Reply to: Underestimating digital media harm

Amy Orben,
Emmanuel College, University of Cambridge, MRC Cognition and Brain Sciences Unit, University of Cambridge

Andrew K Przybylski
Oxford Internet Institute, University of Oxford, Department of Experimental Psychology, University of Oxford

In their commentary, Twenge et al.[1] wrongly conclude that the small associations between adolescent well-being and digital technology use presented in Orben and Przybylski[2] are the product of "analytical decisions that resulted in lower effect sizes". While some concerns raised were already addressed in the original paper's supplementary materials, we reject the other concerns by providing additional analyses based on the study data.

Twenge et al.[1] rightly identify Specification Curve Analysis[3] as a powerful tool for exploratory research. SCA highlights how seemingly inconsequential, though equally valid, decisions taken during secondary data analysis can yield divergent results;[4] its value is rooted in providing transparency and context. That understood, Twenge et al.'s[1] analyses, and with it the commentary's references to Kelly et al.[5], are difficult to interpret because the analysis methods of both the commentary and the paper were not consistently reported, not pre-specified prior to data access, and both presented only a subset of defensible analyses that could have been reported.

Twenge et al.[1] suggest the approach taken in Orben and Przybylski[2] minimised the effects reported and proposed alternative analyses they believed would yield larger estimates. For example, they question our selection of a wide range and combination of outcome variables. The supplement in Orben and Przybylski[2] (Supplementary Table 6[2]) demonstrates doing so is prudent because researchers (including the commentators[6]) routinely treat novel combinations of individual survey items as outcome variables instead of using validated aggregate scales presented in the available surveys. Such flexibility extends further to the selection of validated scales; for example, Kelly et al.[5] reported results from only one of the nine well-being measures present in the dataset. While Twenge et al.[1] conclude that the results of this study are "not a case of 'cherry picking'", there is little solid basis for this judgement: no rationale is provided for using only this single operationalisation of well-being.

As researchers routinely alternate between using individual items, unique averages, and subsets of validated scales as outcomes in their studies, we therefore included all three of these analysis strategies in our SCA approach. Instead of eliminating methodological flexibility, our SCA therefore made the pronounced influence of selective reporting of a subset of results visible. That understood, and in line with the initial request of the commentary[1], analyses present in the supplementary materials of our original paper report

results of an SCA using only aggregate scales and Supplementary Figure 5[2] demonstrates our considering individual items did not change the conclusions of our main analyses.

Similarly, another request made by the commentary was directly addressed in our original paper where we detailed that we did not include the hourly technology use scales present in the American MTF dataset; these questions were only asked in conjunction with a low quality one-item life satisfaction measure. These measures were however included in an alternate SCA which was presented in Supplementary Figure 7[2].

The commentator's argued that the addition of certain analyses would have increased the negative correlations found in Orben and Przybylski[2]. To empirically assess the effects of the proposed modifications we re-ran our SCA (see Methods). The results (Figure 1, Table 1) illustrate that the median association and effect size (β = - 0.051 [-0.072, -0.031], percent variance explained = 0.3% [0.2, 0.6]) are not significantly different from those in our original SCA (median β = - 0.032, percent variance explained = 0.4%). Wearing glasses was still more negatively associated with well-being in adolescents than digital technology use ($\beta_{glasses} = -0.061$ vs $\beta_{technology} = -0.051$). Further, the requested self-harm measure did not produce a significantly more negative association than the complete SCA (median β = - 0.054 [-0.064, -0.046]). In line with our most recent work[7], females showed a more negative association than males (females: median β = - 0.069 [-0.074, -0.065], males: median β = - 0.037 [-0.041, -0.032]) and technology use predicted nearly one half of one percent of the variability in the well-being of girls.

Further, by applying non-linear modelling methodology as requested (Table 1), we found that the average association between digital technology use and well-being was slightly positive for low-intensity users (median effect = 0.018, bootstrapped 95% CI [0.001, 0.018]) and slightly negative for high-intensity users (median effect = -0.080, bootstrapped 95% [-0.094, -0.076]). This finding conceptually and statistically replicates a previously uncovered pattern of small non-monotonic relationships linking technology use to well-being outcomes[8]. It is important to note that the strength of this association is probably not practically significant: less than 1% of variance in adolescent well-being is accounted by either these positive or negative trends. With this understood, we strongly caution against approaches like those proposed by the commentary authors such as comparing pairs of extreme values. Picking arbitrary cut-offs is likely to yield spurious findings[9].

Finally, Twenge et al.[1] assert that percentage of variance explained is not a useful metric for evaluating the size of an effect. Firstly, in Orben and Przybylski[2] we included a wide variety of effect size measurements, leaving it to the reader to select the measure they wish to interpret. Furthermore, unlike Twenge and colleagues[1], we believe that r^2 is a useful effect size measure in the context of our study as it aims to quantify how influential the predictor of interest is while taking into account other predictors of the outcome that were not measured or included in the analyses. Comparing mean differences, as proposed by Twenge et al.[1], requires assuming there are no influential third factors that could be influencing both the amount of technology used and the well-being of adolescents (e.g. socioeconomic status or parenting practices). For example, if we were to compare adolescents who listened to five hours or more of music a day (instead of focusing on technology), we would find that they

have 16% lower well-being than those who listened to only ½ an hour or less – on this basis we might infer, incorrectly, that music has a negative impact on well-being. Instead, we presented comparison specifications (Extended Data 1) to enable readers to gauge, in more detail, the extent to which technology effects are practically significant. These analyses showed, for example, that wearing glasses had a greater negative association with well-being than using digital technologies. Lastly, we have since publication of our original article expanded our effect size approach. For example, we have recently anchored effects to thresholds of subjective awareness[10,11], finding that adolescents might have to use screens for upwards of 11 hours and 14 minutes each day before they would be able to perceive any negative effects.

To conclude, on balance we are optimistic that SCA is a powerful tool to facilitate constructive and evidence-based dialogue. We presented new analyses underscoring its value – we tested the additional specifications requested by the commentary authors and the results did not materially affect the conclusions of our original study. By adopting more transparent approaches to analysing data, and sharing how results are affected by different analytical methods we hope the debate about teens and screens will avoid further ideological entrenchment.

Methods

We chose the UK Millennium Cohort Study for our new SCA presented in this reply as it was the dataset also used by Kelly et al.[5]. The new analyses included all the additional specifications requested by the commentators[1]: (1) only validated well-being (sub-)scales, (2) a new self-harm outcome measure (an earlier version of the commentary[1] questioned why this outcome was not included in our initial analyses, but it was removed during peer review), (3) removed control variables that the commentary authors took issue with (seen as potential moderators, we believed they were potential confounders), (4) added another specification which only includes ethnicity as a demographic-only control variable, and (5) examined gender separately. We could not adjust our analyses to add the commentary's request to include the improved measures available in the MTF dataset as suitable measures were not present in the MCS dataset.

We analysed the non-linear associations using an interrupted regressions approach. We calculated separate regression models for both low- and high-intensity users — adolescents below and above the median in terms of levels of engagement with digital technology (Table 1).

The code used to analyse the relevant data can be found on the Open Science Framework (https://osf.io/byqm5/?view_only=f15a442c7225477eb8378c1e717fbd8a). For brevity, we have not included all sample size measurements in this manuscript, but they can be found in the data files on the OSF as well.

The data that support the findings of this reply are available from 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/?view_only=a24cc2ddeceb42bd85dbf3b2babc9a16).

Extended Data
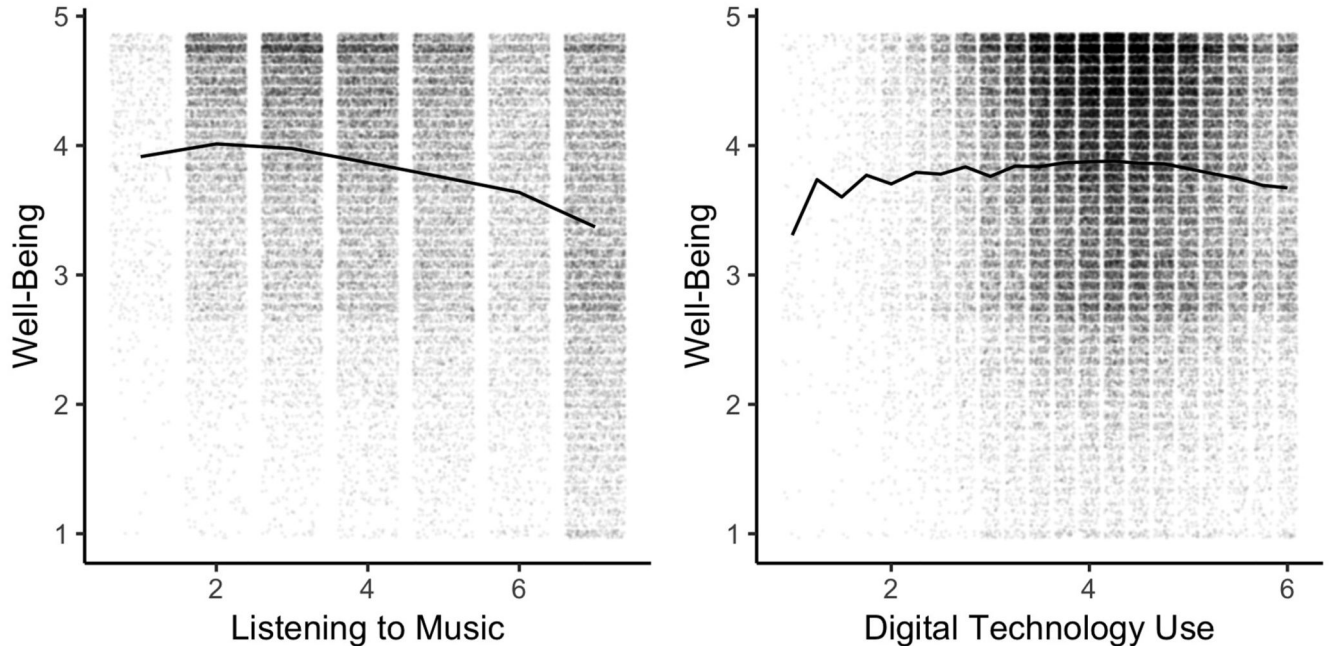
Extended Data 1.
Plots of the variables listening to music and mean digital technology use against the mean of all well-being measures in the survey. The small points are individual participants, while the lines are mean values.

References

1. Twenge JM. Reply to: Adolescent technology and well-being. Nat Hum Behav.

2. Orben A, Przybylski AK. The association between adolescent well-being and digital technology use. Nat Hum Behav. 2019; 3:173–182. [PubMed: 30944443]

3. Simonsohn U, Simmons JP, Nelson LD. Specification Curve: Descriptive and inferential statistics on all reasonable specifications. SSRN Electronic Journal. 2015; doi: 10.2139/ssrn.2694998

4. Silberzahn R, et al. Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. Adv Methods Pract Psychol Sci. 2018; 1:337–356.

5. Kelly Y, Zilanawala A, Booker C, Sacker A. Social Media Use and Adolescent Mental Health: Findings From the UK Millennium Cohort Study. EClinicalMedicine. 2019; 6:59–68. [PubMed: 31193561]

6. Twenge JM, Joiner TE, Rogers ML, Martin GN. 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. 2017; 6:3–17.

7. Orben A, Dienlin T, Przybylski AK. Social media's enduring effect on adolescent life satisfaction. Proc Natl Acad Sci U S A. 2019; 116:10226–10228. [PubMed: 31061122]

8. Przybylski AK, Weinstein N. A Large-Scale Test of the Goldilocks Hypothesis. Psychol Sci. 2017; 28:204–215. [PubMed: 28085574]

9. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ. 2006; 332:1080. [PubMed: 16675816]

10. Orben A, Przybylski AK. Screens, Teens, and Psychological Well-Being: Evidence From Three Time-Use-Diary Studies. Psychol Sci. 2019; 30:682–696. [PubMed: 30939250]

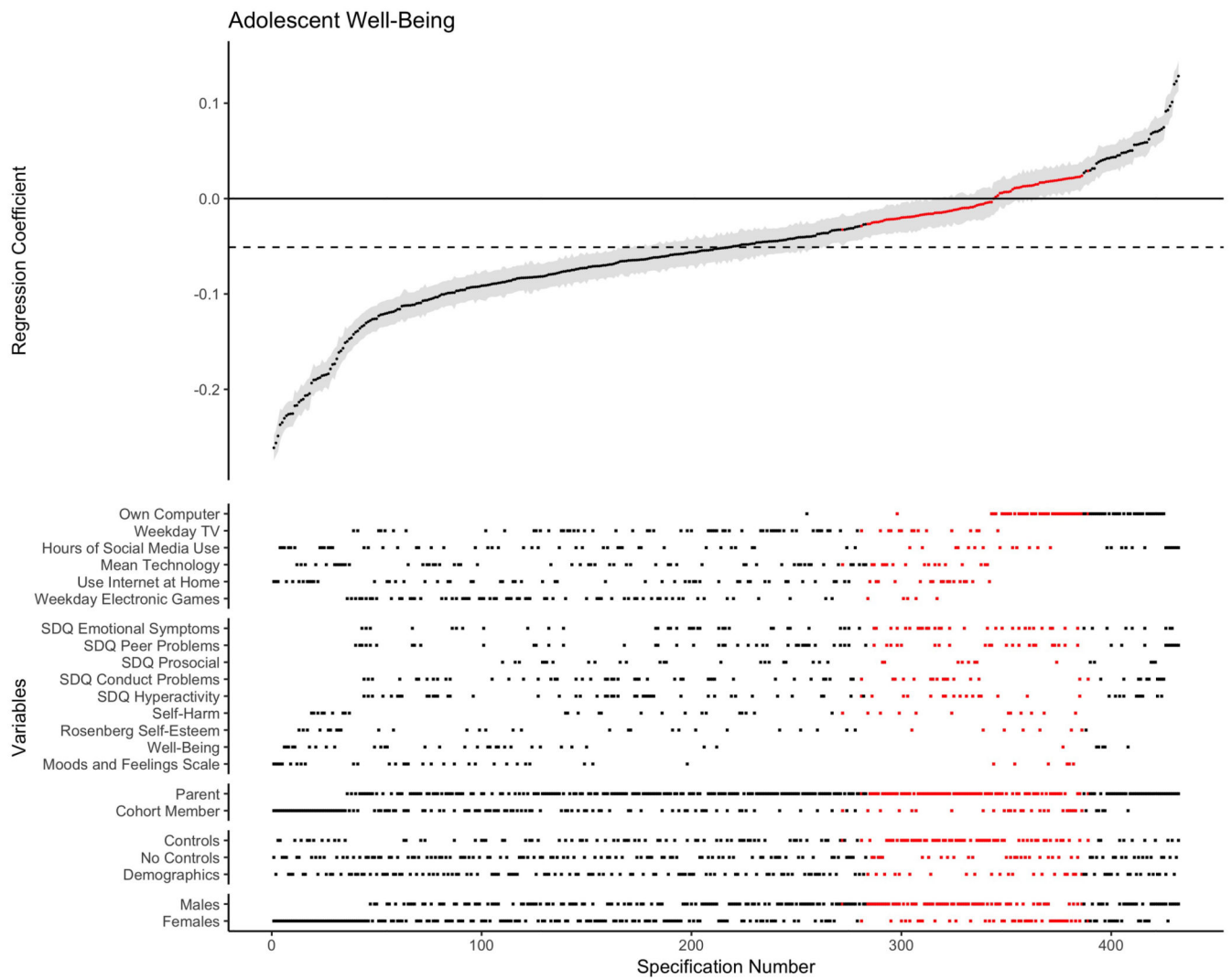11. Orben A. Teens, Screens and Well-Being: An Improved Approach. 2019

Figure 1.
Specification Curve Analysis of MCS showing associations between prespecified well-being subscales and technology for males and females. The red squares indicate specifications that were non-significant, while the black squares represent specifications that were significant. The error bars represent the Standard Error. The dotted line indicates the median Regression Coefficient present in the SCA and is included to ease interpretation of the analysis.

Table 1

Results of the additional SCA analysis implementing the changes suggested by the commentary. The table shows the results both for the complete SCA and specific subgroups. The Confidence Intervals are created by bootstrapping (500 iterations).

Group	Percentage Significant	Median Effect	Median Eta-Squared	Median Number	Median Standard Error
All	75.5 [58.3, 86.8]	-0.051 [-0.072, -0.031]	0.003 [0.002, 0.006]	5272	0.014
Males	69.4 [65.7, 74.1]	-0.037 [-0.041, -0.032]	0.002 [0.002, 0.003]	5248	0.014
Females	81.5 [79.2, 85.0]	-0.069 [-0.074, -0.065]	0.005 [0.005, 0.006]	5295.5	0.014
Controls	58.3 [54.9, 66.0]	-0.033 [-0.039, -0.028]	0.002 [0.002, 0.002]	3491.5	0.016
No Controls	82.6 [79.2, 86.8]	-0.060 [-0.065, -0.054]	0.004 [0.004, 0.005]	5587	0.013
Demographics	85.4 [80.6, 88.2]	-0.058 [-0.065, -0.055]	0.004 [0.004, 0.005]	5272	0.014
Self-Harm	66.7 [61.1, 80.6]	-0.054 [-0.064, -0.046]	0.003 [0.002, 0.004]	5317	0.014
Low Intensity Users	37.6 [31.9, 51.4]	0.018 [0.001, 0.018]	0.001 [0.001, 0.002]	2638	0.020
High Intensity Users	76.4 [72.2, 83.3]	-0.080 [-0.094, -0.076]	0.007 [0.006, 0.009]	2638	0.019