nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-018-0506-1

The association between adolescent well-being and digital technology use

Amy Orben[1*] and Andrew K. Przybylski[1,2]

The widespread use of digital technologies by young people has spurred speculation that their regular use negatively impacts psychological well-being. Current empirical evidence supporting this idea is largely based on secondary analyses of large-scale social datasets. Though these datasets provide a valuable resource for highly powered investigations, their many variables and observations are often explored with an analytical flexibility that marks small effects as statistically significant, thereby leading to potential false positives and conflicting results. Here we address these methodological challenges by applying specification curve analysis (SCA) across three large-scale social datasets (total $n = 355{,}358$) to rigorously examine correlational evidence for the effects of digital technology on adolescents. The association we find between digital technology use and adolescent well-being is negative but small, explaining at most 0.4% of the variation in well-being. Taking the broader context of the data into account suggests that these effects are too small to warrant policy change.

The idea that digital devices and the Internet have an enduring influence on how humans develop, socialize and thrive is a compelling one[1]. As the time spent by young people online has doubled in the past decade[2], the debate about whether this shift negatively impacts children and adolescents is becoming increasingly heated[3]. A number of professional and governmental organizations have therefore called for more research into digital screen-time[4,5], which has led to household panel surveys[6,7] and large-scale social datasets adding measures of digital technology use to those already assessing psychological well-being[8]. Unfortunately, findings derived from the cross-sectional analysis of these datasets are conflicting; in some cases negative associations between digital technology use and well-being are found[9,10], often receiving much attention even when correlations are small. Yet other results are mixed[11] or contest previously discovered negative effects when re-analysing identical data[12]. One high-quality, pre-registered analysis of UK adolescents found that moderate digital engagement does not correlate with well-being, but very high levels of usage possibly have small negative associations[13,14].

There are at least three reasons why the inferences drawn by behavioural scientists from large-scale datasets might produce divergent findings. First, these datasets are mostly collected in collaboration with multidisciplinary research councils and are characterized by a battery of items meant to be completed by postal survey, face-to-face or telephone interview[6–8]. Though research councils engage in public consultations[15], the pre-tested or validated scales common in clinical, social or personality psychology are often abbreviated or altered to reduce participant burden[16,17]. Scientists wishing to make inferences about the effects of digital technology using these data need to make numerous decisions about how to analyse, combine and interpret the measures. Taking advantage of these valuable datasets is therefore fraught with many subjective analytical decisions, which can lead to high numbers of researcher degrees of freedom[18]. With nearly all decisions taken after the data are known, these are not apparent to those reading the published paper highlighting only the final analytical pathway[19,20].

The second possible explanation for conflicting patterns of effects found in large-scale datasets is rooted in the scale of the data analysed. Compared to the laboratory- and community-based samples typical of behavioural research (mostly $<1{,}000$)[21], large-scale social datasets feature high numbers of participant observations (ranging from 5,000 to 5,000,000)[6–8]. This means that very small co-variations (for example, $r < 0.01$) between self-report items will result in compelling evidence for rejecting the null hypothesis at alpha-levels typically interpreted as statistically significant by behavioural scientists (that is, $P < 0.05$). Thirdly, it is important to note that most datasets are cross-sectional and therefore provide only correlational evidence, making it difficult to pinpoint causes and effects. Thus, large-scale datasets are simultaneously attractive and problematic for researchers, peer reviewers and the public. They are a resource for testing behavioural theories at scale but are, at the same time, inherently susceptive to false positives and significant but minute effects using the alpha-levels traditionally employed in behavioural science.

Given that digital technology's impact on child well-being is a topic of widespread scientific debate among those studying human behaviour[1] and has real-world implications[4], it is important for researchers to make the most of existing large-scale dataset investments. This makes it necessary to employ transparent and robust analytical practices which recognize that the measures of digital technology use and well-being in large-scale datasets may not be well matched to specific research questions. Furthermore, behavioural scientists must be transparent about how the hundreds of variables and many thousands of observations can quickly branch out into 'gardens of forking paths'[19] with millions, and in some cases billions, of analysis options. This risk is compounded by a reliance on statistical significance, that is using $P < 0.05$ to demarcate 'true' effects. Unfortunately the large number of participants in these designs means that small effects are easily publishable and, if positive, garner outsized press and policy attention[12].

Given that large-scale secondary datasets are increasingly available freely online, it is not possible to convincingly document a scientist's ignorance of the data before analysis[22–24], making hypothesis

[1]Department of Experimental Psychology, University of Oxford, Oxford, UK. [2]Oxford Internet Institute, University of Oxford, Oxford, UK.
*e-mail: amy.orben@psy.ox.ac.uk

Table 1 | Possible specifications (analytical decisions) used to test a simple linear regression between technology use and adolescent well-being in the datasets YRBS, MTF and MCS

Decision	YRBS	MTF	MCS
Operationalizing adolescent well-being	Mean of any possible combination of five items concerning mental health and suicidal ideation	Mean of any possible combination of 13 items concerning depression, happiness and self-esteem	Mean of any possible combination of 24 questions concerning well-being, self-esteem and feelings (cohort members), or mean of any possible combination of 25 questions from the Strengths and Difficulties Questionnaire (caregivers)
Operationalizing technology use	Two questions concerning electronic device use and TV use, or the mean of these questions	Eleven technology use measures concerning the Internet, electronic games, mobile phone use, social media use and computer use, or the mean of these questions	Five questions concerning TV use, electronic games, social media use, owning a computer and using the Internet at home, or the mean of these questions
Which co-variates to include	Either include co-variates or not	Either include co-variates or not	Either include co-variates or not
Other specifications	Either take mean of dichotomous well-being measures, or code all cohort members who answered 'yes' to one or more as 1 and all others as 0		Use well-being measures declared by cohort members or those declared by their caregivers

pre-registration untenable as a general solution to the problem of subjective analytical decisions. In this article we argue that specification curve analysis[25] provides a promising alternative. Briefly, SCA is a tool for mapping the sum of theory-driven analytical decisions that could justifiably have been taken when analysing quantitative data. Researchers demarcate every possible analytical pathway and then calculate the results of each. Rather than reporting a handful of analyses in their paper, they report all results of all theoretically defensible analyses (for previous examples see[25,26] and the Supplementary Methods).

Given the substantial disagreements within the literature, the extent to which children's screen-time may actually be impacting their psychological well-being remains unclear. The present research addresses this gap in our understanding by relying on large-scale data paired with a conservative analytic approach to provide a more definitive and clearly contextualized test of the association between screen use and well-being.

To this end, three large-scale exemplar datasets—Monitoring the Future (MTF), Youth Risk and Behaviour Survey (YRBS) and Millennium Cohort Study (MCS) from the United States of America (MTF, YRBS) and the United Kingdom (MCS)—were selected to highlight the particular strengths and weaknesses of drawing general inferences from large-scale social data and how these can be reconceptualized by SCA[6–8]. Furthermore, we tackle the problem of significant-but-minimal effects in large-scale social data by using the abundance of questions in each dataset to compute comparison specifications; we directly compare the effects of digital technology to those of other activities on psychological well-being (for example, sleep, eating breakfast, illicit drug use), using extant literatures and psychological theory as a guide. This allows us to simultaneously examine the impact of adolescent technology use against real-world benchmarks while modelling and accounting for analytical flexibility.

Results

Identifying specifications. We identified the main analytical decisions that needed to be taken when regressing digital technology use on adolescents' psychological well-being in each dataset (see Table 1). Three hundred and seventy-two justifiable specifications for the YRBS, 40,966 plausible specifications for the MTF and a total of 603,979,752 defensible specifications for the MCS were identified. Although more than 600 million specifications might seem high, this number is best understood in relation to the total possible iterations of dependent (six analysis options) and independent variables ($2^{24}+2^{25}-2$ analysis options) and whether co-variates are included (two analysis options). The number rises even higher, to 2.5 trillion specifications, for the MCS if any combination of co-variates (2^{12} analysis options) is included. Given this, and to reduce computational time, we selected 20,004 specifications for the MCS. To do so, we included specifications of all used measures per se, and any combinations of measures found in the previous literature, and then supplemented these with other randomly selected combinations. More information about selection can be found in the Supplementary material (see Supplementary Table 1).

Implementing specifications. After noting all specifications, the result of every possible combination of these specifications was computed for each dataset. The standardized β-coefficient for the association of technology use with well-being was then plotted for each specification. The number of participants analysed for each specification can be found in Supplementary Figs. 2–4, while the median standardized β, n, partial η^2 and standard error can be found in Table 2. For the YRBS, the median association of technology use with adolescent well-being was $\beta=-0.035$ (median partial $\eta^2=0.001$, median $n=62{,}297$, median standard error $=0.004$; see Fig. 1). From this figure one can discern the analytical choices that influence the size of this effect. When employing electronic device use as the independent variable in the model, the effects were more negative (median $\beta=-0.071$, median partial $\eta^2=0.005$, median $n=62{,}368$, median standard error $=0.004$); when including TV use in the model the effects were less negative and sometimes became non-significant (median $\beta=-0.012$, median partial $\eta^2<0.001$, median $n=62{,}352$, median standard error $=0.004$). Even though the YRBS does not have high-quality control variables, inclusion of these yielded a smaller effect size for the relations of interest (controls: median $\beta=-0.034$, median partial $\eta^2=0.001$, median $n=61{,}525$, median standard error $=0.004$; no controls: median $\beta=-0.035$, median partial $\eta^2=0.001$, median $n=62{,}638$, median standard error $=0.004$).

For the MTF data, a median standardized β value of -0.005 was observed (median partial $\eta^2<0.001$, median $n=78{,}267$, median standard error $=0.003$), a value which fell within the non-significant

Table 2 | Results of SCA for the YRBS, MTF and MCS, both overall and for different technology use variables, parent/adolescent self-report or with/without control variables

Dataset	Median β of SCA	Median partial η^2 of SCA	Median n	s.e.m.
YRBS				
Complete SCA	−0.035	0.001	62,297	0.004
Electronic device use only	−0.071	0.005	62,368	0.004
TV use only	−0.012	<0.001	62,352	0.004
With control variables only	−0.034	0.001	61,525	0.004
Without control variables only	−0.035	0.001	62,638	0.004
MTF				
Complete SCA	−0.005	<0.001	78,267	0.003
Social media use only	−0.031	0.001	102,963	0.003
TV viewing on weekend only	0.008	0.001	115,738	0.003
Using Internet for news only	−0.002	<0.001	115,580	0.003
TV viewing on weekday only	0.002	<0.001	115,783	0.003
With control variables only	0.001	<0.001	72,525	0.003
Without control variables only	−0.013	<0.001	117,560	0.003
MCS				
Complete SCA	−0.032	0.004	7,968	0.010
Own a computer only	−0.003	0.011	7,973	0.010
Weekday electronic games only	0.013	<0.001	7,977	0.010
Hours of social media use only	−0.056	0.009	7,972	0.010
TV viewing on weekday only	−0.043	0.003	7,971	0.010
Use of home Internet only	−0.070	0.006	7,975	0.010
Parent-report well-being only	<0.001	0.003	7,893	0.010
Adolescent-report well-being only	−0.046	0.008	8,857	0.010
With control variables only	−0.005	0.001	6,566	0.011
Without control variables only	−0.068	0.005	11,018	0.010

range of the justifiable specifications (see Fig. 2). This result was surprising, as the MTF had the highest number of observations, making it difficult for even small associations to be flagged as non-significant using traditional alpha-thresholds (that is, $P<0.05$). In Fig. 2, and in our bootstrapping test, we do not include the few specifications of the participants who declared only one well-being measure (for the SCA of all participants, see Supplementary Fig. 5). From the graph it is again possible to discern that even controls of lower standard made the association either less negative or even positive (no controls: median $\beta=-0.013$, median partial $\eta^2<0.001$, median $n=117{,}560$, median standard error $=0.003$; controls: median $\beta=0.001$, median partial $\eta^2<0.001$, median $n=72{,}525$, median standard error $=0.003$). TV viewing at the weekend only had a median positive association with well-being of $\beta=0.008$ (median partial $\eta^2=0.001$, median $n=115{,}738$, median standard error $=0.003$), while social media use had a median negative association with well-being of $\beta=-0.031$ (median partial $\eta^2=0.001$, median $n=102{,}963$, median standard error $=0.003$) although the effect was small, suggesting that technology use operationalized in these terms accounts for less than 0.1% of the observed variability in well-being. Using the Internet for news, and TV viewing on a weekday only, showed mainly very small median associations, with $\beta=-0.002$ (median partial $\eta^2<0.001$, median $n=115{,}580$, median standard error $=0.003$) and $\beta=0.002$ (median partial $\eta^2<0.001$, median $n=115{,}783$, median standard error $=0.003$), respectively. Because previous studies have addressed the association between technology use and well-being using the same dataset[10], in the Supplementary material we include a figure (Supplementary Fig. 6) showing how the specifications of these studies influence their reported results.

Lastly, results from the MCS, the highest-quality dataset we examined, were interesting because the literature provided us with control variables based on extant theory[11] and convergent data from adolescent and caregiver reports. In these data we found a median β value for the association of technology use with well-being of $\beta=-0.032$ (median partial $\eta^2=0.004$, median $n=7{,}968$, median standard error $=0.010$; see Fig. 3). Across the board, if using well-being measures completed by the caregivers, the median association was less negative or more positive (median $\beta<0.001$, median partial $\eta^2=0.003$, median $n=7{,}893$, median standard error $=0.010$), while the opposite was in evidence when considering well-being measures completed by the cohort member (median $\beta=-0.046$, median partial $\eta^2=0.008$, median $n=8{,}857$, median standard error $=0.010$). This pattern of shared co-variation speaks to the idea that correlations between technology use and well-being might be rooted in common method variance, as one single informant declares well-being and technology measures and the association might be driven by other common factors.

To further address the importance of control variables, we plot separate specification curves for MCS analyses with and without controls (see Fig. 4). The association for the uncorrected models had a median β value of −0.068 (median partial $\eta^2=0.005$, median $n=11{,}018$, median standard error $=0.010$). In contrast, the corrected models found a median β value for technology use regressed on well-being of only −0.005 (median partial $\eta^2=0.001$, median $n=6{,}566$, median standard error $=0.011$). Additional SCAs using only pre-specified questionnaires are presented in Supplementary Fig. 7, while further visualizations about how the addition of controls and parent reports affects the reported associations are presented in Supplementary Figs. 8 and 9.

Statistical inferences. The SCAs showed that there is a small negative association between technology use and well-being, but it is not possible to make many analytical statistical inferences because the specifications are not part of the same model and are not independent. A bootstrapping technique was therefore used to run 500 SCA tests on resampled data, where it is known that the null hypothesis is true. Results presented in Supplementary Table 2 indicate that the effects found were highly significant for all three datasets, and all three measures of significance included in our bootstrapped tests. For the three datasets there was no SCA analysis of bootstrapped samples, which resulted in a larger median effect size than that of the original SCA ($P=0.00$, original effect sizes: YRBS median $\beta=-0.035$, MTF median $\beta=-0.005$, MCS median $\beta=-0.032$). Furthermore, there was no bootstrapped SCA with more total or statistically significant specifications of the dominant sign than the original SCA (share of specifications with dominant sign. $P=0.00$; original number: YRBS = 356, MTF = 24,164, MCS = 12,481; share of statistically significant specifications with dominant sign, $P=0.00$; original number: YRBS = 323, MTF = 19,649, MCS = 10,857). This result provides evidence that digital technology use and adolescent well-being could be negatively related at above-chance levels in our data.

Comparison specifications. To put the results of the SCAs into perspective with respect to the broader context of human behaviour as measured in these datasets, we compared specification curves for the mean of the technology use variables in each dataset to other associations that have been shown to relate, or are hypothesized not to relate, to adolescent mental health: binge-drinking, smoking marijuana, being bullied, getting into fights, smoking cigarettes,

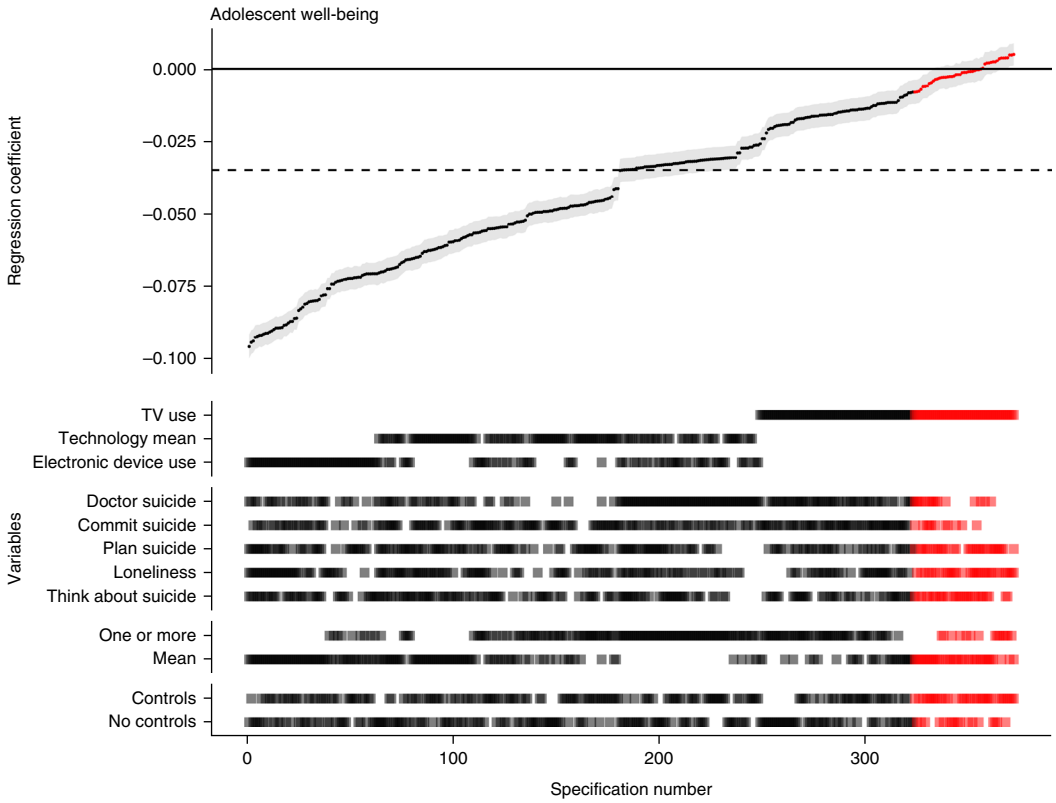

Fig. 1 | Results of SCA of the YBRS. Specification curve analysis showing the range of possible results for a simple cross-sectional regression of digital technology use on adolescent well-being. Each point on the *x* axis represents a different combination of analytical decisions, which are displayed in the 'dashboard' at the bottom of the graph. The resulting standardized regression coefficient is shown at the top of the graph; the error bars visualize the standard error. Red represents non-significant outcomes, while black represents significant outcomes. To ease interpretation, the dotted line indicates the median standardized regression coefficient found in the SCA: $\beta = -0.035$ (median partial $\eta^2 = 0.001$, median $n = 62{,}297$, median standard error = 0.004).

being arrested, perceived weight, eating potatoes, having asthma, drinking milk, going to the movies, religion, listening to music, doing homework, cycling, height, wearing glasses, handedness, eating fruit, eating vegetables, getting enough sleep and eating breakfast. For results see Table 3, Fig. 5 and Supplementary Figs. 10–12.

For the YRBS the association of mean technology use with well-being (median $\beta = -0.049$, median $n = 62{,}166$, partial $\eta^2 = 0.002$, median standard error = 0.004) was exceeded by the association of well-being with being bullied (median $\beta = -0.212$, median $n = 50{,}066$, partial $\eta^2 = 0.044$, median standard error = 0.004), getting into fights (median $\beta = -0.179$, median $n = 62{,}106$, partial $\eta^2 = 0.031$, median standard error = 0.004), binge-drinking (median $\beta = -0.144$, median $n = 62{,}010$, partial $\eta^2 = 0.021$, median standard error = 0.004), smoking marijuana (median $\beta = -0.132$, median $n = 62{,}361$, partial $\eta^2 = 0.018$, median standard error = 0.004), having asthma (median $\beta = -0.066$, median $n = 60{,}863$, partial $\eta^2 = 0.004$, median standard error = 0.004) and perceived weight (median $\beta = -0.050$, median $n = 62{,}752$, partial $\eta^2 = 0.002$, median standard error = 0.004). There is a smaller negative association for eating potatoes (median $\beta = -0.042$, median $n = 61{,}912$, partial $\eta^2 = 0.002$, median standard error = 0.004), eating vegetables (median $\beta = -0.013$, median $n = 62{,}034$, partial $\eta^2 < 0.001$, median standard error = 0.004) and eating fruit (median $\beta = -0.005$, median $n = 62{,}436$, partial $\eta^2 < 0.001$, median standard error = 0.004). There is a smaller positive association for drinking milk (median $\beta = 0.014$, median $n = 60{,}021$, partial $\eta^2 < 0.001$, median standard error = 0.004). Lastly, there is a larger positive association for eating breakfast (median $\beta = 0.116$, median $n = 34{,}010$, partial $\eta^2 = 0.013$, median standard error = 0.006) and getting enough sleep (median $\beta = 0.150$, median $n = 56{,}552$, partial $\eta^2 = 0.022$, median standard error = 0.004).

For the MTF we compared the association of mean technology use (median $\beta = -0.006$, median $n = 102{,}186$, partial $\eta^2 < 0.001$, median standard error = 0.003) to the variables we hypothesized a priori as having no association: going to the movies (median $\beta = 0.064$, median $n = 115{,}943$, partial $\eta^2 = 0.005$, median standard error = 0.003), time spent on homework (median $\beta = 0.020$, median $n = 115{,}225$, partial $\eta^2 = 0.001$, median standard error = 0.003), attending religious services (median $\beta = 0.091$, median $n = 89{,}453$, partial $\eta^2 = 0.010$, median standard error = 0.003) and listening to music (median $\beta = -0.182$, median $n = 49{,}514$, partial $\eta^2 = 0.035$, median standard error = 0.005) all had larger effects. We also examined those we hypothesized as having a more positive association: eating breakfast (median $\beta = 0.170$, median $n = 62{,}330$, partial $\eta^2 = 0.034$, median standard error = 0.004), eating fruit (median $\beta = 0.053$, median $n = 115{,}334$, partial $\eta^2 = 0.003$, median standard error = 0.003), sleep (median $\beta = 0.246$, median $n = 61{,}903$, partial $\eta^2 = 0.070$, median standard error = 0.004) and eating vegetables (median $\beta = 0.115$, median $n = 62{,}072$, partial $\eta^2 = 0.014$, median standard error = 0.004). Lastly we looked at those variables that we hypothesized as having a more negative association: binge-drinking (median $\beta = -0.045$, median $n = 107{,}994$, partial $\eta^2 = 0.002$, median standard error = 0.003), fighting (median $\beta = -0.087$, median $n = 62{,}683$, partial $\eta^2 = 0.008$, median standard error = 0.004), smoking marijuana (median $\beta = -0.056$, median $n = 113{,}611$, partial $\eta^2 = 0.003$, median standard error = 0.003) and smoking cigarettes (median $\beta = -0.103$, median $n = 113{,}424$, partial $\eta^2 = 0.012$, median standard error = 0.003).

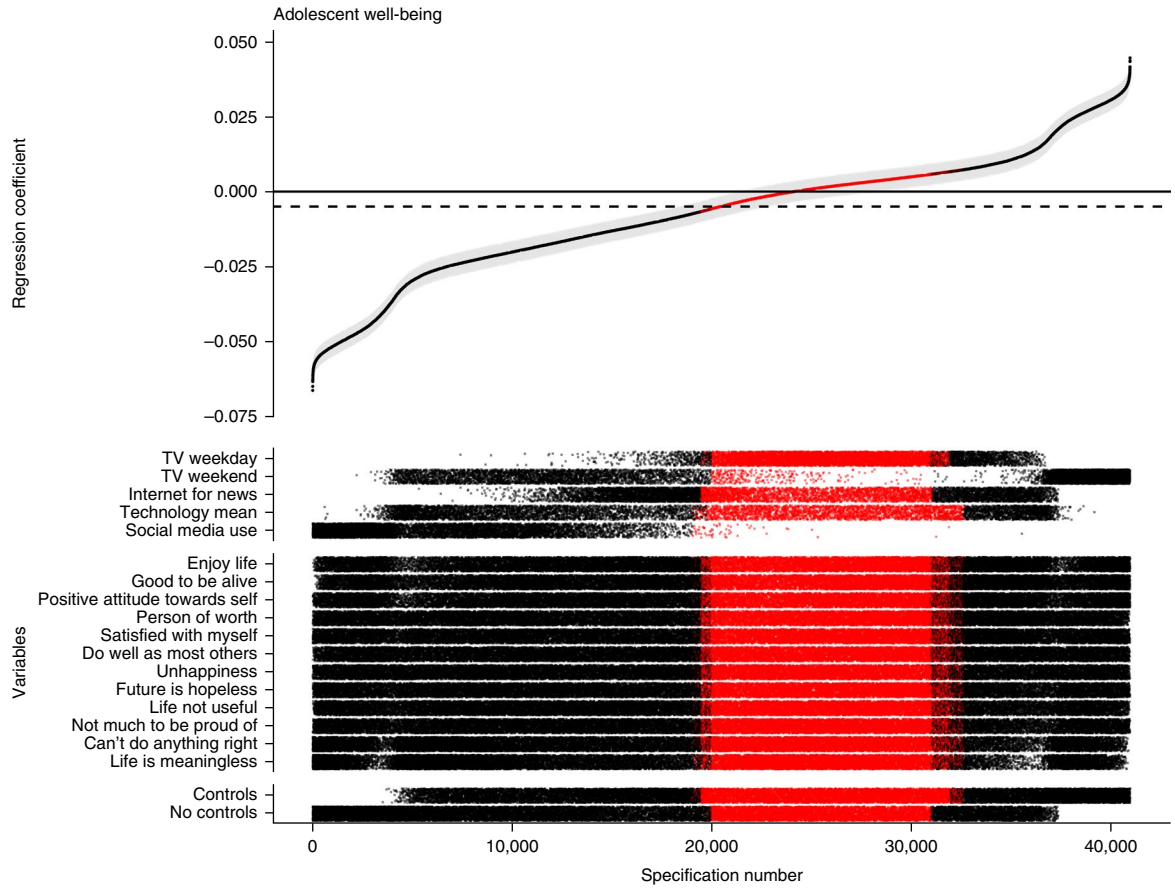

Fig. 2 | Results of SCA for MTF. Specification curve analysis showing the range of possible results for a simple cross-sectional regression of digital technology use on adolescent well-being. Each point on the *x* axis represents a different combination of analytical decisions, which are displayed in the 'dashboard' at the bottom of the graph. The resulting standardized regression coefficient is shown at the top of the graph; the error bars visualize the standard error. Red represents non-significant outcomes while black represents significant outcomes. To ease interpretation, the dotted line indicates the median standardized regression coefficient found in the SCA: $\beta = -0.005$ (partial $\eta^2 < 0.001$, median $n = 78{,}267$, median standard error = 0.003).

For the MCS, mean technology use (median $\beta = -0.042$, median $n = 7{,}964$, partial $\eta^2 = 0.002$, median standard error = 0.010) was compared to amount of sleep (median $\beta = 0.070$, median $n = 7{,}954$, partial $\eta^2 = 0.005$, median standard error = 0.010), eating fruit (median $\beta = 0.056$, median $n = 7{,}960$, partial $\eta^2 = 0.004$, median standard error = 0.010), eating breakfast (median $\beta = 0.140$, median $n = 7{,}964$, partial $\eta^2 = 0.025$, median standard error = 0.010) and eating vegetables (median $\beta = 0.064$, median $n = 7{,}949$, partial $\eta^2 = 0.005$, median standard error = 0.010) that have a priori hypothesized positive associations; being arrested (median $\beta = -0.041$, median $n = 7{,}908$, partial $\eta^2 = 0.002$, median standard error = 0.011), being bullied (median $\beta = -0.208$, median $n = 7{,}898$, partial $\eta^2 = 0.048$, median standard error = 0.010), binge-drinking (median $\beta = -0.043$, median $n = 3{,}656$, partial $\eta^2 = 0.002$, median standard error = 0.015) and smoking marijuana (median $\beta = -0.048$, median $n = 7{,}903$, partial $\eta^2 = 0.003$, median standard error = 0.010) that have a priori hypothesized negative associations; wearing glasses (median $\beta = -0.061$, median $n = 7{,}963$, partial $\eta^2 = 0.005$, median standard error = 0.010), being left-handed (median $\beta = -0.004$, median $n = 7{,}972$, partial $\eta^2 < 0.001$, median standard error = 0.010), bicycle use (median $\beta = 0.080$, median $n = 7{,}974$, partial $\eta^2 = 0.007$, median standard error = 0.010) and height (median $\beta = 0.065$, median $n = 7{,}910$, partial $\eta^2 = 0.005$, median standard error = 0.010) that have no a priori hypothesized associations (Fig. 5).

Discussion

The possibility that the use of digital technology by adolescents has a negative impact on psychological well-being is an important question worthy of rigorous empirical testing. While previous research in this area has equated findings derived from large-scale social data with empirical robustness, the present research highlights deep-seated problems associated with drawing strong inferences from such analyses. To provide a robust and transparent investigation of the effect of digital technology use on adolescent well-being, we implemented SCA with comparison specifications using three large-scale datasets from the United States of America and the United Kingdom.

While we find that digital technology use has a small negative association with adolescent well-being, this finding is best understood in terms of other human behaviours captured in these large-scale social datasets. When viewed in the broader context of the data, it becomes clear that the outsized weight given to digital screen-time in scientific and public discourse might not be merited on the basis of the available evidence. For example, in all three datasets the effects of both smoking marijuana and bullying have much larger negative associations with adolescent well-being (×2.7 and × 4.3, respectively for the YRBS) than does technology use. Positive antecedents of well-being are equally illustrative; simple actions such as getting enough sleep and regularly eating breakfast have much more positive associations with well-being than the average

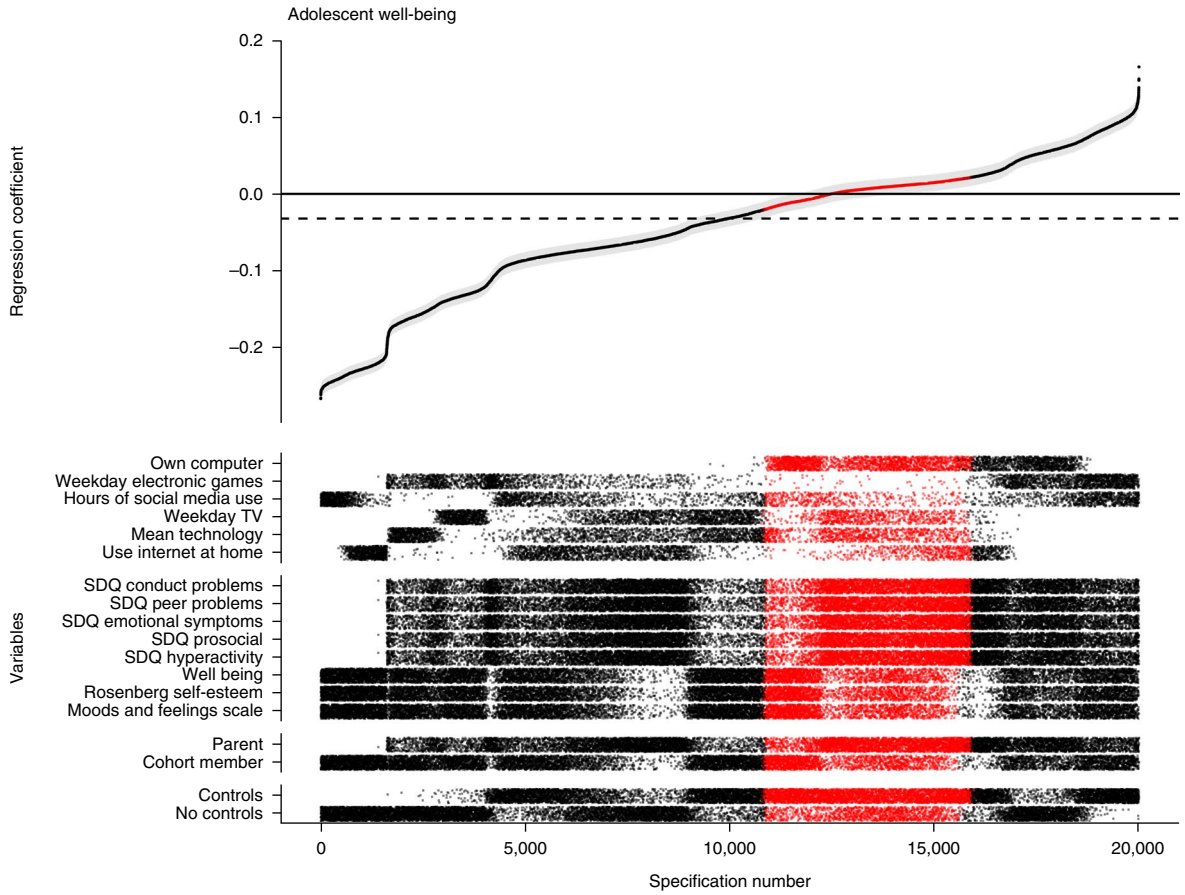

Fig. 3 | Results of SCA for MCS. Specification curve analysis showing the range of possible results for a simple cross-sectional regression of digital technology use on adolescent well-being. Each point on the *x* axis represents a different combination of analytical decisions, which are displayed in the 'dashboard' at the bottom of the graph. The resulting standardized regression coefficient is shown at the top of the graph; the error bars visualize the standard error. Red represents non-significant outcomes while black represents significant outcomes. To ease interpretation, the dotted line indicates the median standardized regression coefficient found in the SCA: $\beta = -0.032$ (partial $\eta^2 = 0.004$, median $n = 7{,}968$, median standard error = 0.010).

impact of technology use (ranging from ×1.7 to ×44.2 more positive in all datasets). Neutral factors provide perhaps the most useful context in which to judge technology engagement effects: the association of well-being with regularly eating potatoes was nearly as negative as the association with technology use (×0.9, YRBS), and wearing glasses was more negatively associated with well-being (×1.5, MCS).

With this in mind, the evidence simultaneously suggests that the effects of technology might be statistically significant but so minimal that they hold little practical value. The nuanced picture provided by these results is in line with previous psychological and epidemiological research suggesting that the associations between digital screen-time and child outcomes are not as simple as many might think[11,13]. This work therefore puts into perspective previous work that used both the YRBS and MTF to highlight technology use as a potential culprit for decreasing adolescent well-being[10], showing the range of possible analytical results and comparison specifications. Our finding that the association between technology use and digital engagement is much smaller than previously put forth has extensive implications for stakeholders and policy makers considering monetary investments into decreasing technology use in order to increase adolescent well-being[27].

Importantly, the small negative associations diminish even further when proper and pre-specified control variables, or caretaker responses about adolescent well-being, are included in the analyses. This finding underlines the importance of considering high-quality control variables, a priori specification of effect sizes of interest and a critical evaluation of the potential role played by common method variance when mapping the effect of digital technology use on adolescent well-being[28]. It is not enough to rely on statistical power to improve scientific endeavour: large-scale social data analysis harbours its own challenges for statistical inference and scientific progress.

This investigation therefore highlights two intrinsic problems confronting behavioural scientists using large-scale social data. First, large numbers of ill-defined variables necessitate researcher flexibility, potentially exacerbating the garden of forking paths problem: for some datasets analysed there were more than a trillion different ways to operationalize a simple regression[19]. Second, high numbers of observations render minutely small associations significant through the default null hypothesis significance testing lens[29]. With these challenges in mind, our approach, grounded in SCA and including comparison specifications, presents a promising solution so that behavioural scientists can build accurate and practically actionable representations of effects found in large-scale datasets. Overall, the findings place into context popular worries about the putative links between technology use and mental health indicators. They underscore the need for open and impartial reporting of small correlations derived from large-scale social data.

Our analyses, however, do not provide a definite answer to whether digital technology impacts adolescent well-being. Firstly, it is important to note that using most large-scale datasets one can only examine cross-sectional correlations links and it is therefore unclear

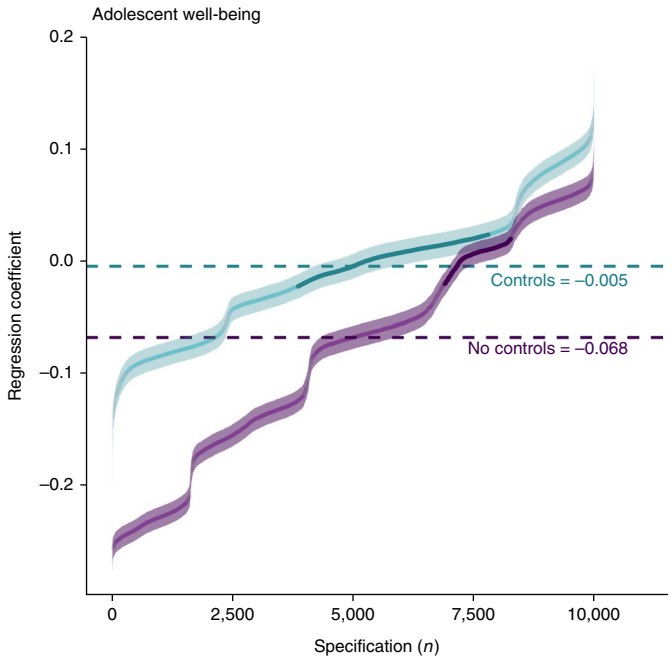

Fig. 4 | Results of SCA for MCS, split by whether controls are included in the analysis. Specification curve analysis showing the range of possible results for a simple cross-sectional regression of digital technology use on adolescent well-being. Each specification number indicates a different combination of analytical decisions. The plot then shows the outcome of the corresponding analysis (standardized regression coefficient) either including control variables (teal, median standardized $\beta = -0.005$, partial $\eta^2 = 0.001$, median $n = 6{,}566$, median standard error = 0.011) or not, including control variables (purple, median standardized $\beta = 0-0.068$, partial $\eta^2 = 0.005$, median $n = 11{,}018$, median standard error = 0.010). The bold parts of the line indicate analyses that did not reach significance ($P < 0.05$). The median standardized regression coefficients for analyses including or not including control variables are denoted by the dashed lines and the error bars represent the standard error.

Table 3 | Comparison specification results

Comparison specifications		YRBS	MTF	MCS
Negative	Binge-drinking	×2.95	×8.10	×1.02
Factors	Marijuana	×2.70	×10.09	×1.14
	Bullying	×4.33	–	×4.92
	Getting into fights	×3.65	×15.58	–
	Cigarettes	–	×18.47	–
	Being arrested	–	–	×0.96
Neutral	Perceived weight	×1.02	–	–
Factors	Potatoes	×0.86	–	–
	Asthma	×1.34	–	–
	Milk	×0.28*	–	–
	Going to movies	–	×11.51*	–
	Religion	–	×16.29*	–
	Music	–	×32.68	–
	Homework	–	×3.57*	–
	Cycling	–	–	×1.88*
	Height	–	–	×1.53*
	Glasses	–	–	×1.45
	Handedness	–	–	×0.10
Positive	Fruit	×0.11	×9.49*	×1.32*
Factors	Vegetables	×0.27	×20.63*	×1.52*
	Sleep	×3.06*	×44.23*	×1.65*
	Breakfast	×2.37*	30.55*	×3.32*

The table shows the size of the effect of comparison variables on adolescent well-being when compared to the size of the effect of technology use (measured using the mean of technology use questions) on adolescent well-being. The values indicate how many times larger the effects of the comparison variables are in comparison to technology use when examining the Youth Risk and Behaviour Survey (YRBS), Monitoring the Future (MTF) and Millennium Cohort Study (MCS) datasets. * Denotes when the effect of the comparison variable on well-being is positive, and therefore in the opposite direction to the effect of technology use. Note: For the YRBS the average effect linking technology to well-being was $\beta = -0.049$; for the MTF the average effect linking technology to well-being was $\beta = -0.006$; for the MCS the average effect linking technology to well-being was $\beta = -0.042$. Please note that these figures may be different from those found in Table 2, because the mean of technology use measures was used in these analyses.

what is driving effects where these are present. We know very little about whether increased technology use might cause lower well-being, whether lower well-being might result in increased technology use or whether a third confounding factor underlies both. Because we are examining something inherently complex, the likelihood of unaccounted factors affecting both technology use and well-being is high. It is therefore possible that the associations we document, and those that previous authors have documented, are spurious.

For the sake of simplicity and comparison, simple linear regressions were used in this study, overlooking the fact that the relationship of interest is probably more complex, non-linear or hierarchical[13]. Many measures used were also of low quality, non-normal, heterogenous or outdated, limiting the generalizability of the study's inferences. As self-report digital technology measures are known to be noisy[30], this could also have led to the effects of technology on well-being being diminished due to low-quality measurement. Lastly, we used null hypothesis significance testing to interpret significance, which is problematic when using such extensive data. To improve partnerships between research councils and behavioural scientists, the implementation of better measurement, and pre-registering of analyses plans, will be crucial.

Whether these are collected as part of multi-laboratory projects or research council-funded cohort studies, large-scale social datasets are an increasingly important part of the research infrastructure in the behavioural sciences. On balance, we are optimistic that these investments provide an invaluable tool for studying technology effects in young people. To realize this promise, we firmly believe that researchers must ground their work and debate in open and robust practices. In the quest for high power, we caution scientists studying technology effects to understand the intrinsic limitations of large-scale data and to implement approaches that guard against researcher degrees of freedom. While pre-registration might be implausible for analyses of open large-scale social data, methodologies such as SCA provide solutions that not only support robust statistical inferences, but also provide a comprehensive way to report the effects found for academia, policy and the public.

Methods

Datasets and participants. This paper's analysis pipeline spans three nationally representative datasets from the United Stats of America and the United Kingdom[6–8], encompassing a total of 355,358, predominately 12- to 18-year-old adolescents surveyed between the years 2007 and 2016. These datasets were selected because they feature measures of adolescents' psychological well-being and digital technology use, and have been the focus of secondary data analysis used to study digital technology effects[10,11,31].

Two of these datasets are based on samples collected in the United States Of America. The first, the YRBS[7] launched in 1990, is a biennial survey of adolescents that reflects a nationally representative sample of students attending secondary schools in that country (years 9–12). The resulting sample from the YRBS was collected from 2007 to 2015 and included 37,402 girls and 37,412 boys, ranging in age from '12 years or younger' to '18 years or older' (median = 16, s.d. = 1.24). The

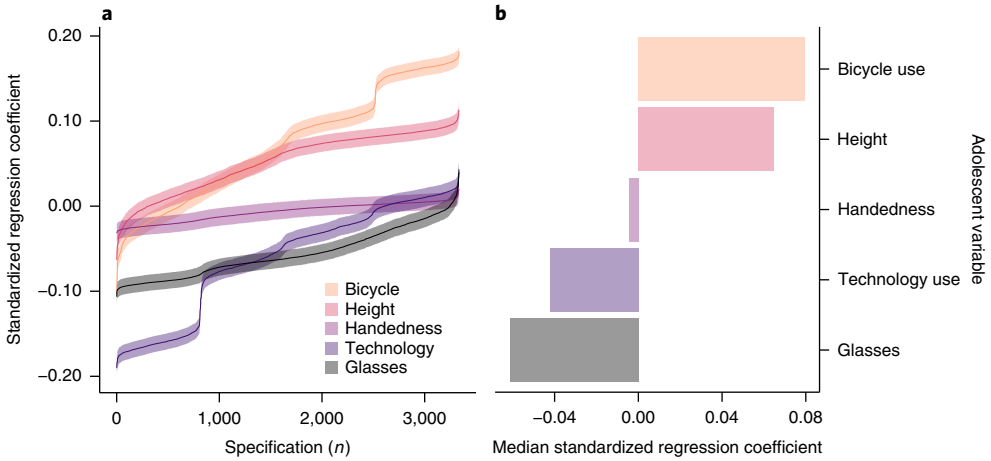

Fig. 5 | Comparison specifications for MCS. Visualization of the comparison specifications hypothesized to have little or no influence on well-being: bicycle use, height, handedness and wearing glasses. This graph shows SCA for both the variable of interest (mean technology use) and the comparison variables; it highlights the range of possible results of a simple cross-sectional regression of the variables of interest on adolescent well-being. Wearing glasses has the most negative association with adolescent well-being (black, median $\beta = -0.061$, median $n = 7{,}963$, partial $\eta^2 = 0.005$, median standard error = 0.010); and more negative than the association of technology use with well-being (purple, median $\beta = -0.042$, median $n = 7{,}964$, partial $\eta^2 = 0.002$, median standard error = 0.010). Handedness (red/purple, median $\beta = -0.004$, median $n = 7{,}972$, partial $\eta^2 < 0.001$, median standard error = 0.010), height of the adolescent (red, median $\beta = 0.065$, median $n = 7{,}910$, partial $\eta^2 = 0.005$, median standard error = 0.010) and whether the adolescent often rides a bicycle (yellow, median $\beta = 0.080$, median $n = 7{,}974$, partial $\eta^2 = 0.007$, median standard error = 0.010) have more positive associations with adolescent well-being than does technology use. **a,** How different analytical decisions (specifications, shown on the *x* axis) lead to different statistical outcomes (standardized regression coefficient, shown on the *y* axis). Each line represents a different variable of interest while the error bars represent the standard error. **b**, The resulting median standardized regression coefficients for those SCAs linking the variables of interest with adolescent well-being.

second US dataset, the MTF[6], was launched in 1975 and is an annual nationally representative survey of approximately 50,000 US adolescents in grades 8, 10 and 12. While the survey includes adolescents in grade 12, many of the key items of interest cannot be correlated in their survey and therefore their data were not included in our analysis. The resulting sample from the MTF was collected from 2008 to 2016, and included 136,190 girls and 132,482 boys, though the exact age of individual respondents was removed from the dataset by study coordinators during anonymization.

The UK dataset under analysis is the MCS[8], a prospective study collected in that country; it follows a specific cohort of children born between September 2000 and January 2001. We see these data as particularly high in quality due to the inclusion of pre-tested measures and extensive documentation, highlighting good data collection and project management practices. The data have an over-representation of minority groups and disadvantaged areas due to clustered stratified sampling. Data in this sample were provided by caregivers as well as adolescent participants. In our analysis, we included only data from primary caregivers and adolescent respondents. The sample under analysis from the MCS comprised 5,926 girls and 5,946 boys who ranged in age from 13 to 15 years (mean 13.77, s.d. 0.45), and 10,605 primary caregivers.

While the omnibus sample of adolescents totals 355,358 teenagers, it is important to note that the sample sizes of the analyses are often smaller, in some cases by an order of magnitude or more. This is due to missing values, but also because in questionnaires such as the MTF, teenagers answered only a subset of questions. More information about what questions were asked together in the MTF can be found in Supplementary Table 3.

Ethical review. Ethical review and approval for data collection for YRBS was conducted and granted by the CDC Institutional Review Board. The University of Michigan Institutional Review Board oversees the MTF. Ethical review and approval for the MCS is monitored by the UK National Health Service London, Northern, Yorkshire and South-West Research Ethics Committees.

Measures. This study focuses on measures of both digital technology use and psychological well-being. Before performing the analysis, all three datasets were reviewed, noting the variables of theoretical interest in each with respect to human behaviour and the effects of technology engagement. Some questions have been modified with successive waves of data collection. In most cases these changes are relatively minor and are noted in the Supplementary materials (Supplementary Table 4). In our ongoing analyses we use the questionnaires in many different constellations and therefore refrain from including reliability measurements. Further details regarding all measures can be found in the Supplementary Note.

Criterion variables: adolescent well-being. All datasets contained a wide range of different questions that concern adolescents' psychological well-being and functioning. We reversed selected measures so that these are all in the same direction, with higher scores indicating higher well-being.

Adolescents were asked five questions related to mental health and suicidal ideation in the YRBS. Three were on a yes–no scale and two were on a frequency scale. In the MTF, participants were asked one of two subsets of self-report questions. The first tranche of participants was asked 13 questions about their mental health: 12 measures uniquely asked to this subset and one completed by all participants in the survey. The 12 items asked only to this subset included a 4-item depressive symptoms scale, which studies state to be "similar to those on the Center for Epidemiologic Studies Depression Scale"[32] and a self-esteem scale created by Rosenberg[33], both of which use a disagree–agree Likert scale. Survey administrators also included two additional negatively worded self-esteem measures and a 1-item measure asking how happy the participants felt.

There are two kinds of psychological well-being indicator included in the MCS: (1) those filled out by the cohort members and (2) those completed by their primary caretakers. The cohort members completed six 7-point agree–disagree measures reflecting their subjective sense of well-being, and twelve 3-point questions tapping into subjective affective states and general mood[34]. Primary caregivers completed the Strengths and Difficulties Questionnaire[35], a well-validated measure of psychosocial functioning, for each adolescent cohort member they took care of (Supplementary Table 5). This questionnaire has been used extensively in schools, homes and clinical settings with adolescents from a wide range of social, ethnic and national backgrounds[36]. It includes 25 questions, five each about pro-social behaviour, hyperactivity or inattention, emotional symptoms, conduct problems and peer relationship problems.

Explanatory variables: adolescent technology use. The YRBS dataset included two 7-point technology use questions. One related to the frequency of electronic device use while the other queried the amount of TV watched on a typical weekday. The MTF asked a variety of technology use measurements. As the questionnaire was split into six parts (with each participant completing only one part), some questions were completed by one subset of adolescents while others by another. One subset answered questions about the frequency of social media use and getting news information from the Internet (5-point scale) and two 7-point questions about the frequency of watching TV on weekends and weekdays. Another group of MTF participants were asked about seven hourly measures of technology use on a 9-point scale. The questions related to using the Internet, playing electronic games, texting on a mobile phone, calling on a mobile phone, using social media, video chatting and using computers for school work. There are, therefore, a total of 11 technology use measures that can be used when analysing the MTF dataset.

In the MCS, the participants were asked five questions concerning technology use. There were four 9-point items relating to the hours per weekday spent watching TV, playing electronic games, using the Internet at home and using social networking sites. There was also one yes–no measure about whether participants owned a computer.

Co-variate and confounding variables. Mirroring previous studies analysing data from the MCS[11], we included sociodemographic factors and maternal characteristics as co-variates in our analyses. These included mother's ethnicity, education, employment and psychological distress (using the K6 Kessler scale), which have previously been found to influence child well-being in studies analysing large-scale data[37,38], including MCS analyses[39]. We also included equivalized household income, whether the biological father was present and number of adolescent's siblings in the household, as these household factors have also been found to affect adolescent well-being[38]. Furthermore, we included parental behavioural factors such as closeness to parents and the amount of time spent by the primary caretaker with the adolescent[40,41]. Addressing previous reports of their influence on child well-being, as co-variates we additionally used parent reports of any adolescent's long-term illness, and the adolescent's own negative attitudes towards school[41,42]. Finally, we included the primary caretaker's word activity score as a measure of current cognitive ability, to control for other environmental factors that could influence child well-being[11].

For both the YRBS and MTF we included all the variables part of the respective questionnaires that conceptually mirrored those co-variates utilized in the MCS. For the YRBS we included the adolescent's race. For the MTF we included ethnicity, number of siblings, mother's education level, whether the mother has a job, the adolescent's enjoyment of school, predicted school grade and whether they feel that they can talk with their parents about problems.

Analytical approach: SCA. The study implements the SCA method to examine the correlation between our explanatory (digital technology engagement) and criterion variables (psychological well-being) using the 3-step SCA approach outlined by Simonsohn et al.[25] and applied in a recent paper by Rohrer et al.[26]. We add a fourth step in order to aid the interpretability of our results in the context of large-scale social data. Details of the SCA method and the corresponding visualizations can be found in the Supplementary Methods. All necessary codes to reproduce these analyses can be found in the Supplementary Software; for details see the Code Availability Statement at the end of the paper.

Identifying specifications. The first step taken was to identify all analysis pathways that could potentially be used to relate technology use and adolescent well-being. Due to the complexity of the original data, we decided to use simple linear regression modelling to draw inferences about technology associations, which left three key analytical decisions: (1) how to measure well-being, (2) how to measure technology use and (3) how to include co-variates (for details about these decisions, and others, see Table 1).

There are a wide variety of questions and questionnaires relating to well-being in each dataset. Many of these items, even if partitioned questionnaires reflect a specific construct, have been selectively reported over the years. It is noteworthy that researchers have not been consistent and have instead engaged in picking and choosing within and between questionnaires (see Supplementary Table 6). These analytical decisions have produced many different possibilities for combining and analysing these measures, making the pre-specified constructs more of an accessory for publication than a guide for analyses. Any combination of the mental health indicators is therefore included in the SCA: the measures by themselves, the mean of the measures in pairs of two, the mean of the measures in threes, etc. up to the mean of all measures.

For the MCS, we included a decision of whether to use well-being questions answered by cohort members or those answered by their caregivers; we did not combine the two. For the YRBS we also included an additional analytical decision of whether to take the mean of the five dichotomous well-being measures or whether to code each participant as '1' who answered yes to one or more of the questions, as this has been done in previous analyses of the data[10]. The Supplementary materials additionally present SCAs that include only pre-specified well-being questionnaires for the MCS (Supplementary Fig. 7); however, these do not allow comparisons of our SCAs to results of previous work that has selectively combined questions from various questionnaires[10]. The next analytical decision related to which technology-use variables to include: where we include all questions concerning technology use in the questionnaires, and their mean, as done by previous studies[10]. The last analytical decision taken was whether to include co-variates in the models. Because of the sheer size of these datasets, there is a combinatorial explosion of different co-variate combinations that could be used in each regression. We therefore analysed regressions either without co-variates or with a pre-specified set of co-variates based on a literature review concerning child well-being and digital technology use[11].

When examining the distributions of the data, many of the variables are highly skewed (for example, the 5-item technology use measures in the MTF) or questionably linear (for example, the 3-item happiness measure in the MTF). We opted to treat these variables as continuous so that our analyses and results would be directly comparable to those of previous studies[10,31]. Data distribution was assumed to be normal throughout the analysis, but was not formally tested for each specification.

Implementing specifications. Next, for each specification defined we ran the appropriate regression and noted the standardized β value for the correlation of technology use with psychological well-being, the corresponding two-sided P value and the partial η^2 were calculated using the R heplots package. List-wise deletion for missing data was used, as this is more efficient in terms of computational time. This assumes that data are missing completely at random, which could easily not be the case. For example, a child's health, academic performance or socio-economic background could change its probability of completing the questionnaire fully, and is likely to bias estimates. It is therefore important to note that this is a potential source of bias, possibly changing the nature or strength of associations found.

To make the results easily interpretable, the specifications were ranked and plotted in terms of ascending standardized β. The median standardized β of all possible specifications provides a general overview of the effect size. Below that plot, we also indicated which set of analytical decisions led to what standardized β. This allows us to visualize which analytical decisions influence the results of the SCA (more details of these plots can be found in the Supplementary Methods).

Statistical inferences. It is then possible to test whether, when considering all the possible specifications, the results found are inconsistent with results when the null hypothesis is true (that is, that technology use and adolescent well-being are unrelated). To do so, a bootstrapping technique put forth by Simonsohn et al.[25] was implemented, creating data where the null hypothesis is true by forcing the null on the data. To create these data, the β-coefficient of the variable of interest from the full regression model, multiplied by the x-variable (technology use), was subtracted from the y-variable (well-being). This created a new set of data points that were then used as the new y-variable, creating datasets where the null hypothesis was known to be true. Participants were then drawn at random—with replacement—from this null dataset, creating bootstrapped null samples on which a new SCA model was run. This was done 500 times. Once we had obtained 500 bootstrapped SCAs, where we knew the null hypothesis to be true, we examined whether the median effect size in the original SCA was significantly different to the median effect size in the bootstrapped SCAs. To do so, we divided the number of bootstrapped datasets with larger median effect sizes than the original SCA by the total number of bootstraps, to find the P value of this test. We repeated this test focusing also on the share of results with the dominant sign, and also the share of statistically significant results with the dominant sign[4].

Comparison specifications. Lastly, these analyses were supplemented by a comparison specifications section, putting into context the effects found in the SCA. To do so, we performed a literature review to select four variables in each dataset that should be positively correlated with psychological well-being, four that should be negatively correlated with psychological well-being and four that should have no or little association with psychological well-being. A SCA was run for each of the variables and the mean of the technology use variables present in the dataset, graphing their specification curves. These methods provide a way for researchers to transparently, openly and robustly analyse large-scale governmental datasets to produce research that accurately depicts associations found in the data for both academia and the public.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability. The code used to analyse the relevant data is provided as Supplementary Software; intermediate analysis files and a live version of the analysis code can be found on the Open Science Framework website (https://osf.io/e84xu/).

Data availability

The data that support the findings of this study are available from the Centre for Disease Control and Prevention (YRBS), Monitoring the Future (MTF) and the UK data service (MCS), but restrictions apply regarding the availability of these data, which were used under licence for the current study and so are not publicly available. Data are, however, available from the relevant third-party repository after agreement to their terms of usage. Information about data collection and questionnaires can be found on the OSF website (https://osf.io/7xha2/).

Received: 5 June 2018; Accepted: 24 November 2018;
Published online: 14 January 2019

References

1. Bell, V., Bishop, D. V. M. & Przybylski, A. K. The debate over digital technology and young people. *BMJ* **351**, h3064 (2015).
2. Children and Parents: Media Use and Attitudes Report. *Ofcom* https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-parents-2017 (2017).

3. Steers, M.-L. N. 'It's complicated': Facebook's relationship with the need to belong and depression. *Curr. Opin. Psychol.* **9**, 22–26 (2016).
4. UK Commons Select Committee. Impact of social media and screen-use on young people's health inquiry launched. *Parliament.uk*. https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/news-parliament-2017/social-media--young-peoples-health-inquiry-launch-17-19/ (2018).
5. Youth Select Committee. *A Body Confident Future* (British Youth Council, 2017).
6. Johnston, L. D., Bachman, J. G., O'Malley, P. M., Schulenberg, J. E. & Miech, R. A. Monitoring the future: a continuing study of American Youth (8th- and 10th-Grade Surveys) https://doi.org/10.3886/ICPSR36799.v1 (2016).
7. Kann, L. et al. Youth risk behavior surveillance – United States, 2015. *MMWR Surveill. Summ.* **65**, 1–174 (2016).
8. University of London, Institute for Education, Centre for Longitudinal Studies. *Millennium Cohort Study: Sixth Survey*, *2015* SN: 8156 (2018).
9. Etchells, P. J., Gage, S. H., Rutherford, A. D. & Munafò, M. R. Prospective investigation of video game use inchildren and subsequent conduct disorder and depression using data from the Avon Longitudinal Study of Parents and Children. *PLoS One* **11**, e0147732 (2016).
10. Twenge, J. M., Joiner, T. E., Rogers, M. L. & Martin, G. N. Increases in depressive symptoms, suicide-related outcomes, and suicide rates among U.S. adolescents after 2010 and links to increased new media screen-time. *Clin. Psychol. Sci.* **6**, 3–17 (2017).
11. Parkes, A., Sweeting, H., Wight, D. & Henderson, M. Do television and electronic games predict children's psychosocial adjustment? Longitudinal research using the UK Millennium Cohort Study. *Arch. Dis. Child.* **98**, 341–348 (2013).
12. Ferguson, C. J. The problem of false positives and false negatives in violent video game experiments. *Int. J. Law Psychiatry* **56**, 35–43 (2018).
13. Przybylski, A. K. & Weinstein, N. A large-scale test of the Goldilocks hypothesis. *Psychol. Sci.* **28**, 204–215 (2017).
14. Ferguson, C. J. Everything in moderation: moderate use of screens unassociated with child behavior problems. *Psychiatr. Q.* **88**, 797–805 (2017).
15. What About Youth Study. *NHS Digital* (National Health Service, 2017).
16. U.S. Department of Health and Human Services, Health Resources and Services Administration, Maternal and Child Health Bureau. *Child Health USA 2014* (U.S. Department of Health and Human Services, 2015).
17. Livingstone, S., Haddon, L., Görzig, A. & Ólafsson, K. *Technical Report and User Guide: The 2010 EU Kids Online Survey* (EU Kids Online, 2011).
18. Silberzahn, R. et al. Many analysts, one data set: making transparent how variations in analytic choices affect results. *Adv. Method Pract. Psychol. Sci.* **1**, 337–356 (2018).
19. Gelman, A. & Loken, E. The garden of forking paths: why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Psychol. Bull.* **140**, 1272–1280 (2014).
20. Simmons, J. P., Nelson, L. D. & Simonsohn, U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22**, 1359–1366 (2011).
21. Marszalek, J. M., Barber, C., Kohlhart, J. & Cooper, B. H. Sample size in psychological research over the past 30 years. *Percept. Mot. Skills* **112**, 331–348 (2011).
22. Chambers, C. D. Registered reports: a new publishing initiative at Cortex. *Cortex* **49**, 609–610 (2013).
23. Munafò, M. R. et al. A manifesto for reproducible science. *Nat. Hum. Behav.* **1**, 0021 (2017).
24. Van't Veer, A. E. Pre-registration in social psychology—a discussion and suggested template. *J. Exp. Soc. Psychol.* **67**, 2–12 (2016).
25. Simonsohn, U., Simmons, J. P. & Nelson, L. D. Specification curve: descriptive and inferential statistics on all reasonable specifications. *SSRN Electron. J.* https://doi.org/10.2139/ssrn.2694998 (2015).
26. Rohrer, J. M., Egloff, B. & Schmukle, S. C. Probing birth-order effects on narrow traits using specification-curve analysis. *Psychol. Sci.* **28**, 1821–1832 (2017).
27. Department of Health and Social Care. Matt Hancock Warns Of Dangers Of Social Media On Children's Mental Health. *Gov.uk*. https://www.gov.uk/government/news/matt-hancock-warns-of-dangers-of-social-media-on-childrens-mental-health (2018).
28. Ferguson, C. J. An effect size primer: a guide for clinicians and researchers. *Prof. Psychol. Res. Pract.* **40**, 532–538 (2009).
29. Lakens, D. & Evers, E. R. K. Sailing from the seas of chaos into the corridor of stability: practical recommendations to increase the informational value of studies. *Perspect. Psychol. Sci.* **9**, 278–292 (2014).
30. Scharkow, M. The accuracy of self-reported Internet use—a validation study using client log data. *Commun. Methods Meas.* **10**, 13–27 (2016).
31. Twenge, J. M., Martin, G. N. & Campbell, W. K. Decreases in psychological well-being among American adolescents after 2012 and links to screen-time during the rise of smartphone technology. *Emotion* https://doi.org/10.1037/emo0000403 (2018).
32. Maslowsky, J., Schulenberg, J. & Zucker, R. Influence of conduct problems and depressive symptomatology on adolescent substance use: developmentally proximal versus distal effects. *Dev. Psychol.* **40**, 1179–1189 (2014).
33. Robins, R. W., Hendin, H. M. & Trzesniewski, K. H. Measuring global self-esteem: construct validation of a single-item measure and the Rosenburg self-esteem scale. *Personal. Soc. Psychol. Bull.* **27**, 151–161 (2001).
34. Angold, A., Costello, E. J., Messer, S. C. & Pickles, A. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *Int. J. Methods Psychiatr. Res.* **5**, 237–249 (1995).
35. Goodman, R., Ford, T., Simmons, H., Gatward, R. & Meltzer, H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br. J. Psychiatry* **177**, 534–539 (2000).
36. Goodman, R. Psychometric properties of the Strengths and Difficulties Questionnaire. *J. Am. Acad. Child Adolesc. Psychiatry* **40**, 1337–1345 (2001).
37. Desai, S., Chase-Lansdale, P. L. & Michael, R. T. Mother or market? Effects of maternal employment on theintellectual ability of 4-year-old children. *Demography* **26**, 545 (1989).
38. Kiernan, K. E. & Mensah, F. K. Poverty, maternal depression, family status and children's cognitive andbehavioural development in early childhood: a longitudinal study. *J. Soc. Policy* **38**, 569 (2009).
39. Mensah, F. K., Kiernan, K. E. & Kiernan, K. Maternal general health and children's cognitive development and behaviour in the early years: findings from the Millennium Cohort Study. *Child Care Health Dev.* **37**, 44–54 (2010).
40. Thomson, E., Hanson, T. L. et al. Family Structure and Child Well-Being: Economic Resources vs. Parental Behaviors. *Social Forces* **73**, 221–242 (1991).
41. Pople, L. & Sharma, N. *Factors Affecting Children's Mental Health over Time* (The Children's Society & Barnardo's, 2018).
42. Cadman, D., Boyle, M., Szatmari, P. & Offord, D. R. Chronic illness, disability, and mental and social well-being: findings of the Ontario Child Health Study. *Pediatrics* **79**, 805–813 (1987).

Acknowledgements

The National Institute on Drug Abuse provided funding for the MTF conducted at the Survey Research Centre in the Institute for Social Research, University of Michigan. The YRBS was collected by the Centres for Disease Control and Prevention. The Centre for Longitudinal Studies, UCL Institute of Education collected the MCS and the UK Data Archive/UK Data Service provided the data. They bear no responsibility for its aggregation, analysis or interpretation. A.O. was supported by a EU Horizon 2020 IBSEN grant. A.K.P. was supported by an Understanding Society Policy Fellowship funded by the Economic and Social Research Council. A.O. and A.K.P.'s funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Thanks are extended to U. Simonsohn, N.K. Reimer and N. Weinstein for their valuable input, and to J.M. Rohrer, U. Simonsohn, J.P. Simmons and L.D. Nelson for code provision. We also acknowledge the use of the University of Oxford Advanced Research Computing facility in carrying out this research: https://doi.org/10.5281/zenodo.22558. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

A.O. conceptualized the study, with regular guidance from A.K.P. A.O. completed the statistical analyses and drafted the manuscript. A.K.P. gave integral feedback in the process.

Competing interests

A.O. has no competing interests. A.K.P. has no competing financial interests; in the last five years A.K.P. has served in an unpaid advisory capacity with the Organization for Economic Co-operation and Development, Facebook Inc., Google Inc. and the ParentZone.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-018-0506-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.O.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Amy Orben

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	N/A (openly available dataset)
Data analysis	We used R as our statistical analysis software. Our customized code can be found on the open science framework and as Supplementary Code

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the Centre for Disease Control and Prevention (YRBS), Monitoring the Future (MTF) and the UK

data service (MCS) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the third party repository after agreement to their terms of usage.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The three secondary datasets used were cohort studies collecting quantitative data
Research sample	The study uses three datasets the YRBS which is a biennial survey of adolescents that reflects a nationally-representative sample of students attending secondary schools in the U.S. (years 9-12), the MTF an annual nationally-representative survey of approximately 50,000 American adolescents in grades 8, 10 and 12 and MCS, a representative study which follows a specific cohort of children born between September 2000 and January 2001.
Sampling strategy	As we analyzed secondary data, we did not determine the sample size of the dataset. We, however, go to great lengths to put effect sizes into perspective to ensure that even the very small effect sizes that become significant using large-scale data are not over-interpreted.
Data collection	Details of the data collection procedure for each study can be found on the relevant third party websites and on the OSF: https://osf.io/e84xu/
Timing	Details are given in the third party documentation
Data exclusions	No data was excluded in these analyses
Non-participation	N/A
Randomization	N/A

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above
Recruitment	The data is nationally representative.