation appropriately addresses concerns about multiple testing while maintaining an emphasis on effect size magnitude and precision.

R2.4 P10 "most" I think maybe need to acknowledge that the twin differences were a bit more mixed.

We have revised the interpretation. Please see our response to R1.3.

R2.5 I understand that this could be personal preference, but I do not find the Sankey diagrams particularly helpful for interpreting the pattern of the trajectories. This might be the best way to represent this data but perhaps it could be a bit clearer?

Thank you. We have now revised the Sankey diagram based on the current pattern of trajectories and inserted the percentage of each sub-category (i.e. constant $\leq 2h$, increased from $\leq 2h$, decreased to $\leq 2h$, constant $> 2h$, increased from $> 2h$, decreased to $> 2h$). Please see **Figure 4** and **Supplementary Figure 12A-B**.

Discussion:

R2.6 P18"... Consistent with these concerns, we observed attenuated associations in discordant DZ and MZ twin comparisons, supporting a role of shared genetic and environmental liability. Nonetheless, the observed associations in our cohort largely persisted within twin pairs cross..."

I think this does represent what you found but wonder if this message is aligned with how you have framed it in abstract and some sections of the results (see earlier comments).

We have revised the Abstract to align with the interpretation. Please see our response to R1.1.

R2.7 I liked the inclusion of the cox analysis and linkage with medical records. I do think though you can emphasise the strengths of symptom based approaches. Something doesn't have to be clinically relevant to have impact on behaviour/wellbeing. Also acknowledging that not everyone with clinically relevant symptoms will approach the health service for treatment.

I thought the discussion was well considered.

We are grateful for this positive and thoughtful suggestion. We have now revised the text in the Discussion to emphasize that symptom measures capture a broader range of internalizing problems than those who presented at the healthcare services, which reads as

"Complementing prior evidence from symptom-based measures, which allowed us to capture a broader spectrum of internalizing problems than those presented at healthcare services, these findings reinforce the clinical relevance of reducing excessive screen use during this key developmental period." (page 26 lines 679-682)

Reviewer #3 (Remarks to the Author):

The present study is a comprehensive, longitudinal and well designed study to investigate associations between screen time and adolescent to adult mental health using a range of methods including cross-sectional, longitudinal, twin comparisons, and time to event analyses. The sample size is large, and the sample is well characterized. This study present a wide range of results using a novel study design and substantially adds to the literature. The design of the study is rigorous. I have several comments for the reviewers to consider.

We thank the reviewer for the considerable time and effort devoted to reviewing our manuscript and for the insightful and constructive feedback.

R3.1 While comprehensive, there are also hundreds of models and results for a reader to wade through. In addition to six figures and table, each of which often includes many screen time measures and several outcomes, there are 28 supplemental tables and 13 supplemental figures. I suggest that the authors pare down this paper to the most novel results - the twin comparisons. For example, cross-sectional associations are not very novel and the results are largely similar to any large study. In my opinion, the paper would have a bigger impact if the authors select the most informative and novel results rather than presenting every possible comparison in the data.

Thank you for the comment. We agree that the volume of results can be overwhelming for readers. We have therefore streamlined the presentation:

1. Restructured RQs to highlight the most novel results on co-twin comparison

After adding the co-twin controlled results on the clinically ascertained outcomes, we have moved the trajectory-based investigation to the final RQ and highlighted the novel findings where twin comparisons were performed. However, we prefer to keep the results on cross-sectional associations because they provide an important foundation for interpreting any co-twin comparisons, as noted in the methodological discussion by Gonggrijp et al². However, we do agree that those findings are largely consistent with the literature; therefore, we trimmed it down as much as we can.

2. Moved one main figure from main text to supplementary information

In particular, the figure presenting within-twin-pair correlations of screen use and internalizing symptoms has been moved from the main figures to the Supplement (former **Figure 2**, now **Supplementary Figure 3**). This ensures that the main figures and tables focus on the key results from each RQ.

We would welcome additional suggestions from the reviewer and the editor.

R3.2 This is especially the case because time spent on digital media is not a very sensitive measure of online exposure. Indeed in my opinion the biggest limitation of the study, one that cannot be overcome with additional analysis, is that the consensus of the field is now that 'time spent' is very limited in terms of the information it provides about risks associated with digital media use. Without information on what the adolescents are doing online, whether they are distressed about their use, or any other more relevant factor, the present study does not add substantially to the literature except that it includes twins which provides more rigorous control for shared genetic and environmental factors, which is a major advantage of the design. Authors divide their measures into 'interactive' and 'passive', but it's not always the case that gaming for example is interactive, or that video watching is passive (e.g. if users are interacting with others on YouTube while watching). These limitations should be taken into account.

Thank you for emphasizing this important point of our exposure measures. We agree that recreational screen time per se is a crude measurement. Digital media content, context, as well as personal motivations, experience and engagement with digital media are likely more informative proxies that are associated with mental health risk than the time duration alone. We also agree that our "interactive" versus "passive" screen activity categories are only approximations, and that many activities (e.g. gaming with limited social interaction or watching YouTube videos while chatting on some platform) blur this distinction. In the revised manuscript, we have now strengthened the limitations to make these issues more explicit, which reads

"...Second, our crude measures of screen time do not capture the complexity of digital media use^{14,15}. Durations of screen exposure across broad activity types do not reflect what adolescents and young adults are engaging online today and in the near future¹⁶, or to what extent their online experiences can be regarded as problematic. The motivations and subjective experiences behind screen use (e.g., gaming for social connection vs. scrolling for emotional escape) are likely central to understanding its link with mental health¹⁷. Our findings should therefore be interpreted with caution as a broad association with time spent on broad domains of screen use, not a direct risk from certain activities. Third, our categorization of "interactive" and "passive" screen activities is only an approximation. Many screen-based behaviors can fall into both categories (e.g., gaming with minimal social interaction, or watching YouTube while simultaneously chatting), and some misclassification is likely" (page 27 lines 716-729)

R3.3 Some additional minor comments are about framing and interpretation. The authors make the case in the introduction that twin comparisons are necessary for this topic to control for shared genetic and environmental factors, but do not make the case that there is substantial evidence for genetic and shared environmental confounding or what these sources of confounding may be. Greater attention to the underlying conceptual model that motivates the twin comparisons would benefit the argument. Aligned with this comment, the authors interpret higher screen time correlations among MZ compared with DZ twins as evidence of genetic contributions. It's not clear how genetic factors would influence the amount of screen time someone engages in, it may be a more plausible scenario that there are violations of the equal environments assumption when it comes to activities that

MZ and DZ twins engage in on a regular basis. The authors could address this point and consider whether there is sufficient literature on the genetic of screen time to make this case.

Thank you for the valuable suggestions on adding the underlying conceptual model to motivate twin comparisons and further clarification on how genetics may influence the individual screen use behavior.

1. We have referenced the literature suggesting genetic confounding and added a short description on how different patterns of attenuation within DZ and MZ twin pairs (no attenuation, full attenuation, or partial attenuation) correspond to scenarios with no familial confounding, only genetic confounding, only shared environmental confounding, or a combination of both explained as part of the confounding. The text reads "In the absence of genetic and shared environmental confounding, associations would be consistent at the population level and within both discordant DZ and MZ twin pairs. If confounding is purely genetic, associations would be null within MZ pairs and partially attenuated within DZ pairs. If confounding is solely due to shared environment, associations would be present only at the population level and fully attenuated within both DZ and MZ pairs. When both partial genetic and shared environmental confounding and a potential causal effect independent of these confounders are present, associations would show progressive but incomplete attenuation from the population level to DZ and MZ pairs²..."(page 3 lines 82-90)
2. We have added the alternative explanation for higher MZ than DZ correlations in the Results (page 11, lines 294-296). The literature on the genetics of screen time and potential sources of genetic and shared familial confounding is addressed in the Discussion, where we contextualize our finding (page 24 lines 574-591).

All in all, this is a very well done study and adds to the literature. I think the paper would be stronger with greater attention to paring down the core set of novel results, and giving additional attention to the plausibility of various arguments.

Thank you again for taking the time to review our paper.

References

1. Benjamini, Y. & Yekutieli, D. The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics* **29**, 1165-1188 (2001).
2. Gonggrijp, B.M.A., van de Weijer, S.G.A., Bijleveld, C.C.J.H., van Dongen, J. & Boomsma, D.I. The Co-Twin Control Design: Implementation and Methodological Considerations. *Twin Research and Human Genetics* **26**, 249-256 (2023).
3. Sjölander, A. & Zetterqvist, J. Confounders, Mediators, or Colliders: What Types of Shared Covariates Does a Sibling Comparison Design Control For? *Epidemiology* **28**, 540-547 (2017).
4. Orben, A. & Przybylski, A.K. The association between adolescent well-being and digital technology use. *Nature Human Behaviour* **3**, 173-182 (2019).
5. Sanders, T. *et al.* An umbrella review of the benefits and risks associated with youths' interactions with electronic screens. *Nature Human Behaviour* **8**, 82-99 (2024).
6. Schoeler, T., Duncan, L., Cecil, C.M., Ploubidis, G.B. & Pingault, J.B. Quasi-experimental evidence on short- and long-term consequences of bullying victimization: A meta-analysis. *Psychol Bull* **144**, 1229-1246 (2018).
7. Baldwin, J.R. *et al.* A genetically informed Registered Report on adverse childhood experiences and mental health. *Nature Human Behaviour* **7**, 269-290 (2023).
8. Page, A., Peeters, G. & Merom, D. Adjustment for physical activity in studies of sedentary behaviour. *Emerg Themes Epidemiol* **12**, 10 (2015).
9. Forte, C. *et al.* Associations between screen-time, physical activity and depressive symptoms differ based on gender and screen-time mode. *Eur Child Adolesc Psychiatry* **32**, 2313-2322 (2023).
10. Kjellenberg, K., Ekblom, O., Ahlen, J., Helgadóttir, B. & Nyberg, G. Cross-sectional associations between physical activity pattern, sports participation, screen time and mental health in Swedish adolescents. *BMJ Open* **12**, e061929 (2022).

11. Khan, A., Lee, E.Y., Rosenbaum, S., Khan, S.R. & Tremblay, M.S. Dose-dependent and joint associations between screen time, physical activity, and mental wellbeing in adolescents: an international observational study. *Lancet Child Adolesc Health* **5**, 729-738 (2021).
12. Rutter, L.A. *et al.* Social Media Use, Physical Activity, and Internalizing Symptoms in Adolescence: Cross-sectional Analysis. *JMIR Ment Health* **8**, e26134 (2021).
13. Chen, Q., Dai, W., Li, G. & Ma, N. The impact of screen time changes on anxiety during the COVID-19 pandemic: Sleep and physical activity as mediators. *Sleep and Biological Rhythms* **20**, 521-531 (2022).
14. Martín-Rojas, R., García-Morales, V.J., Garrido-Moreno, A. & Salmador-Sánchez, M.P. Social Media Use and the Challenge of Complexity: Evidence from the Technology Sector. *Journal of Business Research* **129**, 621-640 (2021).
15. Brinberg, M. *et al.* The Idiosyncrasies of Everyday Digital Lives: Using the Human Screenome Project to Study User Behavior on Smartphones. *Comput Human Behav* **114**(2021).
16. Mansfield, K.L. *et al.* From social media to artificial intelligence: improving research on digital harms in youth. *The Lancet Child & Adolescent Health*.
17. Tang, S., Werner-Seidler, A., Torok, M., Mackinnon, A.J. & Christensen, H. The relationship between screen time and mental health in young people: A systematic review of longitudinal studies. *Clinical Psychology Review* **86**, 102021 (2021).

We thank the editor and reviewers for evaluating our manuscript and providing valuable suggestions to improve the quality of our work. In this revision, we have responded to the comments point-by-point and highlighted the corresponding changes in the manuscript (STR_screentime_depanx_Main_20260305_track.docx). We have included a clean manuscript (STR_screentime_depanx_Main_20260305_clean.docx).

We have made substantial changes to the manuscript following the helpful comments and suggestions. In particular, we have made the following changes in this revision:

1. **Refocusing the presentations to emphasize the co-twin control design:** we revised the structure of the manuscript to more clearly highlight the co-twin control analyses, which represent the most novel aspect of the study (see R3.1). Specifically, we 1) added co-twin control models to the time-to-event analyses of clinically ascertained outcomes (see R1.9), and 2) removed the analyses based on screen-use trajectories, as these did not allow for co-twin comparisons. These changes align the research questions more closely with the co-twin design and allow the manuscript to more clearly communicate the quasi-causal insights provided by co-twin comparisons.
2. **Clarifying interpretation through cohort vs. co-twin comparisons:** the revised structure allows us to directly contrast effect estimates obtained at the cohort level with those from co-twin comparisons (see R1.4, R1.10). This contrast is informative because the degree to which associations attenuate in the co-twin models provides insight into the potential contribution of shared familial factors (genetic and environmental), while any remaining association may indicate a residual effect consistent with a potential causal relationship.
3. **Strengthening statistical rigor through multiple testing correction:** We now clearly present which results remain significant after correcting for multiple testing (see R2.3) and have revised the Abstract, Results and Discussion to ensure that our interpretation and conclusions are based primarily on findings that survive multiple testing correction. This change improves the transparency and robustness of the reported findings.

Thanks to the helpful suggestions from the reviewers and editors, we believe the novelty, rigor, and overall clarity of the manuscript have been significantly strengthened.

Editor Comments:

In particular, please address the following (as well as all other reviewer comments):

(1) Causal claims should be reserved for only the specific results of the co-twin control analyses that hold after correction for multiple testing, and such interpretations must acknowledge the limitations of the co-twin control method, which cannot rule out alternative explanations for observed associations.

In line with the editor’s recommendation, we have substantially reduced causal language throughout the manuscript. Given the notable attenuation of effect estimates in the co-twin control analyses, we have largely removed causal claims and instead interpret the findings more cautiously as evidence mostly consistent with potential familial confounding. Where the co-twin design is discussed, we explicitly acknowledge the methodological limitations and emphasize the need for independent replication in other samples and study designs.

(2) Ensure that all interpretations and conclusions accurately reflect the pattern of results. Non-statistically significant results in the twin comparisons should be described as such, and interpreted as showing no evidence of an effect (rather than consistent with cross-sectional results).

We have revised text throughout the manuscript to cautiously interpret the non-statistically significant results of the co-twin analyses after correction for multiple comparisons.

(3) Reviewer 2 mentions the need to correct for multiple comparisons, and we ask that you include such corrections. Please note that in cases where both corrected and uncorrected results are presented, all interpretations of the results must be based on corrected (rather than uncorrected) tests.

As per Reviewer 2's suggestion, we applied multiple testing corrections across all analyses (see R2.3).

(4) Carry out the additional analyses recommended by Reviewer 1, including the use of additional control variables and the direct calculation of effect attenuation.

In response to these valuable suggestions, we conducted new sensitivity analyses in which we controlled for measures of physical activity and peer bullying (R1.11) and presented the effect attenuation comparing effects estimated from all twins versus those from co-twin control analyses (R1.4, R1.10).

(5) Ensure that your underlying conceptual model and the limitations of your dataset are entirely clear, as highlighted by Reviewer 3.

We have added the underlying conceptual model and clearly stated the limitations now (see R3.2, R3.3).

(6) We appreciate Reviewer 3's suggestion to reduce this paper to presenting only the co-twin comparisons. While we agree that the paper would benefit from being more streamlined, we ask that you do not remove the correlational results entirely. However, we agree that it may increase the readability and reach of the work if you move some of these to the SI, to ensure that your main text is focused on the most informative and novel results.

We have now moved the figure describing twin correlations of screen time and internalizing problems (previously Figure 2) into supplementary information (now Supplementary Figure 3), removed the RQ based on screen use trajectories that we could not perform co-twin comparison due to limited statistical power from the manuscript, leaving with one figure/table per RQ in the main text.

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author): *It was great to read this very interesting and well-conducted study examining the relationship between screen time and internalising problems. The study used a large longitudinal sample of Swedish twins and conducted multiple analyses, including co-twin control methods to strengthen causal inference. The paper was very well written, methodologically sound, and the figure in the introduction providing the study overview was excellent. Below I provide some suggestions for improvement.*

We thank the reviewer for the considerable time and effort devoted to reviewing our manuscript and for the insightful and constructive feedback provided.

Abstract:

R1.1. It would be helpful to state an effect size or range of effect sizes, if possible.

Thank you. The sentences have been revised accordingly and now read as "Cross-sectional associations between longer screen time and elevated internalizing symptoms at ages 15 and 18 were observed (β ranging from 0.25-0.38 when comparing heavy use [>6 hours/day] to restrictive use [≤ 2 hours/day]), with notable sex differences". (page 2 lines 36-39)

Results Section "Cross-sectional associations between recreational screen use and internalizing problems": R1.2. For the sentence beginning "at age 9" it would be helpful to state that this is in the full sample/phenotypic analyses, to clarify to the reader. It would also be helpful to briefly describe the effect sizes in the text (e.g.,

“standardised effect sizes ranging from 0.05-0.14”). Similarly, for the sentence beginning “at ages 15 and 18” it would also be beneficial to state that this is in phenotypic analyses.

Thanks for the suggestion. These sentences have been revised as follows:

“At age 9, we observed associations with depressive and anxiety symptoms across different types of screen use, when comparing the most frequent use (daily) to restrictive use (≤ 2 times/week) in the full sample (standardized β ranging from 0 to 0.12, **Supplementary Table 17**).” (page 5 lines 131-133) and

“At ages 15 and 18, the association analyses between screen time and internalizing problems in all twins demonstrated a dose-response pattern, with longer screen time per day corresponding to higher symptom scores (standardized β ranging from 0.21 to 0.38 when comparing heavy use [>6 hours/day] to restrictive use [≤ 2 hours/day], **Supplementary Table 17**).” (page 5 lines 138-141)

R1.3. For the paragraph on discordant twin comparisons at ages 15 and 18, the text seems misleading, as it says “most associations remained significant within discordant twin pairs, where shared genetic confounding and familial environment factors were implicitly accounted for”. However, in Figure 3D, the results were non-significant for 3 out of 4 screen time exposures at age 15. I therefore suggest rewording this to specify that the associations remained significant at age 18, while at age 15, most associations became non-significant in MZ twins.

We revised the paragraph on page 6 lines 146-156 to reflect cautious interpretation of these results: “In discordant twin comparisons at ages 15 and 18 (**Figure 2D-F**), we observed attenuation of effect estimates within DZ pairs and a more pronounced attenuation within MZ pairs. At age 15, the associations for both passive and interactive screen activities remained significant in discordant DZ pairs (range of standardized β in heavy use compared to restrictive use: 0.22-0.32; $p < 0.0013$ after Benjamini-Yekutieli correction for multiple testing). However, in discordant MZ pairs, the estimates (range of standardized β : 0.02-0.21; **Supplementary Table 17**) were attenuated by 32-95% compared to those from the full sample. At age 18, we only observed significant associations within DZ pairs between interactive screen use and depressive symptoms (range of standardized β : 0.27-0.28; $p < 0.0013$); the corresponding estimates in MZ pairs (range of standardized β : 0.18-0.26; **Supplementary Table 17**) were attenuated by 16-45% compared to the estimates in the full sample.”

R1.4. So that readers can understand the extent of genetic and environmental confounding, it would be helpful to calculate a % reduction in effect size from the phenotypic to the MZ co-twin control analyses.

Thank you for the great suggestion. We have now added a column with percentage of effect size change for cross-sectional and cross-age associations, as well as for time-to-develop clinical condition analyses (see **Supplementary Tables 17, 19, 22, and Table 1**). For example, at age 15, the effect of heavy use [>6 hours/day] on depressive/anxiety symptoms was attenuated by 32-95% from all twins to the MZ co-twin comparison; while the effect attenuation was more modest at age 18 (e.g., heavy use effect by interactive screen activities was attenuated by 9-61%).

Discussion

R1.5. Related to the above point, I suggest rephrasing the sentence “Cross-sectional associations were strongest at ages 15 and 18 and more pronounced among girls, even after adjusting for prior symptoms and shared genetic and familial environmental factors” given that the twin-findings were largely non-significant at age 15.

We have rephrased this sentence, which now reads “Cross-sectional associations were strongest at ages 15 and 18 and more pronounced among girls (RQ1).” (page 9 lines 251-252)

R1.6. Similarly, the sentence “the observed associations in our cohort largely persisted within twin pairs cross sectionally in mid-to-late adolescence” should be rephrased to specify this is mainly in late adolescence, not mid adolescence.

We have revised this sentence accordingly, which reads “...Consistent with previous longitudinal studies^{32,33,37}, our cross-age and longitudinal analyses confirmed that heavy use on interactive screen activities during adolescence (from age 15 in our data) was associated with elevated internalizing problems and clinical outcomes at later ages in all twins”. (page 10 lines 292-295)

R1.7. The sentence “These findings suggest that increased screen use may independently contribute to internalizing problems beyond shared familial factors” seems a bit of an overinterpretation, as many associations were fully attenuated in the co-twin control analyses. E.g., in longitudinal co-twin analyses, only 3 of 8 analyses were statistically significant (Figure 4). I would recommend rephrasing this sentence to reflect a more cautious and nuanced interpretation of the findings.

Thank you for the recommendation. We have rephrased this sentence accordingly, which reads

“In our data, parent-reported screen use at age 9 was not significantly associated with internalizing symptoms at age 15; however, daily gaming at this age, compared with restricted use, was associated with a modestly increased risk of depression and anxiety disorders (aHR=1.38 in both sexes and 1.56 in boys) through age 22”. (page 10 lines 295-298)

R1.8. Sentence “Together, these findings, along with previous literature⁵², emphasize the specific need of focusing on clinically vulnerable subgroups and developing targeted interventions (for example, coping strategies for negative feedback or social comparison) to prevent severe outcomes, including the rising incidence of self-harm among youth^{14,53}”. As self-harm wasn’t examined here, it doesn’t seem justified as an implication of this study.

Thanks for noticing this. We have now removed the self-harm part of the sentence, which now reads “... to prevent worsening internalizing symptoms and related adverse outcomes”.

R1.9. Regarding the paragraph on effect sizes being too small to warrant policy change, I would suggest two changes here. First, it doesn’t seem appropriate to discuss the effect sizes of the analysis on clinically ascertained outcomes, as these results don’t represent causal estimates. Instead, I would recommend comparing the effect sizes from the co-twin control analysis (i.e., the most likely causal estimates) to effect sizes of other developmental risk factors from other quasi-experimental work. For example, meta-analyses of quasi-experimental studies have estimated the causal effects of bullying victimisation (Cohen’s $d=0.27$; Schoeler et al., 2018), childhood maltreatment (Cohen’s $d=0.24$; Baldwin et al., 2023) and parental anxiety ($r=0.13$; Ahmadzadeh et al., 2021) on internalising problems. This would provide a helpful comparison, a bit like in the Orben et al (2019) paper on screen time (although these effect sizes were unadjusted).

Thanks for the great suggestion on effect size comparisons and the potential causal interpretations.

To strengthen potential causal interpretation of the results from survival analysis of clinically ascertained outcomes, we have extended the follow-up period and performed co-twin control analyses on these clinical outcomes (results were added to **Table 1** and **Supplementary Table 22**). Similar to the results in cross-sectional and cross-age associations, the estimates from co-twin control comparison on clinically ascertained outcomes were attenuated compared to the estimates from the full sample but the risk remained elevated. Take interactive screen use at age 15 for example, the hazard ratios for clinically ascertained depressive or anxiety disorders were only partially attenuated, with 40-62% reduction in log(HR) in the co-twin comparisons (e.g., for weekday heavy use: aHR 3.37, 95% CI 1.98-5.73 in all twins compared to aHR 1.65, 1.07–2.56 in within-twin pairs); whereas the results for passive screen use were mostly null.

Regarding the comparison of the effect sizes with other known risk factors, we have z-standardized our internalizing problem measures before fitting all models (mean 0, SD 1) and the current fixed-effect regression coefficients β are standardized effect sizes, similar as the within-pair Cohen's d. For direct comparison with Orben et al (2019)¹, we need to assume linear dose-response relationship: the standardized β on internalizing symptoms were indeed small (ranging from 0.03 to 0.05) per additional hour of interactive screen use at age 18 (Supplementary Table 20). These estimates were broadly in line with those reported by Orben et al¹, where effect sizes ranging (-0.03 to -0.07) on psychological well-being by electronic device/social media use. Accordingly, we have substantially revised the text in the paragraph, which reads as:

"...Consistent with previous literature¹³, we found relatively small effect sizes associated with per-hour increase, while the impact on internalizing symptoms was most pronounced among those in the heavy use group and among adolescent girls. For example, our co-twin-controlled estimates (approximately 0.03–0.05 SD higher internalizing symptoms per additional hour of interactive screen use at age 18) are comparable to those reported by Orben et al. across three large cohorts in the US and the UK⁴³. However, research have shown that the impact of screen use can be nonlinear^{46,47}. In our example within age-18 discordant MZ twins, the weekend heavy use on interactive screen activities was nominally associated with 0.26 SD higher symptoms compared to restrictive use. It is worth noting that this estimate is within the range of those reported for established risk factors for internalizing problems, such as bullying victimization⁴⁸ and childhood maltreatment⁴⁹ (typically 0.2–0.3 SD). Furthermore, symptom elevations do not necessarily translate into clinical conditions¹⁵ ..." (page 11 lines 315-326)

R1.10. Second, it is noted "Assuming a similar degree of effect attenuation due to shared familial factors (as observed in symptom-based co-twin analyses), these results suggest that high screen engagement during adolescence have clinically meaningful implication". This is a fair point, but it could be made more convincing and evidence-based if you (i) calculated the degree of effect attenuation in symptom co-twin control analyses (i.e., on average across outcomes), and (ii) applied it to the clinically ascertained outcomes, to estimate the likely effect on clinically ascertained outcomes if those estimates accounted for familial confounding.

Thank you. In connection to your prior points, we have added the relative percentage of effect size reduction in the co-twin controlled survival analyses on clinically ascertained outcomes directly (**Table 1**).

R1.11. As a final point, it is stated in the limitations section "Fourth, residual confounding from non-shared factors may have influenced the findings. For example, physical activity, peer relationship, and social environments during adolescence could displace screen use and may vary within twin pairs." However, is it not possible to control for these factors in CATSS data? To my knowledge, there are measures of physical activity, peer bullying, and social environments (e.g., family, school, peers) available.

The reviewer was right that crude measures of physical activity (at 9, 15, and 18 yr), peer bullying (at 15 yr), and aspects of the school environment (at 9 yr) are available in our cohort (distributions added in **Supplementary Tables 1-15**). We, however, chose not to include these measures in our primary analyses, due to conceptual reasons and incomplete data (detailed below). The new sensitivity analyses controlling for these measures at the prior wave (as more plausible confounders and with better data coverage than the current wave) led to only modest reductions in the estimates and did not materially change our conclusions.

There are two main reasons why we did not include these variables as time-specific covariates in our primary models.

1. Conceptual/potential causal role of physical activity, peer bullying and social environment
Physical activity, peer relationships and school/family environment are plausibly related to both screen use and internalizing problems. However, their precise roles are not straightforward. Prior work on sedentary behavior and physical activity has highlighted that physical activity can potentially act as a confounder, an intermediary (mediator), or an effect modifier, and that adjusting for physical activity without a clearly specified causal role can lead to biased estimates of the exposure of interest (screen use) rather than reducing

confounding² [see **Table R1** below with directed acyclic graph illustrations]. In particular, if physical activity and screen use are just separate activities within the same 24-hour activity scale, and physical activity is neither a confounder nor an effect modifier, then we should not adjust for physical activity when estimating the effect of screen use. Similar considerations apply to peer bullying and the school environment, which may themselves be partially shaped by earlier internalizing problems or screen use. However, we consider it might be more appropriate to adjust for these variables at a prior time point, when they are more plausibly confounders of subsequent screen use and later internalizing problems. For example, in our cross-sectional analyses, we consider physical activity and bullying victimization at age 15 as a potential confounder of the association between screen use and internalizing problems at age 18. We have clarified this rationale in the Methods and Limitations sections.

2. Missing data on physical activity and social environment and potential risk of selection bias
- For the variables recommended by the reviewer, there are substantial percentages with missing data once we require information at multiple time points (compared to parental sociodemographic factors). When physical activity, peer bullying and school environment are all included as time-varying covariates across ages, >50% of participants have missing data on at least one of these measures (see **Supplementary Tables 1-15**). Adjusting for them in the cross-age or screen use trajectory analyses would either (a) substantially reduce the sample size and increase the risk of selection/collider bias, or (b) require extensive multiple imputation under strong assumptions about the missingness mechanism and the causal structure. Given these concerns, we judged it more transparent to keep the primary models as they are and to acknowledge residual confounding from such non-shared factors as a limitation.

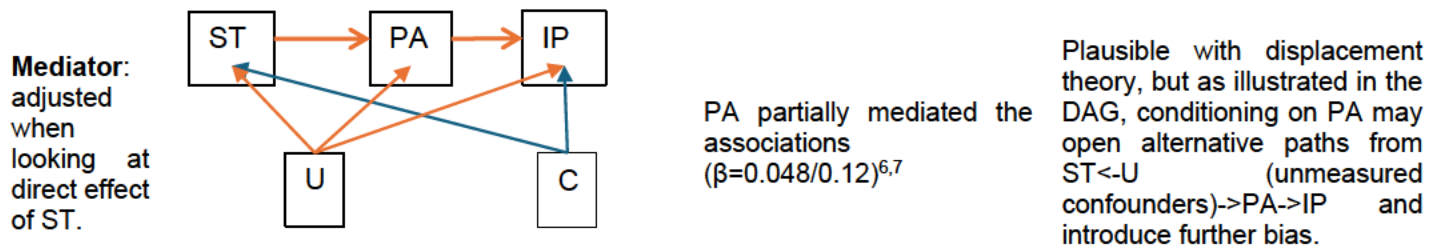
In response to the reviewer’s and editors’ helpful suggestion, we have conducted sensitivity analyses in which we additionally adjusted for physical activity and peer bullying at the immediately preceding wave, where these data were available to the majority of the participants and conceptually plausible as confounders (i.e. cross-sectional associations at 18 and cross-age associations at 18 and 24 while controlling for physical activity and peer bullying at 15). These adjustments had only a modest impact on the estimates and did not materially change our conclusions. We describe these analyses in the Methods and present the results in **Supplementary Tables 23-24**.

Overall, we have revised the methods and limitations, to a) explicitly acknowledge that these measures are available in CATSS with descriptive statistics, b) clarify why we did not include them as covariates in main analyses (see page 15 lines 500-506), and (c) added the new sensitivity analyses (see page 9, lines 240-246, **Supplementary Tables 23-24**) when applicable to consider physical activity and peer bullying at prior time points.

Table R1. Causal assumptions underlying physical activity (PA), screen time (ST) and internalizing problems (IP) with selected literature.

How we were considered	PA	Simplified DAG	Results from literature	Why can it be problematic?
Confounder (C) or two exposure variables: mutually adjusted in the model		<pre> graph LR STPA[ST/PA] --> IP PAST[PA/ST] --> IP C --> STPA C --> PAST C --> IP </pre>	Both ST and PA were significantly associated with internalizing problems. ³⁻⁵	Alternative causal pathways are very likely too, as illustrated below.
Effect modifier: stratified, or interaction model		<pre> graph LR ST --> IP P --> ST P --> C C --> IP </pre>	Effect of phone time at all PA levels but not consistent with other ST ³ . Some combinations of lower ST and higher PA levels seem to have positive	Biologically plausible, but it assumes PA does not cause ST, or vice versa, which is questionable.

effect on mental wellbeing in adolescents.⁵



Reviewer #2 (Remarks to the Author):

Thank you for inviting me to review this manuscript. I have expertise in the methods utilised and the main outcome of interest (adolescent mental health).

The manuscript aims to investigate the relationship between recreational screen use and anxiety and depression. The topic is of importance as anxiety and depression are increasing in young people and screen use, in particular social media, has been associated with these outcomes. As the authors note, much, but not all, previous research has been correlational in nature and there a cause – effect relationship has yet to be established. In the current analysis, the authors utilise CATSS a twin registry based in Sweden. They also use linked medical records to explore this relationship. They use the discordant twin approach and longitudinal design to try and get closer to understanding the causal relationship.

The three main RQs are well defined and they use suitable statistical methods to answer the questions. There are a lot of analyses but overall, I think the authors do well to presenting these clearly and summarising the key points. My assessment of this manuscript is very positive.

I have some comments below that the authors can consider if invited to re-submit.

We thank the reviewer for the considerable time and effort devoted to reviewing our manuscript and for the insightful and constructive feedback provided.

Abstract:

R2.1 “Associations between longer screen time and elevated internalizing symptoms during adolescence persisted in co-twin comparisons, supporting a potential causal link” Is this too strong? This is related to a general point I make in the result section. Not all associations were significant at the level of MZ discordant pairs. I wonder if the key finding is that there are likely to be familial influences on this relationship (given the attenuating associations) however there is also some evidence for causal relationship. I know that sometimes a simpler message is good but want to make sure that you are happy that this is exactly what your results are implying.

Thank you for the thoughtful comment. Carefully considering all reviewers’ comments, we have now revised the sentence to be more specific:

“Cross-sectional associations between longer screen time and elevated internalizing symptoms at ages 15 and 18 were observed (β ranging from 0.25-0.38 when comparing heavy use [>6 hours/day] to restrictive use [≤ 2 hours/day]), with notable sex differences”. (see page 2 lines 36-39)

Methods:

R2.2 Can you give a bit more detail about the discordant twin analysis and give a reference for the approach taken? There are a 2 main ways that this can be done (to my knowledge) and be good to be explicit.

Thank you for the suggestion for further clarifications. We now clarified in the manuscript that “We estimated within-pair associations in discordant MZ and/or DZ twins using fixed-effect linear regression and stratified Cox regression models, conditional on the family strata. Models were further adjusted for sex (if opposite-sex twin pair), and childhood adverse experiences or prior depression/anxiety symptoms. As described by Sjölander & Zetterqvist⁸, this exposure-outcome contrast measurement can be interpreted as a potential causal effect within families, provided the effect is (approximately) constant across families.” (see page 16 lines 526-532)

Results:

R2.3 Did you or have you considered correcting for multiple testing?

We thank the reviewer for raising the important issue of multiple testing. We are aware that the volume of our analyses can increase the risk of false positive findings.

In response to the reviewer's and the editor's suggestion on correction for multiple testing, we now report all p-values, applied the Benjamini–Yekutieli procedure⁹ for multiple test correction, which controls the FDR under arbitrary dependence (with global significance threshold $p < 0.00133$, described in the Methods), presented alongside effect estimates and confidence intervals. We have also clarified in the Results and Discussion which findings remain robust after correction.

R2.4 P10 "most" I think maybe need to acknowledge that the twin differences were a bit more mixed.

We have revised the interpretation. Please see our response to R1.3.

R2.5 I understand that this could be personal preference, but I do not find the Sankey diagrams particularly helpful for interpreting the pattern of the trajectories. This might be the best way to represent this data but perhaps it could be a bit clearer?

Thank you. In the revised manuscript we have removed the research question and analyses on screen-use trajectories, and consequently the Sankey diagrams and associated text are no longer included. This comment is therefore no longer applicable to the revised paper.

Discussion:

R2.6 P18"... Consistent with these concerns, we observed attenuated associations in discordant DZ and MZ twin comparisons, supporting a role of shared genetic and environmental liability. Nonetheless, the observed associations in our cohort largely persisted within twin pairs cross..."

I think this does represent what you found but wonder if this message is aligned with how you have framed it in abstract and some sections of the results (see earlier comments).

We have revised the Abstract to align with the interpretation. Please see our response to R1.1.

R2.7 I liked the inclusion of the cox analysis and linkage with medical records. I do think though you can emphasise the strengths of symptom based approaches. Something doesn't have to be clinically relevant to have impact on behaviour/wellbeing. Also acknowledging that not everyone with clinically relevant symptoms will approach the health service for treatment.

I thought the discussion was well considered.

We are grateful for this positive and thoughtful suggestion. We have now revised the text in the Discussion to emphasize that symptom measures capture a broader range of internalizing problems than those who presented at the healthcare services, which reads as

“This study has several notable strengths, including a large population-based twin cohort with longitudinal assessments and symptom-based measures that capture impacts on wellbeing across general population, particularly individuals who may not seek clinical care. ...” (page 11 lines 338-340)

Reviewer #3 (Remarks to the Author):

The present study is a comprehensive, longitudinal and well designed study to investigate associations between screen time and adolescent to adult mental health using a range of methods including cross-sectional, longitudinal, twin comparisons, and time to event analyses. The sample size is large, and the sample is well characterized. This study present a wide range of results using a novel study design and substantially adds to the literature. The design of the study is rigorous. I have several comments for the reviewers to consider.

We thank the reviewer for the considerable time and effort devoted to reviewing our manuscript and for the insightful and constructive feedback.

R3.1 While comprehensive, there are also hundreds of models and results for a reader to wade through. In addition to six figures and table, each of which often includes many screen time measures and several outcomes, there are 28 supplemental tables and 13 supplemental figures. I suggest that the authors pare down this paper to the most novel results - the twin comparisons. For example, cross-sectional associations are not very novel and the results are largely similar to any large study. In my opinion, the paper would have a bigger impact if the authors select the most informative and novel results rather than presenting every possible comparison in the data.

Thank you for the comment. We agree that the volume of results can be overwhelming for readers. We have therefore streamlined the presentation:

1. Restructured RQs to highlight the most novel results on co-twin comparison
After adding the co-twin controlled results on the clinically ascertained outcomes, we have removed the trajectory-based investigation from the RQs and highlighted the novel findings where co-twin comparisons were performed. However, we prefer to keep the results on cross-sectional associations because they provide an important foundation for interpreting any co-twin comparisons, as noted in the methodological discussion by Gonggrijp et al¹⁰. However, we do agree that those findings are largely consistent with the literature; therefore, we trimmed it down as much as we can.
2. Moved one main figure from main text to supplementary information
In particular, the figure presenting within-twin-pair correlations of screen use and internalizing symptoms has been moved from the main figures to the Supplement (former **Figure 2**, now **Supplementary Figure 3**). This ensures that the main figures and tables focus on the key results from each RQ.

We would welcome additional suggestions from the reviewer and the editor.

R3.2 This is especially the case because time spent on digital media is not a very sensitive measure of online exposure. Indeed in my opinion the biggest limitation of the study, one that cannot be overcome with additional analysis, is that the consensus of the field is now that ‘time spent’ is very limited in terms of the information it provides about risks associated with digital media use. Without information on what the adolescents are doing online, whether they are distressed about their use, or any other more relevant factor, the present study does not add substantially to the literature except that it includes twins which provides more rigorous control for shared genetic and environmental factors, which is a major advantage of the design. Authors divide their measures into ‘interactive’ and ‘passive’, but it’s not always the case that gaming for example is interactive, or that video

watching is passive (e.g. if users are interacting with others on YouTube while watching). These limitations should be taken into account.

Thank you for emphasizing this important point of our exposure measures. We agree that recreational screen time per se is a crude measurement. Digital media content, context, as well as personal motivations, experience and engagement with digital media are likely more informative proxies that are associated with mental health risk than the time duration alone. We also agree that our “interactive” versus “passive” screen activity categories are only approximations, and that many activities (e.g. gaming with limited social interaction or watching YouTube videos while chatting on some platform) blur this distinction. In the revised manuscript, we have now strengthened the limitations to make these issues more explicit, which reads

“...Second, our crude measures of screen time do not capture the complexity of digital media use^{50,51}. Durations of screen exposure across several broad activity types do not reflect what adolescents and young adults are engaging in online today and in the near future¹⁰, or to what extent their online experiences can be regarded as problematic. The motivations and subjective experiences behind screen use (e.g., gaming for social connection vs. scrolling for emotional escape) are likely central to understanding its link with mental health⁵². Our findings should therefore be interpreted with caution as a broad association with time spent on broad domains of screen use, not a direct risk from certain activities. Third, our categorization of “interactive” and “passive” screen activities is only an approximation. Many screen-based behaviors can fall into both categories (e.g. gaming with minimal social interaction, or watching YouTube while simultaneously chatting), and some misclassification is likely” (pages 11-12 lines 345-356)

R3.3 Some additional minor comments are about framing and interpretation. The authors make the case in the introduction that twin comparisons are necessary for this topic to control for shared genetic and environmental factors, but do not make the case that there is substantial evidence for genetic and shared environmental confounding or what these sources of confounding may be. Greater attention to the underlying conceptual model that motivates the twin comparisons would benefit the argument. Aligned with this comment, the authors interpret higher screen time correlations among MZ compared with DZ twins as evidence of genetic contributions. It's not clear how genetic factors would influence the amount of screen time someone engages in, it may be a more plausible scenario that there are violations of the equal environments assumption when it comes to activities that MZ and DZ twins engage in on a regular basis. The authors could address this point and consider whether there is sufficient literature on the genetic of screen time to make this case.

Thank you for the valuable suggestions on adding the underlying conceptual model to motivate twin comparisons and further clarification on how genetics may influence the individual screen use behavior.

1. We have referenced the literature suggesting genetic confounding and added a short description on how different patterns of attenuation within DZ and MZ twin pairs (no attenuation, full attenuation, or partial attenuation) correspond to scenarios with no familial confounding, only genetic confounding, only shared environmental confounding, or a combination of both explained as part of the confounding. The text reads “In contrast, the discordant twin analyses account for genetic and environmental factors shared within twin pairs; thus, the attenuation of associations relative to the full sample suggests a role of these shared factors in the observed relationships²³. If confounding is purely genetic, associations would be null within MZ pairs and partially attenuated within DZ pairs. If confounding is solely due to shared environment, associations would be present only at the population level and fully attenuated within both DZ and MZ pairs. When both partial genetic and shared environmental confounding and a potential causal effect independent of these confounders are present, associations are expected to show progressive but incomplete attenuation from the population level to DZ and MZ pairs²³”.(page 10 lines 275-283)
2. We have added the alternative explanation for higher MZ than DZ correlations in the Results, which reads “ ...Across all ages and measures, correlations were stronger within MZ than within DZ twins, suggesting genetic influences and/or greater similarity in the rearing environment for shared screen activities within MZ twins...” (page 5, lines 121-123)

The literature on the genetics of screen time and potential sources of genetic and shared familial confounding is addressed in the Discussion, where we contextualize our findings. The text reads "... In our cross-sectional analysis, the modest attenuation observed in DZ pairs and the further attenuation in MZ pairs are consistent with partial genetic and shared environmental liability underlying the youth screen use-mental health link. These findings align with previous genetically informed studies^{16,17}. Notably, although associations were often substantially attenuated and only nominally significant, most effect estimates from the discordant MZ and combined DZ-MZ analyses remained above the null. Since these analyses accounted for shared genetic and familial environmental factors and further adjusted for (some) non-shared factors, this pattern may suggest a possibility of a residual causal effect; however, replication in independent twin samples is warranted". (page 10 lines 283-291).

All in all, this is a very well done study and adds to the literature. I think the paper would be stronger with greater attention to paring down the core set of novel results, and giving additional attention to the plausibility of various arguments.

Thank you again for taking the time to review our paper.

References

1. Orben, A. & Przybylski, A.K. The association between adolescent well-being and digital technology use. *Nature Human Behaviour* **3**, 173-182 (2019).
2. Page, A., Peeters, G. & Merom, D. Adjustment for physical activity in studies of sedentary behaviour. *Emerg Themes Epidemiol* **12**, 10 (2015).
3. Forte, C. *et al.* Associations between screen-time, physical activity and depressive symptoms differ based on gender and screen-time mode. *Eur Child Adolesc Psychiatry* **32**, 2313-2322 (2023).
4. Kjellenberg, K., Ekblom, O., Ahlen, J., Helgadóttir, B. & Nyberg, G. Cross-sectional associations between physical activity pattern, sports participation, screen time and mental health in Swedish adolescents. *BMJ Open* **12**, e061929 (2022).
5. Khan, A., Lee, E.Y., Rosenbaum, S., Khan, S.R. & Tremblay, M.S. Dose-dependent and joint associations between screen time, physical activity, and mental wellbeing in adolescents: an international observational study. *Lancet Child Adolesc Health* **5**, 729-738 (2021).
6. Rutter, L.A. *et al.* Social Media Use, Physical Activity, and Internalizing Symptoms in Adolescence: Cross-sectional Analysis. *JMIR Ment Health* **8**, e26134 (2021).
7. Chen, Q., Dai, W., Li, G. & Ma, N. The impact of screen time changes on anxiety during the COVID-19 pandemic: Sleep and physical activity as mediators. *Sleep and Biological Rhythms* **20**, 521-531 (2022).
8. Sjölander, A. & Zetterqvist, J. Confounders, Mediators, or Colliders: What Types of Shared Covariates Does a Sibling Comparison Design Control For? *Epidemiology* **28**, 540-547 (2017).
9. Benjamini, Y. & Yekutieli, D. The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics* **29**, 1165-1188 (2001).
10. Gonggrijp, B.M.A., van de Weijer, S.G.A., Bijleveld, C.C.J.H., van Dongen, J. & Boomsma, D.I. The Co-Twin Control Design: Implementation and Methodological Considerations. *Twin Research and Human Genetics* **26**, 249-256 (2023).