PERSPECTIVES

PHOENIX CENTER FOR ADVANCED LEGAL & ECONOMIC PUBLIC POLICY STUDIES

Assessing the Effects of Social Media Use on Youth Mental Health: An Examination of Selection Bias

George S. Ford, PhD[*]

November 16, 2023

An ever-growing body of academic research points to social media as a source of declining youth mental health.[1] This research is motivating legislative efforts in the U.S. and abroad to regulate some aspects of social media use by young people.[2] While the research on the topic is plentiful, other researchers and practitioners question whether the evidence is sufficiently robust to support legislation (or other government intervention).[3]

One criticism of the existing literature is that most of the empirical evidence supporting a linkage between social media use and mental health is based on cross-sectional analysis, thus leading to questions about whether the measured correlations from such data have a causal interpretation. This concern is typically framed as the question—does social media use cause depression, or does depression cause social media use? Cross-sectional data offer no answers.

In this PERSPECTIVE, I employ longitudinal data to better understand this common and legitimate criticism of cross-sectional studies in this area. To do so, I estimate the (plausibly) causal treatment effect of social media use on youth mental health by the Difference-in-Differences ("DID") estimator, and then compare this effect to the (potentially biased) treatment effect estimated from cross-sectional analysis. The difference in the two effect sizes measures *selection bias* present in the cross-sectional analysis, which is a type of estimation bias frequently mentioned in the criticisms of the literature on social media use and mental health. If selection bias is present, then the estimated relationship between social media use and mental health is not equal to the true relationship, even in large samples. The presence of selection bias is a very serious problem.

I find evidence of selection bias in the cross-sectional approach to quantifying the effects of social media use on mental health. The selection bias is sometimes positive, sometimes negative, and sometimes zero, and differs by gender and by mental health outcome.

Using a popular dataset on mental health outcomes from the United Kingdom, I find evidence of selection bias in the cross-sectional approach to quantifying the effects of social media use on mental health. The selection bias is sometimes positive, sometimes negative, and sometimes zero, and differs by gender and by mental health outcome.

The Empirical Problem

While there are a host of criticisms of the literature on social media's effects on youth mental health, a key complaint is that much of the evidence is based on cross-sectional data. Under most conditions, cross-sectional analysis only permits the measurement of the correlation between variables; this correlation may or may not represent a causal effect. That is, a correlation may be a biased measure of the true causal effect.

> ***If selection bias is present, then the estimated relationship between social media use and mental health is not equal to the true relationship, even in large samples. The presence of selection bias is a very serious problem.***

Bias in statistical estimates refers to a systematic departure of an estimate of a parameter from the true value of the population parameter. Sources of bias are myriad, but a principal type of bias mentioned in criticisms of social media studies is selection bias. Selection bias occurs when the sample used for analysis is not representative of the target population. There are many other types of bias relevant to this field of research, including reporting bias, where survey respondents provide inaccurate responses when self-reporting depressive symptoms or social media use.[4] Here, I focus on selection bias.

Potential Outcomes

Say we observe a sample of people, some using social media (the treated group) and some not using it (the control group). Social media use is a "treatment," and this treatment may have effects on mental health (or other outcomes). Following Angrist and Pischke (2009), let Y_i be the outcome of interest for individual i and D_i indicate whether that individual receives the treatment ($D_i = 1$).[5] The observed outcome may be written as potential outcomes,

$$\begin{aligned} Y_i &= \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases} \\ &= Y_{0i} + (Y_{1i} - Y_{0i})D_i \ . \end{aligned} \tag{1}$$

Note that $Y_{1i} - Y_{0i}$ is the causal effect of interest, which equals the change in the outcome for individual i between the treated and untreated state. This effect may differ across individuals, but research often aims to estimate the average effect from the sample.

Sources of bias are myriad, but a principal type of bias mentioned in criticisms of social media studies is selection bias. Selection bias occurs when the sample used for analysis is not representative of the target population.

In cross-sectional analysis, we cannot observe the change in outcomes for treated respondents. Instead, the best we can do is to observe the difference in means between individuals who are treated and those that are untreated, which is not the difference that measures the causal effect (the difference for a given individual, or the average of such differences among many individuals). We have no real interest in comparing the treated and control groups; the control group is merely a stand-in for the treated group in the untreated state. A chosen control group may be a good proxy for the treated in the untreated state, or it may not be, and this latter possibility is the source of selection bias.[6]

The observed difference in cross-sectional analysis may be written as,

$$\begin{aligned} & E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0] = \\ & \quad E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 1] \\ & \quad + E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0] \ , \end{aligned} \tag{2}$$

where the left-hand side is what we observe from the data. This observed difference consists of two parts: the first difference on the right-hand side is the true treatment effect (the difference in outcomes for treated respondents) and the second difference on the right-hand side is selection bias which is the difference in the outcomes between the treated group ($E[Y_{0i}|D_i = 1]$) and control group ($E[Y_{0i}|D_i = 0]$) in the untreated state. To summarize, we have,

$$TE = OD - SB\,, \tag{3}$$

where TE is the true treatment effect, OD is the observed difference (between groups), and SB is selection bias. Thus, any observed difference in cross-sectional analysis differs from the true treatment effect by the selection bias term, SB.[7]

Some Numerical Examples

Some numerical examples illustrate the issue. Say we have cross-sectional data on social media use and mental health. We observe that respondents that do not use social media have a depression symptom rate of 0.20, while social media users have a depression symptom rate of 0.30. The observed difference is $OD = 0.10$. Concluding that social media increases depressive symptoms by 0.10 ($TE = OD$), as is often done in the literature, assumes $SB = 0$.

This assumption may or may not be valid. Underlying the claim of a causal effect of 0.10 is the assumption that absent social media use, social media users would have a depressive symptom rate of 0.20, like the non-users, so by Equation (3) we have,

$$0.10 = (0.30 - 0.20) - (0.20 - 0.20). \tag{4}$$

If the selection bias term is zero, then the observed difference equals the true treatment effect. Thus, with cross-sectional data, a plausibly causal estimate requires that the treated and untreated groups have the same mean outcome in an untreated state. Without knowledge of the no-treatment outcomes for the treated group we have no idea whether this equality is true or not.

As an alternative, say that absent social media use social media users have a depressive rate of 0.30. That is, people with poorer mental health are more likely to use social media. If so, then the observed 0.10 difference in the cross section is entirely explained by the difference in mental health absent social media use. In this case, by Equation (3), the treatment effect is zero and selection bias accounts for the entire observed difference,

$$0.00 = (0.30 - 0.20) - (0.30 - 0.20). \tag{5}$$

Or, if the outcome is 0.25 for users in an untreated state, then the bias accounts for half of the observed cross-sectional difference, and the true treatment effect is 0.05,

$$0.05 = (0.30 - 0.20) - (0.25 - 0.20). \tag{6}$$

An obvious advantage of panel data is that the outcomes for the treated and untreated may be observed in both the untreated and treated states (though necessarily over time), making it possible to measure the relevant components of Equations (2) or (3).

Equations (4)-(6) are the Difference-in-Differences ("DID") estimators. The DID estimator has good theoretical properties and is increasing the go-to approach in empirical research, but there are a few things to consider. With cross-sectional data, the mean outcome of the control group serves as a proxy for the outcome of the treated group in an untreated state. With panel data, we do not need a proxy for the untreated outcome for the treated since it is observed directly, but necessarily at different times, so now the control group's role is to account for the temporal difference in the measurement of outcomes. The assumption is that the changes in outcomes over time for the control group are the same as those for the treated group in the absence of the treatment; this is the parallel paths (or common trends) assumption, which is untestable and must be made plausible by some means.

Data

The Understanding Society survey is a large-scale, publicly-available, longitudinal household panel survey conducted in the United Kingdom in which the same individuals and households are interviewed repeatedly over time.[8] The survey began in 2009 and has been conducted annually since. It is managed by the Institute for Social and Economic Research ("ISER") at the University of Essex. The survey collects data through face-to-face interviews, self-completion questionnaires, and computer-assisted methods. A portion of the survey is dedicated to young people aged 10 through 18, and data are available biannually, so new respondents enter, and existing respondents exit, the survey. The data on youth are available biannually over the 2009 through 2019 period.

The survey includes a question on the number of hours spent on social media websites on weekdays with a five-level categorical response: (1) no hours; (2) less than one hour; (3) one-to-three hours; (4) four-to-seven hours; and (5) seven-or-more hours. For all years, most of the respondents use social media services, which is not ideal for our purposes, so several restrictions and modifications to the data are applied.

First, the DID method requires social media use to be zero in the first period for all respondents. Since very few respondents used no social media at all, and very few respondents report using the highest usage level (7+ hours), following McNamee, Mendolia and Yerokhin (2021) the five-scale responses are collapsed to three groups: (1) very low use (less than one hour); (2) moderate use (1-3 hours); and (3) high use (four or more hours).[9] The "untreated" state is defined as the very-low usage level.

Second, only respondents that have very-low usage in at least one wave that is followed by at least one more wave where use may be any level are included in the sample. Two periods are retained for each respondent, a first period of very-low use and a second period of any use level (but never lower than the low-use category). For respondents who always have very-low usage, the first two periods available are retained, though most of the respondents have only two-periods of data. At the end, there is a balanced sample of respondents having two periods of data, the first period always being very-low use and the second period separated by only one wave. There are 506 female respondents for and 566 male respondents.

This recasting of the use variable is summarized in Table 1 for both genders. In the pre-treatment period, there are 1,316 respondents all in the very-low treatment group, which is defined as the untreated state. In the post treatment period, there are 658 respondents in the control group, 516 respondents in the moderate-use group, and 142 respondents in the high-use group.

Table 1. Recasting Social Media Use

Raw Usage Level	3 Cat.	Pre	Post
None	Very Low	256	97
Less than 1 hour		1,060	561
1-3 hours	Moderate		516
4-7 hours	High		113
7 or more hours			29

The outcome of interest is the score from the Scoring Strengths and Difficulties Questionnaire ("SDQ") for emotional intelligence (and its components is a separate analysis). The SDQ Emotional Scale provides an indication of emotional well-being and can help identify individuals who may be experiencing emotional problems or are at risk of developing emotional disorders. The SDQ is commonly used in research, clinical practice, and educational settings to screen for emotional and behavioral difficulties, monitor progress, and identify individuals who may require further assessment or support.

The SDQ Emotional Score is based on five questions with three categorical responses: (1) not true (value 0); (2) somewhat true (value 1); and (3) certainly true (value 2). The five questions making up the score include: (1) "I often feel unhappy, down-hearted, or tearful"; (2) "I worry a lot. I am often worried or scared"; (3) "I often have headaches, stomachaches, or other physical problems"; (4) "I am nervous in new situations. I easily lose confidence"; and (5) "I have many fears. I am easily scared." The composite SDQ score is the sum of these values, so it has a range of 0 to 10.

As is common, the SDQ score is dichotomized at a clinically-relevant value (≥ 5), but the raw SDQ Score is also analyzed, though changes in the mean score may be less interesting since the mean is well below the clinically relevant value (2.8 for females and 2.3 for males).

Empirical Model

The effects of social media are estimated using a Difference-in-Differences ("DID") estimator. There are three levels of use: (1) very-low use; (2) moderate use; and (3) high use. There are two treatment levels: moderate (labeled *M* for any respondent that is a moderate user) and high (labeled *H*). There are likewise two periods per respondent. In the first-period, all respondents were very-low users. In the second-period (indicated by the dichotomous indicator *P*), use can be at any level, and it is this variation that permits the estimation of the treatment effect.

Coarsened Exact Matching ("CEM") is used to construct a 1:1 matched sample based on the survey wave, the respondent's gender, and the survey weight. For the dichotomous variables the matches are exact. Since the sample is matched (including the survey weight), the survey weights are unused in estimation and confounders are ignored.

The 2x2 DID model is,

$$\begin{aligned} Y_{it} &= \delta M_{it} \cdot P_{it} + \lambda H_{it} \cdot P_{it} + \alpha_0 \\ &\quad + \alpha_1 M_{it} + \alpha_2 H_{it} + \alpha_3 P_{it} + \varepsilon_{it} \end{aligned} \tag{7}$$

where *M* indicates moderate use, *H* indicates high use, *P* indicates the treatment period, and ε is a disturbance term.[10] The coefficients of primary interest are δ and λ, which are the DID estimates. As the criticisms of cross-sectional studies focus heavily on differences in outcome in the untreated state between the treated and untreated, also of interest are the α_1 and α_2 coefficients, which measure differences in mental health outcomes in the untreated state (the first period, and the source of selection bias).

While the sample is panel data, the first period may be ignored to conduct a cross-sectional analysis. The regression model for the treated period (*P* = 1) is,

$$Y_{it} = \tilde{\delta} M_{it} + \tilde{\lambda} H_{it} + \tilde{\alpha}_0 + \nu_{it}\,, \tag{8}$$

which permits a comparison of the δ and λ coefficients to the $\tilde{\delta}$ and $\tilde{\lambda}$ coefficients to quantify the bias. From the analysis above, the expectation is that $\tilde{\delta} = \delta + \alpha_1$ and $\tilde{\lambda} = \lambda + \alpha_2$.

Results

Before turning to the regression estimates, consider the mean outcomes in the pre-treatment period across the treatment levels (in the treatment period). All treatment levels are very-low use in pre-treatment period, so the table provides the means of the dichotomized SDQ Score for the treatment levels in the treatment period, thus providing a measure of selection bias. The outcome is the dichotomized SDQ Score. Results are summarized in Table 2.

Table 2. Outcomes by Usage Level

	Pre	Post	Diff.
Females			
Very Low	0.218	0.305	0.087
Moderate	0.205	0.299	0.094
High	0.260	0.519	0.259
Males			
Very Low	0.156	0.127	-0.029
Moderate	0.138	0.135	-0.003
High	0.110	0.233	0.123

For females, the means are comparable for very-low and moderate use in both periods, as are the changes in the means. The mean is higher in the pre-treatment period for the high-use group by 0.042 [= 0.260 – 0.218], a sizable difference. Thus, respondents that become high-users in the second period have higher depressive symptoms in the first period—a source of bias. The increase in the mean outcome is much larger for the high-use group than for the control group (0.259 versus 0.087) or the moderate use group (0.259 versus 0.094), suggesting a relationship between social media use and mental health.

By Equation (3), the treatment effect for the high-use group relative to the control group is,

$$0.172 = (0.519 - 0.305) – (0.260 – 0.218), \qquad (9)$$

whereas the cross-sectional effect is,

$$0.214 = (0.519 - 0.305), \qquad (10)$$

with the cross-section effect being larger than the DID effect by 0.042.

For males, the to-be-treated groups have lower pre-treatment outcomes. Thus, the selection bias is negative, and the cross-sectional effects will be biased downward. Comparing the high-use and control groups, the DID estimator is,

$$0.174 = (0.231 - 0.128) – (0.092 – 0.163), \qquad (11)$$

whereas the cross-sectional effect is,

$$0.103 = (0.231 - 0.128), \qquad (12)$$

with the cross-section effect being much smaller than the DID effect (-0.071). Selection bias is material and negative.

To get all the results, a Linear Probability Model ("LPM") is used to estimate the coefficients because non-linear models are not well-suited for DID analysis, and since all the regressions are dichotomous the LPM will provide nearly identical effects as will Logit or Probit.[11] Results are summarized in Table 3 for females. The t-statistics are robust to heteroscedasticity. The regression models are statistically significant at the 1% level.

Table 3. Regression Results, Females

		Cross
	DID	Section
δ: $M{\cdot}P$	0.007	-0.006
	(0.13)	(-0.16)
λ: $H{\cdot}P$	0.173**	0.214***
	(2.05)	(3.39)
α_1: M	-0.013	
	(-0.38)	
α_2: H	0.042	
	(0.75)	
α_3: P	0.088**	
	(2.52)	
Constant	0.218***	0.305***
	(9.43)	(11.04)
Obs.	1,284	642
F-Stat	6.55***	6.36***

Stat. Sig. *** 1% ** 5% * 10%

From the DID model for females, moderate use has little effect on the outcome; the δ coefficient of 0.007 is small and statistically no different from zero. The coefficient λ on high social media use is larger at 0.173, and the coefficient is statistically different from zero but only at the 5% level. There is some evidence here of an effect of social media use on mental health for females at a high use level. The α_3 coefficient indicates a sizable increase in the dependent variable between periods (0.087 on a base of 0.218), so depressive symptoms are rising over time. Thus, younger people are getting more depressed over time irrespective of social media use, a fact often ignored in the legislative debate and suggests there is more to the rise in youth depression than social media use.

Note the sizes of the α_1 and α_2 coefficients (also see Table 2). The differences in the controls and the to-be-treated moderate group is relatively small (-0.013), so the two groups have similar mental health prior to the treatment. The α_2 coefficient, in contrast, is large in the untreated state, with a coefficient of 0.042, and though the coefficient is not statistically different from zero the selection bias is still there.

Turning to the cross-sectional estimates, the coefficient on moderate use is essentially zero (-0.006). The bias is negative but small, though neither the DID nor cross-sectional coefficient is statistically significant. Alternately, the cross-sectional coefficient for the high use group is 0.214 [= 0.173 + 0.042] and statistically different from zero at the 1% level. The bias is somewhat large (24%). The DID model provides some evidence of an effect, but the cross-section model provides stronger evidence of an effect, but the effect size is too large due to selection bias.

The results for males are summarized in Table 4. Note that neither regression model is statistically significant even at the 10% level, so the model does not improve the prediction of the outcome. As such, there is no evidence social media use affects mental health (in these models). The DID coefficients are 0.021 and 0.174 for moderate and high use levels, the latter being statistically different from zero at the 5% level. High use reduces the mental wellbeing of males, though the models have little explanatory power.

Table 4. Regression Results, Males

	DID	Cross Section
δ: $M{\cdot}P$	0.021	0.001
	(0.52)	(0.04)
λ: $H{\cdot}P$	0.174**	0.103*
	(2.49)	(1.84)
α_1: M	-0.020	
	(-0.67)	
α_2: H	-0.071*	
	(-1.70)	
α_3: P	-0.036	
	(-1.31)	
Constant	0.163***	0.128***
	(8.08)	(7.00)
Obs.	1,348	674
F-Stat	1.35	1.76

Stat. Sig. *** 1% ** 5% * 10%

In the pre-treatment period, male moderate users have a smaller mean outcome than very-low users (α_1 = -0.02) and the difference for high users is larger (α_2 = -0.071). These differences carry over to the cross-sectional results; the bias is negative. The coefficient from the cross-sectional analysis for the moderate-use level is 0.001, a small and statistically insignificant effect. At a high-level of use, the cross-section coefficient is 0.103, which is much smaller than the DID coefficient and statistically significant only at the 10% level. Selection bias is material. This analysis may point to a possible reason why the effect sizes for males are often reported as being smaller than that for females.

What can we make of these results? The potential for biased coefficients in cross-sectional analysis is a real concern, though the bias may be positive or negative. For females, the cross-section effect was 24% too large, and for males was 41% too small. Such bias may lead to misleading conclusions, especially about the “strength” of the result in terms of size and statistical significance.

Raw SDQ Score

The prior results are based on the dichotomized SDQ Score (≥ 5). Tables 5 and 6 summarize the results where the raw SDQ Score (range 0 to 10) is the dependent variables.

Table 5. Regression Results, Females

	DID	Cross Section
δ: *M·P*	-0.004 (-0.01)	-0.226 (-1.11)
λ: *H·P*	1.012** (2.47)	1.145*** (3.67)
α_1: *M*	-0.222 (-1.25)	
α_2: *H*	0.132 (0.50)	
α_3: *P*	0.598*** (3.38)	
Constant	2.907*** (25.47)	3.505*** (25.85)
Obs.	1,284	642
F-Stat	10.49***	9.33***
Stat. Sig. *** 1% ** 5% * 10%		

For females, moderate social media use has no meaningful effect on the SDQ Score (-0.004), but high use increases the score by 1.012 units (about 33%, on average). The α_1 and α_2 coefficients are not very large. The cross-section coefficient for high use is 1.145, which is biased slightly upward by about 13%.

Figure 1. Change in SDQ E, Females

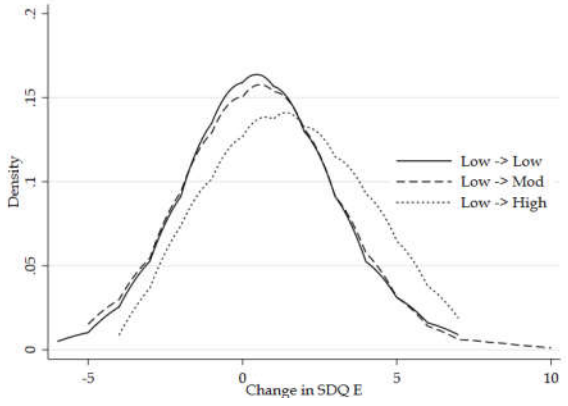


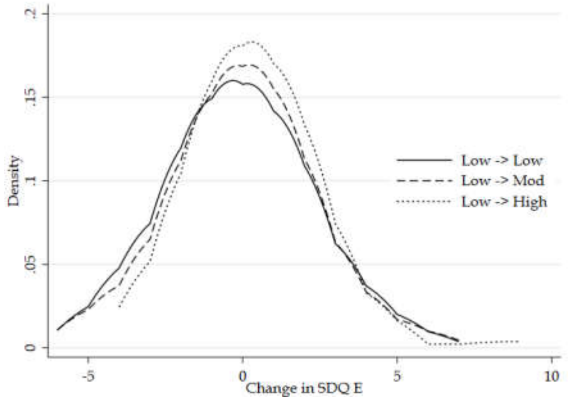
Figure 1 provides the kernel density function for the change in the SDQ Score between periods for each level of social media use in the treated period. While the distributions in the control and moderate use groups are comparable, the distribution for the high-use is shifted in the positive direction.

Table 6. Regression Results, Males

	DID	Cross Section
δ: *M·P*	0.081 (0.34)	0.059 (0.34)
λ: *H·P*	0.491 (1.21)	0.340 (1.17)
α_1: *M*	-0.023 (-0.14)	
α_2: *H*	-0.151 (-0.53)	
α_3: *P*	-0.0122 (-0.77)	
Constant	2.320*** (21.15)	2.199*** (19.36)
Obs.	1,348	674
F-Stat	0.36	0.69
Stat. Sig. *** 1% ** 5% * 10%		

For males, neither of the models is statistically significant, and none of the coefficients are statistically different from zero except for the constant term. These models are largely uninformative; social media use has no apparent effect on the mean SDQ Score.

Figure 2. Change in SDQ E, Males



The distributions of the change in SDQ Score by usage group are illustrated in Figure 2. While there is a slight positive shift at higher usage levels, the distributions are comparable.

Breakdown of the SDQ Score

Next, the SDQ score is broken into its component questions to examine the possible sources of the higher SDQ scores. The responses are dichotomized so that any affirmative response is 1.0 (0 otherwise). The condensed results for females are summarized in Table 7.

Table 7. Regression Results
Females

Outcome	Coef	DID	Cross Section
Unhappy	δ: *M·P*	-0.045	-0.056
	λ: *H·P*	0.260***	0.217***
Worry	δ: *M·P*	0.074	-0.022
	λ: *H·P*	0.149*	0.144***
Headaches	δ: *M·P*	0.083	0.025
	λ: *H·P*	0.152*	0.145**
Nervous	δ: *M·P*	-0.017	-0.051
	λ: *H·P*	0.043	0.047
Fearful	δ: *M·P*	0.041	0.038
	λ: *H·P*	0.153*	0.190***

Stat. Sig. *** 1% ** 5% * 10%

Several of the DID estimates are statistically significant for the high-use group including unhappiness, worry, headaches (and other ailments), and fearfulness. Sadness is the strongest result both in magnitude and precision. In each case, the cross-section coefficient is likewise statistically significant but often at a higher level. The bias may be positive or negative (as in the *fearful* response).

Table 8. Regression Results
Males

Outcome	Coef	DID	Cross Section
Unhappy	δ: *M·P*	0.021	0.009
	λ: *H·P*	0.088	0.086
Worry	δ: *M·P*	0.006	0.043
	λ: *H·P*	0.140	0.054
Headaches	δ: *M·P*	-0.004	-0.028
	λ: *H·P*	0.136	0.055
Nervous	δ: *M·P*	0.030	0.017
	λ: *H·P*	0.011	0.005
Fearful	δ: *M·P*	0.052	0.009
	λ: *H·P*	0.063	0.018

Stat. Sig. *** 1% ** 5% * 10%

For males, summarized in Table 8, none of the DID nor cross-sectional coefficients are statistically different from zero, and most are small. For the most part, the biases in the cross-section results are negative.

Summary

While this analysis is limited in several ways, it does support the concerns regarding cross-sectional analysis of the link between social media use and youth mental health. I find the bias in the estimated effects from cross-sectional data may be large or small, may be positive or negative, and may vary by the outcome of interest.

I stress, however, that these estimates are specific to the empirical approach and the data used. Nevertheless, I believe the results confirm that cross-sectional analysis could produce biased estimates of the effects of social media use on mental health and, as a result, may mislead policymakers about the effects of social media on mental health.

I can make no claim about the generalizability of these findings to other studies, datasets, or models. What I do show is that bias is a legitimate concern, though it may not be important in specific cases (though which cases are largely unknown). The inability to quantify bias in a particular cross-sectional study is worrisome, but the presumption should be that the estimates are, in fact, different than the true effect.

Certainly, the youth mental health crisis is an important issue, but questions regarding whether there exists a sufficiently robust body of evidence today to justify regulating social media services cannot be ignored.

Conclusion

The use of social media services by young people has been linked to worsened mental health. While the results of empirical studies are mixed and the literature is subject to several valid criticisms, this research continues to serve as the basis for legislative efforts at the state and federal level.

One identified problem with the research is that it relies heavily on relationships estimated from cross-sectional data, which may fail to provide a valid estimate of the causal effect between social media use and mental health. In this PERSPECTIVE, I describe one source of bias and look for its presence. I find the bias in the estimated effects from cross-sectional data may be large or small, may be positive or negative, and may vary by the outcome of interest. Thus, some suspicion regarding the empirical evidence on the effects of social media use on mental health is warranted. More research on the presence of selection (and other types of) bias is encouraged.

Certainly, the youth mental health crisis is an important issue, but questions regarding whether there exists a sufficiently robust body of evidence today to justify regulating social media services cannot be ignored.

NOTES:

[*] **Dr. George S. Ford** ***is the Chief Economist of the Phoenix Center for Advanced Legal and Economic Public Policy Studies. The views expressed in this*** **PERSPECTIVE** ***do not represent the views of the Phoenix Center or its staff. Dr. Ford may be contacted at*** ford@phoenix-center.org.

[1] J. Haidt and J. Twenge, *Social Media and Mental Health: A Collaborative Review*, Unpublished Manuscript (Ongoing) (available at: http://tinyurl.com/SocialMediaMentalHealthReview); *Social Media and Youth Mental Health, The U.S. Surgeon General's Advisory*, U.S. Surgeon General (2023); C. Miller, *Does Social Media Use Cause Depression?*, Child Mind Institute (April 27, 2023) (available at: https://childmind.org/article/is-social-media-use-causing-depression); M. Brown, *Does Social Media Cause Depression?*, PSYCHCENTRAL (March 7, 2022) (available at: https://psychcentral.com/depression/does-social-media-cause-depression); M. Doucleff, *The Truth About Teens, Social Media and the Mental Health Crisis*, All Things Considered: National Public Radio (April 25, 2023).

[2] *See, e.g.*, S. 1409, Kids Online Safety Act (text available at: https://www.congress.gov/118/bills/s1409/BILLS-118s1409is.pdf); Protecting Kids on Social Media Act (available at: https://www.schatz.senate.gov/imo/media/doc/protecting_kids_on_social_media_act_2023.pdf); *Fact Sheet: Biden-Harris Administration Announces Actions to Protect Youth Mental Health, Safety & Privacy Online*, The White House (23, 2023) (available at: https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/23/fact-sheet-bidenharris-administration-announces-actions-to-protect-youth-mental-health-safety-privacyonline/?source=email).

[3] *See, e.g.*, P.M. Valkenburg, *Social Media Use and Well-Being: What We Know and What We Need to Know*, 45 CURRENT OPINION IN PSYCHOLOGY 101294 (2022); L. Denworth, *Social Media Has Not Destroyed a Generation*, SCIENTIFIC AMERICAN (November 1, 2019); A. Brown, *The Statistically Flawed Evidence That Social Media Is Causing the Teen Mental Health Crisis*, REASON (March 3, 2023) (available at: https://reason.com/2023/03/29/the-statistically-flawed-evidence-that-social-media-is-causing-the-teen-mental-health-crisis).

[4] S.C. Boyle, S. Baez, B.M. Trager, J.W. LaBrie, *Systematic Bias in Self-Reported Social Media Use in the Age of Platform Swinging: Implications for Studying Social Media Use in Relation to Adolescent Health Behavior*, 10 INT. J. ENVIRON. RES. PUBLIC HEALTH (August 10, 2022) (available at: https://pubmed.ncbi.nlm.nih.gov/36011479).

[5] J.D. Angrist and J. Pischke, MOSTLY HARMLESS ECONOMETRICS (2009) at Ch. 2.

[6] Selection bias is a form of omitted variables bias, so in some cases selection bias may be attenuated by including covariates in the regression that account for differences in the groups (conditioning on covariates) making the treatment "as good as randomly assigned."

[7] With random assignment of the treatment, the selection bias term is plausibly zero; there is no reason to expect the outcomes will be different absent the treatment if the treatment is randomly assigned. The observed difference equals the true effect ($OD = TE$) when selection bias is absent ($SB = 0$). Treatment assignment in observational data is typically non-random—people choose their treatment level. If the processes that determine the choice of treatment level and the outcome are meaningfully related, then selection bias becomes an issue.

[8] Data available at: https://ukdataservice.ac.uk. The dataset identifier is 6614.

[9] P. NcNamee, S. Mendolia, and O. Yerokhin, *Social Media Use and Emotional and Behavioral Outcome in Adolescence: Evidence from the British Longitudinal Data*, 41 ECONOMICS AND HUMAN Biology 10099 (2021).

[10] While this equation could be estimated as a two-way fixed effects model, I choose not to do so that the α coefficients are available.

[11] M. Lechner, *The Estimation of Causal Effects by Difference-in-Difference Methods*, 4 FOUNDATIONS AND TRENDS IN ECONOMETRICS 165-224 (2010).