



Specification curve analysis shows that social media use is linked to poor mental health, especially among girls

Jean M. Twenge^{a,*}, Jonathan Haidt^b, Jimmy Lozano^c, Kevin M. Cummins^d

^a San Diego State University, United States of America

^b New York University–Stern School of Business, United States of America

^c University of California San Diego, United States of America

^d California State University, Fullerton, United States of America

ARTICLE INFO

Keywords:

Media use
Psychological well-being
Social media
Technology use
Mental health

ABSTRACT

An important 2019 paper applied a novel analytic technique called Specification Curve Analysis (SCA) to data from three large-scale community samples to investigate the association between adolescent technology use and mental health/well-being. The paper concluded that an association exists but is tiny, with median betas between -0.01 and -0.04 . This association was reported to be smaller than links between mental health and various innocuous variables in the datasets such as eating potatoes, and therefore to be of no practical significance. The current paper re-ran SCA on the same datasets while applying alternative analytic constraints on the model specification space, including: 1) examining specific digital media activities (e.g., social media) separately rather than lumping all “screen time” including TV together; 2) examining boys and girls separately, rather than examining them together; 3) excluding potential mediators from the list of controls; and 4) treating scales equally (rather than allowing one scale with many subscales to dominate all others). We were able to reproduce the original results with the original configurations. When we used the revised constraints, we found several much larger relationships than previously reported. In particular: among girls, there is a consistent and substantial association between mental health and social media use (median betas from -0.11 to -0.24). These associations were stronger than links between mental health and binge drinking, sexual assault, obesity, and hard drug use, suggesting that these associations may have substantial practical significance as many countries are experiencing rising rates of depression, anxiety, and suicide among teenagers and young adults.

1. Introduction

There is a mental health mystery, especially in the United States and United Kingdom: Why did rates of depression, anxiety, loneliness, dissatisfaction with life, self-harm, suicide attempts, and suicides begin rising among adolescents around 2012 (Burstein et al., 2019; Cybulski et al., 2021; Kalb et al., 2019; Keyes et al., 2019; Marquez & Long, 2021; Mercado et al., 2017; Mojtabai et al., 2016; Morgan et al., 2017; Patalay & Gage, 2019; Plemmons et al., 2018; Ruch et al., 2019; Spiller et al., 2019; Twenge, Cooper, et al., 2019; Twenge et al., 2021)? And why were these increases often larger among girls than among boys?

There is a prime suspect in this mystery: digital media, especially social media such as Instagram, Facebook, and Snapchat (Luby & Kertz, 2019; Spiller et al., 2019). The timing fits: By 2012 most Americans owned a smartphone (Pew Research Center, 2021), American teens

became heavy users of social media just before 2012 (Twenge, Cooper, et al., 2019; Twenge, Martin, & Spitzberg, 2019), and social media became more engaging after the 2009 introduction of “like” buttons and “retweet” or “share” buttons (Haidt & Rose-Stockwell, 2019). Use of social media in particular can also help explain the sex difference, since girls spend far more time on social media than do boys (Twenge & Martin, 2020). Dozens of studies show a clear association between social media use and poor mental health, particularly for girls (Kelly et al., 2019; Lin et al., 2016; Twenge & Farley, 2021). In addition, more than a dozen experiments have randomly assigned participants to reduce their social media use (or not). Most demonstrated an improvement in mood or well-being (e.g., Hunt et al., 2018; Tromholt, 2016; for a review, see Haidt & Twenge, 2021, section 4). Radtke et al. (2021) report mixed results in their review of 23 “digital detox” experiments, but when their analysis is limited to the 13 studies examining reduced social media

* Corresponding author at: Department of Psychology, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-4611, United States of America.
E-mail address: jtwenge@sdsu.edu (J.M. Twenge).

(rather than smartphone) use, the majority found mental health benefits.

Of course, the steady rise in adolescent depression since 2012 could be caused by any number of social, economic, or technological trends; the link to rising social media use could be coincidental. Several other possible culprits have been suggested, such as the global financial crisis that began in 2008 and brought economic insecurity to many families. However, the U.S. economy steadily improved between 2012 and 2019, even as adolescent depression rates increased, which is the opposite of what one might expect if economic stressors explained population-level changes in mental health. Other possible causes such as opioid use—which did increase in prevalence during this time period—affected adults more than adolescents and was not uniform in its impact across regions or across social classes in the way rising technology use was (Twenge, 2020).

Strikingly, several studies and literature reviews in the past few years have concluded that technology use is *not* contributing to rising mental health problems, on the basis that technology use and mental health are not associated in a meaningful way among individuals. The most important and widely cited such paper was published in January 2019. In this paper, Orben and Przybylski (2019a; hereafter, O&P) used a new analytic approach which they argued was less vulnerable to the analytic and interpretive errors that are a common challenge in large-scale epidemiological studies. O&P noted that previous researchers had significant leeway in choosing which variables to examine when analyzing large public datasets, and thus they sought to provide a more rigorous and comprehensive analysis. They analyzed data from more than 350,000 adolescents across three datasets using Specification Curve Analysis (SCA), a statistical technique that can include the summarization of millions of possible analyses by including varying analytical specifications (Simonsohn et al., 2020). SCA is an ensemble analytic approach where many individual models are estimated and their results are aggregated to provide one synthetic estimate. Each specification can vary based on how the exposure variables, covariates, and outcome variables are composited and entered into the model. This technique, O&P contend, can “provide a more definitive and clearly contextualized test of the association between screen use and well-being” (p. 174) and is “a promising solution” (p. 179) as it can account for many different analytical decisions, often known as “researcher degrees of freedom” (p. 180). The end result of their analyses combining all effects for “screen time” and mental health variables produced effect sizes (median standardized Betas) between -0.01 and -0.04 , which they concluded were of “little practical value” (p. 179) and “too small to warrant policy change” (p. 173). The effect size, they noted, was similar to the correlation (in one of the three datasets) between well-being and eating potatoes.

O&P's study was covered extensively in the press, with headlines including “Screen time may be no worse for kids than eating potatoes” (Baron, 2019) and “Calm down, parents: A rigorous new Oxford study of 350,000 teens shows screen time is about as dangerous as potatoes” (Stillman, 2019). The latter article went on to note, “The panic over kids and screen time is wildly overblown, according to the most rigorous study to date,” telling parents “you can mostly stop worrying about screen time now” and calling previous studies that found significant effects “wrong” and based on “cherry picking the data.” Orben and Przybylski (2019a) has been widely cited in the academic literature as well, including in a comprehensive review of research on digital media use and well-being (Odgers & Jensen, 2020), which concluded, primarily on the basis of O&P's presented evidence, that the associations in this area were “unlikely to be of clinical or practical significance” (p. 336).

So is the debate over? Are O&P's analyses more definitive than previous studies, providing the final word on the question?

In a commentary on Orben and Przybylski (2019a, 2019b), Twenge et al. (2020) argued that SCA does not solve the problem of researcher degrees of freedom. They noted six choices that O&P made in setting up

their analyses, each of which had (or was likely to have) the effect of reducing the effect size found. The four most important were: 1) That O&P did not include specifications for sex-specific effects (i.e., they did not examine girls and boys separately), 2) That O&P focused on “screen time” rather than on social media, 3) that O&P included possible mediators (such as negative attitudes toward school) as controls, and that 4) O&P's effect sizes were diluted by including large numbers of individual survey items instead of focusing on a small number of validated scales.

In a response to this critique, O&P re-ran an SCA and reported finding similarly tiny effects even when they excluded mediator controls and included validated scales only instead of individual survey items (Orben & Przybylski, 2020). However, their reanalysis did not separate associations for TV from newer activities such as social media, and it persisted in weighting one mental health scale with many subscales much more heavily than the others. On the other hand, as O&P rightly pointed out, the analyses in the critique did not actually run SCAs using their code.

The current article takes this step, conducting SCAs using O&P's code to examine empirically based specifications. We report the results of SCA using the same three large datasets of U.S. and U.K. adolescents included in Orben and Przybylski (2019a, 2019b). We first show that we can replicate O&P's findings using their specifications. We then re-run those analyses with four alterations:

1.1. Screen time vs. social media

O&P's primary analyses lumped TV watching in with newer types of digital media such as social media under the umbrella of “screen time.” Given that TV time has declined and digital media time has increased after 2012 when adolescent depression spiked (Twenge, Cooper, et al., 2019; Twenge, Martin, & Spitzberg, 2019), it seems important to separate associations by media activity. In fact, other research has found that links between mental health and screen time are larger for social media and internet use than for TV use (Boers et al., 2019; Twenge & Farley, 2021). Thus, separating results by activity is empirically supported.

1.2. Sex-specific models

O&P's analyses were not structured to allow for effect moderation by sex. Given that increases in depression and self-harm are often larger for girls (e.g., Keyes et al., 2019; Mercado et al., 2017, although Cybulski et al. (2021) found a mixed pattern in the UK) and that associations between social media use and mental health are generally larger for girls (e.g., Kelly et al., 2019), it seems important to examine results separately by sex.

1.3. Control variables

O&P used a list of control variables that included not just demographic factors such as race and income but also mood-related variables such as negative attitudes toward school. Given that mental health is the hypothesized *outcome* variable, control variables related to mood are potential mediators. O&P also included variables such as parental distress, closeness to parents, and school grades that are potential mediators (as it is possible that adolescent technology use could effect parental distress, closeness to parents, and school grades). These variables differ from demographic controls, such as race/ethnicity and parents' education level, that cannot be caused by adolescent technology use. Current advice strongly recommends against including potential mediators as controls, as doing so can result in the model suffering from overadjustment bias that attenuates away the association between the exposure and outcome variable (Rohrer, 2018; Schisterman et al., 2009). If only one specification for control variables is to be used, as O&P also did, it seems wiser to perform analyses using only demographic variables as controls, and excluding potential mediators.

1.4. Treatment of scales

O&P's SCA included random combinations of mental health items, including individual items as well as overall scale scores, based on the justification that previous researchers may have selectively chosen certain items or fished through the millions of possible combinations. Twenge et al.'s (2020) comment on O&P argued for including only the overall score from each validated scale, because such scales were designed to measure a construct, and because a scale that happens to have many items should not swamp the importance of a validated scale with few items in it. After the comment was published, we identified an even more fundamental issue in O&P's statistical code. O&P's SCA of the Millennium Cohort Study (which they called "the highest quality dataset we examined") included 4 validated mental health scales: the parent-report Strengths and Difficulties Questionnaire (SDQ), and multi-item self-report scales measuring self-esteem, depressive symptoms, and life satisfaction. However, O&P included the total SDQ, each of the 5 SDQ subscales, and two combinations of SDQ subscales as separate measures. Thus, 8 of the 11 scales in their analysis – 73% of the data using scales – involved the SDQ, even though the SDQ was only one of four validated scales in the dataset. Therefore, it seems reasonable to perform an analysis in which each of the four measures of mental health counts equally if the intent is to evenly distribute the influence of the specifications among the scales.

We explore each of these issues, and others that appear depending on the dataset, within the same three datasets used in Orben and Przybylski (2019a, 2019b), also using the SCA technique.

2. Study 1: Millennium Cohort Study

The Millennium Cohort Study (MCS) is a nationally representative sample of UK adolescents born in 2000 and 2001. The 2015 wave of data collection included measures of hours/day of use of four screen activities (social media, gaming, internet, and TV) and four measures of mental health (self-esteem, depressive symptoms, life satisfaction, and the parent-report Strengths and Difficulties Questionnaire or SDQ). Analyses included 5926 girls and 5946 boys between 13 and 15 years old.

2.1. Replication of O&P

In their SCA, O&P reported that the median standardized beta for technology use and the mental health measures in the MCS was -0.032 ; this analysis includes betas combined across boys and girls, all screen activities, all measures of mental health as both scales and individual items, and with and without control variables. We first attempt to replicate this result and then we examine the question relevant to our hypothesis: How does the median Beta change when analyses zero in on girls and social media, exclude potential mediators as control variables,

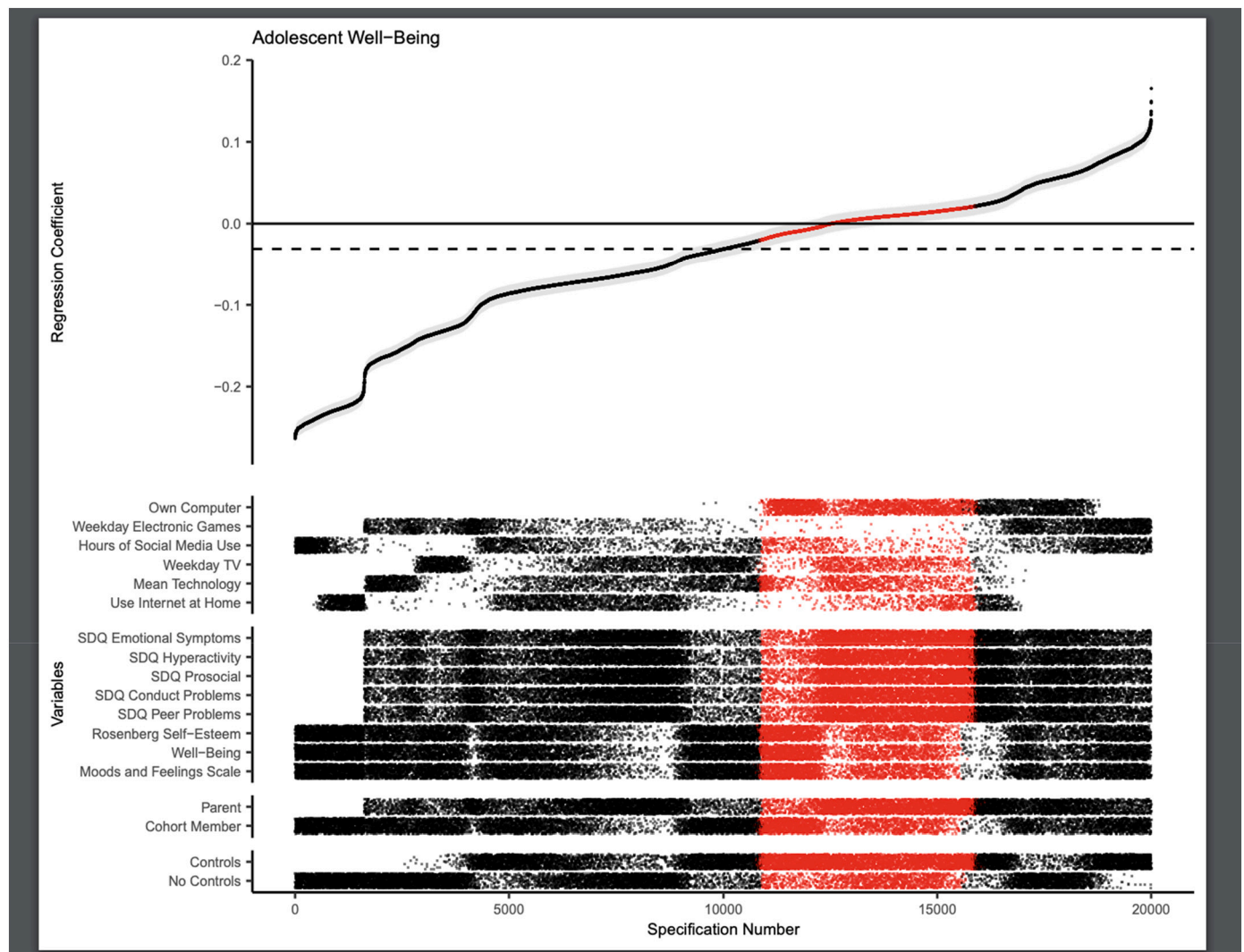


Fig. 1. Replication of SCA on the Millennium Cohort Study, equivalent to Fig. 3 of Orben and Przybylski (2019a, 2019b).

and treat all mental health scales equally?

We relied on [Orben and Przybylski's \(2019a, 2019b\)](#) R code (posted on OSF) to perform an SCA on the MCS dataset, and were able to replicate their SCA (see [Fig. 1](#)). The squiggly line at the top is composed of dots, each of which represents the beta found in one analysis performed using the specifications of the dots immediately below it. The dashed line just below the zero line shows that the median beta is indeed low—a mere -0.032 .

2.2. An alternative set of specifications

Having successfully replicated O&P's analysis, we then set out to determine if theoretically defensible alternative choices for the limits and structure of the specification space sampled in the SCA would produce different results. As detailed in the introduction, these differences include 1) separate analyses for girls and boys, 2) separate analyses by screen activity (social media, internet, gaming, TV), 3) analyses including demographic (non-mediator) controls only and excluding controls that could be mediators (e.g., closeness with parent, time spent with parent, parent distress, and negative attitudes toward school), and 4) treating each of the four mental health scales equally. The code for these analyses is available at: https://github.com/healthykids/SCA_Replication

2.3. Method

We re-ran our SCA analyses using only demographic (non-mediator) controls: family income, natural father present, child's ethnicity (white vs. non-white), primary caregiver education, primary caregiver's employment, number of siblings in household, child's long-standing illness (yes or no), primary caregiver's vocabulary word score, and child age.

2.4. Results and discussion

The SCA for technology use and mental health produces quite different results when our four modifications are implemented (see [Tables 1a and 1b](#)). To be consistent with O&P, all mental health measures are coded such that higher numbers mean better mental health, so negative correlations show that spending more hours on an activity is associated with worse outcomes.

For example, the top half of the social media column of [Table 1a](#) shows that the median beta for girls, applying no controls, is rather substantial at -0.21 . Adding in demographic controls only reduces the median beta to -0.20 . But when potential mediators are included as controls, the median beta falls to -0.01 . The pattern is nearly identical for internet usage overall, while the associations with harm for gaming and TV are lower.

Table 1a
SCA median betas for associations between screen time and mental health, Millennium Cohort Study, girls.

Girls only	Social media	Internet	Gaming	TV	All screen time
Scales treated equally					
No controls	-0.21	-0.22	-0.09	-0.13	-0.18
Demographic controls only	-0.20	-0.21	-0.08	-0.09	-0.17
O&P controls with potential mediators	-0.01	-0.03	-0.02	0.00	-0.01
O&P scales: 73% SDQ					
No controls	-0.11	-0.09	-0.10	-0.06	-0.10
Demographic controls only	-0.06	-0.07	-0.07	-0.03	-0.07
O&P controls with potential mediators	0.01	0.01	-0.04	0.00	0.00

Table 1b

SCA median betas for associations between screen time and mental health, Millennium Cohort Study, boys.

Boys only	Social media	Internet	Gaming	TV	All screen time
Scales treated equally					
No controls	-0.06	-0.10	-0.09	-0.07	-0.09
Demographic controls only	-0.04	-0.08	-0.06	-0.04	-0.06
O&P controls with potential mediators	0.04	-0.01	-0.04	-0.01	0.00
O&P scales: 73% SDQ					
No controls	-0.03	-0.03	-0.09	-0.05	-0.03
Demographic controls only	-0.01	-0.03	-0.05	-0.02	-0.02
O&P controls with potential mediators	0.03	0.00	-0.03	-0.01	0.01

Similarly, the lower half of [Table 1a](#) shows that when the SDQ is allowed to dominate the analysis, as in O&P's SCAs, the beta for social media for girls falls to -0.11 with no controls, and it falls further, to 0.01 , with O&P's controls. [Table 1b](#) shows that for boys, there are fewer substantial links between screen time (including social media) and mental health.

In response to our previous critique, [Orben and Przybylski \(2020\)](#) performed an SCA including only validated scales (with no individual items or random combinations of items) and reported that the median beta was similar to their original analysis including both items and scales. However, the SDQ constituted 8 of the 11 (73%) scales in their analysis (the total scale, the 5 subscales, and 2 different combinations of subscales). The SDQ, which is a parent-report measure, produces notably lower betas than the 3 other mental health measures, lowering the Betas when the majority of specifications involve the SDQ. In contrast, when all four mental health scales are treated equally, the median beta for girls and social media rises to -0.20 (see [Figs. 2 and 3](#)). In contrast to the large number of Betas close to 0 in O&P's analysis (see [Fig. 1](#)), in this analysis most of the Betas are between -0.25 and -0.15 (see [Fig. 2](#)).

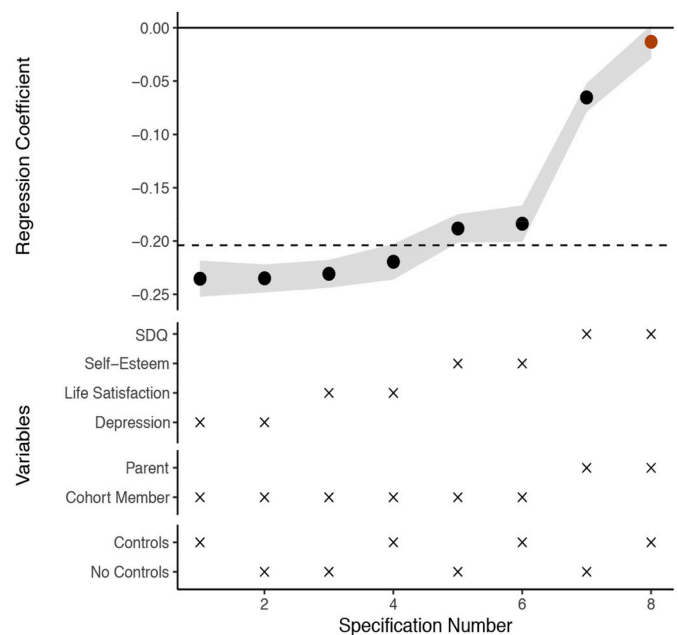


Fig. 2. SCA of social media and mental health among girls, all scales weighted equally, non-mediator controls only
NOTE: There are fewer specifications in this figure than in [Fig. 1](#) because only total scales are included as specifications.

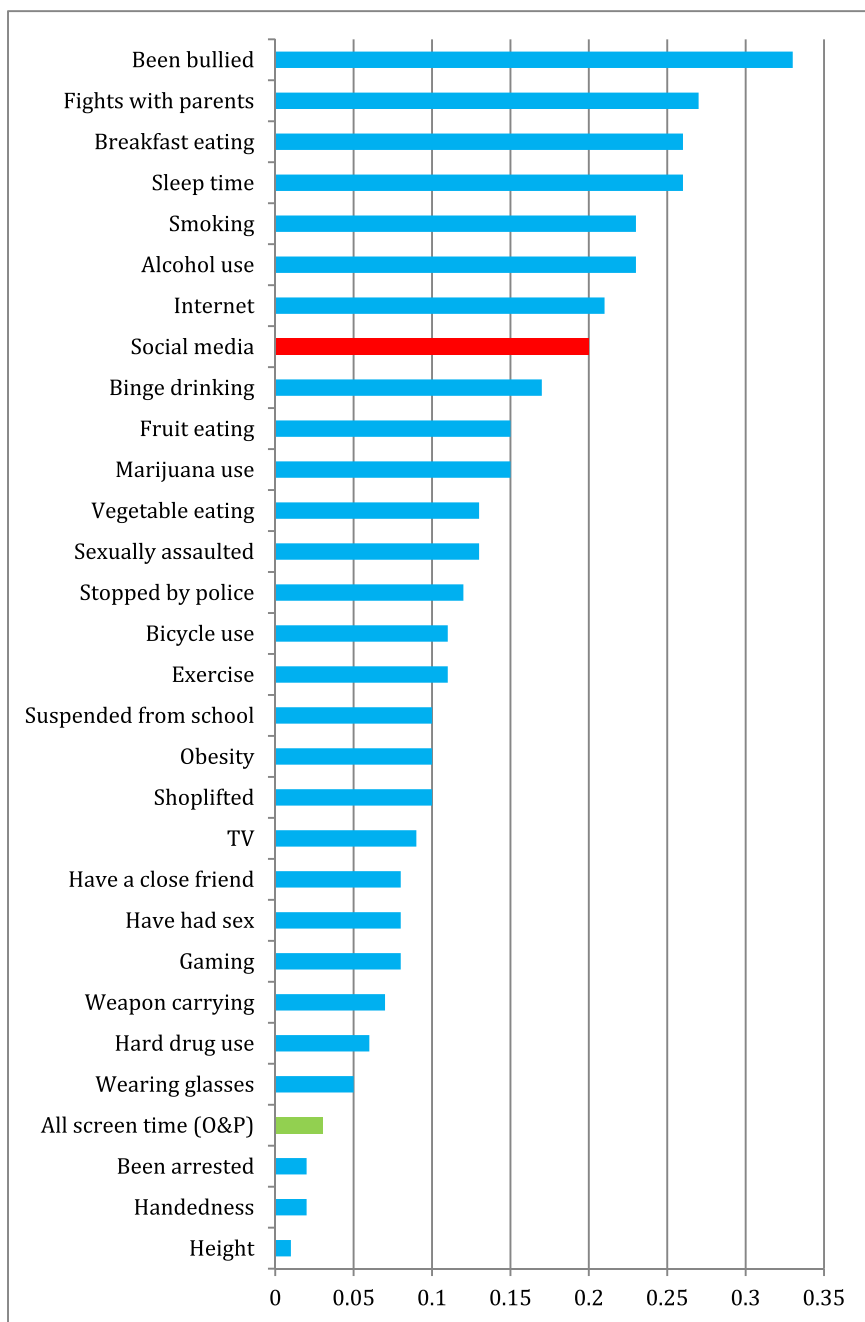


Fig. 3. SCA median betas (absolute values) for associations of well-being, social media, other technology use, and comparison variables, Millennium Cohort Study, girls only.

NOTES: All measures weighted equally and demographic control variables. All screen time (O&P) = median beta for complete SCA from Orben and Przybylski (2019a, 2019b).

Thus, when only demographic variables are included as controls and the four mental health scales are treated equally, effect sizes using SCA are strikingly larger, especially for internet and social media use among girls. However, effect sizes are smaller for TV and gaming, and for boys.

How does a beta of -0.20 for social media use and mental health among girls compare to other behaviors and traits that may be associated with mental health? This is one way to gauge practical importance. O&P concluded that their associations (e.g., a standardized Beta of -0.032) were comparable to the association for “eating potatoes” and therefore they were not practically important. However, a beta of -0.20 is larger than the associations between mental health and binge drinking, sexual assault, marijuana use, use of drugs other than marijuana, early sexual activity, having a close friend, being suspended from school, exercise, obesity, and carrying a weapon in an SCA with the same specifications (Fig. 3).

Thus, the association between social media use and well-being

among girls appears to be practically important. (Note: we are *not* arguing that social media is as dangerous as binge drinking or drug use for a single user at a single time. We are just pointing out that in this dataset, knowing the number of hours that a girl uses social media each week gives you about as much ability to predict her level of mental health problems as does knowing how often she binge drinks or uses certain drugs.) The link between internet use and well-being among girls also appears to be practically important by this standard.

3. Study 2: Monitoring the Future

The Monitoring the Future (MtF) project has collected a nationally representative sample of U.S. 8th and 10th graders every year since 1991. Survey waves since 2009 have included questions about digital media, with those after 2013 asking about use in hours per week. O&P reported a median beta of -0.005 for an SCA of screen time and mental

health in MtF 2009–2015.

However, Orben and Przybylski's (2019a, 2019b) analysis excluded a measure of social media use in hours per week, instead using only a measure of social media use from "never" to "almost every day" which has very low variance because 75% of adolescents in this dataset reported using social media almost every day. The choice of social media engagement measure used by O&P may have had a substantial impact on their results. It seems theoretically defensible, and prudent, to examine the measures asking about hours a week as well.

Similar to their analysis of the MCS, O&P's analysis of the MtF dataset also included several potential mediator variables (enjoyment of school, predicted school grades, and talking to parents about problems). Following the recommendations by Rohrer (2018) and Schisterman et al. (2009), these variables were excluded from the specifications in the current SCAs. These are variables that could potentially be in the causal pathway between digital media use and mental health.

Thus, as with our analysis of the MCS, we examined effect sizes using demographic controls only. Also, we differentiated social media from internet, gaming, texting and video chat use in hours per week. These items were asked of the same participants as general happiness ("Taking all things together, how would you say things are these days—would you say you're very happy, pretty happy, or not too happy these days?").

3.1. Method

With only a single well-being item, the SCA was composed of a single specification, and all individual betas can be reported. Demographic control variables were race (Black, Hispanic, and white), grade, mother's education, mother's employment, and presence of siblings. Participants were 16,374 boys and 17,377 girls in 8th and 10th grade, 2013–2016.

3.2. Results

The left halves of Tables 2a and 2b show analyses of the items that O&P reported, using the digital media items that had a "never" to "every day" response format as well as TV in hours a day. Like O&P, we find only very small associations. The right half of these tables shows our analyses of (non-TV) digital technology use using items in the "hours per week" format, which O&P did not examine. When social media use is measured in hours and only demographic variable controls are used, the beta for happiness is -0.10 among girls (see first column of Table 2a). Including potential mediators as controls, as O&P did, cuts the beta to -0.03 . As in the MCS, effect sizes for boys are smaller than for girls (see Table 2b). However, effect sizes for general internet use and happiness are larger than those for social media for both boys and girls.

How does this effect size for girls compare to other outcomes? Social media use measured in hours is more strongly linked to happiness among girls than selling drugs, using heroin, binge drinking, getting into fights, and stealing larger items. In this dataset, social media is about as closely (negatively) linked to happiness as is using marijuana, damaging school property, skipping school, and running away from home (see Fig. 4). In addition, the association between general internet use and happiness among girls is even larger, exceeding associations for running

away, damaging school property, and stealing smaller items and equaling the association between happiness and marijuana use.

Once again, the association between social media (and internet) use and happiness among girls seems practically important, at least when social media use is measured in hours and the association is not attenuated by using mediators as controls.

4. Study 3: Youth Risk Behavior Surveillance System

The Youth Risk Behavior Surveillance System (YRBSS) is a survey administered by the US Centers for Disease Control to a nationally representative sample of U.S. 9th to 12th graders every other year. Since 2007, it has included a question on "electronic device use" mentioning computers and gaming consoles. In 2013 the survey modified the electronic device use question to also mention smartphones and tablets. The survey also includes one question on daily TV watching. The survey does not have any questions that focus specifically on social media usage. O&P analyzed data 2007–2015 and reported a median beta of -0.035 for the complete SCA of the YRBSS, for the association of screen time and mental health.

However, O&P's SCA lumped together TV use with electronic device use. Although YRBSS does not separate out social media use, we were able to take two steps to bring the analysis closer to testing our hypothesis about social media: 1) examining electronic device use separately from TV, and 2) examining only the years after 2013, when phones and tablets were added to the existing item on electronic device use.

4.1. Method

Unlike in the other two datasets, O&P did not include controls that were likely to be mediators; the only control they included was race (white vs. non-white). To ensure that overly minimal controls were not an issue, we included a more thorough control for race (using Hispanic, Black, and other race as separate variables) and controlled for grade (9th, 10th, 11th, or 12th as a continuous variable).

The YRBSS included five items measuring mental health: depressive symptoms, suicidal ideation, suicide planning, suicide attempts, and injurious suicide attempts. These were treated equally in the SCA with no combinations included. Participants were all those in the 2013 and 2015 waves: 14,304 boys and 14,026 girls. The code for these analyses is available at: https://github.com/healthykids/SCA_Replication

4.2. Results

Our SCA produced a median beta of -0.11 for girls and -0.09 for boys for electronic device use. O&P's beta (-0.035) was likely lower as it included TV, which produces lower betas (see Tables 3a and 3b). Configurations of control variables had little impact on effect sizes, likely because in this dataset Orben and Przybylski (2019a, 2019b) did not include possible mediators as controls.

In their SCA combining electronic device use with TV, O&P concluded that screen time was linked to mental health to about the

Table 2a
Kinds of screen time and general happiness, standardized betas, Monitoring the Future, girls.

O&P items	Digital media in hrs/week											
	Girls only	Week end TV hours	Weekday TV hours	Internet news never/every day	Social media never/every day	Median	Social media hours	Internet hours	Gaming hours	Texting hours	Video chat hours	Median
No controls		-0.01	-0.02	-0.02	-0.03	-0.02	-0.11	-0.13	-0.10	-0.07	-0.07	-0.10
Demographic only		-0.01	0.00	-0.02	-0.03	-0.02	-0.10	-0.13	-0.08	-0.06	-0.07	-0.08
O&P mediator controls		-0.01	0.02	-0.01	0.01	-0.01	-0.03	-0.07	-0.03	0.00	-0.02	-0.03

Table 2b
Kinds of screen time and general happiness, standardized betas, Monitoring the Future, boys.

O&P items						Digital media in hrs/week						
	Boys Only	Weekend TV hours	Weekday TV hours	Internet news never/every day	Social media never/every day	Median	Social media hours	Internet hours	Gaming hours	Texting hours	Video chat hours	Median
No controls		0.00	0.00	0.02	0.03	0.01	-0.02	-0.07	-0.07	-0.01	-0.04	-0.04
Demographic only		0.00	0.00	0.02	0.04	0.01	-0.03	-0.09	-0.07	-0.02	-0.04	-0.04
O&P mediator controls		-0.01	0.01	0.01	0.05	0.01	0.01	-0.07	-0.05	0.01	-0.02	0.01

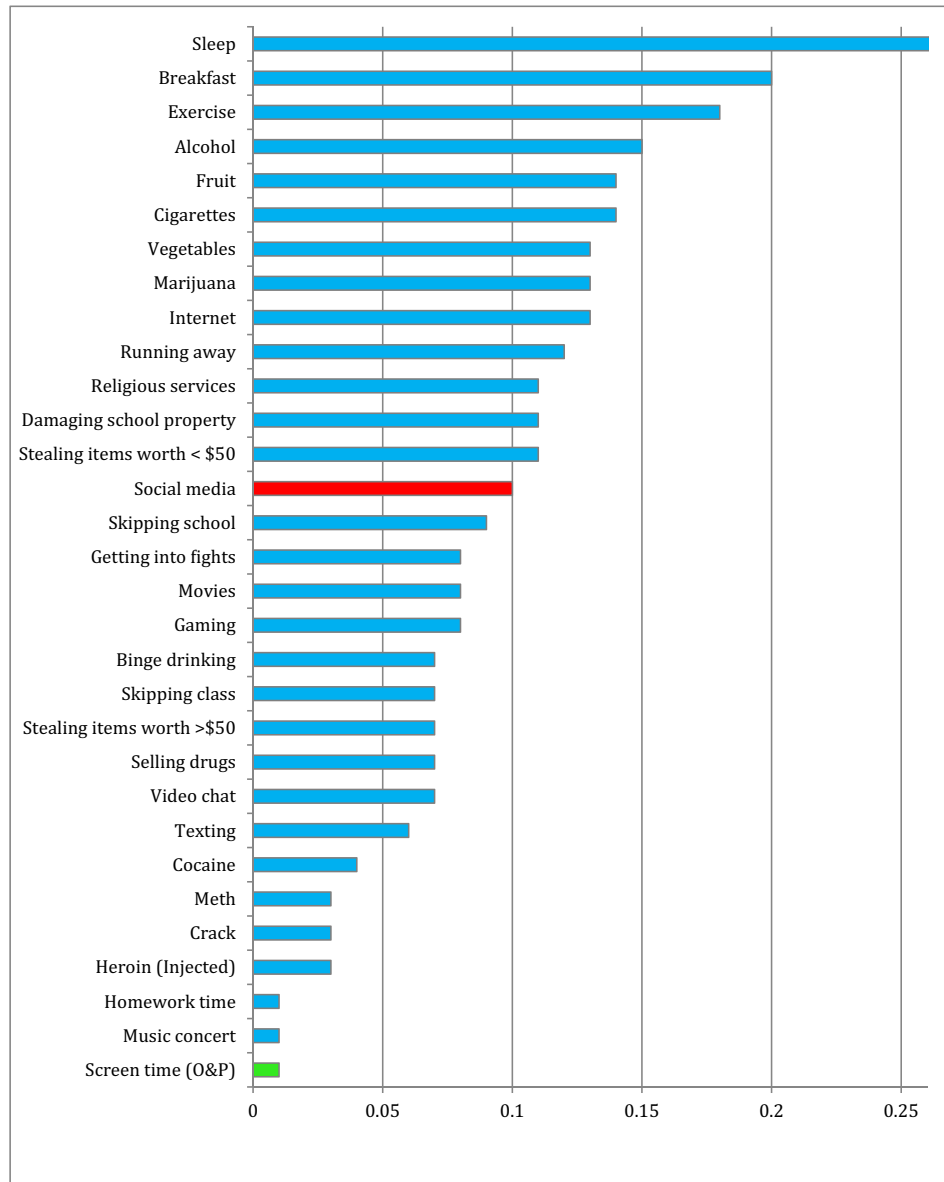


Fig. 4. Betas (absolute values) for happiness, social media use in hours/wk., other technology use in hours/wk., and comparison variables, Monitoring the Future, girls

NOTES: Analyses include demographic controls. Screen time (O&P) = median beta for complete SCA as reported by [Orben and Przybylski \(2019a, 2019b\)](#). with mediator controls and low-variance measure of social media, combining boys and girls and analyses with and without controls.

same degree as was eating potatoes. But when we eliminated TV and examined electronic device use separately, we found that for girls electronic device use is more strongly linked to mental health than injectable drug use, obesity, exercise, having four or more sexual

partners, and having sex before age 13, and is similar to heroin use (see [Fig. 5](#)). The link between electronic device use and mental health in this dataset appears practically important – even though electronic device use includes gaming and other activities that we hypothesize are not as

Table 3a

Median betas, SCA, screen time (hours/day) and mental health, Youth Risk Behavior Surveillance System, 2013–2015, girls.

	Electronic devices	TV
Girls		
No controls	-0.11	0.00
O&P controls	-0.11	0.00
Additional controls	-0.11	0.00

Table 3b

Median betas, SCA, screen time (hours/day) and mental health, Youth Risk Behavior Surveillance System, 2013–2015, boys.

	Electronic devices	TV
Boys		
No controls	-0.09	-0.01
O&P controls	-0.09	-0.01
Additional controls	-0.09	-0.01

strongly linked to mental health issues as social media.

5. General discussion

An important empirical article in 2019 used an advanced statistical technique (Specification Curve Analysis) and reported that the associations between screen time and mental health problems among teens in

the UK and USA was tiny—roughly the same size as for eating potatoes—and therefore “too small to warrant policy change” (Orben & Przybylski, 2019a, p. 173). We were able to replicate these findings in Study 1 when we applied the same analytic constraints. Median Betas from O&P’s SCAs for total screen time are included in Fig. 6 as the green bars.

We suspected that larger relationships would be found using the same analytic framework if different constraints on the specifications were made, constraints that we thought were at least as reasonable and warranted. When we re-analyzed those same three datasets using these alternative constraints, we found much larger relationships, shown in Fig. 6 as the red bars. In particular, social media use (as opposed to all screen time) for girls (as opposed to all teens) showed much larger relationships with poor mental health, particularly when we removed possible mediators (such as unhappiness at school) from the set of variables controlled for, and when we treated all of the mental health scales equally.

The effect sizes we found are comparable to other factors one would expect to be linked to mental health among adolescents, including binge drinking, sexual assault, obesity, and drug use including heroin use. We also found that general internet use was associated with poor mental health outcomes for girls and sometimes for boys, but this finding should be interpreted with caution as “internet use” is such a heterogeneous category, and a large portion of it may be social media use, particularly for girls. Nevertheless, these findings suggest that general internet use is also associated with negative mental health outcomes at meaningful levels.

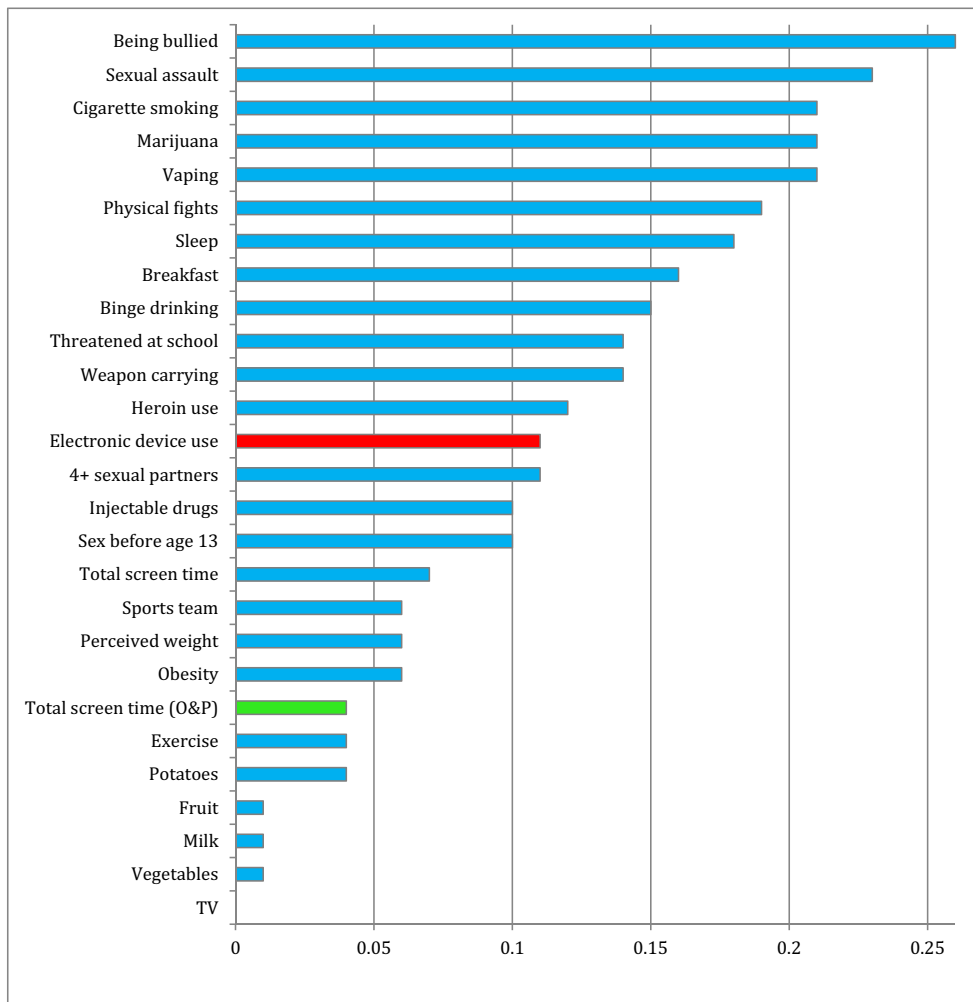


Fig. 5. SCA median betas (absolute values) for mental health, electronic device use, TV use, and comparison variables, Youth Risk Behavior Surveillance System, girls, 2013–2015

NOTE: Analyses include additional controls. Total screen time (O&P) = median beta for complete SCA, combining boys and girls and analyses with and without controls.

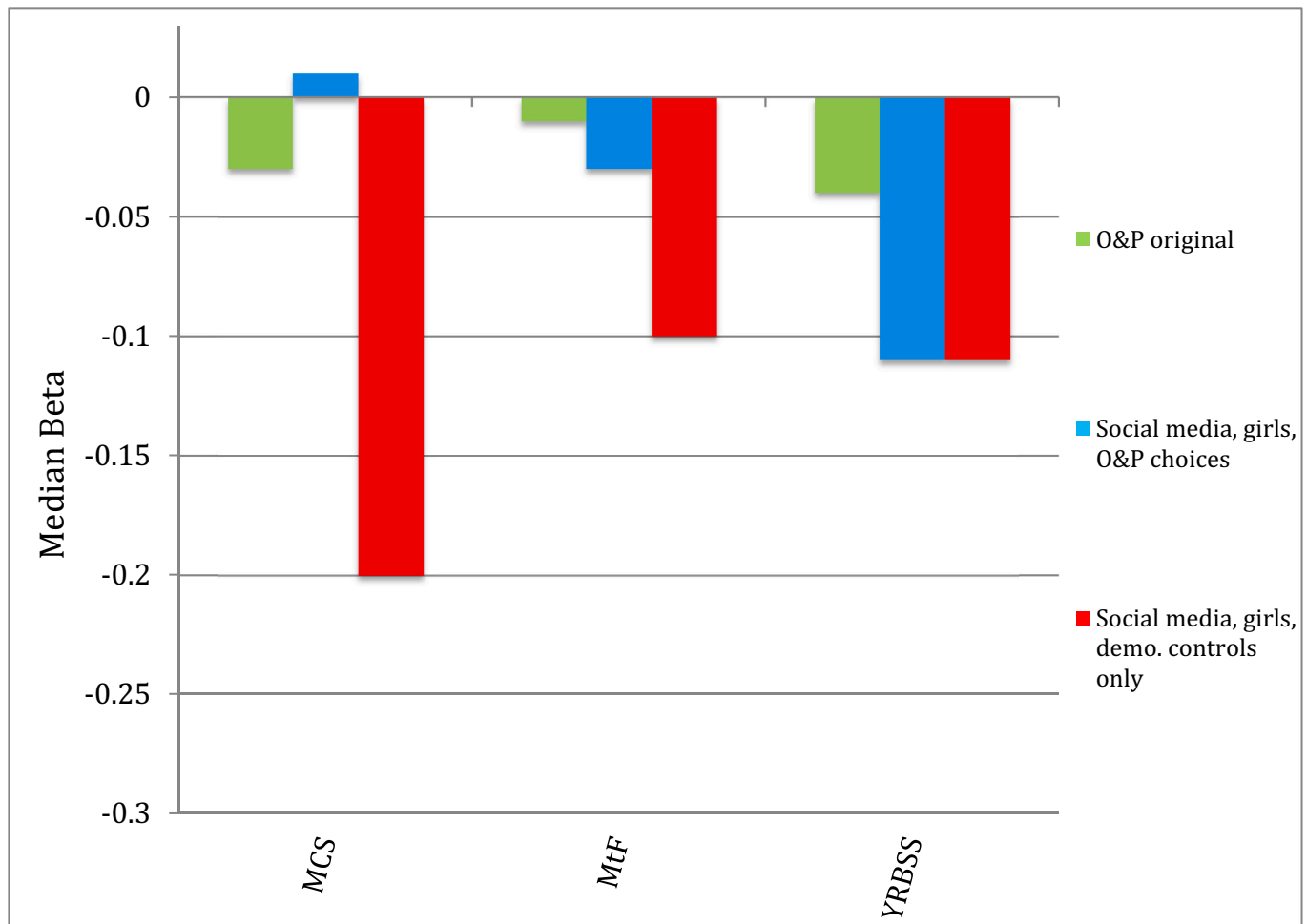


Fig. 6. Median Betas for the association between technology use and mental health given different specifications, three datasets

NOTE: The green bars report O&P's original findings, including both boys and girls, all screen activities including TV, potential mediators as controls, both items and scales, and (in MCS) scales representing 73% of the data. The blue bars report median betas for: MCS: girls only, social media use, all of O&P's controls including potential mediators, scales only, SDQ is 73% of data. MtF: girls only, social media use using the "never" to "every day" item, all of O&P's controls including potential mediators. YRBSS: girls only, electronic device use, controls for race (as white vs. nonwhite). The red bars report median betas for: MCS: girls only, social media use, demographic controls only, scales only, scales treated equally. MtF: girls only, social media use in hours per day, demographic controls only. YRBSS: girls only, electronic device use, controls for race (as individual variables) and grade.(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The question of what parents and policymakers should do hangs on two questions, one empirical, the other theoretical. The empirical question is whether the size of the relationship is less than $r = 0.05$ (as is reported by O&P), or more than $r = 0.10$, as we found in our SCAs. There is an emerging consensus that whatever the relationships are for "screen time" and mental health, the relationships for *social media* and mental health are larger, and are generally above $r = 0.10$ in absolute value. Orben (2020) reached this conclusion as well. In a "narrative review of reviews" of studies, she noted that "The associations between social media use and well-being therefore range from about $r = -0.15$ to $r = -0.10$ " (p. 409). However, she referred to these correlations as "small," and this brings us to the theoretical question.

Is a correlation of 0.10 or 0.15 between social media use and mental health problems "small"? If so, is it too small to warrant any action by parents or policymakers? Many social psychologists have long followed the rule of thumb proposed by Cohen (1988), who suggested that we refer to correlations of 0.10 and above as "small," and that we not consider a correlation to be "moderate" until it reaches 0.30, and not "large" until it reaches 0.50. These standards have been harshly criticized in recent years, however, because the fields of personality and social psychology, along with public health and epidemiology, rarely

produce reliable correlations above 0.30 (see Funder & Ozer, 2019). Human beings and societies are complicated; behavior has multiple causes, and small effects (around $r = 0.10$, or smaller) are not just ubiquitous, they are "the indispensable foundation for a cumulative psychological science," as Gotz et al. (2022) put it in the title of their article.

Public health interventions that correlate with their intended outcome at $r = 0.10$ have an enormous impact when applied over a large population, cumulatively over many years. Both Funder and Ozer (2019) and Gotz et al. (2022) recommend interpreting correlations not in relation to an arbitrary universal standard but by benchmarking them against other known relationships in the same domain, and they offer many in the ballpark of $r = 0.10$, including calcium intake and bone mass in premenopausal women ($r = 0.08$). As another example, the association between childhood lead exposure and adult IQ is $r = -0.09$ (Reuben et al., 2017). Despite these "small" associations, the public health benefits of calcium supplementation and lead removal are quite large for the subset of the population that is at highest risk. We believe the case is at least as strong for social media exposure ($r = 0.20$ for girls) and its cumulative impact across tens of millions of girls over the decade of their adolescence.

Our findings therefore contradict the advice of one article covering O&P's results which said that parents "can mostly stop worrying about screen time now" (Stillman, 2019). If policy makers and parents are going to stop worrying about all forms of screen time on the basis of small effect sizes, they should also stop worrying about binge drinking, heroin use, marijuana use, and many of the other activities listed in Figs. 3, 4, and 5. However, our results do show that parents concerned about mental health outcomes can potentially worry less about how much time their children spend watching TV and playing video games. They may also be able to worry less about their sons than about their daughters, at least for the activities analyzed in these studies.

We acknowledge that the specific social media platform and the way a teen uses a platform may matter more than the raw number of hours of use per week. We acknowledge that many factors other than effect size are involved in whether an association is practically important. We assume that if we had much better measurement of what teens are doing online, and much better measurement of their mental health outcomes, we would find some correlations that were substantially larger than those we report here. Our point is simply that social media effects should not be dismissed on the basis of the median effect size based on all screen time reported by Orben and Przybylski (2019a, 2019b).

Our results have a more general methodological implication: they demonstrate the limitations of SCA in practice. No analysis can feasibly include, as Orben and Przybylski (2019a, 2019b) claimed in their article, "every possible analytical pathway" and "all results of all theoretically defensible analyses" (p. 174). Like all researchers, O&P made analytical decisions when conducting their SCAs. The findings in this study demonstrate that those decisions can influence the conclusions drawn from the same data. SCA is not a judgement-free analytic approach and requires expert judgement about the range of specifications to be evaluated (Simonsohn et al., 2020). Reviewers and editors must evaluate the justifications for these choices. For example, we found that the analytical decision that was among the most powerful in shrinking effects was O&P's decision in two of the studies to include potential mediators as controls. We argued that items such as "How often do you feel unhappy at school?" should not be treated like race and gender and included in the list of variables to be controlled. We also highlighted how one scale could constitute 73% of the SCA specification space when there are four scales measuring mental health. Simonsohn et al. (2020) have recently suggested that the weighting of the individual models in the specification space could be appropriate. This would also require researchers to apply judgement to determine the appropriate weights. We argue that weighting one scale as 73% of the data is not appropriate or logical, but leave the research community to decide.

Simonsohn et al. (2020) also point out that the specification space is likely to underrepresent all of the valid models that could have been included. In both O&P and our analyses, the models were all first order. It is very likely that many of the associations are curvilinear or threshold structured. Inclusion of higher order functional associations, and interactions, would have greatly increased the number of possible specifications leaving the current subset of analyses as small subset of the bigger specification space. If the idea of SCA is to withhold developing conclusions from an epidemiological study sample until all (or most) of the specifications have been aggregated, neither ours nor O&P's conclusions are valid. Further, deference to the median Beta from a large set of possible valid models still remains to be adequately justified as a method for drawing conclusions. Semken and Rossell (2022) recently found that O&P's SCA obscured important effects by relying on the median Beta rather than separating effects for different media activities, self-report and parent-report scales, and boys versus girls. They concluded, as we do, that these datasets contain substantial associations between digital media use and mental health. Where the research community sees value in SCA analyses, evaluation of the decisions used in the SCA presented here and by O&P will determine what conclusions to draw from the differing analyses of the same data.

Given the magnitude of the association between digital media use

and compromised mental health that we found in all three large datasets we analyzed, it seems plausible that increases in digital media use might be responsible for the substantial increases in adolescent depression and anxiety that began around 2012. However, as these increases in mental health issues took place at the population level, future research should explore how social media changed the collective dynamics of social interaction. When most adolescents opened social media accounts and became daily users in the years around 2012, teen social life changed even for adolescents who spent no or little time on social media. We hope that future researchers will find additional ways to investigate the question with which we opened this article: Why did rates of depression, anxiety, self-harm, and suicide attempts begin rising among adolescents around 2012? For now, social media should remain on the list of possible explanations.

Declaration of competing interest

J. M. Twenge and J. Haidt have received honoraria for presenting research and receive royalties from books. The other authors have no declarations of interest.

References

- Baron, J. (2019, January 14). *Screen time may be no worse for kids than eating potatoes*. Forbes.
- Boers, E., Afzali, M. H., & Newton, N. (2019). Association of screen time and depression in adolescence. *JAMA Pediatrics*, *173*, 853–859.
- Burstein, B., Agostino, H., & Greenfield, B. (2019). Suicidal attempts and ideation among children and adolescents in US emergency departments, 2007–2015. *JAMA Pediatrics*, *173*, 598–600.
- Cohen, J. (1988). *Statistical power in the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cybulski, L., Ashcroft, D. M., Carr, M. J., Garg, S., Chew-Graham, C. A., Kapur, N., & Webb, R. T. (2021). Temporal trends in annual incidence rates for psychiatric disorders and self-harm among children and adolescents in the UK, 2003–2018. *BMC Psychiatry*, *21*.
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*, 156–168.
- Gotz, F. M., Gosling, S. D., & Rentfrow. (2022). Small effects: The indispensable foundation for a cumulative psychological science. *Perspectives on Psychological Science*, *17*, 205–215.
- Haidt, J., & Rose-Stockwell, T. (2019, December 15). Social media is warping democracy. *The Atlantic*.
- Haidt, J., & Twenge, J. M. (2021). Is there an increase in adolescent mood disorders, self-harm, and suicide since 2010 in the USA and UK? A review. in a public Google document linked at. In *Unpublished ongoing literature review* <https://www.theoddl.com/better-mental-health>.
- Hunt, M. G., Marx, R., Lipson, C., & Young, J. (2018). No more FOMO: Limiting social media decreases loneliness and depression. *Journal of Social and Clinical Psychology*, *37*, 751–768.
- Kalb, L. G., Stapp, E. K., Ballard, E. D., Hologue, C., Keefer, A., & Riley, A. (2019). Trends in psychiatric emergency department visits among youth and young adults in the US. *Pediatrics*, *143*(4).
- Kelly, Y., Zilanawala, A., Booker, C., & Sacker, A. (2019). Social media use and adolescent mental health: Findings from the UK millennium cohort study. *Lancet Eclinical Medicine*, *6*, 59–68.
- Keyes, K. M., Gary, D., O'Malley, P. M., Hamilton, A., & Schulenberg, J. (2019). Recent increases in depressive symptoms among U.S. adolescents: Trends from 1991 to 2018. *Social Psychiatry and Psychiatric Epidemiology*, *54*, 987–996.
- Lin, L. Y., Sidani, J. E., Shensa, A., Radovic, A., Miller, E., Colditz, J. B., ... Primack, B. A. (2016). Association between social media use and depression among U.S. young adults. *Depression and Anxiety*, *33*, 323–331.
- Luby, J., & Kertz, S. (2019). Increasing suicide rates in early adolescent girls in the United States and the equalization of sex disparity in suicide: The need to investigate the role of social media. *JAMA Network Open*, *2*(5), Article e193916.
- Marquez, J., & Long, E. (2021). A global decline in adolescents' subjective well-being: A comparative study exploring patterns of change in the life satisfaction of 15-year-old students in 46 countries. *Child Indicators Research*, *24*, 1251–1292.
- Mercado, M. C., Holland, K., & Leemis, R. W. (2017). Trends in emergency department visits for nonfatal self-inflicted injuries among youth aged 10 to 24 years in the United States, 2001–2015. *Journal of the American Medical Association*, *318*, 1931–1933.
- Mojtabai, R., Olfson, M., & Han, B. (2016). National trends in the prevalence and treatment of depression in adolescents and young adults. *Pediatrics*, *138*(6).
- Morgan, C., Webb, R. T., Carr, M. J., Kontopantelis, E., Green, J., Chew-Graham, C. A., & Ashcroft, D. M. (2017). Incidence, clinical management, and mortality risk following self-harm among children and adolescents: Cohort study in primary care. *BMJ*, *359*, Article j4351.

- Orben, A. (2020). Teenagers, screens and social media: A narrative review of reviews and key studies. In *Social psychiatry and psychiatric epidemiology: The International Journal for Research in Social and Genetic Epidemiology and Mental Health Services*.
- Ogders, C. L., & Jensen, M. R. (2020). Adolescent mental health in the digital age: Facts, fears, and future directions. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 61, 336–348.
- Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behavior*, 3, 173–182.
- Orben, A., & Przybylski, A. K. (2019b). *Screens, teens, and psychological well-being: Evidence from three time-use-diary studies*. Psychological Science.
- Orben, A., & Przybylski, A. K. (2020). Reply to: Underestimating digital media harm. *Nature Human Behaviour*, 4, 349–351.
- Patalay, P., & Gage, S. H. (2019). Changes in millennial adolescent mental health and health-related behaviours over 10 years: A population cohort comparison study. *International Journal of Epidemiology*, 48, 1650–1664.
- Pew Research Center. (2021). *Mobile fact sheet (April 7)* <https://www.pewresearch.org/internet/fact-sheet/mobile/>.
- Plemmons, G., Hall, M., Douplik, S., Gay, J., Brown, C., Browning, W., & Williams, D. (2018). Hospitalization for suicide ideation or attempt: 2008-2015. *Pediatrics*, 141(6).
- Radtke, T., Apel, T., Schenkel, K., Keller, J., & von Lindern, E. (2021). *Digital detox: An effective solution in the smartphone era? A systematic literature review*. Mobile Media & Communication.
- Reuben, A., Caspi, A., Belsky, D. W., Broadbent, J., Harrington, H., Sugden, K., Houts, R. M., Ramrakha, S., Poulton, R., & Moffitt, T. E. (2017). Association of childhood blood lead levels with cognitive function and socioeconomic status at age 38 years and with IQ change and socioeconomic mobility between childhood and adulthood. *JAMA*, 317, 1244–1251.
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1, 27–42.
- Ruch, D. A., Sheftall, A. H., Schlagbaum, P., Rausch, J., Campo, J. V., & Bridge, J. A. (2019). Trends in suicide among youth aged 10 to 19 years in the United States, 1975 to 2016. *JAMA Network Open*, 2(5), Article e193886.
- Schisterman, E. F., Cole, S. R., & Platt, R. W. (2009). Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology*, 20, 488–495.
- Semken, C., & Rossell, D. (2022). Specification analysis for technology use and teenager well-being: Statistical validity and a Bayesian proposal. *ArXiv.org*.
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 4(11), 1208–1214.
- Spiller, H. A., Ackerman, J. P., Spiller, N. E., & Casavant, M. J. (2019). Sex- and age-specific increases in suicide attempts by self-poisoning in the United States among youth and young adults from 2000 to 2018. *Journal of Pediatrics*, 210, 201–208.
- Stillman, J. (2019, January). *Calm down, parents: A rigorous new Oxford study of 350,000 teens shows screen time is about as dangerous as potatoes*. Inc.com.
- Tromholt, M. (2016). The Facebook experiment: Quitting Facebook leads to higher levels of well-being. *Cyberpsychology, Behavior, and Social Networking*, 19, 661–666.
- Twenge, J. M. (2020). Why increases in adolescent depression may be linked to the technological environment. *Current Opinion in Psychology*, 32, 89–94.
- Twenge, J. M., & Farley, E. (2021). Not all screen time is created equal: Associations with mental health vary by activity and gender. *Social Psychiatry and Psychiatric Epidemiology*, 56, 207–217.
- Twenge, J. M., & Martin, G. N. (2020). Gender differences in associations between digital media use and psychological well-being: Evidence from three large datasets. *Journal of Adolescence*, 79, 91–102.
- Twenge, J. M., Cooper, A. B., Joiner, T. E., Duffy, M. E., & Binau, S. G. (2019). Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017. *Journal of Abnormal Psychology*, 128, 185–199.
- Twenge, J. M., Martin, G. N., & Spitzberg, B. H. (2019). Trends in U.S. Adolescents' media use, 1976–2016: The rise of digital media, the decline of TV, and the (near) demise of print. *Psychology of Popular Media Culture*, 8, 329–345.
- Twenge, J. M., Haidt, J., Joiner, T. E., & Campbell, W. K. (2020). Underestimating digital media harm. *Nature Human Behavior*, 4, 346–348.
- Twenge, J. M., Haidt, J., Blake, A. B., McAllister, C., Lemon, H., & LeRoy, A. (2021). Worldwide increases in adolescent loneliness. *Journal of Adolescence*, 93, 257–269.