

AMENDED Exhibit 96

PLAINTIFFS' OMNIBUS OPPOSITION TO DEFENDANTS' MOTIONS FOR SUMMARY JUDGMENT

Case No.: 4:22-md-03047-YGR

MDL No. 3047

In Re: Social Media Adolescent Addiction/Personal Injury Products Liability Litigation

Effect of FB on Well-being: A Review of Experimental Evidence in the Literature Cited by Haidt & Twenge (2019)

Our assessment of the studies on the impact of FB on well-being

We reviewed the articles that Professors Jonathan Haidt and Jean Twenge (2019) singled out as providing evidence that FB products have a negative impact on well-being among female teens. Based on our reading of these studies we conclude that it would be difficult to draw any strong conclusions based on the available empirical evidence. The reasons for this conclusion are the following:

The studies have significant flaws in design and analysis that undermine their validity:

Contrived interventions (e.g., Ozimek & Bierhoff, 2019 asked participants to log on to FB and spend time writing down how they compare to the first 5 people that first showed up in their newsfeed)

No checks if people complied with the treatment (e.g., Tromholt, 2016, and Vanman et al. 2018, ask participants to stop using FB for a set number of days, but they do not check whether people actually followed through with the study's request.)

Expectations and placebo effects (We do not know the extent to which findings are driven by a participant's belief of what should happen as a result of the intervention in contrast to what actually happens. For example, in Allcott et al. 2019, the effects of FB on well-being are driven by people who thought FB was bad for them prior to participating in the study. In Tromholt, 2016, it is very unlikely that people complied with the treatment and so any of the observed differences could be explained by expectation effects.)

Lack of statistical power (e.g., Hunt et al., 2018 and Mosquera et al., 2019 do not have enough participants in the sample to detect realistic effects sizes of FB on well-being)

Lack of robustness checks (All but one of the studies we reviewed present one and only one statistical model supporting the main findings of the paper. Ideally one would like to see robustness checks to assess the findings are not the result of a particular modeling choice).

The studies' findings are inconsistent (e.g., one study finds FB is bad for life satisfaction, but another finds that FB has no effect)

The studies rely on non-representative populations (mostly on undergraduate psychology students) which make it hard to generalize any findings beyond the populations studied - including the youth population.

In our reading, only Hunt et al. (2018), Mosquera et al. (2019), and Allcott et al. (2019) can be taken seriously. However, the results in Hunt et al. (2018) and Mosquera et al. (2019) are quite flimsy: they report findings based on an unrepresentative population (undergraduate students); they do not have sufficient statistical power; they rely on unconventional estimation techniques; and they do not account for the possibility that their findings could be driven by "placebo" effects. Allcott et al. (2019) is in our perspective the best of all the existing studies. But even in this study, it is unclear the extent to which findings are driven (a) by their specific choices in well-being measurement and (b) by "placebo" effects.

We believe it is key that we continue invest in rigorous studies to understand the impact of our products on well-being. It is important to make sure that our products do not harm well-being. Currently we are conducting our own deactivation study to understand the impact of FB and IG on well-being. Our work internal project addresses the limitations we have identified in the external research. This work will be important to engage with external audiences that claim our products are harmful but who do not provide strong empirical evidence to support these claims.

Where do we go from here?

We believe that there is much to learn about the impact of social media on well-being. We recommend these steps:

Use the findings from internal research as a springboard for future research. We are currently finalizing the pilot of our own experiment on the impact of FB on well-being. Our study design addresses the concerns we have with the existing literature. We are confident that our design is of much higher quality than what currently exists externally and will yield insights that can help with product decisions. However, there are two cons of the internal research: (i) it will be just one data point from which it will be hard to make general statements about broader populations and (ii) if we chose to share our findings externally, we have a severe credibility problem. For this reason we propose the next step:

Invest in developing in external research. Our perspective is that external research should directly address the inferential issues we have identified. We believe we should invest in teams that have expertise in the area of well-being and that have the technical capabilities to design and analyze an experiment. Investing in external research with also help with any credibility issues.

Fund studies with participants from different populations. All of the studies we examined are based on undergraduate students in developed countries. Even if the papers we examined were executed without any major flaws, it would be impossible to know whether the findings generalize to people from different ages, nationality, income, gender, etc.

Our detailed assessment of the papers we reviewed

In our perspective, studies attempting to estimate the effect of FB (or social media) on well being area should meet the following criteria: (1) They should have good statistical power and focus on a pertinent population; (2) they should include manipulation checks of a realistic experimental intervention; (3) they should have a solid research design and analysis and; (4) they should address expectation (i.e., placebo) effects and; (5) they should provide consistent findings.

None of the studies we reviewed meet this criteria. We acknowledge that research is hard and that no one study is perfect. However, we find that all of the studies we reviewed have flagrant flaws across most if not all of these dimensions. Below we provide a brief summary of our assessment of these studies we reviewed. The table below summarizes how the studies we reviewed compares to a hypothetical credible research study across the dimensions we propose.

	A	B	C	D	E
1	Study	Population	Sample Size	Percent Female	Check Treatment Compliance
2	Hypothetical ideal study	Teens	1000	100	Yes
3	Mosquera et al. (2019)	Undergraduate students (TAMU)	151	?	Yes
4	Hunt et al. (2018)	Psychology undergraduate students (UPenn)	143	76%	Yes
5	Tromholt (2016)	Adults in Denmark	1095	86%	No
6	Sagioglou and Greitmeyer (2014)	Mturk workers	263	61%	No
7	Yuen et al. (2018)	Psychology undergraduate students	312	79%	Yes
8	Vanman et al. (2018)	Australian undergraduate students	138	63%	No
9	Ozimek & Bierhoff (2019)	Undergraduate students	75	77%	Yes
10	Alcott et al. (2019)	US adults recruited via FB Ads	2,743	57%	Yes

As you can see, in contrast to the hypothetical study, the studies we review focus on populations different from the one of interest (female teens); lack statistical power; rely on unrealistic interventions; do not include manipulation checks; or they fail to assess the robustness of their findings. Additionally, none of the studies account for the possibility that the findings are driven by placebo effects. Finally, you can see that the findings are not consistent across studies 3 out of the 4 metrics the studies have in common.

We went a step further and ranked the studies Haidt and Twenge referenced across the four dimensions we proposed: power, realistic intervention, manipulation checks, and quality of analysis. Based on this exercise we consider that the only studies that can taken seriously are Mosquera et al. (2019) and Hunt et al. (2018). The reason for this is that both of these studies involved to a certain extent realistic interventions and both implemented treatment compliance checks. These two find that restricting the use of FB (or SNS) reduces depression. However, as discussed in detail below, these two studies have significant flaws: they are underpowered and report significant limitation in the analysis. Thus, we have a sense that the findings they report could be the result of: (1) lack of statistical power; (2) flawed executions of the empirical analysis; and/or (3) expectation effects. As a result, our overall conclusion after reading this set of papers, is that we do not know what the effect on FB on well-being is.

If you want to dig deeper into any of the studies we have drafted a small summary for each of them [here](#).

Discussion of main issues we identified in the studies we reviewed

- Inconsistent findings. Even if we took all the experimental papers at face value, the studies report inconsistent findings. For instance, two studies measure the effect of FB on loneliness, but only one finds a negative effect. Four studies measure "affect", but only two find a negative effect. The only consistent finding is that restricting the use of FB (or SNS more generally) reduces signs of depression. However, as we discuss below there are significant issues in the way the authors analyzed the data to arrive at these findings.
- Unrepresentative population. Even if the findings of the effect on FB on well-being were consistent across all studies, most of them rely on an unrepresentative population. Specifically, in all but two of the studies, the participants in the experiments were college-age psychology students in developed countries. Although the majority of the participants in the studies are female. It would be unclear how such findings generalize to the teenage population across the globe.

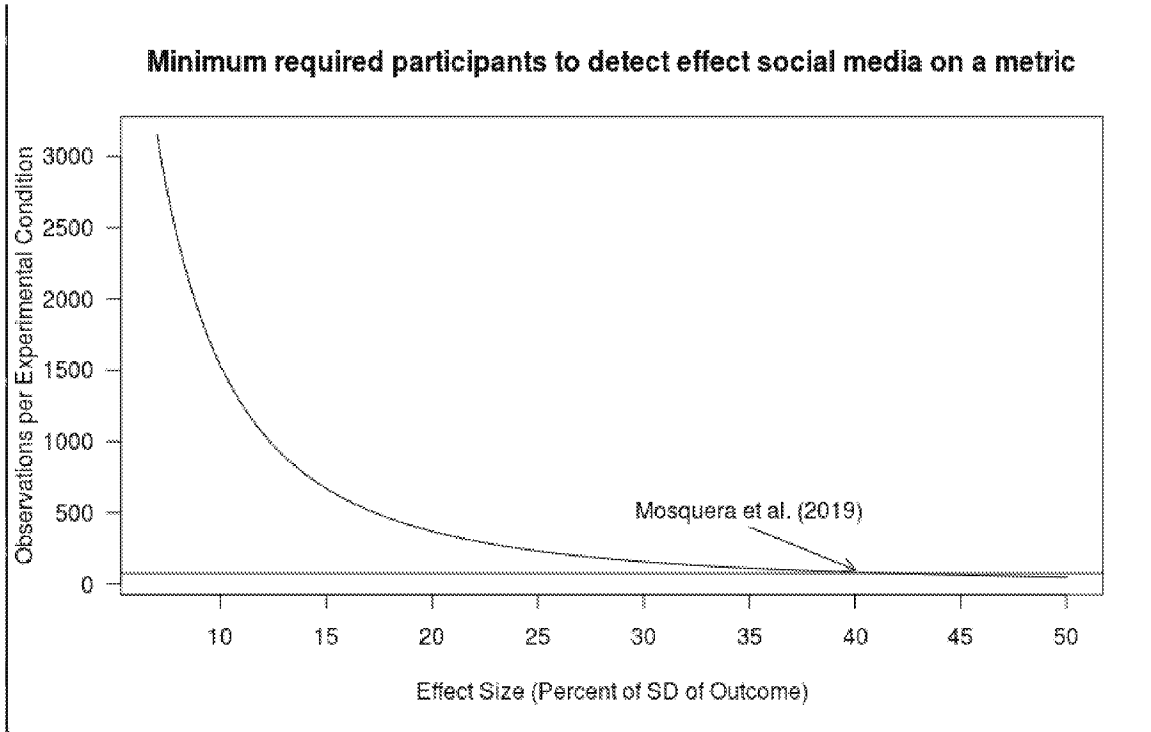
- Most of the studies are underpowered. For studies to yield reliable inferences, they must have sufficient statistical power. This means that studies must have having enough participants so that one ensures a low variability of treatment effect estimates. The standard in the social studies is to implement studies with 80% power. This means that one should have enough observations so that when one runs an experiment, one is able to detect an effect of a given size with 80 probability. However, only one of the studies we reviewed makes any mention of power (but even this one study makes no reference regarding the effect size they are able to detect). However, based on our own calculations one needs at least 1000 study participants to detect an effect size representing 20% of the standard deviation of a generic well-being outcome (Our power calculations are documented in the Appendix below). The danger of running underpowered experiments is finding large artificial effects resulting from the higher variability of estimation with relatively small number of observations.
- Contrived interventions. Even if the findings were consistent, based on representative populations, and sufficiently powered, we find that there is no consistency in the treatment across experiments. 3 out of the 7 studies ask participants to use FB in an artificial way. Here are two stark examples: (1) A study asks participants to go to a lab, log onto FB, and compare themselves with the first five people that pop-up in their newsfeed and; (2) another study asks people to scroll mindlessly the page with their FB profile for a given time period. Neither of these interventions reflect the way in which people actually use our products. For this reason, we think studies restricting the access to FB (or related products) comparing against a group of people who keep using the product as they usually do are better.
- Lack of treatment compliance checks. Unfortunately, 2 of the 4 studies that adopt the more realistic intervention of asking people to restrict their use of FB (or other SNS) do not check whether people actually comply with the treatment they were assigned to. In the absence of these checks, it is unclear whether any significant difference in a well-being metric between treatment group can be attributed to a given intervention. Notably, the only study that is sufficiently powered to detect a realistic effect on well-being (Tromholt, 2016) simply invites people to stop using FB for a week without checking whether people follow through with the invitation. The only "check" the author performed was to ask people in a survey whether they accepted the invitation (only 13% say they did not). This seems a large underestimate of the true proportion. As such, it is difficult to attribute any difference in outcomes to participants actually changes their use of FB.
- Expectation (or "Placebo") Effects. One additional problem is that even if these papers have realistic interventions they do not deal with the problems associated with expectation effects. What does this mean? In medical settings experiments include a placebo test as the control condition. Researchers do this in order to prevent people from injecting their beliefs into the measures used as outcomes of interest (i.e., one may simply report feeling better about a medical ailment simply because one thinks the medicine should have this effect even if the person's underlying medical condition remains the same). In experiments estimating the impact of social media on well-being including such a condition this is not possible, raising the concern that several of the findings could be driven not because of any real effects, but rather because of the effect participants think social media should have on their behavior/health. The prevalence of expectation effects can be exacerbated with (a) the design and (b) the type of participants included in the study. On (a), we suspect that invasive compliance check reminding people what the study is about (by asking people, for example, to take a screenshot of their phone every day to monitor use of social media as in Hunt et al., 2018) is likely to exacerbate expectation effects. On (b) we highlight that most of the studies rely on psychology students as study participants. This population is probably the most attuned about how social media "should" affect specific constructs of well-being.
- Flawed and limited empirical analyses. Further, even in the absence of all the issues identified so far, the studies we reviewed either (1) have significant flaws in the analysis and/or (2) lack robustness checks. The consequence of (1) is that the reported estimates are biased. The consequence of (2) is that the reported results are not very credible. On (1) two studies are notable: Mosquera et al. (2019) and Hunt et al. (2018). Although these are underpowered, they are the best studies we reviewed in terms of design (in the sense that they employ more less realistic intervention and implemented treatment compliance checks). Yet all the results in Hunt et al. (2018) rely on dropping from the sample all people that did not follow the treatment protocol. We know at least since 1996 that this kind of conditions leads to bias of treatment effect estimates. Similarly, Mosquera et al. (2019) implement an mis-specified regression model to estimate their treatment effects. On (2), the vast majority of the studies present findings based on one an only one model specification. This raises questions about the credibility of the findings as it is usual in applied research that the size and significance of estimates will depend on statistical modeling choices, and selection of constructing of metrics to account for any potential confounders (even in the case of experiments).
- No publicly available data. Ideally, as next steps it would be helpful to assess our specific concerns concerns on the methods we identified in these papers. However, as far as we can tell, none of the dataset analyzed in these studies is publicly available.

How much do we know about the effect of social media on well-being among teens?

Answer: Not much.

The story reported here is not very different from the one described in this document when we look at the scientific literature that focuses on teens. Anya Drabkin has written a wonderful [review](#) of the literature on the impact of social media on well-being among the youth. Her work is consistent with several of the points we make here: (1) the findings in the literature are inconsistent; (2) most of the studies in this literature rely on observational data, which prevent us from making causal claims about the impact of FB on the well-being of teens and; (3) we need further research to assess how social media impact different population. An additional important point that Anya's review highlights is that well-being comprises dimensions not limited only to mental health (as implicitly assumed in the articles Haidt and Twenge referenced).

Appendix



```
## Health Outcomes

var.prop <- function(beta.zero, beta.one, sig.zero, sig.one, rho){## Variance of SATE estimator fr
  out <- ((beta.one * sig.one^2) + (beta.zero * sig.zero^2) + (2 * (rho * sig.one * sig.zero)))
  return(out)
}

min_n <- function(sizesd, x, sig.zero){
  out <- ((2.8 * x) / (sizesd * sig.zero))^2
  return(data.frame(nper = round(out / 2, digit = 2), size = sizesd * sig.zero, sizesd = sizesd))
}

happy <- data.frame(answer = c("much better", "somewhat better", "about same", "somewhat worse", "
happy$happy <- ifelse(happy$answer %in% c("much better", "somewhat better", "about same"), 1, 0)

pzero <- sum(happy$freq[happy$happy == 1]) / sum(happy$freq)
sig.zero <- (pzero * (1 - pzero))
tau.true <- seq(0.07, 0.5, 0.001) ## True effect sizes (percent of sd in control condition)
n.vec <- seq(100, 2000, 10) ## Varying sample size

nhealth <- c()
for(tau in 1:length(tau.true)){
  ptau <- pzero + (tau.true[tau] * sqrt(sig.zero))
  sig.one <- ptau * (1 - ptau)
  for(i in 1:length(n.vec)){
    nhealth <- rbind(nhealth,
      min_n(sizesd = tau.true[tau], sqrt(var.prop(beta.zero = 1, beta.one = 1, sig.
  )
}
```