Preregistration

Seeing the Bigger Picture: Exploring Children's Screen Time and Outcomes through Collaborative Data Analysis

Taren Sanders[1], James Conigrave[2], Michael Noetel[3], Rebecca Pagano[1], Chloe Gordon[1], Bridget Booker[1], Chris Lonsdale[1], Aliza Werner-Seidler[4], Leon Straker[5], Dylan Cliff[6]

[1] Australian Catholic University

[2] La Trobe University

[3] University of Queensland

[4] Black Dog Institute

[5] Curtin University

[6] University of Wollongong

Study Information

Title	Seeing the Bigger Picture: Exploring Children's Screen Time and Outcomes through Collaborative Data Analysis
Authors	Taren Sanders, James Conigrave, Michael Noetel, Rebecca Pagano, Chloe Gordon, Bridget Booker, Chris Lonsdale, Aliza Werner-Seidler, Leon Straker, Dylan Cliff Data contributors will be invited to co-author resulting publications (up to two authors per team).

Description

This study will pool and analyze individual-level data from multiple research projects to clarify how screen time affects children's and adolescents' learning, mental health, wellbeing, and behaviour. By uniting data from diverse samples, our team can pinpoint the specific amount of screen use (i.e., the dose) that may lead to either positive or negative outcomes, as well as how these relationships vary by characteristics such as age and gender. By examining the type and/or content of the screen time, we can also get a better understanding of how engaging with screens may impact on children's development. We will invite authors of relevant studies to contribute their de-identified data or share results through secure remote analysis (via DataSHIELD). After harmonising the data, piecewise regression models will be applied to identify thresholds where screen time use shifts from beneficial to harmful. The findings of this IPD meta-analysis will be translated into an evidence toolkit for parents, teachers, and students.

We aim to answer these research questions:

1. What is the impact of screen use on children's learning, cognitive abilities, mental health, wellbeing, and behaviour?
2. Does the relationship between screen use and outcomes vary by different types of screen use (e.g., content or type of device)?
3. Is there a specific screen use duration at which notable harm/benefit becomes apparent, and does it vary by type of screen use?
4. Does the relationship/duration vary by where the screen time occurs (i.e., home vs school)?
5. Does the relationship/duration vary by characteristics of the children (i.e., age, gender, socioeconomic status)?

Hypotheses

We hypothesise the following:

1. [RQ1] Overall screen use will have a small but statistically significant negative association with children's learning, cognitive abilities, mental health, wellbeing, and behaviour.
2. [RQ2] The type of screen time (i.e., the content or type of device) will moderate the relationship between screen use and children's outcomes.

   a. Educational content (i.e., screen time intended to educate children) will have a small-to-moderate positive association with children's learning and cognitive abilities, but no association with mental health, wellbeing, or behaviour.
   b. Non-interactive entertainment content (e.g., television/streaming) will have a small negative association with children's learning, cognitive abilities, mental health, wellbeing, and behaviour.
   c. Interactive entertainment content (e.g., video games) will have a small negative association with children's mental health, wellbeing, and behaviour, but a negligible association with learning and cognitive abilities.
   d. Social media will have a small to moderate negative association with children's mental health and wellbeing, but no association with learning, cognitive abilities, or behaviour.

3. [RQ3] There will be a threshold of screen time at which notable harm/benefit becomes apparent, and this threshold will vary by the type of content.
4. [RQ4] The relationship between screen time and children's outcomes will not be significantly moderated by the location of the screen time (i.e., home vs school), after adjusting for content.
5. [RQ5] The relationship between screen time and children's outcomes will be moderated by characteristics of the children. Specifically:

   a. Age will moderate the relationship between screen time and children's outcomes. However, given the inconsistency in previous findings on this relationship, we do not make a specific prediction about the direction of this moderation. Instead, this hypothesis will be considered exploratory.
   b. Child gender will moderate the relationship between some forms of screen time and children's outcomes: for social media, there will be stronger negative effects for girls' mental health and wellbeing outcomes than boys.
   c. Socioeconomic status will moderate the relationship between screen time and children's outcomes, with children from lower socioeconomic backgrounds experiencing stronger negative effects of screen time on educational, mental health, and behaviour outcomes.

Design Plan

Study type

Other {» Note: this is a dropdown «}

This will be an individual participant data (IPD) meta-analysis.

Blinding

No blinding is involved in this study.

Study design

We will use an observational research design, using pooled data from multiple studies. We will include both cross-sectional and longitudinal studies in the pooled analysis.

Randomization

There is no randomisation involved in this study.

Sampling Plan

Existing data

Registration prior to accessing the data. As of the date of submission, the data exist, but have not been accessed by you or your collaborators. Commonly, this includes data that has been collected by another researcher or institution. {» Note: This is a dropdown «}

Explanation of existing data

We will be collating datasets from multiple existing studies on children's screen time and outcomes of interest. The data will be de-identified, and shared with the research team either through secure transfer of data files or through secure remote analysis (using DataSHIELD). The research team may contribute their data to the pooled analysis, and therefore have prior knowledge of these data. But, as the final analysis will be based on the pooled data, this prior knowledge does not meaningfully affect the nature of the analysis. We will include a sensitivity analysis that excludes these datasets to ensure that the results are robust to this prior knowledge.

Data collection procedures

Data collection for this project will occur in two stages: one for identifying potential datasets and another for collating and harmonising the data.

Identifying datasets

We will identify potentially relevant datasets in two ways:

1. We will examine the included studies of relevant meta-analyses, using our recent umbrella review (Sanders et al., 2023) to identify these meta-analyses.
2. Where these meta-analyses are dated, or where a relevant meta-analysis is not identified, we will conduct a rapid review of the literature to identify relevant studies.

Dataset eligibility criteria

To be included in the pooled analysis, datasets must meet the following criteria:

1. Have quantitatively measured screen time exposure. Given the increasing evidence that the content of screen time is an important factor in determining impact, we will only include studies that have a disaggregate measure of screen time (i.e., they have measured the content [e.g., social media], or the type as a proxy for content[e.g., TV as 'video']).
2. Have quantitatively measured at least one outcome related to children's learning, cognitive abilities, mental health, wellbeing, or behaviour. {» This is probably too broad. Do we want to a priori pick outcomes for these? «}
3. Have a mean sample age younger than 18 years. That is, a sample who are predominantly pre-school or school-aged children and adolescents. If a mean study age is not available, we will use the midpoint of the age range.

Prioritising datasets

We expect that the process of harmonising and collating data will be very time-consuming, and the time required to complete this process will grow linearly with the number of datasets included. Therefore, we may not be able to include all datasets that are identified, and instead need to prioritise datasets that are most

likely to add value. To do this, we will calculate the expected value of each dataset based on the following criteria:

1. The size of the sample.
2. The extent to which the dataset provides underrepresented outcomes.
3. The extent to which the dataset provides underrepresented age groups.
4. The recency of the dataset, with more recent data considered more valuable.

We will calculate the value of each dataset (i) as:

$$\text{Value}_i = \underbrace{\alpha \, \ln\big(N_i + 1\big)}_{\text{sample size}} + \underbrace{\beta \, O_i}_{\text{outcome need}} + \underbrace{\gamma \, A_i}_{\text{age need}} + \underbrace{\delta \, S_i}_{\text{synergy}} + \underbrace{\varepsilon \, R_i}_{\text{recency}},$$

where:

- N_i is the sample size of dataset i, with a logarithmic transformation to dampen the influence of very large samples.
- O_i (**Outcome Need**) quantifies how underrepresented the dataset's outcome is in our overall pool. For instance:

  $$O_i = \frac{1}{1 + \text{coverage}(\text{outcome}_i)},$$

  where $\text{coverage}(\text{outcome}_i)$ is the total number of participants (across the currently included datasets) that measure the same outcome. A larger value for O_i means that the outcome is more underrepresented.
- A_i (**Age Need**) captures how underrepresented the dataset's age distribution is in set of included datasets. We will calculate this based on the dataset's mean age μ_i and standard deviation σ_i by following this approach:

  1. Maintain a coverage table, $\text{Cov}(a)$, for each relevant age (or bin) a of datasets already included.
  2. Approximate dataset i's age distribution as a normal curve around μ_i with SD σ_i.

3. Compute a weighted coverage:

$$\text{CovWeighted}_i = \sum_a \text{Cov}(a)\, w_i(a),$$

where

$$w_i(a) = \frac{\exp\left(-\frac{(a-\mu_i)^2}{2\,\sigma_i^2}\right)}{\sum_x \exp\left(-\frac{(x-\mu_i)^2}{2\,\sigma_i^2}\right)}.$$

4. Define

$$A_i = \frac{1}{1 + \text{CovWeighted}_i}.$$

This ensures A_i is larger when the dataset's mean (and spread) falls in underrepresented ages.

- S_i (**Synergy**) captures the additional value when a dataset contributes *both* an underrepresented outcome *and* an underrepresented age group.

$$S_i = O_i \times A_i.$$

- R_i (**Recency**) quantifies the fact that we consider newer data to be more valuable value added by more recent data. We define R_i as:

$$R_i = \frac{\text{year}_i - \text{year}_{\min}}{\text{year}_{\max} - \text{year}_{\min}},$$

where year_i is the data collection year of dataset i, and $\text{year}_{\min}$ and $\text{year}_{\max}$ are the earliest and most recent data collection years, respectively, in our pool of candidate datasets. This scales R_i between 0 (oldest) and 1 (newest).

Finally, α, β, γ, δ, and ε are weights reflecting how strongly we prioritise each component:

- α emphasises sample size,
- β underrepresented outcomes,
- γ underrepresented age groups,
- δ the synergy between outcome and age coverage,
- ε dataset recency.

We will initially set these weights to $\alpha = 2$, $\beta = 1$, $\gamma = 1$, $\delta = 2$, and $\varepsilon = 1$, but may adjust these based on relative importance as data is collected.

Note that since O_i and A_i are undefined in the initial state (as there are no included datasets), our initial calculation for value will leave out these terms and only be based on N_i and R_i. We will then rank-order datasets based on their value, and work through the list in order of value, updating the values each time a new dataset is added. We will continue this process until we reach a point where the time required to harmonise and collate the data is no longer feasible.

Collating and harmonising data

Once datasets are identified, we will contact the corresponding authors of these studies to invite them to participate. Authors who agree to participate will be asked to sign a letter of agreement, which will outline the terms of data sharing. We will submit these letters of agreement to the lead institution's Human Research Ethics Committee for approval.

Once ethics approval has been granted, we will ask authors to provide their de-identified data. We will give authors two options for sharing their data:

1. Securely sharing the de-identified raw data files with us directly. This method is less work for contributors but requires them to have ethical approval that allows for data sharing.
2. Setting up a DataSHIELD server, an open-source solution to federated analysis where the individual-level data can be analysed remotely but without risking disclosure. Analysis code is sent from a central machine to each of the servers, and only non-disclosive summary statistics are returned. This software allows for IPDs to be conducted without accessing the data directly, which can meet the requirements of many ethics boards.

Before conducting the analysis, we will harmonise the data to ensure variables are consistent across datasets. We will follow a process used in other federated analyses (Pinot De Moira et al., 2021). To ensure that the process is feasible and that the data can be harmonised in a meaningful way, we will pilot the harmonisation process on a subset of datasets that we can directly access. We will then ask data

contributors who are using DataSHIELD to harmonise their data in the same way. To validate that this has happened correctly, we will provide a script to contributors that will check that the data matches expectations. This harmonised data can then be added to a DataShield server for analysis.

Sample size

Given the volume of research on children's screen time, we anticipate that we will be able to include a large number of datasets in the pooled analysis, and therefore are not concerned about the number of participants.

Sample size rationale

Given that we expect to recruit a very large sample (>10,000 participants), we are not concerned about statistical power.

Stopping rule

We will create a version of the dataset in March 2026 to be used for this registered analysis. However, as we intend to conduct further analyses on this dataset in the future, we will allow additional datasets to be added to the analytical sample after this date.

Variables

Manipulated variables

There are no manipulated variables in this study.

Measured variables

The exact variables that will be measured will depend on the datasets that are included in the pooled analysis, and the extent to which they are able to be harmonised. However we expect to include at least the following variables:

Measures of screen use

While there is no consensus or standard tool for measuring screen, several survey tools have gained popularity in the literature. {» For team to consider: should we be including studies that use time use diaries? This would make harmonising more difficult, but lots of studies have used MARCA etc as their measure. «} These include the Screen Based Media Use Scale (Houghton et al., 2015), and Youth Risk Behavior Survey (Schmitz et al., 2004), and time use diary methods such as the

Multimedia Activity Recall for Children and Adolescents (Ridley, Olds, & Hill, 2006). From these, we can predict some of the measures we expect to be included in the pooled dataset.

- **Total screen time**: As an aggregated measure of screen time. We expect most studies to have already calculated this value, but if not, we will calculate it as the sum of time spent on different devices or types.
- **Video/e-games**: Time spent playing video/e-games.
- **Television**: Time spent watching television.
- **Mobile device**: Time spent on mobile devices, such as smartphones and tablets.
- **Social media**: Time spent on social media.
- **Computer**: Time spent on laptop or desktop general-purpose computers.
- **Educational time**: Time spent on using devices for educational purposes, such as to complete homework.

We will harmonise all measures of screen use to a common unit (average hours per day). In addition, we will record the tool used and include it as a covariate to adjust for any systematic differences across tools.

Note that we will not include measures which only indicate 'problematic' screen use, or have only a dichotomous measure of screen use (e.g., 'meets guidelines' or 'does not meet guidelines').

Outcome measures

{» Outcomes are the part of this I am most concerned about. We need to balance the extent to which we can meaningfully combine measures, with the extent to which we can reasonably expect to find datasets. E.g., if we limit 'beahviour' to the SDQ, we may not find enough datasets. But if we include all measures of behaviour, we may not be able to meaningfully combine them. Thoughts on how we address this are welcome. «} We will include a range of outcome measures related to children's learning, cognitive abilities, mental health, wellbeing, and behaviour. After identifying datasets, we will examine the measures used in these datasets and determine which measures can be harmonised and have sufficient data before contacting authors.

The below outline some of the measures we expect to be included.

Main Outcome	Specific Outcomes	Example Measures
Learning	General education	Standardised test scores
	Numeracy	Grades
	Literacy	Reading comprehension
Cognitive abilities	Executive function	Inhibitory control
	Cognitive function	Working memory
		Cognitive flexibility
		Attention
Mental health	Anxiety	Multidimensional Anxiety Scale for Children
	Depression	Children's Depression Inventory
	Emotions	Emotion Regulation Questionnaire for Children and Adolescents
Behaviour	Aggression	Child Behaviour Checklist
	Self-regulation	Strengths and Difficulties Questionnaire
	Prosocial behaviour	Behavior Rating Inventory of Executive Function
Wellbeing	Self-perceptions	General Self-Efficacy Scale
	Positive mental health	Students' Life Satisfaction Scale
		Pediatric Quality of Life Inventory TM

Covariates and moderators of effects

We will also ask authors to provide data on a range of covariates and moderators that may influence the relationship between screen time and children's outcomes, if they were measured in the study. These include:

- Child demographics, such as age, gender, and ethnicity.

- Socioeconomic status, such as parental education and income.
- The regional location of the child, such as urban or rural.
- Location of screen time, such as home or school.
- The country of data collection.
- The year of data collection.

Indices The nature of this study makes it hard to predict which measures will be combined in an index, beyond aggregated total screen time. However, we will publish a codebook which includes the variables and how to create them as part of the harmonisation process, which will be prior to analysis.

Analysis Plan

Statistical models We will address the research questions using piecewise regression. This approach allows us to identify thresholds where screen time use notably shifts from beneficial to harmful. We will examine the extent to which these thresholds vary by the type or content of screen time, and by the characteristics of the children.

Transformations The nature of this study, where data collection is outside of our control, makes it hard to predict which measures and what transformations or re-codes will be necessary for harmonisation. However, we will publish a codebook which includes the variables and how to create them as part of the harmonisation process, which will be prior to analysis.

Inference criteria Given the large number of outcomes, exposures, and potential moderators in this IPD meta-analysis, we will control the risk of false discoveries using a tiered false discovery rate (FDR) adjustment procedure.

Following the approach outlined in Bartik et al. (2024) and Guess et al. (2023), we will organize hypotheses into nested tiers based on conceptual importance and relatedness. We will apply the FDR adjustment described by Benjamini & Hochberg (1995) cumulatively at each level by accounting not only for the number of comparisons within a tier but also for the number of higher-level comparisons leading into it.

For each individual hypothesis, the total number of comparisons (k_{total}) used to compute the adjusted q-value will reflect:

- k_0: the number of primary outcome families tested,
- $k_{1,f}$: the number of content or moderator components within the relevant family,
- $k_{2,f}$: the number of specific items or interaction terms within that component.

Thus, each q-value reflects the **full testing burden up to that point**, not just the number of tests in the immediate family.

The tiered structure of hypotheses is as follows:

Tier	Description	Examples	Correction from
Tier 1: Primary Outcomes	Main effects of total screen time on key outcomes	Learning, mental health, wellbeing, cognitive ability, behaviour	Tier 1 only
Tier 2: Content-Specific Effects	Effects of screen content/type on each outcome	Educational content, Social media, Video games, Non-interactive entertainment	Tier 1 + Tier 2
Tier 3: Dose-Response / Thresholds	Identification of thresholds where screen time effects change	Piecewise regression estimating turning points per outcome/content	Tier 1 + Tier 2 + Tier 3
Tier 4: Moderation Analyses	Moderation by child characteristics or context	Age, Gender, SES, Home vs School	Tier 1 + Tier 2 + Tier 4
Tier 5: Exploratory / Secondary	Additional exploratory analyses, interactions, sensitivity tests	3-way interactions, less theory-driven tests	Tier 1 + Tier 2 + Tier 5

For each hypothesis test, we will compute FDR-adjusted q-values using the Benjamini & Hochberg (1995), adjusting for the cumulative number of comparisons.

This ensures that finer-grained or exploratory analyses are appropriately penalised for preceding layers of multiple testing. We will report both the original p-values and the adjusted q-values.

Data exclusion

For the primary analyses we will exclude data which exceeds 3 absolute deviations from the median, which is generally more robust than standard deviations from the mean (Leys, Ley, Klein, Bernard, & Licata, 2013). We will also exclude implausible values, such as screen time values that exceed 24 hours per day. In all cases, we will include sensitivity analyses that include these data points to ensure that the results are robust to these exclusions.

Missing data

We will assume that missing screen time, covariate, and moderator data are missing at random. We will impute missing values using multiple imputation by chained equations to provide 50 imputed datasets. We will use Rubin's rule to combine the results from all imputed datasets into one single set of results. As with missing data, we will include sensitivity analyses that do not impute missing data to test if the results are robust to this assumption.

Exploratory analyses (optional)

We are not registering any exploratory analyses. If interesting patterns emerge from the data, we will report these but explicitly note that they are exploratory.

Other

Other (Optional)

Not applicable.

References

Bartik, A. W., Rhodes, E., Broockman, D. E., Krause, P. K., Miller, S., & Vivalt, E. (2024). *The impact of unconditional cash transfers on consumption and household balance sheets: Experimental evidence from two US states*. National Bureau of Economic Research.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., ... Tucker, J. A. (2023). Reshares on social media amplify political news but do not detectably affect beliefs or opinions. *Science*, *381*(6656), 404–408. doi:10.1126/science.add8424

Houghton, S., Hunter, S. C., Rosenberg, M., Wood, L., Zadow, C., Martin, K., & Shilton, T. (2015). Virtually impossible: Limiting Australian children and adolescents daily screen based media use. *BMC Public Health*, *15*(1), 5. doi:10.1186/1471-2458-15-5

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. doi:10.1016/j.jesp.2013.03.013

Pinot De Moira, A., Haakma, S., Strandberg-Larsen, K., Van Enckevort, E., Kooijman, M., Cadman, T., ... Nybo Andersen, A.-M. (2021). The EU Child Cohort Network's core data: Establishing a set of findable, accessible, interoperable and re-usable (FAIR) variables. *European Journal of Epidemiology*, *36*(5), 565–580. doi:10.1007/s10654-021-00733-9

Ridley, K., Olds, T. S., & Hill, A. (2006). The Multimedia activity recall for children and adolescents (MARCA): Development and evaluation. *International Journal of Behavioral Nutrition and Physical Activity*, *3*(1), 10. doi:10.1186/1479-5868-3-10

Sanders, T., Noetel, M., Parker, P., Del Pozo Cruz, B., Biddle, S., Ronto, R., ... Lonsdale, C. (2023). An umbrella review of the benefits and risks associated with youths' interactions with electronic screens. *Nature Human Behaviour*,

8(1), 82–99. doi:10.1038/s41562-023-01712-8

Schmitz, K. H., Harnack, L., Fulton, J. E., Jacobs, D. R., Gao, S., Lytle, L. A., & Van Coevering, P. (2004). Reliability and Validity of a Brief Questionnaire to Assess Television Viewing and Computer Use by Middle School Children. *Journal of School Health*, *74*(9), 370–377. doi:10.1111/j.1746-1561.2004.tb06632.x