The association between adolescent well-being and digital technology use

In Press at Nature Human Behaviour

Please note that figures are located at the end of the manuscript.

The association between adolescent well-being and digital technology use

Amy Orben*

Department of Experimental Psychology, University of Oxford

Andrew K. Przybylski

Oxford Internet Institute, University of Oxford

Department of Experimental Psychology, University of Oxford

Correspondence concerning this article should be addressed to Amy Orben, Department of Experimental Psychology, University of Oxford, New Radcliffe House, Oxford OX2 6GH.

Email: amy.orben@psy.ox.ac.uk

Orcid ID: orcid.org/0000-0002-2937-4183

Abstract

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 analytic 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 across three large-scale social datasets (n_{tot} = 355,358) to rigorously examine correlational evidence for digital technology affecting 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.

Re-Evaluating the Relation between Digital Technology Use and Adolescent Well-Being

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 young people spend 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 found negative effects when re-analysing identical data[12]. A 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 has small negative associations[13,14].

There are at least three reasons why the inferences behavioural scientists draw 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 digital technology's effects 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, they 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 very small covariations (e.g. r's < .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 (i.e. p's < .05). Thirdly, it is important to note that most datasets are cross-sectional and therefore only provide 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[22] and has real-world implications[23], it is important for researchers to make the most of existing large-scale dataset investments. This makes it necessary to employ transparent and robust analytic 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. Further, 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, i.e. using $p < .05$, to demarcate "true" effects. Unfortunately the large number of participants in these designs means 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[24–26], making hypothesis preregistration untenable as a general solution to the problem of subjective analytical decisions. Specification Curve Analysis (SCA)[27] provides a promising alternative. Briefly, SCA is a tool for mapping the sum of theory-driven analytic decisions that could have been justifiably taken when analysing quantitative data. Researchers demarcate every possible analytical pathway and then calculate the results of each one. Instead of reporting a handful of analyses in their paper, they report all results of all theoretically defensible analyses (for previous examples see [27,28] 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, Youth Risk and Behaviour Survey and Millennium Cohort Study) from the US and the UK were selected to highlight the particular strengths and weaknesses of drawing general inferences from large-scale social data and how they can be reconceptualised by SCA [6–8]. Further, 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 the effects of other activities on psychological well-being (e.g. 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 analytic 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). 372 justifiable specifications for YRBS, 40,966 plausible specifications for MTF, and a total of 603,979,752 defensible specifications for 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 (6 analysis options) and independent variables ($2^{24} + 2^{25} - 2$ analysis options) and whether covariates are included or not (2 analysis options). The number rises even higher to 2.5 trillion specifications for MCS if any combination of covariates (2^{12} analysis options) is included. Given this, and to reduce computational time, we selected 20,004 specifications for MCS. To do so, we included specifications of all used measures by themselves, and any combinations of measures found in the previous literature and then supplemented them with other randomly selected combinations. More information about selection can be found in the supplementary materials (see Supplementary Table 1).

Implementing Specifications.

After noting down all specifications, the result of every possible combination of these specifications was computed for each dataset. The standardised β coefficient for technology uses' association with well-being was then plotted for each specification. The number of

Table 1: Possible specifications (analytical decisions) to test a simple linear regression between technology use and adolescent well-being in the Youth Risk and Behaviour Survey (YRBS), Monitoring the Future (MTF) and Millennium Cohort Study (MCS) datasets.

Decision	YRBS	MTF	MCS
Operationalising adolescent well-being	Mean of any possible combination of five items to do with mental health and suicidal ideation	Mean of any possible combination of 13 items to do with depression, happiness and self-esteem	Mean of any possible combination of 24 questions about well-being, self-esteem and feelings (cohort members) or mean of any possible combination of 25 questions from the strength and difficulties questionnaire (caregivers)
Operationalising technology use	2 questions about electronic device use and TV use, or the mean of these questions	11 technology use measures about the internet, electronic games, mobile phone use, social media use and computer use, or the mean of these questions	5 questions concerning TV use, electronic games, social media use, owning a computer and using internet at home, or the mean of these questions
Which covariates to include	Either include covariates or not	Either include covariates or not	Either include covariates 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 filled out by cohort members or those filled out by their caregivers

participants analysed for each specification can be found in Supplementary Figure 1-3, the median standardised β, n, partial η^2 and standard error can be found in Table 2. For YRBS, the median association of technology use with adolescent well-being was $\beta = -.035$ (median partial $\eta^2 = .001$, median n = 62,297, median standard error = .004, see Figure 1). From the figure one can discern the analytical choices that influence the size of this effect. When using electronic device use as the independent variable in the model, the effects were more negative (median $\beta = -.071$, median partial $\eta^2 = .005$, median n = 62,368, median standard error = .004), while when including TV use in the model the effects were less negative and sometimes become non-significant (median $\beta = -.012$, median partial $\eta^2 < .001$, median n = 62,352, median standard error = .004). Even though YRBS does not have high quality control variables, including them yielded a smaller effect size for the relations of interest (controls: median $\beta = -.034$, median partial $\eta^2 = .001$, median n = 61,525, median standard error = .004; no controls: median $\beta = -.035$, median partial $\eta^2 = .001$, median n = 62,638, median standard error = .004).

For the MTF data, a median standardised β of -.005 was observed (median partial $\eta^2 < .001$, median n = 78,267, median standard error = .003), a value which fell into the non-significant range of the justifiable specifications (see Figure 2). This result was surprising, as MTF had the highest number of observations, making it difficult for even small associations to be flagged as non-significant using traditional alpha thresholds (i.e., $p < .05$). In Figure 2, and our bootstrapping test, we do not include the few specifications of the participants that only filled in one well-being measure (to see the SCA of all participants, see Supplementary Figure 4). 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 $\beta = -.013$, median partial $\eta^2 < .001$, median n = 117,560, median standard error = .003; controls: median $\beta = .001$, median partial $\eta^2 < .001$, median n = 72,525, median standard error = .003). TV

Table 2: Results of Specification Curve Analysis for the Youth Risk and Behaviour Survey (YRBS, United States), Monitoring the Future Survey (MTF, United States) and Millennium Cohort Study (MCS, United Kingdom), both overall and for different technology use variables, parent/adolescent self-report or with/without control variables.

Dataset	Median β of Specification Curve Analysis	Median partial η^2 of Specification Curve Analysis	Median n	Median Standard Error
YRBS				
Complete Specification Curve Analysis	-.035	.001	62,297	.004
Electronic Device Use Only	-.071	.005	62,368	.004
TV Use Only	-.012	< .001	62,352	.004
With Control Variables Only	-.034	.001	61,525	.004
Without Control Variables Only	-.035	.001	62,638	.004
MTF				
Complete Specification Curve Analysis	-.005	< .001	78,267	.003
Social Media Use Only	-.031	.001	102,963	.003
TV Viewing On Weekend Only	.008	.001	115,738	.003
Using Internet for News Only	-.002	< .001	115,580	.003
TV Viewing on Weekday Only	.002	< .001	115,783	.003
With Control Variables Only	.001	< .001	72,525	.003
Without Control Variables Only	-.013	< .001	117,560	.003
MCS				
Complete Specification Curve Analysis	-.032	.004	7,968	.010
Own a Computer Only	-.003	.011	7,973	.010
Weekday Electronic Games Only	.013	< .001	7,977	.010
Hours of Social Media Use Only	-.056	.009	7,972	.010
TV Viewing on Weekday Only	-.043	.003	7,971	.010
Use of Internet of Home Only	-.070	.006	7,975	.010
Parent-Report Well-Being Only	<.001	.003	7,893	.010
Adolescent-Report Well-Being Only	-.046	.008	8,857	.010
With Control Variables Only	-.005	.001	6,566	.011
Without Control Variables Only	-.068	.005	11,018	.010

viewing on the weekend had a median positive association with well-being of $\beta = .008$ (median partial $\eta^2 = .001$, median n = 115,738, median standard error = .003), while social media use had a median negative association with well-being of $\beta = -.031$ (median partial $\eta^2 = .001$, median n = 102,963, median standard error = .003), though the effect was small suggesting that technology use operationalised 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 showed mainly very small median associations, $\beta = -.002$ (median partial $\eta^2 < .001$, median n = 115,580, median standard error = .003) and $\beta = .002$ (median partial $\eta^2 < .001$, median n = 115,783, median standard error = .003) respectively. As previous studies have addressed the association between technology use and well-being using the same dataset[10], we include a figure showing how these study's specifications influence their reported results in the supplementary materials (see Supplementary Figure 5).

Lastly, results from 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 of technology use's association with wellbeing of $\beta = -.032$ (median partial $\eta^2 = .004$, median n = 7,968, median standard error = .010, see Figure 3). Across the board, if using well-being measures completed by the caregivers, the median association was less negative or more positive (median $\beta < .001$, median partial $\eta^2 = .003$, median n = 7,893, median standard error = .010), while the opposite was in evidence when considering well-being measures completed by the cohort member (median $\beta = -.046$, median partial $\eta^2 = .008$, median n = 8,857, median standard error =. 010). This pattern of shared covariation speaks to the idea that correlations between technology use and well-being might be rooted in common method variance, as one single informant fills out 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 Figure 4). The association for the uncorrected models had a median β of -.068 (median partial η^2 = .005, median n = 11,018, median standard error = .010). In contrast, the corrected models only found a median β of technology use regressed on wellbeing of -.005 (median partial η^2 = .001, median n = 6,566, median standard error = .011). Additional SCAs using only pre-specified questionnaires are presented in Supplementary Figure 6, further visualisations about how adding controls and parent-report affects the reported associations are presented in Supplementary Figures 7 and 8.

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 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 analysing bootstrapped samples which resulted in a larger median effect size than the median effect size of the original SCA (p = 0.00, original effect sizes = YRBS median β = - .035, MTF median β = -.005, MCS median β = -.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 compare specification curves for the mean of the technology use variables in each dataset to other associations that have been shown to relate, or are hypothesised not to relate, to adolescent mental health: binge drinking, smoking marijuana, being bullied, getting into fights, smoking cigarettes, 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, Figure 5 and Supplementary Figures 9-11.

For YRBS the association of mean technology use with well-being (median β = -.049, median n = 62,166, partial η^2 = .002, median standard error = .004) was exceeded by well-being's association with being bullied (median β = -.212, median n = 50,066, partial η^2 = .044, median standard error = .004), getting into fights (median β = -.179, median n = 62,106, partial η^2 = .031, median standard error = .004), binge drinking (median β = -.144, median n = 62,010, partial η^2 = .021, median standard error = .004), smoking marijuana (median β = -.132, median n = 62,361, partial η^2 = .018, median standard error = .004), having asthma (median β = -.066, median n = 60,863, partial η^2 = .004, median standard error = .004) and perceived weight (median β = -.050, median n = 62,752, partial η^2 = .002, median standard error =.004). There is a smaller negative association for eating potatoes (median β = -.042, median n = 61,912, partial η^2 = .002, median standard error = .004), eating vegetables (median β = -.013, median n = 62,034, partial η^2 < .001, median standard error = .004) and

Table 3: Comparison Specification results: The table shows the size of the effect of comparison variables on adolescent-wellbeing 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.

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

Note. For the YRBS the average effect linking technology to well-being was: $\beta = -.049$. For the MTF the average effect linking technology to well-being was: $\beta = -.006$. For the MCS the average effect linking technology to well-being was: $\beta = -.042$. Please note that these numbers can be different from those found in Table 2 as the mean of technology use measures was used in these analyses.

eating fruit (median β = -.005, median n = 62,436, partial η^2 < .001, median standard error = .004). There is a smaller positive association for drinking milk (median β = .014, median n = 60,021, partial η^2 < .001, median standard error = .004). Lastly, there is a larger positive association for eating breakfast (median β = .116, median n = 34,010, partial η^2 = .013, median standard error = .006) and getting enough sleep (median β = .150, median n = 56,552, partial η^2 = .022, median standard error = .004).

For the MTF we compare the association of mean technology use (median β = -.006, median n = 102,186, partial η^2 < .001, median standard error = .003) to the variables we hypothesised *a priori* to have no association: going to the movies (median β = .064, median n = 115,943, partial η^2 = .005, median standard error = .003), time spent on homework (median β = .020, median n = 115,225, partial η^2 = .001, median standard error = .003), attending religious services (median β = .091, median n = 89,453, partial η^2 = .010, median standard error = .003) and listening to music (median β = -.182, median n = 49,514, partial η^2 = .035, median standard error = .005) all had larger effects. We also examined those we hypothesised to have a more positive association: eating breakfast (median β = .170, median n = 62,330, partial η^2 = .034, median standard error = .004), eating fruit (median β = .053, median n = 115,334, partial η^2 = .003, median standard error = .003), sleep (median β = .246, median n = 61,903, partial η^2 = .070, median standard error = .004), and eating vegetables (median β = .115, median n = 62,072, partial η^2 = .014, median standard error = .004). Lastly we looked at those variables that we hypothesised to have a more negative association: binge drinking (median β = -.045, median n = 107,994, partial η^2 = .002, median standard error = .003), fighting (median β = -.087, median n = 62,683, partial η^2 = .008, median standard error = .004), smoking marijuana (median β = -.056, median n = 113,611, partial η^2 = .003, median standard error = .003) and smoking cigarettes (median β = -.103, median n = 113,424, partial η^2 = .012, median standard error = .003).

For MCS, mean technology use (median β = -.042, median n = 7,964, partial η^2 = .002, median standard error = .010) was compared to amount of sleep (median β = .070, median n = 7,954, partial η^2 = .005, median standard error = .010), eating fruit (median β = .056, median n = 7,960, partial η^2 = .004, median standard error = .010), eating breakfast (median β = .140, median n = 7,964, partial η^2 = .025, median standard error = .010) and eating vegetables (median β = .064, median n = 7,949, partial η^2 = .005, median standard error = .010) that have a priori hypothesised positive associations; being arrested (median β = -.041, median n = 7,908, partial η^2 = .002, median standard error = .011), being bullied (median β = -.208, median n = 7,898, partial η^2 = .048, median standard error = .010), binge drinking (median β = -.043, median n = 3,656, partial η^2 = .002, median standard error = .015) and smoking marijuana (median β = -.048, median n = 7,903, partial η^2 = .003, median standard error = .010) that have a priori hypothesised negative associations; wearing glasses (median β = -.061, median n = 7,963, partial η^2 = .005, median standard error = .010), being left-handed (median β = -.004, median n = 7,972, partial η^2 < 0.001, median standard error = .010), bicycle use (median β = .080, median n = 7,974, partial η^2 = .007, median standard error = .010) and height (median β = .065, median n = 7,910, partial η^2 = .005, median standard error = .010) that have no *a priori* hypothesised associations (Figure 5).

Discussion

The possibility that adolescents' digital technology use 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 Specification Curve Analysis (SCA) with comparison specifications using three large-scale datasets from the US and UK.

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 (2.7x and 4.3x respectively for YRBS) than technology use does. Positive antecedents of well-being are equally illustrative; simple actions like getting enough sleep and regularly eating breakfast have much more positive associations with well-being than the average impact of technology use (ranging from 1.7x to 44.2x more positive in all datasets). Neutral factors provide perhaps the most useful context to judge technology engagement effects: the association of well-being with regularly eating potatoes was nearly as negative as the association with technology use (0.9x, YRBS) and wearing glasses was more negatively associated with well-being (1.5x, MCS).

With this in mind, the evidence simultaneously suggests technology effects might be statistically significant but so minimal that they hold little practical value. The nuanced picture these results provide are in line with previous psychological and epidemiological research suggesting the associations between digital screen time and child outcomes are not as simple as many might think[11,13]. This work therefore puts previous work that used the YRBS and MTF to highlight technology use as a potential culprit for decreasing adolescent well-being[10] into perspective, showing the range of possible analytical results and comparison specifications. The 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 [29].

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 role that common method variance may play when mapping the effect of digital technology use on adolescent well-being[30]. 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 NHST lens[31]. 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 popular worries about the putative links between technology use and mental health indicators into context. 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 what is driving effects where present. We know very little about whether more technology use might cause lower well-being, whether lower well-being might cause more technology use or whether a third confounding factor underlies both. As we are examining something inherently complex, the likelihood of unaccounted factors affecting both technology use and wellbeing 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 generalisability of the study's inferences. As self-report digital technology measures are known to be noisy[32], this could have also led to the effects of technology on well-being being diminished due to low-quality measurement. Lastly, we used NHST to interpret significance, which is problematic when using such extensive data. To improve, partnerships between research councils and behavioural scientists to better measurement, and pre-registering of analyses plans, will be crucial.

Whether they are collected as part of multi-lab 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 these investments provide an invaluable tool for studying technology effects in young people. To realise this promise, we firmly believe 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 implemented approaches that guard against researcher degrees of freedom. While preregistration might be implausible for analyses of open large-scale social data, methodologies like Specification Curve Analyses provide solutions that don't 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 US and the UK [6–8], encompassing a total of 355,358, predominately 12 to 18-year-old, adolescents surveyed between the years of 2007 and 2016. These datasets were selected because they feature measures of adolescents' psychological well-being, digital technology use, and have been the focus of secondary data analysis to study digital technology effects[10,11,33].

Two of these datasets are based on samples collected in the United States. The first, the Youth Risk and Behaviour Survey (YRBS)[7] launched in 1990, 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 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, *sd* = 1.24). The second U.S. dataset, Monitoring the Future (MTF)[6], launched in 1975 and is an annual nationally-representative survey of approximately 50,000 American 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 was 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 U.K. dataset under analysis was the Millennium Cohort Study (MCS)[8], a prospective study collected in the U.K.; it follows a specific cohort of children born between September 2000 and January 2001. We see this data as particularly high in quality due to its inclusion of pre-tested measures and extensive documentation, highlighting good data collection and project management practices. The data has an over-representation of minority groups and disadvantaged areas due to clustered stratified sampling. Data in this sample is provided by caregivers as well as adolescent participants. In our analysis, we only include data from the primary caregivers and adolescent respondents. The sample under analysis from the MCS was comprised of 5,926 girls and 5,946 boys who ranged in age from 13 to 15 (m = 13.77, sd = .45) and 10,605 primary caregivers.

While the omnibus sample of adolescents is 355,358 teenagers in total, 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 like MTF teenagers only answered a subset of questions. More information about what questions were asked together in 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 MTF. Ethical review and approval for the MCS is monitored by the U.K. National Health Service (NHS) London, Northern, Yorkshire and South-West Research Ethics Committees.

Measures

This study focuses on measures of both digital technology use and psychological well-being. Prior to performing the analysis, all three datasets were reviewed, noting the variables of theoretical interest in each with respect to human behaviour and 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 the adolescents' psychological well-being and functioning. We reversed select measures so that they are all in the same direction, with higher scores indicating higher well-being.

Adolescents were asked five items 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 MTF, participants were asked one of two subsets of self-report questions. The first tranche of participants was asked thirteen questions about their mental health: twelve measures uniquely asked to this subset, and one measure completed by all participants in the survey. The twelve items asked only to this subset included a four-item depressive symptoms scale which studies state to be "similar to those on the Center for Epidemiologic Studies Depression Scale (CES-D)[34] and a self-esteem scale created by Rosenberg[35], both use a disagree-agree Likert scale. Survey administrators also included two additional negatively worded self-esteem measures and a one-item measure asking how happy the participants feel.

There are two kinds of psychological well-being indicators 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 seven-point agree-disagree measures reflecting their subjective sense of well-being and twelve three-point questions tapping into subjective affective states and general mood.[36] Primary caregivers completed the Strengths and Difficulties Questionnaire (SDQ)[37], a well-validated measure of psychosocial functioning, for each adolescent cohort member they took care of (Supplementary Table 5). The SDQ has been used extensively in school, home, and clinical settings with adolescents from a wide range of social, ethnic, and national backgrounds[38]. It includes 25 questions, five each about prosocial behaviour, hyperactivity or inattention, emotional symptoms, conduct problems and peer relationship problems.

Explanatory variables: adolescent technology use. The YRBS dataset included two seven-point technology use questions. One was about the frequency of electronic device use, the other questioned 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 only filling in one part), some questions were filled out by one subset of adolescents, while other questions were filled out by another. One subset answered questions about frequency of social media use and getting information about news from the internet (five-point scale) and two seven-point questions about frequency of watching TV on the weekend and weekday. Another group of MTF participants were asked seven hourly measures of technology use on a nine-point scale. The questions asked about using the internet, playing electronic games, texting on a cell phone, calling on a cell phone, using social media, video chatting and using computers for school. There are, therefore, a total of eleven 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 eight-point items tapping hours per weekday spent watching TV, playing electronic games, spent using the internet at home and using social networking sites. There was also one yes-no measure about whether participants own a computer.

Covariate and confounding variables. Mirroring previous studies analysing data from the MCS[11], we included sociodemographic factors and maternal characteristics as covariates in our analyses. Those include 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[39,32], including analyses of the

MCS[41]. We also included equivalised household income, whether the biological father is present and number of adolescent's siblings in household, as these household factors have also been found to affect adolescent well-being[42]. Furthermore, we include parental behavioural factors such as closeness to parents and the amount of time the primary caretaker spends with the adolescent[43,44]. Addressing previous reports of their influence on child well-being, we additionally use parent reports of any adolescent's long-term illness, and the adolescent's own negative attitudes towards school as covariates[45,46]. 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 YRBS and MTF we included all variables part of the respective questionnaires that conceptually mirrored those covariates utilized in the MCS. For YRBS we included the adolescent's race. For 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 like they can talk with their parents about problems.

Analytic Approach: Specification Curve Analysis

The study implements a Specification Curve Analysis examining the correlation between our explanatory (digital technology engagement) and criterion variables (psychological well-being) using the three-step SCA approach outlined by Simonsohn and colleagues[27] and applied in a recent paper by Rohrer and colleagues[28]. We add a fourth step in order to aid interpretability of our results in the context of large-scale social data. Details of the SCA method and the corresponding visualisations can be found in the Supplementary Methods. All the necessary code 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 the possible analysis pathways that could 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 covariates (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 reflecting 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 analytic 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 MCS, we included a decision of whether to use well-being questions answered by cohort members or those answered by their caregivers, we do not combine the two. For 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 which include only pre-specified well-being questionnaires for MCS (Supplementary Figure 6), however these do not allow comparisons of our SCAs to results of previous work that has selectively combined questions from various datasets[10]. The next analytical decision is what 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 analytic decision taken is whether or not to include covariates in the models. Because of the sheer size of these datasets there is a combinatorial explosion of different covariate combinations that could be used in each regression. We therefore analysed regressions either without covariates or with a pre-specified set of covariates 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 (e.g. the 5-item technology use measures in MTF) or questionably linear (e.g. 3-item happiness measure in MTF). We opted to treat these variables as continuous so that our analyses and results would be directly comparable with those of previous studies[10,33]. Data distribution was assumed to be normal throughout the analysis but is not formally tested for each specification.

Implementing Specifications. Next, for each specification defined we ran the appropriate regression, and noted the standardised β of technology uses' correlation with psychological well-being, the corresponding two-sided p value and the partial η^2 calculated using the R *heplots* package. Listwise deletion for missing data was used as this is more efficient in terms of computational time. This assumes that data is missing completely at random, which could easily not be the case. For example, a child's health, academic performance or socioeconomic 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 standardised β. The median standardised β of all the possible specifications provides a general overview of the effect size. Below that plot, we also indicated which set of analytical decisions led to what standardised β. This allows us to visualise what 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 (i.e. that technology use and adolescent well-being are unrelated). To do so, a bootstrapping technique put forth by Simonsohn et al.[27] was implemented, creating data where the null hypothesis is true by forcing the null on the data. To create this data, the beta-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 is run. This was done 500 times. Once we obtained 500 bootstrapped SCAs, where we knew the null hypothesis was 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 that have larger median effect sizes than the original SCA by the total number of bootstraps to find the *p* value of this test. We repeat 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[23].

Comparison Specifications. Lastly, these analyses were supplemented with a comparison specifications section, putting into context the effects found in the SCA. To do so, we performed a literature review to choose four variables in each dataset that should be positively correlated with psychological well-being, four variables 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 the academy and the public.

Code Availability Statement

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 (https://osf.io/e84xu/).

Data Availability Statement

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 relevant third-party repository after agreement to their terms of usage. Information about data collection and questionnaires can be found on the OSF (https://osf.io/7xha2/).

References

1. Bell, V., Bishop, D. V. M. & Przybylski, A. K. The debate over digital technology and young people. *BMJ* **351,** h3064 (2015).
2. Ofcom. Children and Parents: Media Use and Attitudes Report 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. *Journal of the Association for Consumer Research* (2017). doi:10.1086/691462
5. British Youth Council. A Body Confident Future. (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). (2016). doi:10.3886/ICPSR36799.v1
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, C. for L. S. Millennium Cohort Study: Sixth Survey, 2015. SN: 8156 (2017). doi:http://doi.org/10.5255/UKDA-SN-8156-2
9. Etchells, P. J., Gage, S. H., Rutherford, A. D. & Munafò, M. R. Prospective Investigation of Video Game Use in Children 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. National Health Service. What About Youth study - NHS Digital. (2017). Available at: https://digital.nhs.uk/data-and-information/areas-of-interest/public-health/what-about-youth-study. (Accessed: 5th November 2018)
16. U.S. Department of Health and Human Services, Health Resources and Services Administration, M. and C. H. B. *Child Health USA 2014*. (2014).
17. Livingstone, S., Haddon, L., Görzig, A. & Ólafsson, K. *Technical Report and User Guide: The 2010 EU Kids Online Survey*. (2011).

18. Silberzahn, R. *et al.* Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Adv. Methods Pract. Psychol. Sci.* 251524591774764 (2018). doi:10.1177/2515245917747646

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. Bell, V., Bishop, D. V. M. & Przybylski, A. K. The debate over digital technology and young people. *BMJ* **351,** h3064 (2015).

23. UK Commons Select Committee. Impact of social media and screen-use on young people's health inquiry launched. *Journal of the Association for Consumer Research* (2017). doi:10.1086/691462

24. Chambers, C. D. Registered Reports: A new publishing initiative at Cortex. *Cortex* **49,** 609–610 (2013).

25. Munafò, M. R. *et al.* A manifesto for reproducible science. *Nat. Hum. Behav.* **1,** 0021 (2017).

26. van 't Veer, A. E. Pre-registration in social psychology—A discussion and suggested template. *J. Exp. Soc. Psychol.* **67,** 2–12 (2016).

27. Simonsohn, U., Simmons, J. P. & Nelson, L. D. Specification Curve: Descriptive and inferential statistics on all reasonable specifications. *SSRN Electron. J.* 1–18 (2015). doi:10.2139/ssrn.2694998

28. 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).

29. Department of Health and Social Care. Matt Hancock warns of dangers of social media on children's mental health. *gov.uk* (2018). Available at: https://www.gov.uk/government/news/matt-hancock-warns-of-dangers-of-social-media-on-childrens-mental-health. (Accessed: 8th November 2018)

30. Ferguson, C. J. An effect size primer: A guide for clinicians and researchers. *Prof. Psychol. Res. Pract.* **40,** 532–538 (2009).

31. 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).

32. Scharkow, M. The Accuracy of Self-Reported Internet Use—A Validation Study Using Client Log Data. *Commun. Methods Meas.* **10,** 13–27 (2016).

33. 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* (2018). doi:10.1037/emo0000403

34. 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).

35. 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).

36. 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).

37. 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,** 534–9 (2000).

38. Goodman, R. Psychometric Properties of the Strengths and Difficulties Questionnaire. *J. Am. Acad. Child Adolesc. Psychiatry* **40,** 1337–1345 (2001).

39. Desai, S., Chase-Lansdale, P. L. & Michael, R. T. Mother or Market? Effects of Maternal Employment on the Intellectual Ability of 4-Year-Old Children. *Demography* **26,** 545 (1989).

40. Kiernan, K. E. & Mensah, F. K. Poverty, Maternal Depression, Family Status and Children's Cognitive and Behavioural Development in Early Childhood: A Longitudinal Study. *J. Soc. Policy* **38,** 569 (2009).

41. 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 Studyc ch_1150 44..54. (2010). doi:10.1111/j.1365-2214.2010.01150.x

42. Kiernan, K. E. & Mensah, F. K. Poverty, Maternal Depression, Family Status and Children's Cognitive and Behavioural Development in Early Childhood: A Longitudinal Study. *J. Soc. Policy* **38,** 569 (2009).

43. Thomson, E., Hanson, T. L. & McLanahan, S. S. *Family Structure and Child Well-Being: Economic Resources vs. Parental Behaviors*. Social Forces* **73,** (McLanahan & Booth, 1991).

44. The Children's Society & Barnardo's. *Factors affecting children's mental health over time*. (2018).

45. 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,** (1987).

46. The Children's Society & Barnardo's. *Factors affecting children's mental health over time*. (2018).

Acknowledgements

The National Institute on Drug Abuse provided funding for MTF conducted at the Survey Research Centre in the Institute for Social Research, University of Michigan; YRBS was collected by the Centres for Disease Control and Prevention; The Centre for Longitudinal Studies, UCL Institute of Education collected MCS and the UK Data Archive/UK Data Service provided the data; They bear no responsibility for its aggregation, analysis, or interpretation.

AO was supported by a EU Horizon 2020 IBSEN Grant; AKP was supported by an Understanding Society Fellowship funded by the Economic and Social Research Council. AO's and AKP's funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Thank you to U. Simonsohn, N. K. Reimer and N. Weinstein for their valuable input and 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 (ARC) facility in carrying out this research: http://dx.doi.org/10.5281/zenodo.22558.

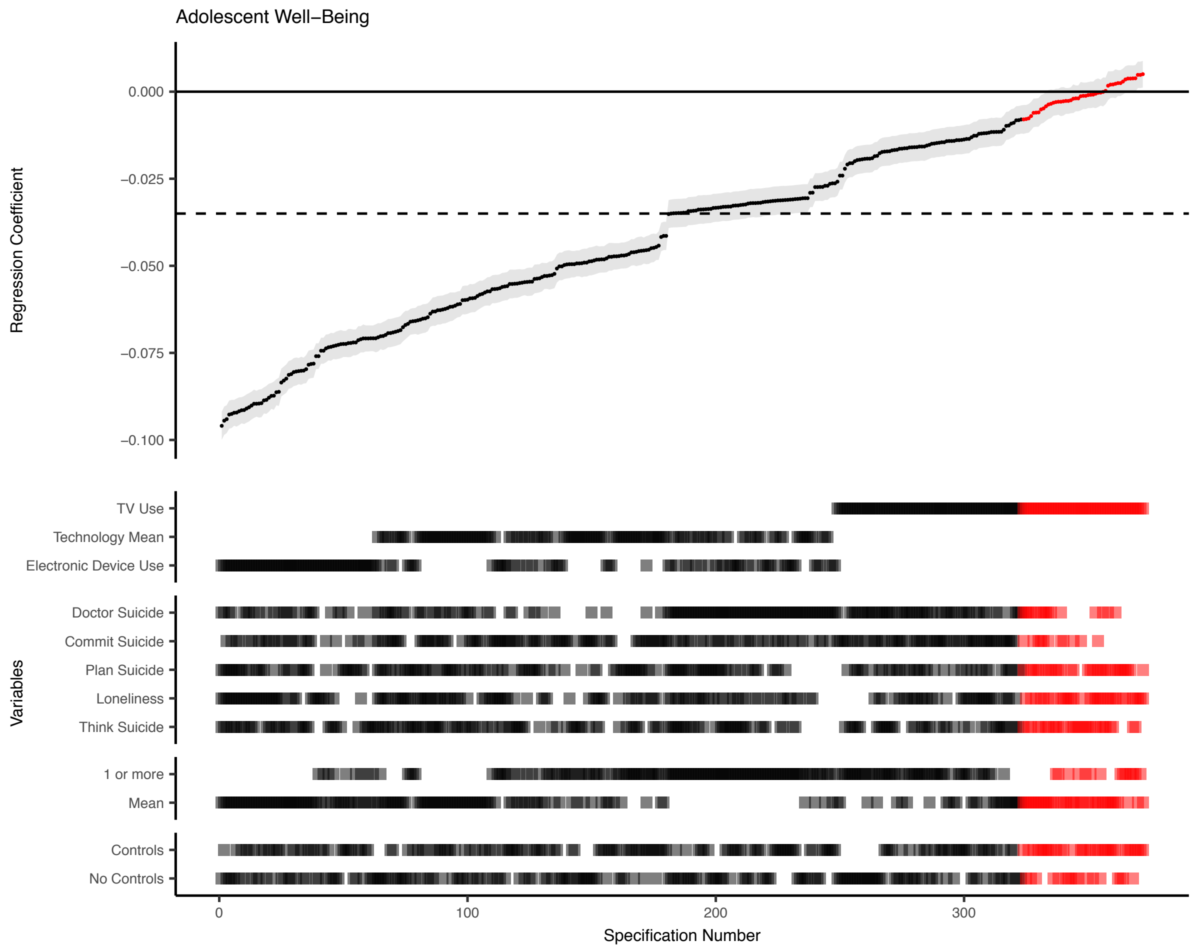

Figure 1: Results of Specification Curve Analysis of the Youth Behaviour and Risk Survey

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 standardised regression coefficient is shown at the top of the graph; the error bars visualise the standard error. Red represents non-significant outcomes, while black represents significant outcomes. To ease interpretation, the dotted line indicates the median standardised regression coefficient found in the Specification Curve Analysis: β = -.035 (median partial η^2 = .001, median n = 62,297, median standard error = .004)

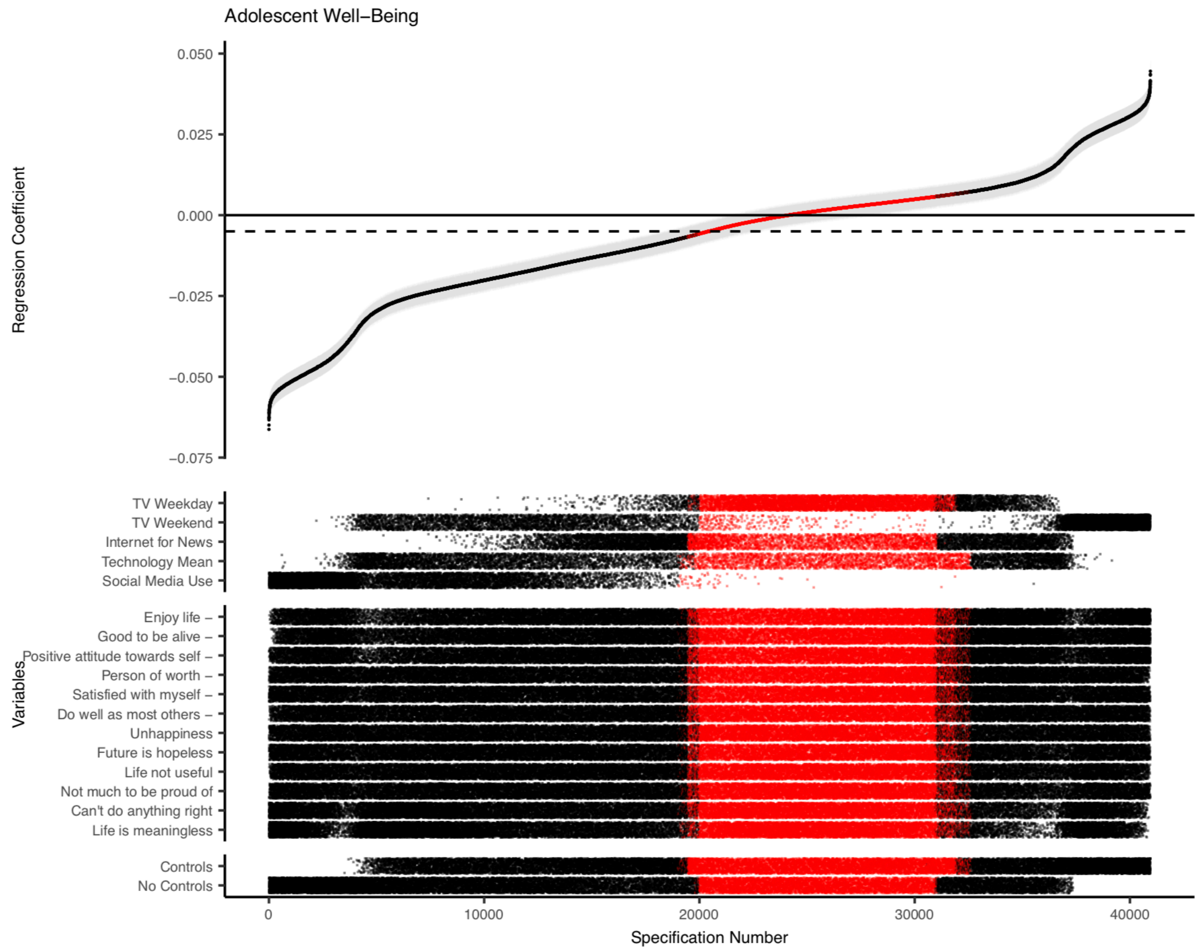

Figure 2: Results of Specification Curve Analysis of the Monitoring the Future study

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 standardised regression coefficient is shown at the top of the graph; the error bars visualise the standard error. Red represents non-significant outcomes, while black represents significant outcomes. To ease interpretation, the dotted line indicates the median standardised regression coefficient found in the Specification Curve Analysis: β = -.005 (partial $\eta^2 < .001$, median n = 78,267, median standard error = .003)

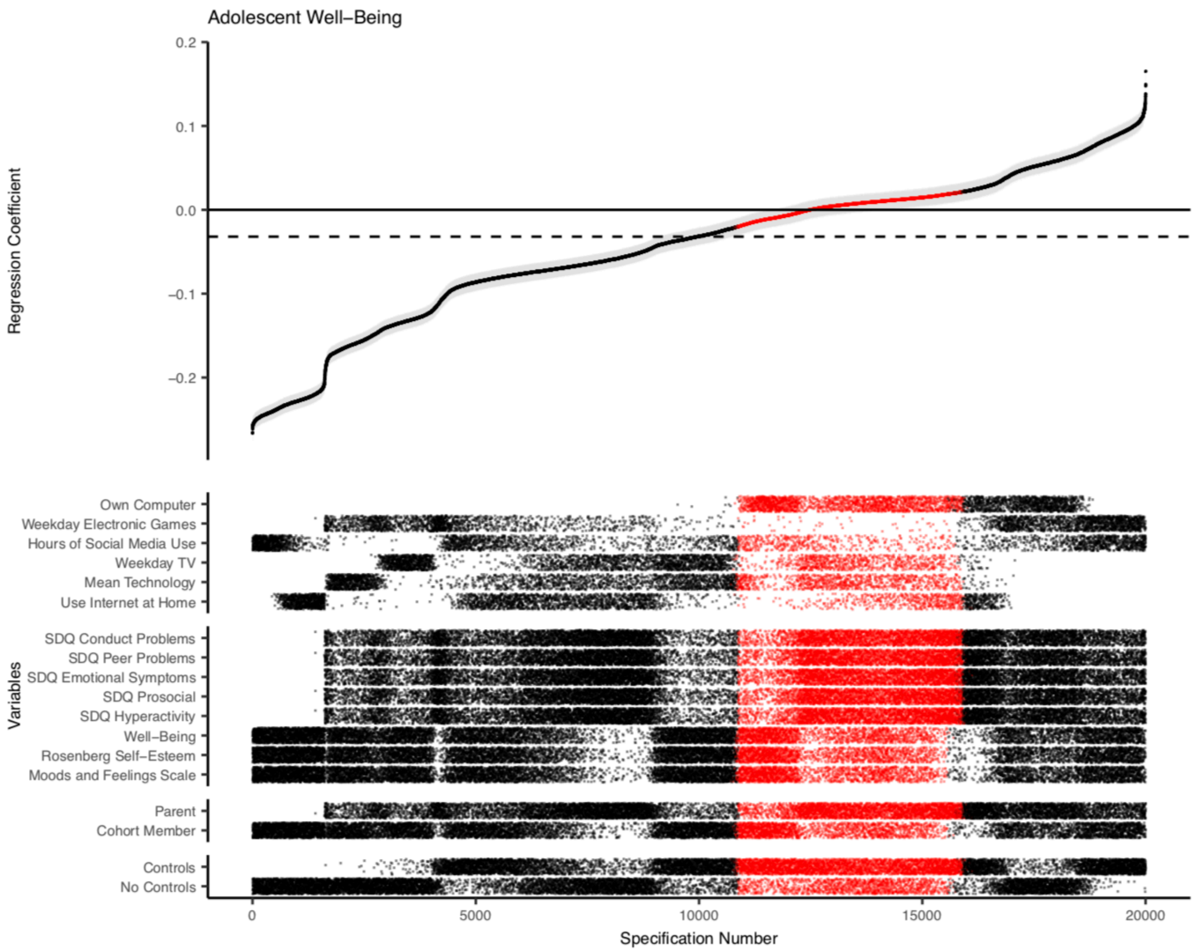

Figure 3: Results of Specification Curve Analysis of the Millennium Cohort Study

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 standardised regression coefficient is shown at the top of the graph; the error bars visualise the standard error. Red represents non-significant outcomes, while black represents significant outcomes. To ease interpretation, the dotted line indicates the median standardised regression coefficient found in the Specification Curve Analysis: β = -.032 (partial η^2 = .004, median n = 7,968, median standard error = .010)

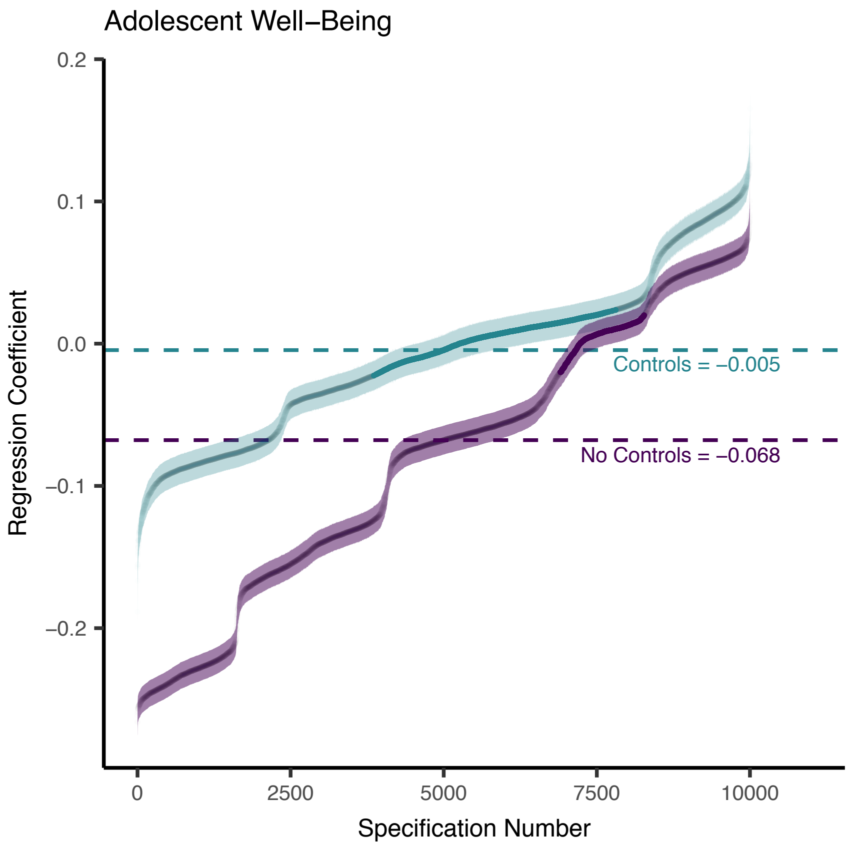

Figure 4: Results of Specification Curve Analysis of the Millennium Cohort Study split by whether controls are included in the analysis or not

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 (standardised regression coefficient) either including control variables (teal, median standardised β = -0.005, partial η^2 = .001, median n = 6,566, median standard error = .011) or not including control variables (purple, median standardised β = -0.068, partial η^2 = .005, median n = 11,018, median standard error = .010). The bolded parts of the line indicate analyses that did not reach significance ($p < 0.05$). The median standardised regression coefficients for analyses including or not including control variables are shown using the dashed lines and the error bars visualise the standard error.

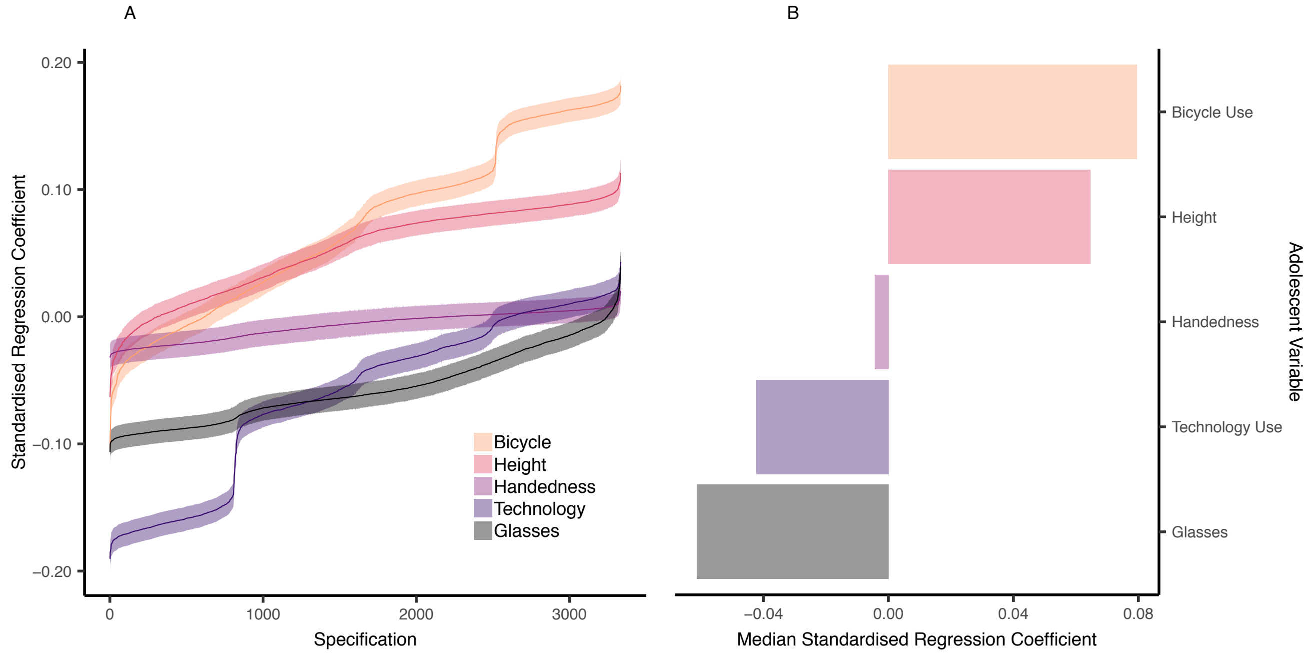

Figure 5: Comparison Specifications of the Millennium Cohort Study

Visualisation of the Comparison Specifications hypothesised to have little or no influence on well-being: bicycle use, height, handedness and wearing glasses. This graph shows Specification Curve Analyses 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* β = -.061, median n = 7,963, partial η^2 = .005, median standard error = .010), more negative than the association of technology use with well-being (purple, *median* β = -.042, median n = 7,964, partial η^2 = .002, median standard error = .010). Handedness (red/purple, *median* β = -.004, median n = 7,972, partial η^2 < 0.001, median standard error = .010), height of the adolescent (red, *median* β = .065, median n = 7,910, partial η^2 = .005, median standard error = .010) and whether the adolescent often rides a bicycle (yellow, *median* β = .080, median n = 7,974, partial η^2 = .007, median standard error = .010) have more positive associations with adolescent well-being than technology use does.

Panel A shows how different analytical decisions (Specifications, shown on the x-axis) lead to different statistical outcomes (Standardised Regression Coefficient, shown on the y-axis). Each line represents a different variable of interest, the error bars represent the standard error. Panel B visualises the resulting Median Standardised Regression Coefficients for those Specification Curve Analyses linking the variables of interest with adolescent well-being.