Social Media and Mental Health

By Luca Braghieri, Ro'ee Levy, and Alexey Makarin*

We provide quasi-experimental estimates of the impact of social media on mental health by leveraging a unique natural experiment: the staggered introduction of Facebook across U.S. colleges. Our analysis couples data on student mental health around the years of Facebook's expansion with a generalized difference-in-differences empirical strategy. We find that the roll-out of Facebook at a college had a negative impact on student mental health. It also increased the likelihood with which students reported experiencing impairments to academic performance due to poor mental health. Additional evidence on mechanisms suggests the results are due to Facebook fostering unfavorable social comparisons.
JEL: D12, D72, D90, I10, L82, L86
Keywords: Social media, mental health, Facebook, depression, social comparisons

In 2021, 4.3 billion people—more than half of the world population—had a social media account, and the average user spent around two and a half hours per day on social media platforms (GWI, 2021; We Are Social, 2021). Very few technologies since television have so dramatically reshaped the way people spend their time and interact with others.

As social media started gaining popularity in the mid-2000s, the mental health of adolescents and young adults in the United States began to worsen (Patel et al., 2007; Twenge et al., 2019). For instance, the total number of individuals aged 18–23 who reported experiencing a major depressive episode in the past year increased by 83% between 2008 and 2018 (NSDUH, 2019). Similarly, over the same time period, suicides became more prevalent and are now the second leading cause of death for individuals

15–24 years old (National Center for Health Statistics, 2021). Although the ultimate causes of these trends are still largely unknown, scholars have hypothesized that the diffusion of social media might be an important contributing factor (Twenge et al., 2019). In fact, concerns about a potential negative effect of social media on mental health have become so prominent that the U.S. Senate held a committee hearing about the topic in late 2021. Well-identified causal evidence, however, remains scarce.

In this paper, we provide quasi-experimental estimates of the impact of social media on mental health by leveraging a unique natural experiment: the staggered introduction of Facebook across U.S. colleges in the mid-2000s. Coupling survey data on college students' mental health collected in the years around Facebook's expansion with a generalized difference-in-differences empirical strategy, we find that the introduction of Facebook at a college had a negative impact on student mental health. We also find that, after the introduction of Facebook, students were more likely to report that poor mental health negatively affected their academic performance. Finally, we present an array of additional evidence suggesting that the results are consistent with Facebook enhancing students' abilities to engage in unfavorable social comparisons.

The early expansion of Facebook across colleges in the United States is a particularly promising setting to investigate the effects of social media use on the mental health of young adults. Facebook was created at Harvard in February 2004, but it was only made available to the general public in September 2006. Between February 2004 and September 2006, Facebook was rolled out across U.S. colleges in a staggered fashion. Upon being granted access to Facebook's network, colleges witnessed rapid and widespread Facebook penetration among students (Wilson, Gosling and Graham, 2012; Brügger, 2015). The staggered and sharp introduction of Facebook across U.S. colleges provides a source of quasi-experimental variation in exposure to social media that we can leverage for causal identification.

We employ two main datasets in our analysis: the first dataset specifies the dates in which Facebook was introduced at 775 U.S. colleges; the second consists of the universe of answers to seventeen consecutive waves of the National College Health Assessment (NCHA), the most comprehensive survey about student mental and physical health available at the time of Facebook's expansion.

Our analysis relies on a generalized difference-in-differences research design, where one of the dimensions of variation is the college a student attends, and the other dimension is whether the student took the survey before or after the introduction of Facebook at her college. Under a parallel trends assumption, the college by survey-wave variation generated by the sharp but staggered introduction of Facebook allows us to obtain causal estimates of the introduction of Facebook on student mental health. Our empirical strategy allows us to rule out various confounding factors: first, college-specific differences fixed in time (e.g., students at more academically demanding colleges may have worse baseline mental health than students at less demanding colleges); second, differences across time that affect all students in a similar way (e.g., certain macro-economic fluctuations); third, mental health trends affecting colleges in different Facebook expansion groups differentially, but smoothly (e.g., colleges where Facebook was rolled out earlier may

be on different linear trends in terms of mental health than colleges where Facebook was rolled out later).[1] We also address recent econometric concerns with staggered difference-in-differences research designs by showing robustness to the use of a variety of alternative estimators.[2] Lastly, we complement the difference-in-differences strategy with a specification that exploits variation in length of exposure to Facebook across students within a college and survey wave, and that, therefore, does not rely on our baseline college-level parallel trends assumption for identification.

Our main finding is that the introduction of Facebook at a college had a negative effect on student mental health. Our index of poor mental health, which aggregates all the relevant mental health variables in the NCHA survey, increased by 0.085 standard deviation units as a result of the introduction of Facebook. As a point of comparison, this magnitude is around 22% of the effect of losing one's job on mental health, as reported in a meta-analysis by Paul and Moser (2009). We further benchmark our results against a clinically-validated depression scale: the PHQ-9 (Kroenke, Spitzer and Williams, 2001). The effect of the introduction of Facebook on our index of poor mental health is equivalent to a two-percentage-point increase in the share of students suffering from depression according to the PHQ-9 over a baseline of 25%. Lastly, we perform a back-of-the-envelope calculation to determine what fraction of the increased prevalence of severe depression among college students over the last two decades can be explained by the introduction of Facebook. Under a set of relatively strong assumptions, we calculate that the introduction of Facebook accounts for approximately 24% of such increase.

We highlight three additional results. First, the negative effects on mental health are strongest for students who, based on immutable characteristics such as gender and age, are predicted to be most susceptible to mental illness. For those students, we also observe a significant increase in depression diagnoses, take-up of psychotherapy for depression, and use of anti-depressants. Second, in the short-to-medium run, the negative effects of Facebook on mental health increase with length of exposure to the platform. Third, students reported suffering some negative downstream effects as a result of their worsened mental health conditions. Specifically, after the introduction of Facebook, students were more likely to report that their academic performance was negatively affected by conditions related to poor mental health.

What explains the negative effects of Facebook on mental health? The pattern of results is consistent with Facebook increasing students' ability to engage in unfavorable social comparisons. Two main pieces of evidence bear on this conclusion. First, we find that the effects are particularly pronounced for students who might view themselves as comparing unfavorably to their peers, such as students who live off-campus—and therefore are more likely to be excluded from on-campus social activities—students of lower socioeconomic status, and students not belonging to fraternities/sororities. Second, we show that the introduction of Facebook directly affected the students' beliefs about their peers' social lives and behaviors. Consistent with the content on Facebook at the

---

[1]The last confounding factor in the list is taken into account in a specification that includes linear time trends at the Facebook-expansion-group level.

[2]See Roth et al. (2022) for a recent overview.

time, changes in perceptions are limited to alcohol as opposed to other drugs. As far as other channels are concerned, we do not find significant evidence that the negative effects of Facebook on mental health are due to disruptive internet use. We also rule out several alternative mechanisms, such as reduced stigma about mental illness and direct effects on drug use, alcohol consumption, and sexual behaviors.

The results presented in this paper, which rely on the staggered roll-out of Facebook across U.S. college campuses in 2004-2006, should be interpreted with caution for several reasons. First, our estimates cannot speak directly to the effects of social media features— e.g., news pages—that were introduced after the time period we analyze. Similarly, our estimates cannot speak directly to the possibility that years of experience with the platform might teach users ways to mitigate the negative effects on mental health.[3] Second, despite being the core component of most mental health diagnoses, self-reports may still suffer from measurement error due to recall bias, lack of incentives, and social image concerns.[4] Finally, we note that our estimates are local to college students, a population of direct interest in the discussion about the recent worsening of mental health trends among adolescents and young adults. Nevertheless, future research should test whether social media has a similar effect on the mental health of other demographic groups.

Aside from these caveats, our findings are in line with the hypothesis that social media has a negative impact on mental health and played a role in the increase in mental illness among adolescents and young adults over the past two decades. Of course, our results do not imply that the overall welfare effects of social media are necessarily negative. Such calculation would require estimating the effects of social media along various other dimensions; furthermore, they would require taking into account potential positive effects, such as a reduction in the cost of connecting with friends and family across a distance. Nevertheless, we believe our results will be informative to social media users and policymakers alike.

This paper contributes to the literature by providing the most comprehensive causal evidence to date on the effects of social media on mental health. The three closest papers to ours—Allcott et al. (2020); Mosquera et al. (2020), and Allcott, Gentzkow and Song (2021)—feature experiments that incentivize a randomly-selected subset of participants to reduce their social media use.[5] Those studies find negative effects of social media use on self-reported well-being, and Allcott, Gentzkow and Song (2021) shows evidence of digital addiction. Our findings complement the aforementioned literature in five main ways. First, our mental health outcome variables are more comprehensive and detailed than the ones in the experimental papers above. Specifically, our outcome variables include eleven questions related to depression—covering symptoms, diagnoses, take-up of psychotherapy, and use of anti-depressants—and various questions related to other

---

[3]One of our specifications— Equation (4)—can look at up to two-and-a-half years of experience with the platform and finds that the effects, if anything, increase in the short-to-medium term. Longer-term effects, however, could be quite different.

[4]The effects on academic performance are also self-reported and could suffer from similar issues.

[5]For correlational evidence on the link between social media and mental health, see Lin et al. (2016); Dienlin, Masur and Trepte (2017); Berryman, Ferguson and Negy (2018); Kelly et al. (2018); Bekalu, McCloud and Viswanath (2019); Twenge and Campbell (2019).

mental health conditions, ranging from seasonal affect disorder to anorexia. By contrast, the three experimental studies above measure broadly-defined subjective well-being and include only one question that relates directly to a mental health condition listed in the Diagnostic and Statistical Manual of Mental Disorders (DSM-V). Second, rather than studying the partial equilibrium effects of paying isolated individuals to reduce their social media use, our estimates capture the general equilibrium effects of introducing social media in an entire community.[6] Such general equilibrium effects are arguably particularly important for technologies like social media that exhibit strong network externalities. Third, our analysis is less prone to experimenter demand, Hawthorne, and income effects.[7] Fourth, the experiments above study fairly short-term disruptions in social media use, ranging from one to twelve weeks; conversely, we can estimate effects up to several semesters after the introduction of Facebook at a college. Fifth, our study specifically targets the population—young adults—that experienced the most severe deterioration in mental health in recent decades and studies it around the time in which those mental health trends began to worsen. Focusing on young adults is arguably important for two additional reasons: first, because early adulthood may be a particularly vulnerable time as far as mental health is concerned (Kessler et al., 2007); second, because early adulthood is an age in which individuals often make critical life decisions.

This paper also relates to the rapidly-growing literature in economics about the determinants and consequences of mental illness (Ridley et al., 2020). Research on the determinants of mental illness showed that unconditional cash transfers, in-utero exposure to the death of a maternal relative, unemployment shocks, and economic downturns can affect mental health (Paul and Moser, 2009; Haushofer and Shapiro, 2016; Persson and Rossin-Slater, 2018; Golberstein, Gonzales and Meara, 2019). Donati et al. (2022) provide quasi-experimental evidence that access to high-speed Internet increased the incidence of mental disorders among young adults in Italy. We contribute to this strand of the literature by focusing on social media, which many consider to be an important driver of the recent rise in depression rates among adolescents and young adults (Twenge, 2017; Twenge et al., 2019). Studies focusing on the consequences of mental illness found that better mental health is associated with fewer crimes, increased parental investment in children, and better labor market outcomes (Blattman, Jamison and Sheridan, 2017; Biasi, Dahl and Moser, 2019; Baranov et al., 2020; Shapiro, 2021). We complement this literature by showing that, after the introduction of Facebook, students were more likely to report experiencing impairments to academic performance as a result of poor mental health.

Lastly, this paper contributes to an emerging literature exploiting the expansion of

---

[6]There are likely substantial endogenous adjustments of one's social media use to one's peers' social media use, as well as spillover effects on one's mental health due to one's peers' social media use. We employ the term 'general equilibrium effects' to indicate that our estimates capture such indirect effects, as well as any direct effects.

[7]In the case of the experiments listed above, subjects in the treatment group are paid to reduce their social media use and are therefore not blind to treatment status, which might give rise to experimenter demand effects. In addition, the mere fact of being observed (e.g., via daily text messages asking participants how they feel) might affect subjects' behaviors independently of treatment status, giving rise to general Hawthorne effects. Lastly, incentive payments might directly affect self-reported well-being and confound interpretation. An additional issue with social media experiments is that they often screen participants who do not meet certain criteria and, therefore, employ rather selected samples. For instance, the main sample analyzed in Allcott et al. (2020) includes participants who reported using Facebook more than 15 minutes per day and were willing to accept $102 to deactivate their Facebook accounts for a month.

social media platforms to study the effects of social media on a variety of outcomes. The empirical strategy adopted in this paper is closely related to the one in Armona (2019), who leverages the staggered introduction of Facebook across U.S. colleges to study labor market outcomes more than a decade later. Enikolopov, Makarin and Petrova (2020) and Fergusson and Molina (2020) exploit the expansion of the social media platform VK in Russia and of Facebook worldwide, respectively, to show that social media use increases protest participation. Bursztyn et al. (2019) and Müller and Schwarz (2020) exploit the expansion of VK and Twitter, respectively, and find that social media use increases the prevalence of hate crimes.[8] A unique feature of our setting is that it allows us to measure the effects of the sharp roll-out of the biggest social media platform in the world at a time in which very few close substitutes were available.

The remainder of the paper is organized as follows: Section I provides some background on mental health and on Facebook's early expansion; Section II describes the data sources used in the analysis and presents descriptive statistics; Section III discusses the empirical strategy; Section IV presents the results; Section V explores mechanisms; Section VI discusses potential implications of the results; Section VII concludes.

I. Background

Mental Health   Mental illnesses, such as depression, anxiety, bipolar disorder, and schizophrenia, are disturbingly common and can be highly debilitating. According to the Global Burden of Disease Study, around a billion people in the world suffered from mental disorders in 2017, with depression and anxiety-related disorders as the leading conditions (James et al., 2018). In the U.S., around 1 in 5 adults experiences some form of mental illness each year, and 1 in 20 experiences serious mental illness (NAMI, 2020). Mental health conditions can have severe adverse effects, hampering people's ability to work, study, and be productive. According to the WHO's Global Burden of Disease, mental illness is the most burdensome disease category in terms of total disability-adjusted years for adults younger than 45, and depression is one of the most taxing conditions (WHO, 2008; Layard, 2017).

Alarmingly, the last two decades witnessed a worsening of mental health trends in the United States, especially among adolescents and young adults (Twenge et al., 2019). As shown in Appendix Figure A.1, self-reported episodes of psychological distress and depression have risen substantially over the past fifteen years, with the highest growth rate among young adults. Similarly, both self-reports of suicidal thoughts, plans, or attempts and actual suicides have increased considerably among that demographic group. Since the timing of the divergence in mental health trends between young adults and older generations roughly coincides with wider adoption of social media, various scholars have hypothesized the two phenomena might be related (Twenge, 2017; Twenge et al., 2019).

A Brief History of Facebook's Expansion and Initial Popularity   Facebook was created at Harvard in February 2004 and was rolled out gradually to other colleges in the

---

[8]Additional research on social media and political outcomes includes Enikolopov, Petrova and Sonin (2018), Fujiwara, Müller and Schwarz (2021), and Levy (2021). For a detailed overview, see Zhuravskaya, Petrova and Enikolopov (2020).

U.S. and abroad over the subsequent two-and-a-half years. The staggered nature of the roll-out was enforced by requiring users to be in possession of verified email addresses (e.g., addresses ending in @harvard.edu). The roll-out of Facebook across U.S. colleges was not random: as shown in the descriptive statistics section, more selective colleges were granted access to Facebook relatively earlier than less selective colleges. The staggered nature of the expansion is arguably due to three factors: first, scale constraints due to limited server capacity (Kirkpatrick, 2011); second, Facebook's willingness, at least at the outset, to maintain a flavor of exclusivity; third, Facebook's desire to strengthen network effects by keeping the fraction of users who likely knew each other offline artificially high (Aral, 2021).

Even in its infancy, Facebook was extremely popular. Upon being granted access to the platform, colleges witnessed rapid and very widespread adoption among students.[9] To get a sense of the early adoption rates among college students, we matched data provided by Facebook on the number of users at each of the first 100 colleges that were granted access to the platform with IPEDS (Integrated Postsecondary Education Data System) data on the number of full-time undergraduate students at those colleges (U.S. Department of Education, 2005; Traud, Mucha and Porter, 2012). Appendix Figure A.2 presents a histogram of the number of users per 100 undergraduate students at those colleges and shows that, in September 2005, there were on average 86 Facebook users for every 100 undergraduate students. This result is consistent with Facebook's statement that, across all the colleges with access to the platform as of September 2005, approximately 85% of students had a Facebook profile (Arrington, 2005).[10]

Not only was Facebook immediately popular, usage was also quite intense. In early 2006, close to three-quarters of users logged into the site at least once a day, and the average user logged in six times a day (Hass, 2006). As of early 2006, Facebook was the ninth most visited website on the Internet, despite not yet being open to the general public (Hass, 2006).

---

[9]According to a description by Kirkpatrick (2011): "*within days, [Facebook] typically captured essentially the entire student body, and more than 80 percent of users returned to the site daily*" (p. 88).

[10]Various smaller-scale studies using survey and/or Facebook data and focusing mostly on undergraduate students confirm the high adoption rates in 2005–2006. Specifically, those studies show that, at the colleges in which they were administered, 82%–94% of students had a Facebook account (Stutzman, 2006; Kolek and Saunders, 2008; Lampe, Ellison and Steinfield, 2008). While women may have been more likely than men to join Facebook, Facebook usage was very common across demographic groups (Kolek and Saunders, 2008).

II. Data Sources & Descriptive Statistics

A. Data Sources

Our analysis relies primarily on two data sources. The first data source specifies the dates in which Facebook was introduced at 775 U.S. colleges. The second consists of the universe of answers to seventeen consecutive waves of the National College Health Assessment (NCHA) survey—the largest and most comprehensive survey on college students' mental and physical health at the time of Facebook's expansion.

Facebook Expansion Dates Data The Facebook Expansion Dates dataset was assembled in two steps: for the first 100 colleges that received Facebook access, we rely on introduction dates collected and made public in previous studies (Traud, Mucha and Porter, 2012; Jacobs et al., 2015). For the remaining 675 colleges in the dataset, we obtained Facebook introduction dates using the Wayback Machine, an online archive that contains snapshots of various websites at different points in time and that allows users to visit old versions of those websites. Specifically, between the spring of 2004 and the spring of 2005, the front page of Facebook's website was regularly updated to show the most recent set of colleges that had been given access to the platform.[11] As an example, Appendix Figure A.3 shows the front page of Facebook as of June 15th 2004, recovered via the Wayback Machine. As shown in the figure, Facebook was open to 34 colleges at that point in time.

Armed with a time-series of snapshots of the front page of Facebook's website, it is possible to reconstruct tentative dates in which Facebook was rolled out at each college. Specifically, the roll-out date at a certain college should be between the date of the first snapshot in which the college is listed and the date of the previous snapshot. When the distance between the snapshots is more than one day, we consider the first date in which a college is listed on Facebook's front page as the introduction date.

Since the Wayback Machine took snapshots of Facebook's website at a high temporal resolution, our imputation procedure for the introduction dates is rather precise. For the months in which our introduction dates rely on the Wayback Machine—September 2004 to May 2005—the average number of days between consecutive snapshots is one and a half. Therefore, on average, our imputed introduction dates should be within two days of the actual introduction dates.

Appendix Table A.32 lists the colleges in the Facebook Expansion Dates dataset and the date in which Facebook was rolled out at each of them.

National College Health Assessment Data Our second main data source consists of more than 430,000 responses to the National College Health Assessment (NCHA) survey, a survey administered to college students on a semi-annual basis by the American College Health Association (ACHA). The NCHA survey was developed in 1998 by a

---

[11]Beginning with the fall of 2005, Facebook started listing the colleges that had access to the platform on a separate page that is snapshotted too infrequently to allow us to extract meaningful introduction dates. Therefore, our Facebook Introduction Dates dataset ends after the spring of 2005.

team of college health professionals with the purpose of obtaining information from college students about their mental and physical health. Specifically, the NCHA survey inquires about demographics, physical health, mental health, alcohol and drug use, sexual behaviors, and perceptions of these behaviors among one's peers.

As far as mental health is concerned, the NCHA survey includes both questions about symptoms of mental illness and questions about take-up of mental healthcare services. We emphasize that reliance on self-reported symptoms is part of standard medical practice in the domain of mental health (Chan, 2010). Specifically, according to the official diagnostic manual of the American Psychiatric Association (DSM-V), the diagnosis of many mental health disorders including depression relies almost exclusively on patients' self-reports of symptoms such as difficulty sleeping, anhedonia, fatigue, feelings of worthlessness and guilt, diminished ability to think or concentrate, and recurrent suicidal thoughts (APA, 2013). In fact, self-administered questionnaires inquiring about depression symptoms have been shown to predict medical diagnoses with accuracy rates up to 90% (Kroenke and Spitzer, 2002).[12]

The NCHA dataset includes the universe of responses to all NCHA survey waves administered between the spring of 2000 and the spring of 2008, the longest stretch of time around Facebook's early expansion in which the content of the survey remained constant.[13] Colleges included in the NCHA dataset administered the survey to randomly selected classrooms, randomly selected students, or all students. The average response rate across the survey waves for which we have such information is 37% (ACHA, 2000–2019). In order to assuage concerns about the possibility that the introduction of Facebook affected the composition of students who participated in the survey, Appendix Tables A.3 and A.10 show that, along the demographic characteristics elicited in the NCHA survey, there are no meaningful compositional changes following the introduction of Facebook. Throughout our analysis, we limit our sample to full-time undergraduate students.[14]

The NCHA dataset is an unbalanced panel, in which colleges drop in and out. Specifically, every college in the U.S. can voluntarily select into any wave of the NCHA survey and is not required to keep administering the survey in subsequent years. To account for compositional changes to the panel, our preferred specification includes college fixed effects.

The NCHA survey does not include any questions on social media use; therefore, it is not possible for us to determine whether a particular survey respondent had a Facebook account. It is, however, possible to determine whether the college attended by the survey respondent had Facebook access at the time in which the respondent took the survey. In order to protect the privacy of the institutions that participate in the NCHA survey while still allowing us to carry out the analysis, the ACHA kindly agreed to provide us with a customized dataset that includes a variable indicating the semester in which Facebook

[12]Section III, Appendix B, and Appendix C discuss our symptom measures in detail and present an array of exercises to validate them.

[13]Between 1998 and 2000, the survey was being fine-tuned and changed considerably across survey waves; similarly, after the spring of 2008, the survey underwent a major revision that substantially limits comparability to previous waves.

[14]Graduate students also had access to the Facebook platform, but take-up was a lot smaller among graduate students than among undergraduates (e.g., Acquisti and Gross, 2006).

was rolled out at each college. Specifically, the ACHA adopted the following procedure: i) merge our dataset containing the Facebook introduction dates to the NCHA dataset; ii) add a variable listing the semester in which Facebook was rolled out at each college;[15] iii) strip away any information that could allow us to identify colleges (including the specific date in which Facebook was introduced at each college).

B. Descriptive Statistics

Appendix Tables A.1 and A.2 present college-level and student-level descriptive statistics for colleges in different Facebook expansion groups.[16] Appendix Table A.1 shows that colleges in earlier Facebook expansion groups are more selective in terms of test scores, larger, more likely to be on the East Coast, and have more residential undergraduate programs than colleges in later Facebook expansion groups. Panel A of Appendix Table A.2, which averages student-level variables available in the NCHA dataset across the different Facebook expansion groups, shows that colleges in earlier Facebook expansion groups enroll students from relatively more advantaged economic backgrounds. Lastly, Panel B of Appendix Table A.2 shows that students in colleges that received Facebook relatively earlier have worse baseline mental health outcomes than students attending colleges in later Facebook expansion groups.[17] The baseline differences across Facebook expansion groups may lead one to wonder about the plausibility of the parallel trends assumption in this setting; we address concerns related to parallel trends in Section III.

Appendix Table A.1 also shows the number of colleges in the NCHA dataset that received Facebook access in each semester between the Spring of 2004 and the Fall of 2005. Other than the Spring of 2004, when Facebook was first introduced, the fraction of colleges in the NCHA dataset that received Facebook access in each semester is fairly equally distributed across the remaining introduction semesters.

III. Empirical Strategy

Construction of the Primary Outcome Variables In order to mitigate concerns about cherry-picking outcome variables, we consider all the questions in the NCHA survey that are related to mental health and that inquire about a respondent's recent past

---

[15]For the set of colleges that appear both in our introduction-date dataset and the NCHA survey, the ACHA listed the semesters corresponding to the introduction dates in our dataset. For the set of colleges that appear only in the NCHA dataset, we list the Fall of 2005 as the semester in which Facebook was introduced at those colleges. Such imputation is sensible in virtue of the fact that our introduction-date dataset ends after the spring semester of 2005 and that, by the end of 2005, the vast majority of U.S. colleges had been granted access to Facebook. As shown in Appendix A, the results are robust to dropping those colleges altogether.

[16]Appendix Table A.1 was constructed by merging our Facebook Expansion Dates dataset with data from IPEDS. We cannot directly provide college-level summary statistics using the NCHA dataset, because most college-level information in the NCHA was stripped away for privacy reasons.

[17]The differences in baseline mental health across Facebook expansion groups are particularly stark when comparing the first Facebook expansion group to the other groups; among the other groups the differences are more muted. In Appendix A, we present and discuss a robustness check showing that our results do not significantly change when we drop colleges in each expansion group in turn or when we interact college-level baseline mental health with survey-wave fixed effects.

(e.g., "Within the last school year, how many times have you felt so depressed that it was difficult to function?").

To impose structure on our analysis and assuage concerns about multiple hypothesis testing, we group the individual mental health variables into nested families and combine them into indices. The coarsest level of analysis combines all the mental health questions (*index of poor mental health*); a second level of analysis splits symptoms of mental illness (*index symptoms poor mental health*) and self-reported take-up of depression-related services (*index depression services*) into separate families; a third level of analysis splits the symptoms of mental illness into depression-related symptoms (*index of depression symptoms*) and symptoms related to other mental health conditions (*index symptoms other mental health conditions*); the finest level comprises the individual variables themselves.

The index of depression symptoms includes questions that inquire as to whether a student exhibited various symptoms of depression such as feeling hopeless, overwhelmed, exhausted, very sad, debilitatingly depressed, seriously considered committing suicide, or attempted suicide. The index of symptoms of other mental health conditions includes questions that inquire as to whether a student experienced issues related to anorexia, anxiety disorder, bulimia, and seasonal affect disorder. The overall index of symptoms of poor mental health encompasses both sets of symptoms.

The index of depression services requires a slightly more detailed discussion due to a peculiarity in the way the questions were structured. Specifically, the NCHA survey asked three questions about depression-related services: i) whether the student was diagnosed with depression within the year prior to taking the survey, ii) whether the student was in therapy for depression at the time in which she took the survey, and iii) whether the student was on anti-depressants at the time in which she took the survey. The NCHA survey asked those questions only to students who had given an affirmative answer to a previous question inquiring as to whether they had ever been diagnosed with depression. Therefore, the variables related to the three questions above should be interpreted as "having ever received a depression diagnosis" plus "having received a depression diagnosis in the last year", or "being in therapy for depression", or "taking anti-depressants." Under this interpretation, we can safely impute zeros to the three questions about depression-related services for students who gave a negative answer to the question about whether they had ever been diagnosed with depression.

Our indices are constructed as follows: first, we orient all variables that compose an index in such a way that higher values always indicate worse mental health outcomes; second, we standardize those variables using means and standard deviations from the pre-period; third, we take an equally-weighted average of the index components, excluding from the analysis observations in which any of the components are missing; fourth, we standardize the final index. This way, our indices are essentially *z*-scores.[18]

Appendix Table A.31 lists all the variables used in our analysis, describes their construction in detail, and includes the exact wording of the questions in the NCHA survey that each variable is based on.

---

[18]In Appendix A, we show that are results are unchanged if we construct the indices in other ways, for instance as described in Anderson (2008).

Validation of the Primary Outcome Variables   We validate the NCHA survey questions that form the basis of our primary outcome variables both internally and externally. We validate the questions about symptoms of mental illness internally by relating them to self-reported mental healthcare diagnoses within our dataset. Appendix B presents an array of validation exercises suggesting that the questions about symptoms of mental illness in the NCHA survey are indeed highly predictive of mental illness diagnoses.

We validate the NCHA survey questions externally by conducting an original survey on more than 500 college students. Our survey contained both the questions from the NCHA survey that feature in the construction of our index of poor mental health and the questions from canonical depression and generalized anxiety disorder screeners—the PHQ-9 and GAD-7 respectively—known to be highly predictive of medical diagnoses (Kroenke, Spitzer and Williams, 2001; Spitzer et al., 2006). Appendix Figures A.14-A.15 show that our index of poor mental health is strongly correlated with the PHQ-9 and GAD-7 scores (correlation coefficients of 0.66 and 0.61 respectively). The validation exercise is described in detail in Appendix C.

Construction of the Treatment Indicator   The construction of our treatment indicator is straightforward but for a minor caveat. A respondent to the NCHA survey is considered treated if, at the time the respondent took the survey, Facebook was available at her college and not treated otherwise. The caveat relates to the fact that we cannot determine whether or not a respondent was treated when the semester in which she took the survey coincides with the semester in which Facebook was rolled out at her college. For most of the analysis, we disregard such observations. In Appendix A, we show that the results do not substantially change depending on whether we consider those respondents treated, untreated, or whether we assign them a treatment status of 0.5.

Identification Strategy   The primary goal of this paper is to identify the causal impact of social media on mental health. A naive correlation may be plagued by severe endogeneity concerns and, therefore, cannot credibly be given a causal interpretation. Examples of such endogeneity concerns include reverse causality (e.g., depressed individuals could use social media more) and omitted variable bias (e.g., the end of a romantic relationship might lead to both worse mental health outcomes and more free time to spend on social media).

To obtain estimates that can be more credibly interpreted as causal, we leverage the sharp and staggered roll-out of Facebook across U.S. colleges in the years 2004 through 2006. Under a set of assumptions described below, the quasi-experimental variation generated by the staggered Facebook roll-out allows us to estimate the causal impact of social media on mental health using a generalized difference-in-differences strategy. The strategy compares the before-after difference in outcomes between students in colleges where Facebook was introduced and students in colleges that did not change their Facebook status between the two periods.

As a baseline specification, we estimate the following two-way fixed-effect (TWFE) model:

$$(1) \qquad Y_{icgt} = \alpha_g + \delta_t + \beta \times \text{Facebook}_{gt} + \mathbf{X}_i \cdot \gamma + \mathbf{X}_c \cdot \psi + \varepsilon_{icgt},$$

where Y_{icgt} represents an outcome for individual i who participated in survey wave t and attends college c that belongs to expansion group g; α_g (or sometimes α_c) indicates expansion-group (or college) fixed effects; δ_t indicates survey-wave fixed effects; Facebook_{gt} is an indicator for whether, in survey wave t, Facebook was available at colleges in expansion group g; $\mathbf{X}_i$ and $\mathbf{X}_c$ are vectors of individual-level and college-level controls, respectively. We estimate Equation (1) using OLS and cluster standard errors at the college level.

To the extent that, in the absence of the Facebook roll-out, the mental health outcomes of students attending colleges in different Facebook expansion groups would have evolved along parallel trends, and assuming college-level average treatment effects are homogeneous across treated colleges and over time, the coefficient of interest β identifies the average treatment effect on the treated (ATT) of the introduction of Facebook on student mental health.

Under the assumptions from the previous paragraph, the two-way fixed-effect (TWFE) model allows us to rule out various concerns that could otherwise impair our ability to interpret the results as causal. First, we can rule out that the results are driven by time-invariant differences in mental health across colleges. Specifically, one could worry that more selective colleges recruit wealthier students who may have better (or worse) baseline mental health outcomes. By including Facebook-expansion-group or, depending on the specification, college fixed effects we can rule out such concerns. Second, we can rule out that our results are driven by mental health outcomes evolving over time in a way that is common across students at different colleges. For instance, certain macroeconomic fluctuations might influence all students' job prospects in a similar way, and, in turn, affect their mental health. Survey-wave fixed effects allow us to rule out such concerns.

One may worry about the plausibility of the parallel trends assumption in our setting—that is, one might worry that colleges belonging to different Facebook expansion groups might be on different mental health trends. We address this concern in four ways. First, we estimate a fully dynamic version of Equation (1) and check for potential pre-trends. Second, we explore the existence of pre-trends by estimating a fully dynamic version of the alternative estimators introduced in De Chaisemartin and d'Haultfoeuille (2020), Borusyak, Jaravel and Spiess (2021), Callaway and Sant'Anna (2021), and Sun and Abraham (2021). Third, to the extent that the trends are linear, we would be able to account for them in a robustness check that includes expansion-group-level linear time trends. Fourth, we present results using a specification that does not rely on our baseline college-level parallel trends assumption. In particular, we present results using a specification that includes college×survey-wave fixed effects and that compares students within the same college–survey-wave who were exposed to Facebook for different lengths of time based on the year in which they entered college. These strategies, explored in detail in later sections, should assuage concerns about violations of the parallel trends

assumption in our setting.

Limitations of TWFE models and suggested remedies   Although TWFE regressions similar to Equation (1) are the workhorse models for staggered adoption research designs, they have been shown to deliver consistent estimates only under relatively strong assumptions about homogeneity in treatment effects (De Chaisemartin and d'Haultfoeuille, 2020; Borusyak, Jaravel and Spiess, 2021; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021). Specifically, as shown in Goodman-Bacon (2021), the treatment effect estimate obtained from a TWFE model is a weighted average of all possible 2×2 difference-in-differences comparisons between groups of units treated at different points in time. If treatment effects are homogeneous across treated groups and across time, the TWFE estimator is consistent for the average treatment effect on the treated (ATT). Conversely, if treatment effects are heterogeneous across groups or time, the TWFE estimator does not deliver consistent estimates for the ATT.

We address concerns about the reliability of TWFE estimator by replicating our results using the robust estimators introduced in De Chaisemartin and d'Haultfoeuille (2020), Borusyak, Jaravel and Spiess (2021), Callaway and Sant'Anna (2021), and Sun and Abraham (2021). By shutting down the 2×2 difference-in-differences comparisons between newly-treated and already-treated units, the robust estimators deliver consistent estimates even in the presence of heterogeneous treatment effects across time and/or treated units.

IV.   Results

A.   Baseline Results

Baseline Estimates   Table 1 presents estimates of β in Equation (1) on our overall index of poor mental health and shows that the introduction of Facebook at a college had a negative impact on student mental health. The first column in the table shows results for our simplest specification, which includes only Facebook-expansion-group fixed effects, survey-wave fixed effects, and an indicator for post-Facebook introduction. In the second column, we also include individual- and college-level control variables. In the third column, we replace Facebook-expansion-group fixed effects with college fixed effects to account for the changing composition of our sample. In the fourth column, we add expansion-group-level linear time trends, in order to take into account the possibility that colleges belonging to different Facebook expansion groups might be on different linear mental-health trends. Our results are fairly stable across specifications. The point estimates decrease but remain significant at the 5 percent level when college fixed effects and Facebook-expansion-group-level linear time trends are included.

The effect size on the index of poor mental health in our preferred specification, namely the one that includes college rather than Facebook-expansion-group fixed effects and that does not include linear time trends, is 0.085 standard deviation units. The effect above is estimated on the entire population of students taking the NCHA survey, which

Table 1: Baseline Results: Index of Poor Mental Health

	Index of Poor Mental Health			
	(1)	(2)	(3)	(4)
Post Facebook Introduction	0.137	0.124	0.085	0.077
	(0.040)	(0.022)	(0.033)	(0.032)
Observations	374,805	359,827	359,827	359,827
Survey Wave FE	✓	✓	✓	✓
FB Expansion Group FE	✓	✓		
Controls		✓	✓	✓
College FE			✓	✓
FB Expansion Group Linear Time Trends				✓

Notes: This table explores the effect of the introduction of Facebook at a college on student mental health. Specifically, it presents estimates of coefficient β from Equation (1) with our index of poor mental health as the outcome variable. The index is standardized so that, in the pre-period, it has a mean of zero and a standard deviation of one. Column (1) estimates Equation (1) without including controls; column (2) estimates Equation (1) including controls; column (3)—our preferred specification—replaces Facebook-expansion-group fixed effects with college fixed effects; column (4) includes linear-time trends estimated at the Facebook-expansion-group level. Our controls consist of: age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. Column (2) also includes indicators for geographic region of college (Northeast, Midwest, West, South); such indicators are omitted in columns (3) and (4) because they are collinear with the college fixed effects. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.31. Standard errors in parentheses are clustered at the college level.

includes both students who did and who did not sign-up for a Facebook account after Facebook was made available at their college. Therefore, the point estimate captures both the direct effect of Facebook on students who joined the platform and the indirect effect of Facebook on students who did not join the platform, but whose peers did. Although we cannot separate these two channels in the absence of data on an individual's Facebook use, we note that it is unlikely that our results are primarily driven by students who did not have a Facebook account.[19]

In order to help build intuition about the magnitude of our baseline effects, we provide a few benchmarks. First, the magnitude of our baseline effect corresponds to approximately 84% of the difference in the index of poor mental health between students in our sample with and without credit card debt. Second, we benchmark the magnitude of our estimates against the effect of a sudden unemployment spell on mental health. Comparing our estimates to the most closely related ones in a meta-analysis by Paul and Moser (2009), we find that the impact of introducing Facebook at a college on mental health is around 22% of the effect of job loss.[20] Third, we benchmark our results against the canonical PHQ-9 and GAD-7 mental health scales. We use data from the validation

---

[19]As discussed in Section I, the average penetration rate of Facebook at each college was around 85%. Therefore, an effect concentrated solely among students who did not join the platform would have to be implausibly large—approximately 0.57 standard deviations in our main specification—to be consistent with our baseline effect.

[20]Paul and Moser (2009) analyze studies estimating various aspects of mental health including symptoms of distress, depression, anxiety, psychosomatic symptoms, subjective well-being, and self-esteem. The estimates from Paul and Moser (2009) that can most credibly be interpreted as causal and hence be compared to our estimates are those that rely on quasi-experimental variation in job loss due to factory closures and mass layoffs.

survey mentioned in Section III and discussed in detail in Appendix C to determine how to weigh the variables contained in our index of poor mental health in a way that best predicts an indicator for having depression according to the PHQ-9 and an indicator for having generalized anxiety disorder according to the GAD-7. Next, we apply these weights to the NCHA sample to predict whether a student taking the NCHA survey would be classified as having depression or generalized anxiety disorder according to the PHQ-9 and GAD-7. Appendix Table A.30 shows that the introduction of Facebook increased by two percentage points the fraction of students whom, according to our prediction, the PHQ-9 and GAD-7 would classify as having depression or generalized anxiety disorder. The two percentage point increase corresponds to a 9% increase over the pre-period mean of 25% for depression and a 12% increase over the pre-period mean of 16% for generalized anxiety disorder.[21]

As a final benchmark, we leverage additional assumptions to compare our results to long-run mental health trends. The effect we find on the share of students who suffered from severe depression at least once in the last year is approximately 24% of the increase in that share between 2000 and 2019.[22] This number can be interpreted as the fraction of the increase in the prevalence of severe depression among college students that is explained by Facebook. Such calculation relies on strong assumptions and should therefore be interpreted with caution. Specifically, we assumed that: i) Facebook utilization rates among college students did not change substantially after 2004-2005; ii) the effects of Facebook did not change over time; iii) Facebook does not have cumulative effects.[23]

Figure 1 presents results on our individual outcome variables and shows that most of the dimensions of mental health in our dataset were negatively affected by the introduction of Facebook.[24] For all but one of the mental health outcomes from Figure 1, the point estimates are positive, which indicates worsened mental health. The conditions that appear to be most affected are depression and anxiety-related disorders, while the point estimates on anorexia and bulimia are close to zero.[25] The effect on severe depression is similar in magnitude to the effect observed in Allcott et al. (2020) on whether a respondent

---

[21]This exercise is discussed in more detail in Appendix C.

[22]Data on the prevalence of severe depression among students comes from ACHA reports containing aggregate statistics about mental health (ACHA, 2000–2019). Since the wording of the question inquiring about severe depression changed in 2008 and caused a clear series break, we calculate the trend in depression by regressing the share of severely depressed students on year dummies, on whether the survey was conducted in the spring or fall, and on whether the survey contained the new wording. We define the trend in depression as the point estimate of the 2019 fixed effect dummy. According to our calculation, the share of severely depressed students increased by approximately 12 percentage points between 2000 and 2019. Based on our main specification, the introduction of Facebook at a college increased the share of students who reported suffering from severe depression at least once in the past year by 2.96 percentage points (p-value<0.05). Hence, the effect of the introduction of Facebook is approximately 24% (2.96/12.15) of the increase in depression rates between 2000 and 2019.

[23]Subsection IV.C, which shows that the negative effects of Facebook on mental health become stronger with longer exposure to the platform, already casts some doubt on assumption iii).

[24]Appendix Table A.4 provides regression results for the individual mental health variables in both normalized (standard deviation) units and un-normalized (original) units. The table also provides unadjusted *p*-values and "sharpened" False Discovery Rate (FDR)-adjusted *q*-values following the procedure of Benjamini, Krieger and Yekutieli (2006), as outlined by Anderson (2008). The *p*-values are appropriate for readers with a priori interest in a particular outcome; the *q*-values adjust the inference for multiple hypotheses testing.

[25]Similar patterns can be observed in Figure A.5 which is a version of Figure 1 with expansion-group-specific linear trends.
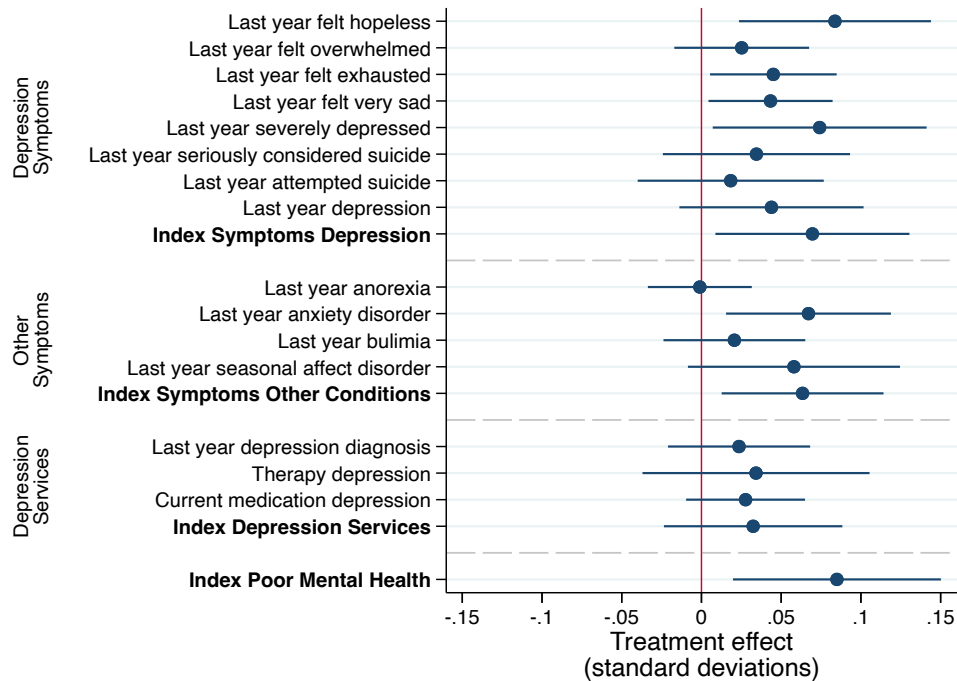
Figure 1: Effects of the Introduction of Facebook on Student Mental Health

Notes: This figure explores the effects of the introduction of Facebook at a college on all our mental-health outcome variables and on the related indices. Specifically, it presents estimates of coefficient β from Equation (1) using our preferred specification, namely the one including survey-wave fixed effects, college fixed effects, and controls. The outcome variables are our overall index of poor mental health, the individual components of the index, and three sub-indices: the index of depression symptoms, the index of symptoms of other mental health conditions, and the index of depression services. All outcomes are standardized so that, in the pre-period, they have a mean of zero and a standard deviation of one. Our controls consist of: age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. The reason why the point estimate on an index might be relatively large compared to the point estimates on each of the components of the index is that averaging across the index components reduces noise and, as a consequence, might increase the effect size measured in standard deviation units. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.31. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

felt depressed in the past month (0.07 vs. 0.09 standard deviations, respectively). This striking similarity is consistent with the possibility that the effects of the introduction of Facebook on depression are due primarily to direct use rather than general equilibrium effects. Having said that, the substantive differences between the studies, including the time period, target population, and empirical strategy, call for caution when drawing conclusions from such comparison.

The bottom section of Figure 1 also presents suggestive evidence that the introduction of Facebook at a college might have increased the extent to which students took up depression-related services. For all three items comprising the index of depression services—receiving an official depression diagnosis, going to therapy for depression, and taking anti-depressants—the point estimates are positive, though not significant at

conventional levels.[26] Finding a more muted average effect on depression-related services than on depression symptoms is arguably in line with intuition, in that an increase in symptoms of poor mental health induces the marginal student, rather than the average student, to take up mental healthcare services.[27] In Section IV.B below, we show that students who, based on immutable baseline characteristics, are predicted to be most susceptible to mental illness—and therefore more likely to be on the margin of receiving a depression diagnosis—are indeed significantly more likely to take up depression-related services after the introduction of Facebook.

Event Study Figures   In order to test for parallel trends and study the dynamics of treatment effects, we estimate an event study-version of the TWFE model with indicators for distance to/from the introduction of Facebook. Specifically, we estimate the following specification:

$$(2) \qquad\qquad Y_{igt} = \alpha_g + \delta_t + \beta_k \times \sum_{k=-8}^{5} D_{k(gt)} + \varepsilon_{igt},$$

where Y_{igt} is our index of poor mental health and $D_{k(gt)}$ is set of indicator variables that take value one if, for expansion group g in survey wave t, the introduction of Facebook was k semesters away. When estimating the model using OLS, we treat students who took the survey in the semester just before Facebook was rolled out at their college as the omitted category and compare them to students who took the NCHA survey in other semesters.

As discussed in Sun and Abraham (2021), the fully-dynamic version of the TWFE model in Equation (2) estimated using OLS delivers consistent estimates only under relatively strong assumptions regarding treatment effect homogeneity. In order to allow for heterogeneity in treatment effects across time and treated units, we also present the event study figures generated by a set of recently-proposed estimators that are robust to treatment effect heterogeneity (De Chaisemartin and d'Haultfoeuille, 2020; Borusyak, Jaravel and Spiess, 2021; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021).

Figure 2 presents the event-study figures and shows that the estimates are consistent with the parallel trends assumption: independently of the estimator used, the coefficients on the semesters prior to the introduction of Facebook at a college are all close to zero and exhibit no discernible pre-trends.[28] Figure 2 also sheds light on the dynamics of treatment effects: all the recently-developed robust estimators show treatment effects that

---

[26]Note that, given the low average take-up of these services, the estimates represent large increases over the baseline mean. For anti-depressants and psychotherapy, the point estimates represent an increase of about 13% and 20% over the baseline mean, respectively.

[27]The argument above relies on the baseline propensity to experience mental illness likely being normally distributed in the population (Plomin, Haworth and Davis, 2009) and the intuition that only individuals above a certain threshold in the right tail of the distribution experience sufficiently severe symptoms to seek out mental healthcare services.

[28]Appendix Figure A.4 shows the TWFE OLS estimates of a version of equation (2) that considers each of the first three Facebook expansion groups in turn and compares it to the last Facebook expansion group. These figures are constructed at the yearly level to reduce noise arising from the smaller number of observations. Consistent with Facebook having a negative impact on student mental health, in all the pairwise comparisons, all the estimates in the post-period are positive and most are statistically significant while the estimates in the pre-period are not statistically different from zero.
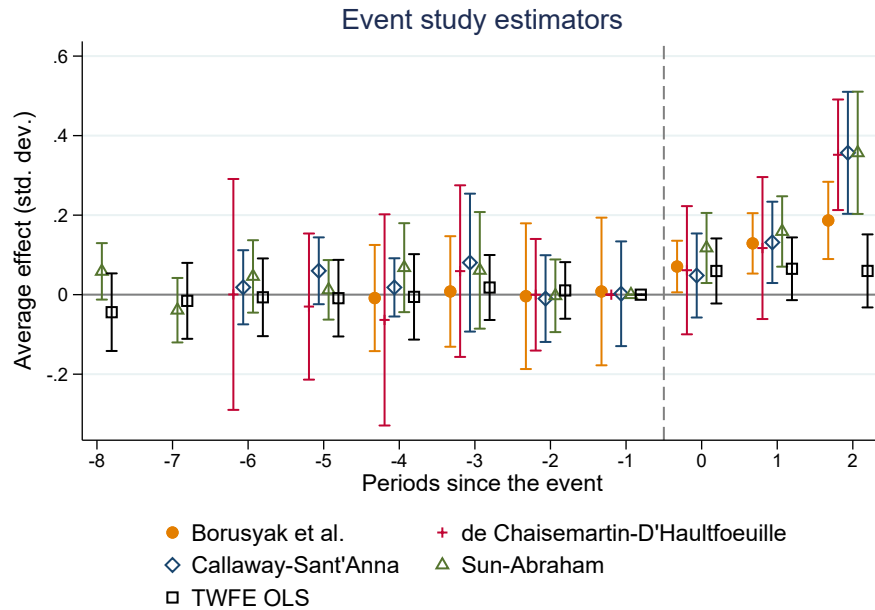
Figure 2: Effects of Facebook on the Index of Poor Mental Health Based on Distance to/from Facebook Introduction

Notes: This figure overlays the event-study plots constructed using five different estimators: a dynamic version of the TWFE model—Equation (2)—estimated using OLS (in black with square markers), Sun and Abraham (2021) (in green with triangle markers), Callaway and Sant'Anna (2021) (in blue with diamond markers), De Chaisemartin and d'Haultfoeuille (2020) (in red with cross markers), and Borusyak, Jaravel and Spiess (2021) (in orange with circle markers). The outcome variable is our overall index of poor mental health. The time variable is the survey wave and the treatment group variable is given by the semester in which the college attended by the student was granted Facebook access. The figure displays only two post-periods because the estimation of additional post periods would require employing already-treated units as controls for newly-treated units. In the presence of heterogeneous dynamic treatment effects, such comparisons would bias the estimation and, therefore, they are shut down by all the newly-introduced robust estimators. As a result, the maximum number of post-periods that can be estimated robustly is two. For the Borusyak, Jaravel and Spiess (2021) estimator, we estimate four pre-periods since estimating more pre-periods dramatically increases the standard errors in the pre-period (Borusyak, Jaravel and Spiess, 2021, p. 24). Similarly, for the estimator by De Chaisemartin and d'Haultfoeuille (2020), the maximum number of pre-periods that can be estimated in our panel is only five. In order to estimate the standard errors for the $t+2$ estimate, the De Chaisemartin and d'Haultfoeuille (2020) estimator includes controls for age and age squared. For appropriate estimation of the coefficients on $t=-8$ and $t=-7$ using the Sun and Abraham (2021) estimator, we include data from additional pre-periods, even though, in those pre-periods, we do not observe all four Facebook expansion groups (Sun and Abraham, 2021, p. 13). For a detailed description of the outcome and treatment variables, see Appendix Table A.31. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

increase over time in the post-periods.[29] The increase in treatment effects over time could be explained by: i) higher adoption rates at a college over time; ii) higher intensity of usage at the individual level over time; iii) the effects becoming stronger as a function of length of exposure to the platform. Given the evidence presented in Section I on the rapid and widespread penetration of Facebook at each college and evidence that intensity

---

[29]Contrary to the recently-developed robust estimators, the OLS estimator shows a relatively flat trend in the post-period. This is likely because, in the case of dynamically-increasing treatment effects, the OLS estimator, which uses already-treated units as controls for newly-treated units, exhibits a downward bias.

of usage did not increase substantially over time (Stutzman, 2006; Lampe, Ellison and Steinfield, 2008), we tentatively lean in favor of the length-of-exposure explanation. We further study the effects of differential length of exposure to Facebook at the individual level in Section IV.C

B. Heterogeneity

Heterogeneity by Predicted Susceptibility to Mental Illness In order to study whether the introduction of Facebook at a college led students on the margin of a depression diagnosis to take up depression-related services, we proceed in two steps: first, we estimate a LASSO to identify individuals who, based on baseline immutable characteristics, are more susceptible to mental illness. Second, we show heterogeneous treatment effects based on our LASSO-predicted measure of susceptibility to mental illness.

The LASSO prediction is generated as follows: first, we construct an indicator that equals one if a student has ever been diagnosed with a mental health condition. Second, we consider a set of immutable individual-level characteristics (age, year in school, gender, race, an indicator for U.S. citizenship, and height), generate all two-way interactions between these characteristics, and generate second- and third-order monomials of each characteristic. Third, we implement a LASSO procedure in the pre-period to predict our indicator for ever having been diagnosed with a mental health condition based on the immutable individual-level characteristics and functions thereof described above.

In order to test the quality of the prediction, we plot our measure of predicted susceptibility to mental illness against our index of poor mental health. Appendix Figure A.7 shows a strong relationship between the index of poor mental health and our predicted measure of susceptibility to mental illness.

Armed with our LASSO prediction, we can study how the introduction of Facebook at a college affected students across the mental-illness-susceptibility spectrum, and whether it induced students who are more likely to be on the margin of a depression diagnosis to seek out depression-related services such as psychotherapy. The upper left panel of Figure 3 presents the estimated effects on the index of poor mental health across quintiles of our LASSO-predicted measure of susceptibility to mental illness.[30] As shown in the figure, the effects of the introduction of Facebook on symptoms of poor mental health tend to be stronger for individuals with a higher baseline risk of developing mental illness.[31]

---

[30] Specifically, we estimate the following modification of Equation (1):

$$(3) \qquad Y_{icgt} = \alpha_c + \delta_t + \beta_q \times \text{Facebook}_{gt} \times \text{MHSusceptQ}_i + \zeta \times \text{MHSusceptQ}_i + \mathbf{X}_i \cdot \gamma + \mathbf{X}_c \cdot \psi + \varepsilon_{icgt},$$

where MHSusceptQ_i are the quintiles of $i's$ predicted susceptibility to mental illness. Figure 3 presents the estimates of β_q. Appendix Table A.5 presents these estimates in a table form, together with p-values for comparisons between the first quintile and other quintiles.

[31] We note that, for predicting baseline susceptibility to mental illness, the stock variable of 'having ever been diagnosed' with a mental illness is arguably more relevant than the flow variable of having exhibited a certain symptom in the past year, because the former captures information covering a longer time span. Appendix Figure A.10 examines robustness of our results to an alternative measure of susceptibility to mental illness based on a LASSO regression predicting whether a respondent's index of poor mental health is in the top 10% of the pre-period sample. The results are qualitatively similar. As shown in a corresponding Table A.6, the coefficient for the top quintile remains statistically different from the
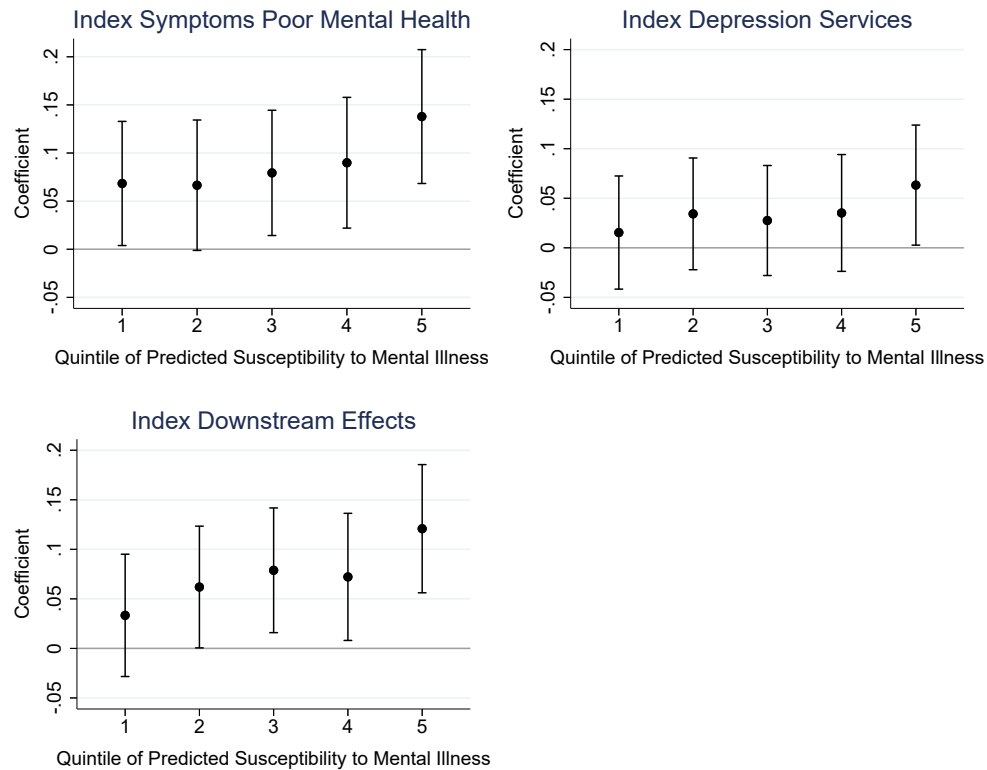
Figure 3: Heterogeneous Effects by Predicted Susceptibility to Mental Illness

Notes: This figure explores the extent to which the effects of the introduction of Facebook at a college are heterogeneous depending on students' predicted susceptibility to mental illness. Specifically, it presents the estimates from equation (3) in which our indicator for post-Facebook introduction is interacted with a set of indicators for belonging to each quintile of a LASSO-predicted measure of susceptibility to mental illness. The outcome variable in the top-left panel is our index of symptoms of poor mental health; the outcome variable in the top-right panel is our index of depression services; the outcome variable in the bottom-left panel is our index measuring whether students reported that conditions related to poor mental health negatively affected their academic performance. All indices are standardized so that, in the pre-period, they have a mean of zero and a standard deviation of one. The estimates (also displayed in Table A.5) are obtained using our preferred specification, namely the one including survey-wave fixed effects, college fixed effects, and controls. Our controls consist of: age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. For a detailed description of the outcome, treatment, interaction, and control variables, see Appendix Table A.31. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

The upper right panel of Figure 3 shows that the introduction of Facebook on the take-up of depression-related services exhibits a similar pattern. We find weak positive effects throughout the distribution of predicted susceptibility to mental illness, though for most quintiles the point estimates are fairly small and not statistically significant. The effects become more pronounced for individuals in the top quintile; in particular, the point estimate on the top quintile is relatively large in magnitude (0.063 standard deviations)

---

coefficient for the bottom quintile at the 10% level for all outcomes.

and four times as large as the point estimate on the bottom quintile. As indicated in column (2) of Table A.5, the difference between the coefficients for the top and the bottom quintiles is significant at the 1% level. These results suggest that, indeed, students who are predicted to be most susceptible to mental illness—and therefore more likely to seek mental healthcare due to a worsening in symptoms—are more likely to take up depression-related services such as psychotherapy for depression and anti-depressants as a result of the introduction of Facebook.

Other dimensions of heterogeneity   Appendix Figure A.6 estimates heterogeneous effects across several baseline characteristics. Consistent with surveys showing that women use social media more often and are more likely to report using Facebook for longer than they intend, we find suggestive evidence that the results are larger among women (Thompson and Lougheed, 2012; Lin et al., 2016).[32] We also find stronger effects on non-Hispanic whites, and a weaker effect on international students, younger students, and first-years.

C.   Effects Based on Length of Exposure to Facebook

The effects of the introduction of Facebook estimated thus far leverage variation that occurs at the college–survey-wave level. Our dataset also features variation at the college–survey-wave–year-in-school level that we can leverage to study the effects of length of exposure to Facebook at the level of individual students. For instance, in the early spring of 2006, a freshman at Harvard would have been exposed to Facebook for one full semester, whereas a senior at Harvard would have been exposed for more than three full semesters.

In order to study the effects of length of exposure to Facebook at the level of individual students, we estimate a version of Equation (1) with individual-level treatment intensity. In this alternative specification, we include a student-level treatment component that equals the number of semesters that the student had access to Facebook given: i) the college the student attends; ii) the survey wave the student participated in; and iii) the year in which the student started college. Specifically, we estimate the following equation:

$$(4) \qquad Y_{icgt} = \alpha_c + \delta_t + \sum_{k=0}^{5} \beta_k \times \text{Semesters}_{k(ict)} + \mathbf{X}_i \cdot \gamma + \varepsilon_{icgt},$$

where $\text{Semesters}_{k(ict)}$ is a set of indicators that equal one if student i at college c in survey-wave t had access to Facebook for k semesters. The number of treated semesters is calculated as $k = \text{FB}_{gt} \times [t - \max\{\tau_i, \tau_c\}]$; t represents time in semesters; τ_c represents the semester in which Facebook was introduced at college c attended by student i; τ_i represents the semester in which student i started studying at college c; and, as before,

---

[32]Furthermore, baseline prevalence of depression is found to be higher among women, across different nations, cultures, and age groups (Nolen-Hoeksema and Hilt, 2008; Albert, 2015; Salk, Hyde and Abramson, 2017). Thus, the slightly stronger effects among women are also consistent with studies showing that women are more likely to be affected by certain mental illnesses.
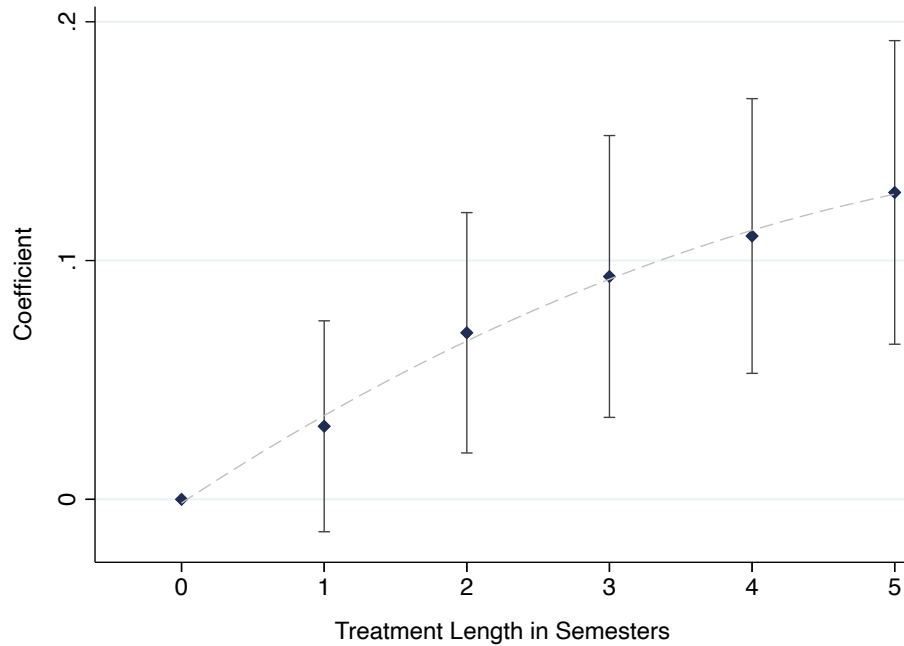
Figure 4: Effect on Poor Mental Health by Length of Exposure to Facebook

Notes: This figure explores the effects of length of exposure to Facebook on our index of poor mental health by presenting estimates of Equation (4). The index is standardized so that, in the pre-period, it has a mean of zero and a standard deviation of one. The dashed curve is the quadratic curve of best fit. Our controls consist of: age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. Students who entered college in 2006 might have been exposed to Facebook already in high-school, because, starting in September 2005, college students with Facebook access could invite high-school students to join the platform. Such students are excluded from the regression. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.31. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

FB_{gt} is the indicator function for whether Facebook was available at student i's college c by time t.[33]

Figure 4 displays the estimates of β_k and shows that the negative effects of the introduction of Facebook on mental health worsen the longer students are exposed to Facebook. Appendix Table A.7 presents the results in a regression framework where we assume that the effects grow linearly over time. The table shows that the number of treated semesters has a significant effect on our main index, on symptoms of poor mental health, and on the utilization of depression-related healthcare services.

Since the index of depression services only comprises binary variables that have a straightforward yes-no interpretation, we provide intuition for the magnitude of our

---

[33] Students who entered college in 2006 might have been exposed to Facebook already in high-school, because, starting in September 2005, college students with Facebook access could invite high-school students to join the platform. We exclude cohorts of students who might have been exposed to Facebook in high school from the length-of-exposure analysis. Including them does not meaningfully affect the results.

results by presenting the effects on each component of the index of poor mental health services in original units. Specifically, Appendix Table A.8 shows that being exposed to Facebook for 5 semesters increases the probability that a student is diagnosed with depression by around 32%, the probability that a student is in therapy for depression by around 50%, and the probability that a student is on anti-depressants by around 33%.

D.   Robustness Checks and Alternative Explanations

Robustness Checks   Appendix A describes a battery of exercises that probe the robustness of our estimates. The exercises include various placebo tests on variables that should not be affected by the introduction of Facebook and modified versions of our main specifications that take into account a host of possible concerns related to: i) the construction of our index of poor mental health, ii) the construction of our treatment variable, iii) particular Facebook expansion groups driving the effects, iv) particular variables driving the effects, v) the parallel trends assumption, vi) the level at which standard errors are clustered. We highlight one of our most convincing robustness check, which consists of a variant of the length-of-exposure specification from Section IV.C that includes college by survey-wave fixed-effects. Such specification, which delivers estimates consistent with the hypothesis that longer exposure to Facebook has a negative impact on student mental health, does not rely on our baseline college-level parallel trends assumption for identification.

Stigma as an Alternative Explanation   One might worry that Facebook affected the stigma associated with mental illness and that our results may not reflect an increase in the prevalence of mental illness per se but rather an increase in willingness to discuss it. To formally investigate the role of stigma, we adopt a three-pronged strategy. First, we collected all the college newspaper articles containing the word Facebook published around the time of Facebook's expansion and checked whether any of them mention stigma in relation to mental health. While we do find articles hinting at potential negative effects of Facebook on mental health, we do not find any articles mentioning stigma. Second, we study whether the fraction of missing answers to the mental health questions in the NCHA survey was affected by the introduction of Facebook. If Facebook made people more comfortable discussing mental illness, we would expect to observe fewer missing answers after the introduction of Facebook.[34] Consistent with the effects being driven by increased prevalence of mental illness rather than by stigma, Appendix Table A.18 shows that the prevalence of missing answers was not affected by the introduction of Facebook. Third, in Section V, we present evidence that the introduction of Facebook did not affect the reporting of other stigmatized conditions, such as being a victim of sexual assault or consuming illegal drugs. Furthermore, we find no detectable effects of the introduction of Facebook on eating disorders, even though such conditions are often highly stigmatized (Puhl and Suh, 2015). If reduction in stigma was indeed the driving force behind our

---

[34]Indeed, missing values are more common in the NCHA survey among sensitive questions (Kays, Gathercoal and Buhrow, 2012).

results, it would be surprising not to find similar effects on other stigmatized behaviors and conditions.

E. Downstream Implications of Poor Mental Health

Does the effect of Facebook on mental health have negative downstream repercussions on academic performance? According to the students' reports, the answer is affirmative.

One of the NCHA survey questions inquires as to whether various conditions affected the students' academic performance. The conditions related to mental health are: attention deficit disorder, depression/anxiety disorder/seasonal affect disorder, eating disorders, stress, and sleep difficulties.[35] The main advantage of analyzing these questions is that they trace a pathway from the introduction of Facebook to perceptions of worsened academic performance via poor mental health. It is important to emphasize, however, that we do not directly measure effects on grades, and that we do not rule out potential positive effects of Facebook on students' academic performance due to channels unrelated to mental health, such as improved teamwork.[36]

Figure 5 presents estimates of Equation (1) and shows how the introduction of Facebook affected each of the measures described in the previous paragraph. All the point estimates are positive and the coefficient for an equally-weighted index summarizing them is positive and significant, suggesting that, after the introduction of Facebook, students were more likely to report that their academic performance was impaired as a result of poor mental health. The effect size on the index is 0.067 standard deviation units. Consistent with our evidence suggesting that depression and anxiety-related disorders are the conditions most severely affected by the introduction of Facebook, we find the largest effect on the depression/anxiety-disorder/seasonal-affect-disorder measure. The number of students who reported that those conditions impaired their academic performance increased by three percentage points over a baseline of 13%. Finally, the bottom-left panel in Figure 3 and column (3) of Table A.5 show that the negative effect of poor mental health on self-reported academic performance is especially pronounced among the students who are predicted to be most susceptible to mental illnesses.

V. Mechanisms

Recent scholarship identified two main channels whereby Facebook might directly affect mental health: unfavorable social comparisons (Appel, Gerlach and Crusius, 2016) and disruptive internet use (Griffiths, Kuss and Demetrovics, 2014). Another, albeit

---

[35] According to the DSM V, sleep difficulties are a symptom of depression (APA, 2013). Similarly, stress has been associated with depression (Yang et al., 2015).

[36] The NCHA dataset does include a question inquiring about the students' cumulative GPA, but the effects of the introduction of Facebook on cumulative GPA are small and noisy. This is likely because the answer options to the GPA question are rather coarse (A, B, C, D/F), because cumulative GPA is a stock variable whose value might largely be determined before the introduction of Facebook at a college, and because students might receive grades based on relative rather than absolute performance. We note that, when analyzing questions on how mental health conditions affected academic performance, it is possible to find an effect even if students are graded on a curve. In particular, students' absolute performance and perception thereof can decrease as a result of the introduction of Facebook.
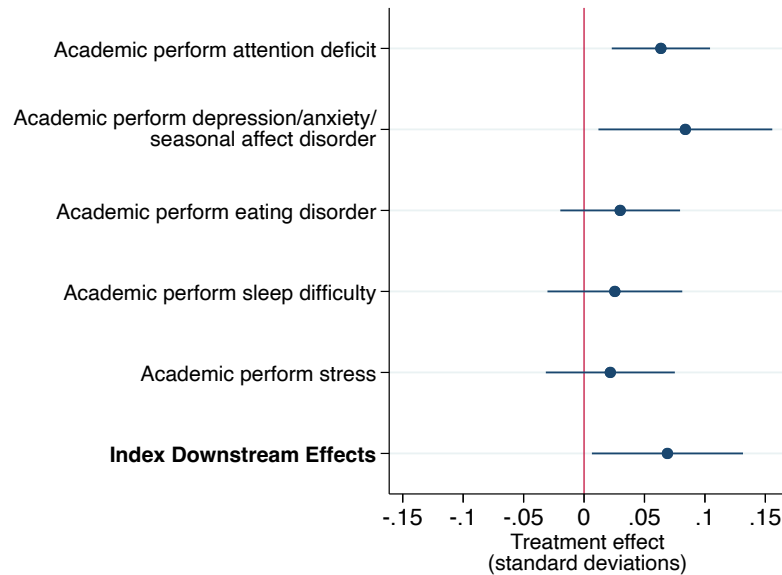
Figure 5: Downstream Effects on Academic Performance

Notes: This figure explores downstream effects of the introduction of Facebook on the students' academic performance. It presents estimates of coefficient β from Equation (1) using our preferred specification, including survey-wave fixed effects, college fixed effects, and controls. The outcome variables are answers to questions inquiring as to whether various mental health conditions affected the students' academic performance and our index of downstream effects. All outcomes are standardized so that, in the pre-period, they have a mean of zero and a standard deviation of one. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.31. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

indirect, possibility is that the introduction of Facebook might lead to behavioral changes that, in turn, affect mental health. We present evidence related to each set of mechanisms in turn. Overall, our evidence is mostly consistent with the unfavorable social comparisons channel.

Unfavorable Social Comparisons   Facebook and other social media platforms make it easier for people to compare themselves to members of their social networks.[37] Such social comparisons, if unfavorable, could be detrimental to users' self-esteem and mental health (Vogel et al., 2014).[38]

Theoretically, the set of individuals who might be negatively affected by social comparisons is unclear. A simple model of social comparisons might posit that individuals compare themselves to the median member of their group along some dimension of interest

---

[37]Indeed, surveys reveal that college students generally used Facebook to learn more about their classmates or about individuals they already knew offline, and used it less often to meet new people (Lampe, Ellison and Steinfield, 2008).

[38]We consider "Fear of Missing Out" (FoMO) as being related to social comparisons, though we recognize that certain features of the phenomenon may not be fully captured by social comparisons. In relation to social media, FoMO refers to the idea that social media platforms might make users more aware of the existence of exciting events that they are missing out on.

(e.g., popularity, wealth, or looks).[39] If social media users are sophisticated, they will be able to extract accurate information from social media platforms about their relative ranking vis-à-vis their peers along the dimension of interest. In that case, we might expect around half of social media users to benefit from social comparisons and about half to suffer from them. Conversely, if social media users are to some extent naive, they will fail to understand that the content that their peers post on social media is likely to be highly curated rather than representative (Appel, Gerlach and Crusius, 2016). In that case, they will systematically underestimate their relative ranking vis-à-vis their peers and, as a result, more than half of them will perceive social comparisons on Facebook as unfavorable.

In this section, we present evidence showing that: i) sub-populations which, in virtue of their baseline characteristics, might be more likely to suffer from social comparisons exhibit larger effects;[40] ii) the introduction of Facebook did not correct the students' misperceptions about their peers' social lives and, in some cases, exacerbated them. The latter piece of evidence is consistent with students exhibiting a degree of naivete in interpreting the information conveyed through social media.

Figure 6 shows that the introduction of Facebook at a college affected more severely the mental health of students who might be more likely to be affected by unfavorable social comparisons. The figure plots estimates of the coefficient on the interaction between our treatment indicator and various moderators in a regression with our index of poor mental health as the outcome variable. Specifically, we consider the following sub-populations of students: i) students who live off-campus and are therefore less likely to participate in on-campus social life; ii) students who have weaker offline social networks as measured by not belonging to a fraternity or sorority organization; iii) students who have lower socioeconomic status as measured by carrying credit card debt or working part-time alongside studying; and iv) students who are overweight. We generate an index of social comparisons based on the above variables and consider, as an additional moderator, an indicator that takes value one if a student is above the median value of the index of social comparisons. All of the point estimates are positive and we find a strong and statistically significant effect on the index, on students living off-campus, and on students with credit card debt. Consistent with the social comparison mechanism, the introduction of Facebook has particularly detrimental effects on the mental health of students who might view themselves as comparing unfavorably to their peers.[41]

To test whether the introduction of Facebook affected the students' beliefs about their peers' social lives, we estimate the impact of the roll-out of Facebook on all survey

---

[39]Individuals could compare themselves to some other percentile of the distribution. The higher the percentile, the larger the set of individuals who would suffer from an increase in the ability to engage in social comparisons.

[40]Such sub-populations are expected to exhibit larger effects independently of whether, in general, social media users are naive or sophisticated.

[41]Of course, we cannot rule out that the sub-populations above exhibit larger effects for reasons other than social comparisons. One concern we *can* rule out is that such sub-populations exhibit larger effects because they have worse baseline mental health. Appendix Figure A.11 shows a version of Figure 6 in which we include as an additional control our treatment indicator interacted with our individual-level LASSO-predicted measure of susceptibility to mental illness. The results are not meaningfully affected.
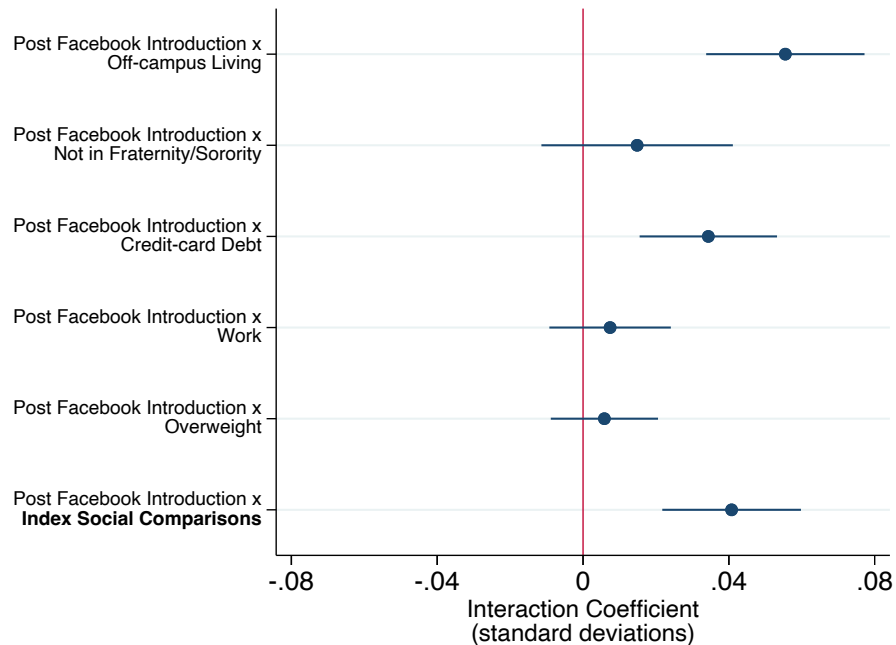
Figure 6: Heterogeneous Effects as Evidence of Unfavorable Social Comparisons

Notes: This figure explores the mechanisms behind the effects of Facebook on mental health. It presents estimates from a version of Equation (1) in which our treatment indicator is interacted with a set of indicators for belonging to a certain sub-population of students. The outcome variable is our overall index of poor mental health. The estimates are obtained using our preferred specification, namely the one including survey-wave fixed effects, college fixed effects, and controls. For a detailed description of the outcome, treatment, interaction, and control variables, see Appendix Table A.31. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

questions that elicit students' perceptions of their peers' drinking behaviors.[42] Specifically, we study the following three sets of beliefs: i) beliefs about the number of alcoholic drinks the typical student has at a party, ii) beliefs about the share of the student population who has had an alcoholic drink in the month before the survey, iii) beliefs about the share of the student population who drinks alcohol on a regular basis. Appendix Table A.19a finds a positive and significant effect on each of the three outcomes above and on an equally-weighted index summarizing the three outcomes. Furthermore, Appendix Table A.20 shows that the effects on perceptions are particularly pronounced for students who live off-campus and who, therefore, have to rely more heavily on social media when estimating their peers' behaviors.[43]

Did Facebook affect beliefs about alcohol consumption because it led students

---

[42]We focus on drinking behavior because alcohol is the most commonly consumed intoxicant among college students and because the NCHA survey includes several questions on drinking-related perceptions.

[43]Appendix Table A.25 provides suggestive evidence that perceptions regarding other students' sexual behavior may have also been affected by the introduction of Facebook. Conversely, Appendix Table A.27 shows that perceptions regarding the usage of illicit substances did not change. Finding effects on the perceptions of alcohol consumption but not on the perceptions of drug consumption is consistent with the fact that drinking and positive references to alcohol were common on Facebook profiles at the time, whereas images of students using drugs were very rare (Watson, Smith and Driver, 2006; Kolek and Saunders, 2008; Morgan, Snelson and Elison-Bowers, 2010)

to actually drink more, or did Facebook affect beliefs without a concurrent increase in drinking behaviors? Appendix Table A.19b shows that the effects on self-reported alcohol usage are substantially smaller than the effects on perceptions, suggesting that the effects on perceptions are unlikely to be driven by a change in actual behavior.[44]

If peers' behaviors did not change, why did Facebook affect perceptions? One option is that baseline perceptions were incorrect and the additional information provided on Facebook corrected such misperceptions. An alternative explanation is that Facebook led students to update their beliefs, but without aligning them more closely to reality. Appendix Table A.21 shows that the introduction of Facebook at a college did not lead students to develop more accurate perceptions about their peers' drinking behaviors and, for one of the outcomes, significantly exacerbated misperceptions. Specifically, the table estimates the effects on the difference between a student's perception of the alcohol consumption of the typical student at her college and the actual typical consumption at the student's college calculated using self-reported alcohol usage in the student's college–survey-wave. The results are consistent with students failing to fully take into account the fact that the content they see on social media is a curated rather than representative portrayal of their peers' lives. Such naivete could lead to distorted beliefs and exacerbate the effects of social comparisons.[45]

Disruptive Internet Use   The second direct channel whereby social media may negatively affect mental health is disruptive internet use (Griffiths, Kuss and Demetrovics, 2014). Specifically, some scholars argue that social media use might disrupt concentration, impair people's ability to focus, and lead to anxiety (e.g., Paul, Baker and Cochran, 2012; Meier, Reinecke and Meltzer, 2016).

We do not find significant evidence supporting the disruptive internet use channel. The main survey question that speaks to disruptive internet use asks students whether the internet or computer games affected their academic performance. Students could answer that the issue affected their academic performance, that they experienced the issue but it did not affect their performance, and that they did not experience the issue. If, after the introduction of Facebook at their college, students found the internet more distracting and had a harder time focusing because of it, we would expect a larger number of students to answer that they experienced the internet or computer games as an issue and that it affected their academic performance. Appendix Table A.22 shows that the share of students experiencing internet or computer games as an issue increased by around 5%, but the effect is not statistically significant.

---

[44]If the introduction of Facebook decreased the stigma related to alcohol consumption, our results about alcohol usage could be biased (see also our discussion of stigma in the context of mental health in Section IV.D). Although we cannot rule out the possibility that changes in stigma due to the introduction of Facebook had an effect specifically on alcohol-related questions, such bias would, if anything, make our results even starker. Specifically, if the introduction of Facebook reduced the stigma around underage drinking, the actual effect on alcohol usage would be smaller than the effect we estimate. Thus, the gap between the changes in usage and changes in perceptions would be even larger than the effect we currently estimate.

[45]Although it is easy to imagine that Facebook users might learn over time how to interpret the content they are exposed to on social media, a recent review of the psychology literature points to social comparisons as a concern that is relevant to this day (Verduyn et al., 2020).

Other Behaviors   The introduction of Facebook at a college might have led students to engage or refrain from engaging in a set of other behaviors that have some bearing on mental health. For instance, the roll-out of Facebook might have popularized illicit drug use.

Appendix Tables A.23, A.24, and A.26 present estimates of the effects of the introduction of Facebook using Equation (1) on various offline behaviors measured in the survey that could plausibly affect mental health. Comfortingly, we do not find any effects on sexual assaults. Similarly, none of the outcomes related to relationships and drug use exhibit significant effects. Combined with the null results on drinking behaviors (Appendix Table A.19b), we do not find much evidence that the introduction of Facebook at a college had meaningful effects on various self-reported behaviors that could have a bearing on mental health.

VI. Discussion

In this section, we elaborate on the extent to which our findings have the potential to inform our understanding of the effects of social media on mental health today.

Over the last two decades, Facebook underwent a host of important changes. Such changes include: i) the introduction of a personalized feed where posts are ranked by an algorithm; ii) the growth of Facebook's user base from U.S. college students to almost three billion active users around the globe (Facebook, 2021); iii) video often replacing images and text; iv) increased usage of Facebook on mobile phones instead of computers; and v) the introduction of Facebook pages for brands, businesses, and organizations. The nature of the variation we are exploiting does not allow us to identify the impact of these features of social media. For instance, our estimates cannot shed light on whether the increased reliance on Facebook for news consumption has exacerbated or mitigated the effects of Facebook on mental health. Similarly, we cannot provide evidence as to whether years of experience with the platform mitigate or exacerbate the effects on mental health.[46]

Despite these caveats, we believe the estimates presented in this paper are still highly relevant today for two main reasons. First, the mechanisms whereby social media use might affect mental health arguably relate to core features of social media platforms that have been present since inception and that remain integral parts of those platforms today. At their core, Facebook and similar platforms are online forums where individuals share information, often about themselves, including pictures, videos, and personal details. Even today, the most common primary reason for using social media is staying in touch with family and friends, in contrast to reading news stories or watching live streams (GWI, 2021). The ease with which one can access information about ones' network, together with the fact that the content posted on social media is generally highly curated, might

---

[46]The effects might be mitigated if, over time, users learn how to interpret the content they are exposed to on Facebook. The effects could be exacerbated if, over time, users become dependent on and potentially even addicted to Facebook (Allcott, Gentzkow and Song, 2021). A change in the social norms around the content that people post on social media might also affect the relationship between Facebook use and mental health.

naturally invite social comparisons. To the extent that the effects of Facebook on mental health at inception were at least partly driven by unfavorable social comparisons, we would expect our findings to still be relevant today.

Second, the mechanisms whereby Facebook use can affect mental health might have been exacerbated rather than mitigated by many of the technological changes undergone by Facebook and related platforms in the last 15 years. Individuals now receive information about their social network directly in their news feeds, and the information is more relevant to them because it is ranked by an algorithm. The content on the platform is richer in that it often includes videos, and it can be accessed at any time or place using a smartphone. These changes might make Facebook even more engaging and might exacerbate the effects on mental health.[47]

VII. Conclusion

In 2021, 4.3 billion individuals had a social media account, accounting for over half the world population and over 90% of internet users (We Are Social, 2021). The repercussions of the rise of social media are thus likely to be far-reaching. In this paper, we leveraged the staggered introduction of Facebook across U.S. colleges to estimate the impact of social media on mental health and found that the introduction of Facebook at a college had a negative effect on student mental health. Our evidence points to unfavorable social comparisons as the leading mechanism.

Overall, our results are consistent with the hypothesis that social media might be partly responsible for the recent deterioration in mental health among teenagers and young adults. It is up to social media platforms, regulators, and future research to determine whether and how these effects can be alleviated.

References

ACHA. 2000–2019. "American College Health Association-National College Health Assessment (ACHA-NCHA) Spring 2000–2019 Reference Group Data Report (Abridged)." *Journal of American College Health*, 54(4): 201.

Acquisti, Alessandro, and Ralph Gross. 2006. "Imagined communities: Awareness, information sharing, and privacy on the Facebook." 36–58, Springer.

Albert, Paul R. 2015. "Why is depression more prevalent in women?" *Journal of Psychiatry & Neuroscience: JPN*, 40(4): 219.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. "The Welfare Effects of Social Media." *American Economic Review*, 110(3): 629–676.

Allcott, Hunt, Matthew Gentzkow, and Lena Song. 2021. "Digital Addiction." *NBER Working Paper*.

---

[47]Of course, some of the changes underwent by social media platforms might push in the opposite direction. For instance, the increased popularity of Facebook might dilute the effects of social comparisons by changing the reference group from one's peers to a broader and more diverse set of individuals.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, 103(484): 1481–1495.

APA. 2013. "Diagnostic and statistical manual of mental disorders: DSM-5."

Appel, Helmut, Alexander L. Gerlach, and Jan Crusius. 2016. "The Interplay Between Facebook Use, Social Comparison, Envy, and Depression." *Current Opinion in Psychology*, 9: 44–49.

Aral, Sinan. 2021. *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health–and How We Must Adapt.* Currency.

Armona, Luis. 2019. "Online Social Network Effects in Labor Markets: Evidence From Facebook's Entry into College Campuses." *Available at SSRN 3381938*.

Arrington, Michael. 2005. "85% of college students use Facebook." https://techcrunch.com/2005/09/07/85-of-college-students-use-facebook.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko. 2020. "Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Controlled Trial." *American Economic Review*, 110(3): 824–59.

Bekalu, Mesfin A, Rachel F McCloud, and K Viswanath. 2019. "Association of social media use with social well-being, positive mental health, and self-rated health: disentangling routine use from emotional connection to use." *Health Education & Behavior*, 46(2_suppl): 69S–80S.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli. 2006. "Adaptive linear step-up procedures that control the false discovery rate." *Biometrika*, 93(3): 491–507.

Berryman, Chloe, Christopher J Ferguson, and Charles Negy. 2018. "Social media use and mental health among young adults." *Psychiatric quarterly*, 89(2): 307–314.

Biasi, Barbara, Michael S Dahl, and Petra Moser. 2019. "Career Effects of Mental Health." *Available at SSRN 2544251*.

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. 2017. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review*, 107(4): 1165–1206.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. "Revisiting Event Study Designs: Robust and Efficient Estimation." *arXiv preprint arXiv:2108.12419*.

Brügger, Niels. 2015. "A brief history of Facebook as a media text: The development of an empty structure." *First Monday*, 12(5).

Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov, and Maria Petrova. 2019. "Social media and xenophobia: evidence from Russia."

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of Econometrics*, 225(2): 200–230.

Chan, David. 2010. "So why ask me? Are self-report data really that bad?" In *Statistical and methodological myths and urban legends*. 329–356. Routledge.

De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review*, 110(9): 2964–96.

Dienlin, Tobias, Philipp K Masur, and Sabine Trepte. 2017. "Reinforcement or displacement? The reciprocity of FtF, IM, and SNS communication and their effects on loneliness and life satisfaction." *Journal of Computer-Mediated Communication*, 22(2): 71–87.

Donati, Dante, Ruben Durante, Francesco Sobbrio, and Dijana Zejcirovic. 2022. "Lost in the Net? Broadband Internet and Youth Mental Health."

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2020. "Social media and protest participation: Evidence from Russia." *Econometrica*, 88(4): 1479–1514.

Enikolopov, Ruben, Maria Petrova, and Konstantin Sonin. 2018. "Social media and corruption." *American Economic Journal: Applied Economics*, 10(1): 150–74.

Facebook. 2021. "Facebook Reports First Quarter 2021 Results." https://bit.ly/3BBVfE3.

Fergusson, Leopoldo, and Carlos Molina. 2020. "Facebook Causes Protests."

Fujiwara, Thomas, Karsten Müller, and Carlo Schwarz. 2021. "The effect of social media on elections: Evidence from the United States." *NBER Working Paper 28849*.

Golberstein, Ezra, Gilbert Gonzales, and Ellen Meara. 2019. "How do economic downturns affect the mental health of children? Evidence from the National Health Interview Survey." *Health economics*, 28(8): 955–970.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, 225(2): 254–277.

Griffiths, Mark D., Daria J. Kuss, and Zsolt Demetrovics. 2014. "Chapter 6 - Social Networking Addiction: An Overview of Preliminary Findings." In *Behavioral Addictions*. , ed. Kenneth Paul Rosenberg and Laura Curtiss Feder, 119–141. San Diego:Academic Press.

GWI. 2021. "Social GWI's Flagship Report on the Latest Trends in Social Media." https://www.gwi.com/reports/social.

Hass, Nancy. 2006. "In your Facebook.com." *The New York Times*.

Haushofer, Johannes, and Jeremy Shapiro. 2016. "The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya." *The Quarterly Journal of Economics*, 131(4): 1973–2042.

Jacobs, Abigail Z., Samuel F. Way, Johan Ugander, and Aaron Clauset. 2015. "Assembling Thefacebook: Using Heterogeneity to Understand Online Social Network Assembly." *WebSci '15*. New York, NY, USA:Association for Computing Machinery.

James, Spencer L, Degu Abate, Kalkidan Hassen Abate, Solomon M Abay, Cristiana Abbafati, Nooshin Abbasi, Hedayat Abbastabar, Foad Abd-Allah, Jemal Abdela, Ahmed Abdelalim, et al. 2018. "Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017." *The Lancet*, 392(10159): 1789–1858.

Kays, Kristina, Kathleen Gathercoal, and William Buhrow. 2012. "Does survey format influence self-disclosure on sensitive question items?" *Computers in Human Behavior*, 28(1): 251–256.

Kelly, Yvonne, Afshin Zilanawala, Cara Booker, and Amanda Sacker. 2018. "Social media use and adolescent mental health: Findings from the UK Millennium Cohort Study." *EClinicalMedicine*, 6: 59–68.

Kessler, Ronald C, G Paul Amminger, Sergio Aguilar-Gaxiola, Jordi Alonso, Sing Lee, and T Bedirhan Ustun. 2007. "Age of onset of mental disorders: a review of recent literature." *Current opinion in psychiatry*, 20(4): 359.

Kirkpatrick, David. 2011. *The Facebook Effect: The Inside Story of the Company That Is Connecting the World.* Simon & Schuster.

Kolek, Ethan A, and Daniel Saunders. 2008. "Online disclosure: An empirical examination of undergraduate Facebook profiles." *Journal of Student Affairs Research and Practice*, 45(1): 1–25.

Kroenke, Kurt, and Robert L Spitzer. 2002. "The PHQ-9: a new depression diagnostic and severity measure."

Kroenke, Kurt, Robert L Spitzer, and Janet BW Williams. 2001. "The PHQ-9: validity of a brief depression severity measure." *Journal of general internal medicine*, 16(9): 606–613.

Lampe, Cliff, Nicole B. Ellison, and Charles Steinfield. 2008. "Changes in Use and Perception of Facebook." *CSCW '08*, 721–730. New York, NY, USA:Association for Computing Machinery.

Layard, Richard. 2017. "The Economics of Mental Health." *IZA World of Labor*.

Levy, Ro'ee. 2021. "Social media, news consumption, and polarization: Evidence from a field experiment." *American Economic Review*, 111(3): 831–70.

Lin, Liu Yi, Jaime E Sidani, Ariel Shensa, Ana Radovic, Elizabeth Miller, Jason B Colditz, Beth L Hoffman, Leila M Giles, and Brian A Primack. 2016. "Association between social media use and depression among US young adults." *Depression and anxiety*, 33(4): 323–331.

Meier, Adrian, Leonard Reinecke, and Christine E Meltzer. 2016. ""Facebocrastina-tion"? Predictors of using Facebook for procrastination and its effects on students' well-being." *Computers in Human Behavior*, 64: 65–76.

Morgan, Elizabeth M, Chareen Snelson, and Patt Elison-Bowers. 2010. "Image and video disclosure of substance use on social media websites." *Computers in human behavior*, 26(6): 1405–1411.

Mosquera, Roberto, Mofioluwasademi Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie. 2020. "The economic effects of Facebook." *Experimental Economics*, 23(2): 575–602.

Müller, Karsten, and Carlo Schwarz. 2020. "From hashtag to hate crime: Twitter and anti-minority sentiment." *Available at SSRN 3149103*.

NAMI. 2020. "National Alliance on Mental Illness (NAMI): Mental Health By the Numbers." https://bit.ly/3hRddco.

National Center for Health Statistics. 2021. "Health, United States, 2019: Table 007." Available from: https://www.cdc.gov/nchs/hus/contents2019.htm.

Nolen-Hoeksema, Susan, and Lori M Hilt. 2008. "Gender Differences in Depression." *Handbook of Depression*, 386.

NSDUH. 2019. "2002–2019 National Survey on Drug Use and Health Final Analytic File (Codebook)." https://bit.ly/3e1d8Sg.

Patel, Vikram, Alan J Flisher, Sarah Hetrick, and Patrick McGorry. 2007. "Mental health of young people: a global public-health challenge." *The Lancet*, 369(9569): 1302–1313.

Paul, Jomon Aliyas, Hope M Baker, and Justin Daniel Cochran. 2012. "Effect of online social networking on student academic performance." *Computers in Human Behavior*, 28(6): 2117–2127.

Paul, Karsten I, and Klaus Moser. 2009. "Unemployment Impairs Mental Health: Meta-Analyses." *Journal of Vocational Behavior*, 74(3): 264–282.

Persson, Petra, and Maya Rossin-Slater. 2018. "Family ruptures, stress, and the mental health of the next generation." *American Economic Review*, 108(4-5): 1214–52.

Plomin, Robert, Claire M. A. Haworth, and Oliver S. P. Davis. 2009. "Common Disorders are Quantitative Traits." *Nature Review Genetics*, 10: 872–878.

Puhl, Rebecca, and Young Suh. 2015. "Stigma and eating and weight disorders." *Current psychiatry reports*, 17(3): 1–10.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel. 2020. "Poverty, depression, and anxiety: Causal evidence and mechanisms." *Science*, 370(6522).

Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. 2022. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." *arXiv preprint arXiv:2201.01194*.

Salk, Rachel H, Janet S Hyde, and Lyn Y Abramson. 2017. "Gender differences in depression in representative national samples: Meta-analyses of diagnoses and symptoms." *Psychological Bulletin*, 143(8): 783.

Shapiro, Bradley. 2021. "Promoting Wellness or Waste? Evidence from Antidepressant Advertising." *American Economic Journal: Microeconomics.*

Spitzer, Robert L, Kurt Kroenke, Janet BW Williams, and Bernd Löwe. 2006. "A brief measure for assessing generalized anxiety disorder: the GAD-7." *Archives of internal medicine*, 166(10): 1092–1097.

Stutzman, Frederic. 2006. "Student life on the Facebook." https://bit.ly/3keTneb.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, 225(2): 175–199.

Thompson, Sharon H, and Eric Lougheed. 2012. "Frazzled by Facebook? An exploratory study of gender differences in social network communication among undergraduate men and women." *College student journal*, 46(1).

Traud, Amanda L, Peter J Mucha, and Mason A Porter. 2012. "Social structure of Facebook networks." *Physica A: Statistical Mechanics and its Applications*, 391(16): 4165–4180.

Twenge, Jean M. 2017. *iGen: Why Today's Super-Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy–and Completely Unprepared for Adulthood–and What That Means for the Rest of Us.* Simon and Schuster.

Twenge, Jean M, A Bell Cooper, Thomas E Joiner, Mary E Duffy, and Sarah G Binau. 2019. "Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017." *Journal of Abnormal Psychology*, 128(3): 185.

Twenge, Jean M, and W Keith Campbell. 2019. "Media use is linked to lower psychological well-being: evidence from three datasets." *Psychiatric Quarterly*, 90(2): 311–331.

U.S. Department of Education, National Center for Education Statistics. 2005. "Integrated Postsecondary Education Data System (IPEDS)." Fall enrollment: gender, attendance status, and level of student.

Verduyn, Philippe, Nino Gugushvili, Karlijn Massar, Karin Taht, and Ethan Kross. 2020. "Social Comparisons on Social Networking Sites." *Current Opinion in Psychology*, 36: 32–37.

Vogel, Erin A, Jason P Rose, Lindsay R Roberts, and Katheryn Eckles. 2014. "Social comparison, social media, and self-esteem." *Psychology of Popular Media Culture*, 3(4): 206.

Watson, Sandy White, Zachary Smith, and Jennifer Driver. 2006. "Alcohol, Sex and Illegal Activities: An Analysis of Selected Facebook Central Photos in Fifty States."

We Are Social. 2021. "Digital 2021 April Global Statshot Report."

WHO. 2008. *The Global Burden of Disease: 2004 Update.* World Health Organization.

Wilson, Robert E, Samuel D Gosling, and Lindsay T Graham. 2012. "A review of Facebook research in the social sciences." *Perspectives on psychological science*, 7(3): 203–220.

Yang, Longfei, Yinghao Zhao, Yicun Wang, Lei Liu, Xingyi Zhang, Bingjin Li, and Ranji Cui. 2015. "The effects of psychological stress on depression." *Current neuropharmacology*, 13(4): 494–504.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political effects of the internet and social media." *Annual Review of Economics*, 12: 415–438.