MATTERS ARISING
https://doi.org/10.1038/s41562-020-0839-4

nature human behaviour

Underestimating digital media harm

Jean M. Twenge[1]✉, Jonathan Haidt[2], Thomas E. Joiner[3] and W. Keith Campbell[4]

ARISING FROM A. Orben & A. K. Przybylski *Nature Human Behaviour* https://doi.org/10.1038/s41562-018-0506-1 (2019)

Orben and Przybylski[1] use a new and advanced statistical technique to run tens of thousands of analyses across three large datasets. The authors conclude that the association of screen time with well-being is negative but "too small to warrant policy change." However, Orben and Przybylski made six analytical decisions that resulted in lower effect sizes, and their conclusions are in stark contrast with the practically important differences identified in other analyses of the same datasets, especially for social media use among girls.

For example, Kelly et al.[2] used the Millennium Cohort Study (MCS)—the same dataset Orben and Przybylski described as "the highest-quality dataset we examined". Kelly et al.[2] found that twice as many heavy users of social media (versus non-users) had clinically relevant symptoms of depression (see Fig. 1). This difference is large enough to justify concern and perhaps policy changes; it is also consistent with the conclusions of many other studies[3].

How can these results be reconciled with those of Orben and Przybylski? Kelly et al.[2] used the standard, validated measure of depressive symptoms as well as standard demographic controls, minimizing the role of researcher degrees of freedom. This is not a case of cherry picking. The discrepancy may instead lie in six choices made by Orben and Przybylski.

The first issue is the consideration of only monotonic effects. Associations between digital media use and well-being are often non-monotonic; in fact, Przybylski himself named this the Goldilocks hypothesis[4]. Associations often follow a J-shaped curve (see Extended Data Fig. 1). Thus, statistics other than linear r may be necessary to capture the effects. One possibility is relative risk. For internet use and self-harm among boys in the MCS cohort with demographic controls, linear $r = 0.06$ (or 0.36% of the variance), but those spending ≥2 h on the internet are 48% more likely to self-harm than those spending <2 h on the internet (relative risk = 1.48; 95% confidence interval = 1.15–1.92; 5.9% versus 8.8%). Those spending ≥7 h on the internet are more than twice as likely to self-harm than those spending between 30 min and 1 h on the internet (relative risk = 2.15; 95% confidence interval = 1.31–3.55; 5.1 versus 10.7%). Perhaps because relative risk can account for non-monotonic patterns, it indicates a much stronger relationship.

The second issue is the aggregation of data across screen time types and gender. The mental health crisis among adolescents that began after 2012 is hitting girls far harder than boys, in multiple countries[5]. Thus, it is vital that researchers pay special attention to girls, and to the types of media that became more popular after 2012. Given that television has existed for 70 years and television watching declined among adolescents after 2012[6], television is not a primary concern. In contrast, social media use became pervasive among teens just before 2012 and is used much more heavily by girls, making it a prime suspect.

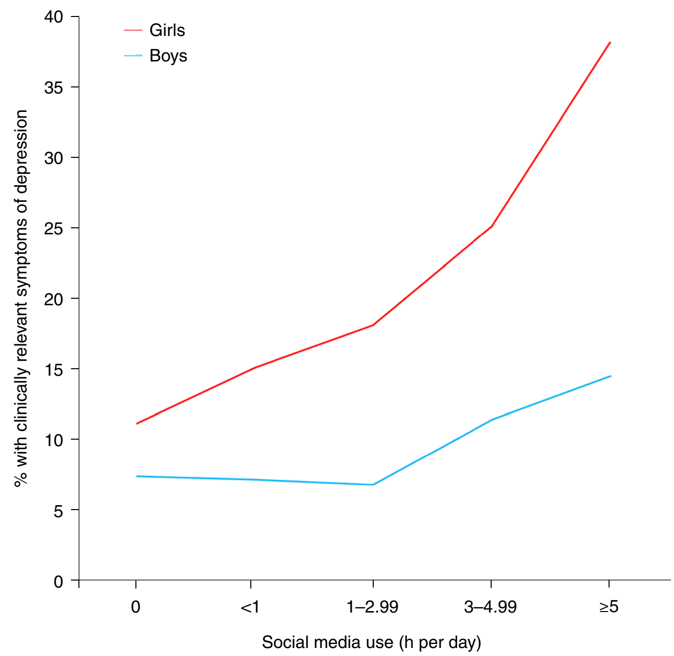


Fig. 1 | Percentage of the MCS cohort with clinically relevant depressive symptoms as a function of social media use. The results are displayed separately by gender.

Most of Orben and Przybylski's comparisons combine all types of screen time, and none separate by gender. Associations between well-being and social media use for girls are considerably stronger than associations with television or for boys (see Fig. 2). In this case, there are theoretical and practical reasons to focus on certain comparisons.

The third issue is the use of individual items. Orben and Przybylski's effect sizes include many individual items, which are lower in internal reliability than multiple-item scales[7] and thus produce lower effect sizes. In addition, scales with more items count more heavily in the analysis—not because they are more important, but because of the arbitrary fact of having more items.

The fourth issue is missing measures. The Monitoring the Future dataset[8] measures digital media use in two ways: (1) on a scale of 'never' to 'almost every day', which has very low variance, as the vast majority of teens now use digital media every day; and (2) in hours per week, which has sufficient variance. Surprisingly, Orben and Przybylski did not include the Monitoring the Future

[1]Department of Psychology, San Diego State University, San Diego, CA, USA. [2]Stern School of Business, New York University, New York, NY, USA. [3]Department of Psychology, Florida State University, Tallahassee, FL, USA. [4]Department of Psychology, University of Georgia, Athens, GA, USA. ✉e-mail: jtwenge@sdsu.edu

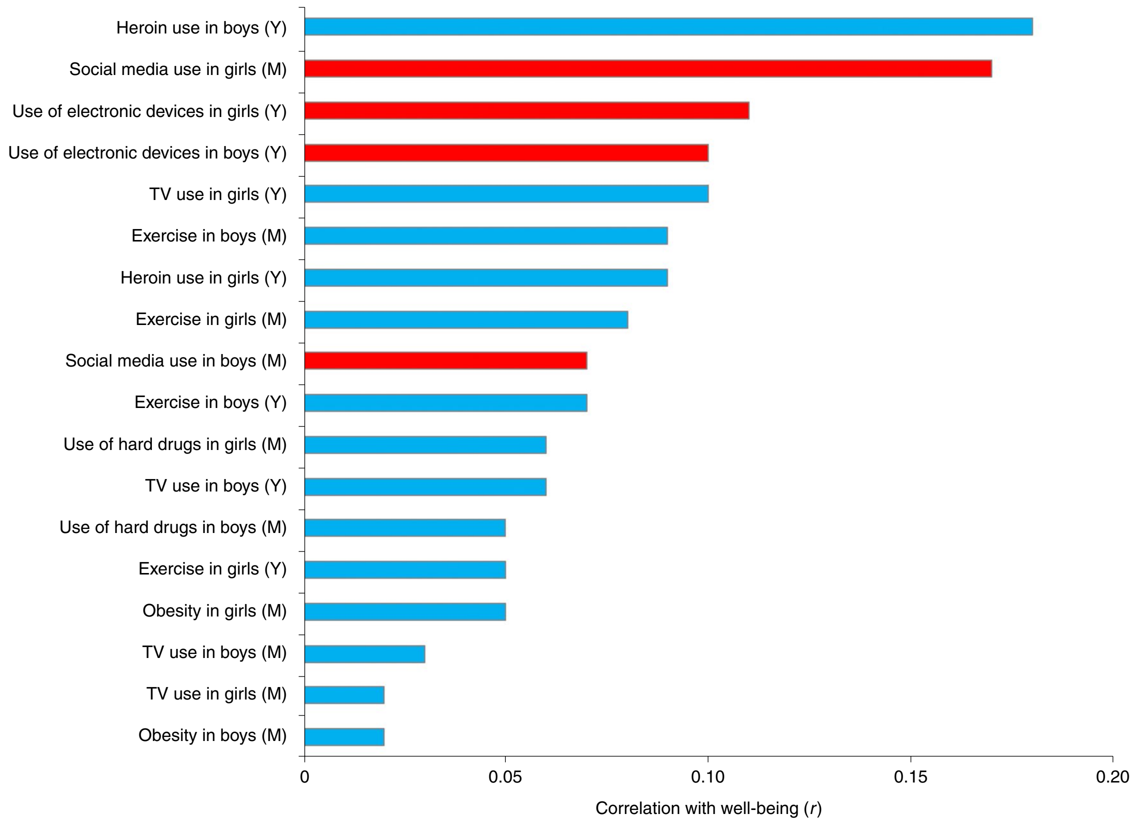


Fig. 2 | Average linear *r* values between well-being and various factors in boys and girls from two datasets. An 'M' or 'Y' in parentheses indicates that the data were from the MCS cohort or 2009–2015 or from the Youth Risk Behavior Surveillance System (YRBSS), respectively. Well-being for the MCS cohort is the average *r* for depression, self-harm, self-esteem, life satisfaction and the Strengths and Difficulties Questionnaire. Well-being for the YRBSS is the average *r* for depression, suicidal thoughts, suicide plans, suicide attempts and having at least one of these suicide risk factors. Both include demographic controls (for the MCS: age, ethnicity, weekly family income, number of children in household, natural father present in household, longstanding illness, parent employment, parent educational attainment and parent's vocabulary score; for the YRBSS: race (black, hispanic, other or white) and grade). Exercise was reverse scored. Obesity was defined using body mass index cut-offs. The bars are coloured red to indicate digital media use and blue for television and comparison variables.

hours-per-week items on non-television digital media (social media, internet use, gaming, texting and video chat); they only included the low-variance items. The low-variance items produce substantially lower linear *r* values. For example, the *r* value for happiness and social media use on the low-variance item is −0.01, compared with −0.09 when measured in hours (Table 2 in ref. [9]). Although Supplementary Fig. 5 in ref. [1] lists these hourly items, it does not report any statistics using them. In addition, Orben and Przybylski do not include the measure of self-harm behaviours included in the MCS.

The fifth issue is the inclusion of controls that may be mediators. Orben and Przybylski control for factors such as negative attitudes towards school, time spent with parents, parent distress and closeness to parents, which could be caused by heavy social media use. The types of factors that should be controlled for include confounding variables, such as demographics. They should not include possible mediators as this might explain away mechanisms of interest[10,11]. In addition, 'negative attitudes towards school' is itself a measure of low well-being (one item is 'How often do you feel unhappy at school?'). Among girls, social media use and the five well-being measures average $r=-0.17$ with demographic controls only, but r drops to −0.05 with the potential mediators included.

The sixth issue is the use of r^2 as an effect size. Whether or not r^2 is useful is debated[12,13], but r^2 clearly produces smaller numbers (for example, $d=0.30$ is equivalent to $r=0.15$ and $r^2=2.3\%$). Especially for lower-base-rate events, r and r^2 may be low even when the difference between groups is large. The polio vaccine explained a tiny 0.0001% of the variance in whether children caught polio, but unvaccinated children were more than three times more likely to catch polio than vaccinated children (relative risk = 3.00)[14]. A major-league baseball player's skill explains <1% of the variance in the outcome of a single at-bat (each time a player faces a pitcher)[15]. The same pattern appears in social media and clinical outcomes, such as the doubling of relative risk for depression found by Kelly et al.[2]—a practically important difference that lies far from Orben and Przybylski's conclusion that these associations are trivial.

Orben and Przybylski included comparisons with other variables, such as wearing glasses and potato eating, as a gauge of practical importance. However, it is unclear how these specific comparison variables were chosen. Choosing different comparison variables might have led to very different conclusions. One could just as easily conclude that social media use is more important for well-being than hard drug use, exercise and obesity (see Fig. 2). For girls, social media use is far more important than exercise, or even than heroin use. Thus, given the potential to choose variables at will, using comparison variables may not be a useful solution for conveying practical importance to readers.

In conclusion, we agree with Orben and Przybylski's warning that researchers are far too free to pick and choose variables and analytical strategies when analysing large correlational datasets. However, their analyses do not solve the problems of non-transparency or researcher degrees of freedom, and we think this is an important topic for discussion. The three datasets analysed by Orben and Przybylski actually contain a warning for parents: heavy use of social media is consistently associated with negative mental health outcomes, at non-trivial levels, especially for girls. Given the rapidly rising rates of anxiety, depression, self-harm and suicide currently affecting girls in the United States and other countries[5,16], for which no one has found a plausible alternative explanation, we believe that these associations should not be dismissed.

Received: 15 March 2019; Accepted: 16 February 2020;
Published online: 17 April 2020

References

1. Orben, A. & Przybylski, A. K. The association between adolescent well-being and digital technology use. *Nat. Hum. Behav.* **3**, 173–182 (2019).
2. Kelly, Y., Zilanawala, A., Booker, C. & Sacker, A. Social media use and adolescent mental health: findings from the UK Millennium Cohort Study. *EClinicalMedicine* **6**, 59–68 (2018).
3. Twenge, J. M. More time on technology, less happiness? Associations between digital media use and psychological well-being. *Curr. Dir. Psychol. Sci.* **28**, 372–379 (2019).
4. Przybylski, A. K. & Weinstein, N. A large-scale test of the Goldilocks hypothesis: quantifying the relations between digital-screen use and the mental well-being of adolescents. *Psychol. Sci.* **28**, 204–215 (2017).
5. McManus, S., Gunnell, D., Cooper, C., Bebbington, P. E., Howard, L. M. & Brugha, T. Prevalence of non-suicidal self-harm and service contact in England, 2000–14: repeated cross-sectional surveys of the general population. *Lancet Psychiat.* **6**, 573–581 (2019).
6. Twenge, J. M., Martin, G. N. & Spitzberg, B. H. Trends in U.S. adolescents' media use, 1976–2016: the rise of digital media, the decline of TV, and the (near) demise of print. *Psychol. Pop. Media Cult.* **8**, 329–345 (2019).
7. Cronbach, L. J. & Meehl, P. E. Construct validity in psychological tests. *Psychol. Bull.* **52**, 281–302 (1955).
8. Johnston, L. D., O'Malley, P. M., Bachman, J. G., Schulenberg, J. E., & Miech, R. A. *Monitoring the Future: National Survey Results on Drug Use, 1975–2016.* (University of Michigan, National Institute on Drug Abuse at the National Institute of Health, 2018).
9. 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* **18**, 765–780 (2018).
10. Schisterman, E. F., Cole, S. R. & Platt, R. W. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* **20**, 488–495 (2009).
11. Rohrer, J. M. Thinking clearly about correlations and causation: graphical causal models for observational data. *Adv. Methods Pract. Psychol. Sci.* **1**, 27–42 (2018).
12. Ferguson, C. J. An effect size primer: a guide for clinicians and researchers. *Prof. Psychol. Res. Pract.* **40**, 532–538 (2009).
13. Funder, D. C. & Ozer, D. J. Evaluating effect size in psychological research: sense and nonsense. *Adv. Methods Pract. Psychol. Sci.* **2**, 156–168 (2019).
14. Rosnow, R. L. & Rosenthal, R. Effect sizes for experimenting psychologists. *Can. J. Exp. Psychol.* **57**, 221–237 (2003).
15. Abelson, R. P. A variance explanation paradox: when a little is a lot. *Psychol. Bull.* **97**, 129–133 (1985).
16. Twenge, J. M. Why increases in adolescent depression may be linked to the technological environment. *Curr. Opin. Psychol.* **32**, 89–94 (2020).

Author contributions

J.M.T. developed the study concept, analysed the data and wrote the first draft of the manuscript. J.H., W.K.C. and T.E.J. provided crucial revisions.

Competing interests

J.M.T., J.H., W.K.C. and T.E.J. have received speaking honoraria and consulting fees from non-profit and for-profit entities for presenting research.

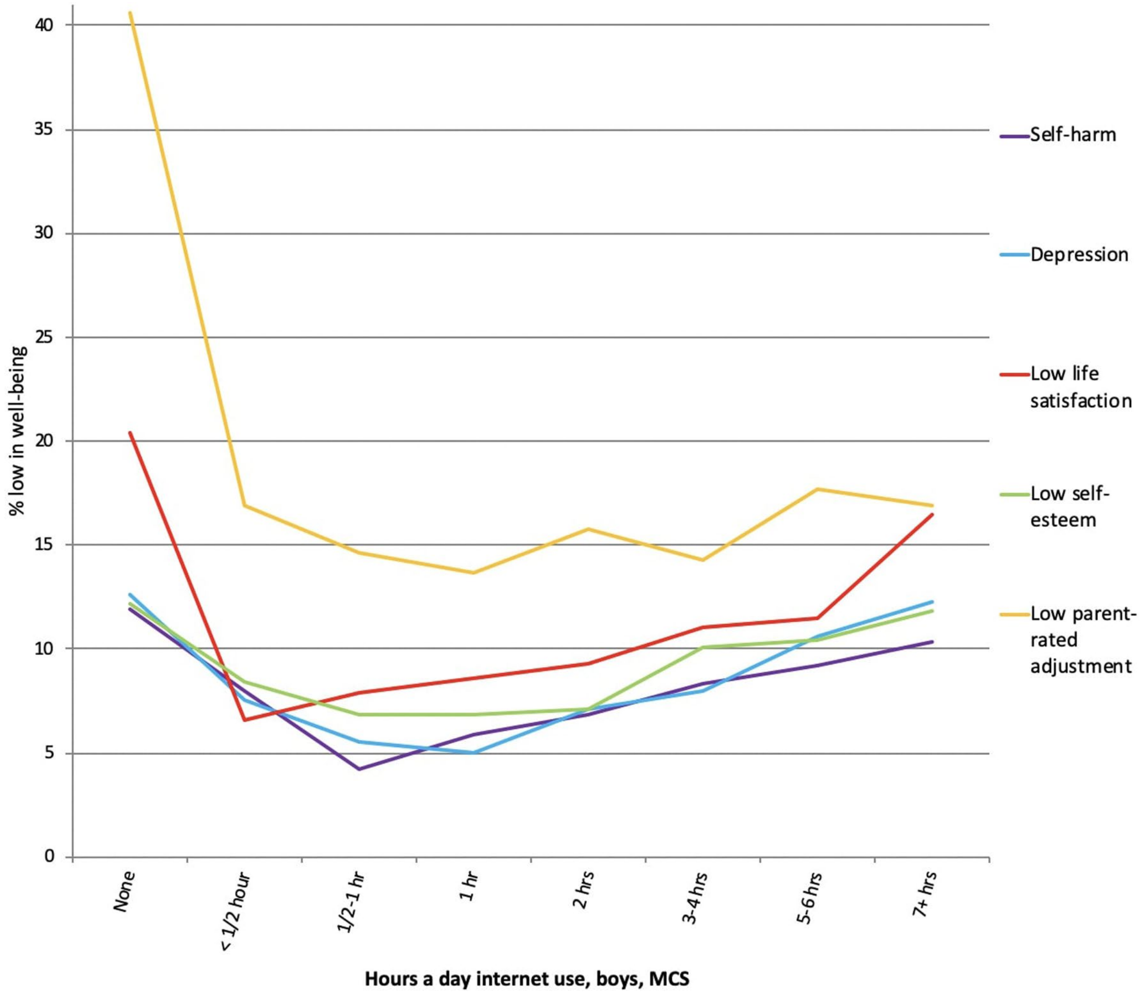
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-020-0839-4.

Correspondence and requests for materials should be addressed to J.M.T.

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

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



Extended Data Fig. 1 | Associations between internet use and well-being indicators, boys, Millennium Cohort Study (includes demographic controls).