A Report on "Specification Curve Analysis Shows that Social Media Use is Linked to Poor Mental Health, Especially Among Girls" by Twenge et al. (2022)

Reviewer 2

February 11, 2026

v2

isitcredible.com

Disclaimer

This report was generated by large language models, overseen by a human editor. It represents the honest opinion of The Catalogue of Errors Ltd, but its accuracy should be verified by a qualified expert. Comments can be made here. Any errors in the report will be corrected in future revisions.

> I am wiser than this person; for it is likely that neither of us knows anything fine and good, but he thinks he knows something when he does not know it, whereas I, just as I do not know, do not think I know, either. I seem, then, to be wiser than him in this small way, at least: that what I do not know, I do not think I know, either.
>
> Plato, *The Apology of Socrates*, 21d

> To err is human. All human knowledge is fallible and therefore uncertain. It follows that we must distinguish sharply between truth and certainty. That to err is human means not only that we must constantly struggle against error, but also that, even when we have taken the greatest care, we cannot be completely certain that we have not made a mistake.
>
> Karl Popper, 'Knowledge and the Shaping of Reality'

Overview

Citation: Twenge, J. M., Haidt, J., Lozano, J., & Cummins, K. M. (2022). Specification Curve Analysis Shows that Social Media Use is Linked to Poor Mental Health, Especially Among Girls. *Acta Psychologica*, 224, 103512.

URL: https://doi.org/10.1016/j.actpsy.2022.103512

Abstract Summary: This paper re-runs a Specification Curve Analysis (SCA) on three large-scale adolescent datasets, applying revised constraints (separating media types, separating sexes, excluding mediators, and treating scales equally) to re-examine the link between technology use and mental health. The revised analysis finds a consistent and substantial association between social media use and poor mental health, particularly among girls, contradicting previous findings that suggested the association was tiny.

Key Methodology: Specification Curve Analysis (SCA) re-run on three large-scale community datasets (Millennium Cohort Study, Monitoring the Future, Youth Risk Behavior Surveillance System) using revised analytic constraints.

Research Question: Does the association between adolescent technology use and mental health remain tiny when Specification Curve Analysis (SCA) is re-run using alternative, theoretically defensible analytic constraints?

Editor's Note

Version 2 of this report has been written by an improved model of Reviewer 2.

Summary

Is It Credible?

This article serves as a direct rebuttal to a prominent 2019 study by Orben and Przybylski (O&P), which utilized Specification Curve Analysis (SCA) to argue that the association between adolescent screen time and mental well-being is negligible—comparable to the association between well-being and eating potatoes. Twenge et al. reanalyze the same three large datasets (Millennium Cohort Study, Monitoring the Future, and Youth Risk Behavior Surveillance System) using the same statistical technique but with altered analytical constraints. Their headline claim is that when the analysis focuses specifically on girls, isolates social media from general screen time, and removes certain control variables, the association becomes "consistent and substantial," with median standardized βs rising from negligible levels to between -0.10 and -0.24 (p. 1). They argue this magnitude is practically significant, comparable to risk behaviors like binge drinking or heroin use (p. 5).

The credibility of these findings rests entirely on the validity of the authors' "revised constraints." The article successfully demonstrates that SCA, despite being designed to minimize researcher bias, remains highly sensitive to subjective analytical choices. However, the authors' decision to exclude specific control variables is the primary driver of the increased effect sizes, and this decision is methodologically contentious. In the Millennium Cohort Study (MCS), the median β for girls' social media use jumps from -0.01 to -0.20 only when "potential mediators" such as negative attitudes toward school, closeness to parents, and school grades are removed from the model (p. 4). The authors justify this exclusion by arguing that controlling for mediators causes overadjustment bias, citing methodological literature. This assumes a specific causal pathway: social media use causes school/family problems, which then cause poor mental health.

This causal assumption is not substantiated by the cross-sectional data available. It is equally plausible that a third variable—such as a difficult home environment or a personality trait like neuroticism—causes both increased social media use and negative attitudes toward school. If these variables are confounders rather than mediators, excluding them introduces omitted variable bias, inflating the observed association. Consequently, the "substantial" link reported may be an artifact of this specific analytical choice. The article acknowledges that "social media should remain on the list of possible explanations" (p. 10), yet the framing throughout strongly implies causality, describing social media as a "prime suspect" in the "mystery" of rising mental health issues (p. 1).

Furthermore, the consistency of the findings across datasets is overstated due to measurement limitations. In the YRBSS dataset, the authors rely on a variable for "electronic device use" as a proxy for social media. As the article notes, this question includes "computers and gaming consoles" (p. 6). While the authors attempt to isolate the relevant years, this remains a heterogeneous measure that conflates social media with gaming and other digital activities, weakening the claim that the specific harm of social media has been isolated across all three studies. Additionally, the critique regarding the weighting of mental health scales in the MCS—where the authors move from O&P's approach to weighting all scales equally—is presented as a correction for validity, preventing one scale (the SDQ) from dominating the specification space simply because it has multiple subscales. While this is a reasonable methodological correction to avoid redundancy, it also mechanically increases the median β by down-weighting the scale that showed the weakest associations (p. 4). Ultimately, while the article effectively challenges the "minimal effects" narrative, it replaces it with a "substantial effects" narrative that is heavily dependent on contestable causal assumptions applied to correlational data.

The Bottom Line

Twenge et al. effectively demonstrate that the link between social media and mental health is highly sensitive to analytical choices, challenging previous claims that the association is universally negligible. However, the "substantial" negative effects they report for girls depend critically on the decision to exclude control variables like school unhappiness and family closeness—a choice that assumes these are outcomes of social media use rather than pre-existing causes. Because the study relies on cross-sectional data that cannot distinguish between these causal pathways, the reported effect sizes likely represent an upper-bound estimate that may be inflated by confounding factors.

Potential Issues

Exclusion of control variables based on an unproven causal model: The article's central finding of a large association between social media use and poor mental health depends critically on the decision to exclude a set of control variables that the original study by Orben and Przybylski had included. The authors label these variables—such as "negative attitudes toward school," "closeness with parent," and "school grades"—as "potential mediators" and argue that controlling for them would lead to overadjustment bias (p. 2). This decision is justified by citing methodological literature that warns against controlling for variables that lie on the causal pathway between an exposure and an outcome. However, this rests on the strong, unproven assumption that the causal pathway is indeed: Social Media Use → School/Family Problems → Poor Mental Health. An equally plausible alternative model is that a third factor, such as a negative home environment or a pre-existing psychological trait, acts as a common cause (a confounder) that leads to both increased social media use and problems at school and home. In the confounding scenario, controlling for these variables would be the correct analytical procedure. While the authors provide a theoretical justification for treating these as mediators, they cannot empirically validate this causal direction with the cross-sectional data used. This single decision is the primary driver of the increase in the reported effect size in two of the three datasets. For example, in the Millennium Cohort Study data for girls, the median β for social media use and mental health is -0.01 when including these controls, but it increases twentyfold to -0.20 when they are excluded (p. 4, Table 1a). This demonstrates that the magnitude of the effect is entirely conditional on the specific causal model assumed by the researcher.

Causal framing of findings from cross-sectional data: The article frames its research using a strong causal narrative that is not fully supported by its cross-sectional data. The introduction presents the rise in adolescent mental health issues as a "mystery" for which social media is the "prime suspect" (p. 1). The conclusion revisits this theme, stating it "seems plausible that increases in digital media use might be responsible for the substantial increases in adolescent depression and anxiety that began around 2012" (p. 10). This framing encourages a causal interpretation of the findings. However, the data from all three studies are correlational and cannot distinguish whether social media use causes poor mental health, poor mental health leads to greater social media use (reverse causation), or unmeasured third variables cause both (confounding). While the authors are careful to use correlational language such as "linked to" and "association" when reporting results, the overarching argumentative structure of the article uses acausal evidence to advance a causal conclusion. The article does acknowledge this limitation by stating that "social media should remain on the list of possible explanations" (p. 10), but the dominant narrative may overstate the certainty of the evidence presented.

Tension between individual-level analysis and a proposed network-level mechanism: The article's conclusion introduces a network spillover hypothesis that creates a tension with its own individual-level analytical approach. The authors suggest that as social media became ubiquitous, "teen social life changed even for adolescents who spent no or little time on social media" (p. 10). If the primary mechanism of harm is a systemic change to the social environment that affects all adolescents regardless of their personal usage levels, then an analysis correlating an individual's hours of use with their individual mental health outcome is likely to be misspecified. This approach implicitly treats low-users as a valid comparison group, but the authors' own hypothesis suggests this group is also affected by the environmental change. Consequently, the individual-level correlation measured in the article may not capture the total effect of social media on adolescent mental health. The article acknowledges this possibility, calling for "future research" to "explore how social media changed the collective dynamics of social interaction" (p. 10), but the theoretical claim highlights a limitation of the study's own design.

Use of an invalid proxy variable for social media in one dataset: The analysis of the Youth Risk Behavior Surveillance System (YRBSS) dataset, one of the three pillars of the article's argument, relies on a proxy variable with questionable construct validity. The authors acknowledge that the YRBSS "does not have any questions that focus specifically on social media usage" (p. 6). Instead, they use a broad measure of "electronic device use," which explicitly includes "computers and gaming consoles" and was later updated to mention "smartphones and tablets" (p. 6). This variable encompasses a wide range of heterogeneous activities, such as homework, gaming, and video streaming, that are distinct from the article's primary construct of interest: social media. Despite this mismatch, the article's abstract and discussion frame the YRBSS results as consistent evidence for the harms of "social media use" specifically. The authors attempt to justify this by arguing that the association is stronger for this variable than for TV and by restricting their analysis to years after 2013 when smartphones were added to the question's wording (p. 6). While this is a reasonable step, it does not resolve the fundamental issue that the variable remains an imprecise proxy, weakening the claim of a consistent finding across three large datasets.

Omission of key psychological confounding variables: The analysis does not control for or discuss major potential psychological confounders, such as pre-existing personality traits. Stable individual differences like neuroticism, low self-esteem, or high rejection sensitivity are established predictors of both a tendency to use social media more heavily and a greater vulnerability to mental health problems. The observed association could therefore be partially or wholly explained by these underlying dispositions rather than by social media use itself. While the article is a reanalysis of existing datasets and is therefore constrained by the variables available within them (which did not include these personality measures), the absence of controls for such fundamental individual differences represents a significant limitation. This leaves a powerful alternative explanation for the core findings unexamined and may lead to an overestimation of the independent effect of social media.

Subjectivity in the application of Specification Curve Analysis: The article critiques the original analytical choices of Orben and Przybylski as arbitrary and effect-attenuating, but replaces them with an alternative set of choices that could also be seen as subjective. For instance, the authors criticize the original analysis of the Millennium Cohort Study for allowing the parent-reported Strengths and Difficulties Questionnaire (SDQ) to constitute "73% of the data using scales" (p. 3). They explain this was because the original analysis included the total SDQ, its five subscales, and two combinations of subscales as separate measures, effectively counting the same instrument eight times (p. 3). Their proposed solution is to weight each of the four mental health scales equally. While this is a valid correction to prevent redundancy in the specification space, the authors also note that "The SDQ... produces notably lower betas than the 3 other mental health measures" (p. 4). Therefore, the decision to weight all scales equally also has the effect of down-weighting the influence of the measure showing the weakest association, which mechanically increases the median β. This highlights how Specification Curve Analysis, a technique designed to mitigate "researcher degrees of freedom," is itself highly sensitive to the researcher's decisions in defining the specification space.

Overstated narrative emphasis on a secondary analytical choice: The article's narrative places heavy emphasis on the critique that the original analysis was flawed because one mental health scale (the SDQ) dominated the specification space. However, the article's own results in Table 1a suggest this factor was less influential than the choice of control variables (p. 4). While the decision to exclude controls has the largest absolute impact on the effect size, the scale weighting decision also has a substantial effect. For example, in the authors' preferred specification with no mediator controls, switching from the original scale weighting to equal weighting nearly doubles the median β for girls from -0.11 to -0.21 (p. 4). The narrative's focus on the scale weighting issue may overstate its relative importance compared to the control variable decision, though both choices interact to produce the largest reported effects.

Potential for unaddressed systematic measurement error: The study's conclusions are based on self-reported measures of technology use, which are known to be imprecise and subject to potential biases. The authors acknowledge this limitation and suggest that better measurement would likely reveal "substantially larger" correlations (p. 10). This assumes that the measurement error is random, which tends to attenuate correlations. However, it does not account for the possibility of systematic measurement error. For example, adolescents who are depressed may be more likely to recall their screen time as being excessive or may ruminate more on their social media use, leading them to over-report their usage. Such a bias would create a spurious association between mental health and reported use that is not reflective of a true causal effect. The article does not address this possibility, and its optimistic assumption that better measurement would strengthen its findings is speculative.

Restriction to linear relationships: The analysis exclusively models a linear relationship between the duration of social media use and mental health outcomes. The article acknowledges that "It is very likely that many of the associations are curvilinear or threshold structured" (p. 10). For instance, it is plausible that light-to-moderate use has a different, perhaps even beneficial, effect compared to heavy use. By not testing for non-linear relationships, the analysis may mischaracterize the nature of the association. A linear model could average out a null or positive effect for a majority of users with a strong negative effect for a small group of very heavy users, resulting in a misleading summary of a moderate negative effect for all. The authors state that this was a practical decision to limit the complexity of the specification space and maintain comparability with O&P (p. 10), but it remains a significant limitation on the interpretation of the findings.

Future Research

Causal modeling of disputed controls: Future work should utilize longitudinal data to explicitly test the directionality between social media use and the disputed control variables (e.g., school attitudes, family closeness). By employing cross-lagged panel models or similar techniques, researchers could determine whether these variables function primarily as mediators (justifying their exclusion) or confounders (requiring their inclusion), thereby resolving the central methodological disagreement between this article and prior analyses.

Non-linear specification curves: Given the authors' acknowledgment that associations are likely "curvilinear or threshold structured" (p. 10), future SCA research should abandon the strict linearity assumption. Analyses should incorporate quadratic terms or threshold models within the specification curve to detect if harms are concentrated among the heaviest users, which would provide a more nuanced understanding of risk than a single linear β coefficient.

Granular measurement of digital behaviors: To address the limitations of broad proxy variables like "electronic device use," future studies must utilize objective logging data or highly specific time-use diaries that distinguish between active social media use, passive scrolling, gaming, and other digital activities. This would prevent the conflation of heterogeneous behaviors and allow for a precise assessment of which specific digital activities are associated with mental health outcomes in girls.

isitcredible.com