Social Media and Mental Health*

Luca Braghieri, Ro'ee Levy, and Alexey Makarin

August 2021

Abstract

The diffusion of social media coincided with a worsening of mental health conditions among adolescents and young adults in the United States, giving rise to speculation that social media might be detrimental to mental health. In this paper, we provide the first 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 increased symptoms of poor mental health, especially depression, and led to increased utilization of mental healthcare services. We also find that, according to the students' reports, the decline in mental health translated into worse academic performance. Additional evidence on mechanisms suggests the results are due to Facebook fostering unfavorable social comparisons.

JEL Codes: D12, D72, D90, I10, L82, L86

*Braghieri: Ludwig Maximilian University of Munich. Email: luca.braghieri@econ.lmu.de. Levy: MIT. Email: roeelevy@mit.edu. Makarin: Einaudi Institute for Economics and Finance (EIEF) and CEPR. Email: alexey.makarin@eief.it. We would like to thank Sarah Eichmeyer for her contributions at the early stages of this project, Mary Hoban for helping us access the NCHA dataset, and Luis Armona for his collaboration in putting together the Facebook expansion dates dataset. We are grateful to Davide Cantoni, Georgy Egorov, Ruben Enikolopov, Amy Finkelstein, Matthew Gentzkow, Luigi Guiso, Jack Mountjoy, Samuel Norris, Petra Persson, Andrea Prat, Maya Rossin-Slater, Frank Schilbach, Sebastian Schweighofer-Kodritsch, Joseph Shapiro, Andrey Simonov, and seminar participants at EIEF, LMU, the Istituto Superiore di Sanità, and the Center for Rationality and Competition for helpful comments. We thank Juan Carlos Cisneros, Valerio Sergio Castaldo, Gleb Kozlyakov, and Meruyert Tatkeyeva for excellent research assistance.

1 Introduction

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 (We Are Social, 2021; GWI, 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).[1] 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). Well-identified causal evidence, however, remains scarce.

In this paper, we provide the first 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 negatively impacted student mental health. We also find that, according to the students' reports, the negative effects on mental health translated into worse 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 the Facebook network, colleges witnessed rapid and widespread Facebook penetration among stu-

[1] Conversely, the mental health trends of older generations remained relatively stable.

dents (Brügger, 2015; Wilson et al., 2012). 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 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 the Facebook 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 outcomes than students at less demanding colleges); second, differences across time that affect all students in a similar way (e.g., certain macroeconomic 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).[2] We also address recent concerns with staggered difference-in-differences research designs by showing robustness to using the estimand and placebo exercises suggested in De Chaisemartin and d'Haultfoeuille (2020). 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 the 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 men-

[2]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.

tal health variables in the NCHA survey, increased by 0.085 standard deviation units as a result of the Facebook roll-out. 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). The mental health conditions driving the results are primarily depression and anxiety-related disorders. We find that the effects are strongest for students who, based on immutable characteristics such as gender and age, are more 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. Finally, we find that, after the introduction of Facebook at their colleges, students reported a worsening of academic performance specifically due to poor mental health. As a placebo check, we show that the introduction of Facebook at a college did not substantially affect the students' physical 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 results are particularly pronounced for students who may 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 who are overweight, students of lower socio-economic 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, especially in relation to alcohol consumption. As far as other channels are concerned, we do not find significant evidence that the negative effects of Facebook on mental health is due to disruptive internet use. We also rule out several alternative mechanisms, such as reduced stigma and direct effects on drug use, alcohol consumption, and sexual behaviors.

Overall, our findings are in line with the hypothesis that social media played a role in the increase in mental illness among adolescents and young adults over the past two decades. Clearly, 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 use along various other dimensions, including externalities on the political domain. 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, 2021), and Mosquera et al. (2020)—feature experiments that incentivize a randomly-selected subset of participants to reduce their social media use.[3] Those studies find small negative effects of social media use on self-reported well-being, and Allcott et al. (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 previous papers. 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 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).[4] 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. 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 effects.[5] 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 two and a half years 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.

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

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

[4]The question asks respondents how often they felt depressed. From a psychometric standpoint, the seven-question scale about depression symptoms featured in the NCHA survey is likely to be more discerning than the single-question scale used in Allcott et al. (2020, 2021), and Mosquera et al. (2020).

[5]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. Since elicitation of subjective well-being relies on self-reports, it is impossible to rule out that, for participants assigned to the treatment group, knowledge of treatment status generates experimenter demand effects. Furthermore, incentive payments might directly affect self-reported well-being and confound interpretation.

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 et al., 2019). 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 et al., 2017; Biasi et al., 2019; Baranov et al., 2020; Shapiro, 2021). We complement this literature by showing that, according to the students' reports, the deterioration in mental health after the introduction of Facebook had negative repercussions on academic performance.[6]

Lastly, this paper contributes to an emerging literature exploiting the expansion of 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 et al. (2020) and Fergusson and Molina (2020) exploit the expansion of 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.[7] 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 2 provides some background on mental health and on Facebook's early expansion; Section 3 describes the data sources used in the analysis and presents descriptive statistics; Section 4 discusses the empirical strategy; Section 5 presents the results; Section 6 explores mechanisms; Section 7 discusses potential implications of the results; Section 8 concludes.

[6] This result also complements papers finding that trauma due to school shootings and police violence has a negative effect on academic performance (Ang, 2021; Cabral et al., 2021).

[7] Additional research on social media and political outcomes includes Enikolopov et al. (2018), Fujiwara et al. (2020), and Levy (2021). For a detailed overview, see Zhuravskaya et al. (2020).

2 Background

2.1 Mental Health

As defined by the World Health Organization (WHO), mental health is "a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community" (WHO, 2018). As such, mental health is considered an integral part of one's overall health status.

Mental illnesses, such as depression, anxiety, bipolar disorder, and schizophrenia, can be extremely debilitating and seriously hamper a person'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 years old, and depression is one of the most taxing conditions (WHO, 2008; Layard, 2017).

Recent estimates show that mental illnesses are also disturbingly common, both in the United States and globally. 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).

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 Figure 1, self-reported episodes of psychological distress and depression have grown 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 such cohorts.[8] Because 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).

[8]Suicide is now the second leading cause of death for individuals 15–24 years old—up from the third most common cause in 1980, overtaking homicides (National Center for Health Statistics, 2021).

2.2 A Brief History of Facebook's Expansion and Initial Popularity

Facebook—originally thefacebook.com—is a social networking platform created by Mark Zuckerberg. The site was launched on February 4$^{\text{th}}$, 2004 and was initially only open to members of the Harvard community. In a sign of things to come, Facebook caught on immediately at Harvard. Within 24 hours, more than 1,000 students had registered, and, by the end of the month, three-quarters of Harvard undergraduates had signed up (Cassidy, 2006).

Following the overwhelming success at Harvard, Facebook gradually expanded to other colleges. In June 2004, Facebook was available at 40 selective U.S. colleges and had 150,000 users. In February of the following year, Facebook was available at 370 colleges and had 2 million active users (Kirkpatrick, 2011, p.111). By September 2005, Facebook supported almost 900 colleges and had 3.85 million users (Cassidy, 2006; Arrington, 2005). Over the next year, Facebook expanded to additional universities, high schools, and selected workplaces until, in September 2006, it opened its membership to anyone in the world above the age of 13. Throughout the expansion period, access to Facebook was restricted by requiring users to be in possession of verified email addresses.[9]

The Facebook roll-out 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 two factors (Kirkpatrick, 2011): first, scale constraints due to limited server capacity; second, Facebook's willingness, at least at the outset, to maintain a flavor of exclusivity.

Even in its infancy, Facebook was extremely popular and usage was intense. Upon being granted access to Facebook, colleges witnessed rapid and very widespread adoption among students.[10] As of September 2005, one-and-a-half years after Facebook first went online, out of all the students at colleges in which Facebook was available, 85% had a Facebook profile (Arrington, 2005).[11] In early 2006, close to three-quarters of users logged into the site at least

[9] For instance, in early February 2004, only individuals in possession of email addresses ending in harvard.edu were granted access to the platform.

[10] According to a description of Facebook's early expansion 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).

[11] 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 (Lampe et al., 2008; Sheldon, 2008; Stutzman, 2006; Kolek and Saunders, 2008).

once a day, and the average user logged in six times a day (Hass, 2006). Finally, 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).

The rapid and widespread adoption of Facebook has important implications for interpreting our results. First, due to network externalities, the effects of social media could in principle be quite different depending on whether adoption is partial or full. The large adoption rates make our setting more similar to today's social media environment, in which most young people have a social media account. Second, dynamic effects, if any, are likely to be driven by differential length of exposure to Facebook rather than to increased take-up rates over time.

3 Data Sources & Descriptive Statistics

3.1 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 the Facebook expansion.

Facebook Expansion Dates Data The Facebook Expansion Dates dataset was assembled in two steps: for the first 100 colleges in the Facebook roll-out schedule, we rely on Facebook introduction dates collected and made public in previous studies (Traud et al., 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 allows users to visit old versions of those websites. Specifically, between the spring of 2004 and spring of 2005, the front page of the Facebook website was regularly updated to show the most recent set of colleges that had been given access to the platform.[12] As an example, Appendix Figure A.1 shows the Facebook front page as of June

[12] 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.

15$^{\text{th}}$ 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 the Facebook 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 the Facebook front page as the introduction date.

Since the Wayback Machine took snapshots of the Facebook 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.[13] Therefore, on average, our imputed introduction dates should be within two days of the actual introduction dates.

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 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. Self-reported symptoms, although relatively uncommon as outcome variables in economics, belong to 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

[13] Whenever the Wayback Machine took multiple snapshots of the Facebook website in a single day, we consider only the first snapshot when constructing our measure of the average number of days between consecutive snapshots.

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 (American Psychiatric Association, 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).[14]

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.[15] Only colleges that administered the survey to randomly selected students, to students in randomly selected classrooms, or to all students are included in the NCHA dataset (ACHA, 2005). The average response rate across the survey waves for which we have such information is 37% (ACHA, 2005, 2006a,b, 2007, 2008, 2009).[16] Throughout our analysis, we limit our sample to full-time undergraduate students.[17]

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, some of our specifications include college fixed effects.

In order to protect the privacy of the institutions that participate in the NCHA survey while still allowing us to carry out our analysis, the ACHA kindly agreed to provide us with a customized dataset that includes a variable indicating the semester in which Facebook was rolled out at each college. The ACHA produced the customized dataset according to the following procedure: i) merge our dataset containing the Facebook introduction dates to their dataset; ii) add a variable listing the semester in which Facebook was rolled out at each college in their dataset;[18] iii) strip away any information that could allow us to identify colleges in the dataset

[14] Sections 4.1 and Appendix A discuss our symptoms measures in detail and present an array of exercises to validate them.

[15] 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.

[16] Appendix Table A.3 shows that, along the demographic dimensions observable in our dataset, the composition of students responding to the survey did not change as a result of the introduction of Facebook at a college.

[17] 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).

[18] For the set of colleges that appear both in our introduction-date dataset and the ACHA dataset, the ACHA listed the semesters corresponding to the introduction dates in our dataset. For the set of colleges that appear only in the ACHA dataset, we list the Fall of 2005 as the semester in which Facebook was introduced at those colleges.

(including the specific date in which Facebook was introduced at each college).

3.2 Descriptive Statistics

Appendix Tables A.1 and A.2 present college-level and student-level descriptive statistics for colleges in different Facebook expansion groups.[19] Appendix Table A.1 shows that colleges in earlier Facebook expansion groups are more selective, larger, more residential, and more likely to be on the East Coast 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 relatively more students from 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 in later Facebook expansion groups.[20] 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 4.3.

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.

Finally, Appendix Table A.3 shows a balance table on the immutable student-level demographic characteristics that we observe in the NCHA survey. As shown in the table, the average composition of the students in our sample along each characteristic is similar in the pre- and post-Facebook introduction periods.

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 Section 5.4, the results are robust to dropping these colleges altogether.

[19] Appendix Table A.1 was constructed by merging our Facebook Expansion Dates dataset to data from the Integrated Postsecondary Education Data System (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.

[20] 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 Section 5.4, 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 and estimate the effects using the remaining expansion groups.

4 Empirical Strategy

In this section, we describe the construction of our primary outcome variables, the construction of our treatment indicator, and our identification strategy.

4.1 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 (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 heath 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.

We internally validate the questions about symptoms of mental illness by relating them to self-reported mental healthcare diagnoses within our dataset.[21] Appendix A 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.

[21] Such questions have also been externally validated by other researchers by benchmarking them against the results of several major nationally-representative surveys (ACHA, 2005).

The index of depression services requires a slightly more detailed discussion in virtue of the way in which the ACHA formulated the questions it comprises. 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 those variables; fourth, we standardize the final index. This way, our indices are essentially z-scores.

Appendix Table A.19 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.

4.2 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 a robustness check described in detail in Section 5.4, 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.

4.3 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 mental health 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:

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

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 at a college 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.[22] 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 macro-economic fluctuations might affect all students' job prospects in a similar way, and, in turn, 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 in a set of placebo exercises suggested by De Chaisemartin and d'Haultfoeuille (2020). 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 such parallel trends assumption to deliver consistent estimates. In particular, we present results using a specification that includes college$\times$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

[22]College-level fixed effects also rule out concerns about the changing composition of the panel.

homogeneity in treatment effects (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2020; Sun and Abraham, 2020; Goodman-Bacon, 2021; Borusyak et al., 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. Depending on the severity and type of heterogeneity, it might even flip the sign of the estimate compared to the true effect.

We address concerns about the reliability of TWFE estimators by replicating our results using the modified difference-in-differences estimator suggested in De Chaisemartin and d'Haultfoeuille (2020), which we refer to henceforth as DCDH estimator. By shutting down the 2×2 difference-in-differences comparisons between newly-treated and already-treated units, the DCDH estimator consistently recovers the average of the treatment effects occurring at the time when a group starts receiving treatment. As discussed in Section 5.4, the point estimates obtained with the DCDH estimator turn out to be virtually identical to the ones produced with the TWFE model.

5 Results

This results section is organized as follows. First, we present and interpret our baseline estimates of the causal effect of social media on mental health. Second, we explore heterogeneity, especially vis-à-vis predicted susceptibility to mental illness. Third, we propose an alternative specification to study the effects of differential length of exposure to Facebook. Fourth, we probe the robustness of our baseline results and rule out alternative explanations. Fifth, we study whether the negative impact of Facebook on mental health had downstream effects on the students' academic performance.

5.1 Baseline Results

In order to test for pre-trends and to explore whether there is a sharp discontinuity on the first semester in which Facebook was introduced at a college, we estimate an event-study version of the TWFE model with indicators for distance to/from the Facebook introduction. Specifically, rather than grouping cohorts before and after the introduction of Facebook at a college in two coarse categories, we allow students to be affected differentially depending on the distance between the semester in which they took the survey and the semester of Facebook introduction at their college.[23] 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.

Figure 2 presents estimates of the coefficients on the indicator variables indexing distance to/from Facebook's introduction; the outcome variable is our overall index of poor mental health. Consistent with the parallel trends assumption, the coefficients on the semesters prior to the introduction of Facebook at a college are all close to zero and exhibit no pre-trends. Furthermore, we observe a sharp discontinuity arising in the first semester after the introduction of Facebook. The presence of a discontinuity is in line with evidence that the take-up of Facebook at a college was rapid and widespread.[24] Appendix Figure A.2 shows that similar patterns arise when splitting symptoms of mental illness and take-up of depression-related healthcare services into separate indices.

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

[23] We estimate the following specification:

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

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.

[24] One might be surprised that the effects materialize already in the first semester after the introduction of Facebook at a college ($t = 1$). We note that, given the dates in which Facebook was introduced at the various colleges in our dataset, the minimum amount of time between the Facebook introduction date at a college and the first day of the subsequent semester is approximately two months. Therefore, the coefficient on $t = 1$ in Figure A.2 captures effects arising at least two months after the introduction of Facebook at a college.

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 effect size on the index of poor mental health in our preferred specification, namely the one that includes college rather than expansion-group fixed effects, is 0.085 standard deviation units. Such magnitude corresponds to approximately 84% of the baseline difference in the index of poor mental health between students in our sample with and without credit card debt. In order to provide additional intuition, we benchmark the magnitude of our estimates against the effect on mental health of a sudden unemployment spell. Comparing our estimates to the most closely related ones in a meta-analysis by Paul and Moser (2009), we find that the effects of introducing Facebook at a college on symptoms of poor mental health is around 22% of the effect of job loss.[25]

Figure 3 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.[26] For all but one of the mental health outcomes from Figure 3, the point estimates are positive, which indicates worsened mental health. The conditions that appear to be most affected are depression and anxiety-related disorders.[27]

The bottom section of Figure 3 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—

[25] 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.

[26] 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 et al. (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.

[27] The reason why some of the times the point estimate on an index is larger than the point estimates on each of the components of the index is that averaging across the components reduces noise. As a consequence, the effects, which are always measured in standard deviation units, are often larger for less noisy variables.

the point estimates are positive, though not significant at conventional levels.[28] 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.[29] In the next section, 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.

5.2 Heterogeneity

In order to study whether the introduction of Facebook at a college led marginal students to take up depression-related services, we proceed in two steps: first, we implement 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 and only 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 the 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 using 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.5 shows a strong relationship, both in and out of sample, between the index of poor mental health and our predicted measure of susceptibility to mental illness.

[28] 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.

[29] The argument above relies on the baseline propensity to experience mental illness likely being normally distributed in the population (Plomin et al., 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.

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 marginal to seek out depression-related services such as psychotherapy. The left panel of Figure 4 presents estimates of β in Equation (1) on the index of symptoms of poor mental health across quintiles of our LASSO-predicted measure of susceptibility to mental illness. 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 higher baseline risk of developing mental illness.

The effects of the introduction of Facebook on the take-up of depression-related services exhibit a similar pattern. The right panel of Figure 4 presents estimates of β in Equation (1) on the index of depression-related services across quintiles of our LASSO-predicted measure of susceptibility to mental illness. We find weak positive effects of the introduction of Facebook on the take-up of depression-related services throughout the distribution of predicted susceptibility to mental illness, though for most quintiles the point estimates are fairly small and not 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 (0.06 standard deviations) and more than twice as large as the point estimate on the bottom quintile. The additional heterogeneity estimate for the top quintile, with the first quintile as an omitted category, is significant at the 5% level.[30] The results suggest that, indeed, students who are predicted to be most susceptible to mental illness—and therefore more likely to be seeking mental healthcare due to a worsening in symptoms—are more likely to take up depression-related services as a result of the introduction of Facebook.

Other dimensions of heterogeneity. Appendix Figure A.4 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).[31] We also find stronger effects on non-Hispanic whites, and a weaker

[30] Looking at Figure 4, it may not seem obvious that the heterogeneity estimate on the top quintile is statistically significant. It is important to notice, however, that the point estimates shown in Figure 4 are the sum of the baseline coefficient and the heterogeneity estimate, not just the heterogeneity estimates.

[31] Furthermore, baseline prevalence of depression is consistently found to be higher among females than among males, across different nations, cultures, and age groups (Nolen-Hoeksema and Hilt, 2008; Salk et al., 2017).

effect on international students, younger students and first-years.

The smaller effects on first years and younger students might be driven by at least two channels. First, consistent with the documented age gradient in the onset of mental illness (Kessler et al., 2007), they might simply reflect heterogeneous effects across age groups. Second, they could be due to differential length of exposure to Facebook across students in different cohorts. To shut down the length-of-exposure channel, we restrict our sample to only include students who took the survey at most one semester after the introduction of Facebook at their college. This way, all students in the restricted sample, regardless of their cohorts, are exposed to Facebook for at most one semester. The last row of Appendix Figure A.4 shows heterogeneous effects on first year students using the restricted sample. The fact that the point estimate from the restricted sample is less negative than that from the full sample suggests both length of exposure and heterogeneity along the age/year-in-school dimension play a role in driving the effects on the full sample. We investigate the effects of length of exposure to Facebook more formally in the next section.

5.3 Effects Based on Length of Exposure to Facebook

Although Figure 2 shows that, at the level of an entire college, the effects of Facebook's introduction on mental health remain fairly stable over time, the effects could be increasing over time at the level of individual students. Such seeming discrepancy might arise because, at the college level, dynamic effects are partly muted by the arrival of new students who are only exposed to Facebook upon entering college. For instance, in the Spring of 2006, a freshman at Harvard would have been exposed to Facebook for only one semester, whereas a senior at Harvard would have been exposed for more than three 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 goes to; ii) the survey wave the student participated in; and iii) the year in which the student started college. Specifically, we estimate the following equation:

$$Y_{icgt} = \alpha_c + \delta_t + \beta \times \mathrm{FB}_{gt} \times [t - \max\{\tau_i, \tau_c\}] + \mathbf{X}_i \cdot \gamma + \varepsilon_{icgt}, \qquad (3)$$

where 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, FB_{gt} is the indicator function for whether Facebook was available at student i's college c by time t.[32] We begin by assuming the effects grow linearly over time and show results using a specification that does not impose such parametric assumption in Appendix Figure A.3.

Table 2 shows that the effects of the introduction of Facebook on our overall index of poor mental health and on our sub-indices grow over the span of time covered by our survey. Specifically, for the average student, being exposed to Facebook for 5 semesters—the maximum length of exposure we observe in our data—leads to a worsening of the index of poor mental health of approximately 0.12 standard deviation units.[33]

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 results by presenting the effects on each component of the index of poor mental health services in original units. Specifically, Appendix Table A.5 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%.

5.4 Robustness Checks and Alternative Explanations

Robustness Checks First, as a placebo test, Table A.6 presents a set of specification checks on our LASSO-predicted measure of susceptibility to mental illness. Since the prediction is based on students' immutable characteristics, it cannot be affected by the introduction of Facebook at a college. In fact, if we did find an effect on this measure, we would worry about differential selection before and after the introduction of Facebook along dimensions that are predictive of mental illness. Comfortingly, the point estimates in Table A.6 are small and not

[32] Cohorts of students who might have been exposed to Facebook in high school are excluded from the regression. Including them does not meaningfully affect the results.

[33] We note that the effects in the length-of-exposure specification could partly be driven by a higher probability of having a Facebook account and/or to higher intensity of usage over time. Given the evidence presented in Section 2.2 on the rapid and widespread penetration of Facebook at each college and evidence that intensity of usage did not increase substantially over time (Lampe et al., 2008; Stutzman, 2006), we find the length-of-exposure channel more plausible.

significant.

As an additional placebo test, Table A.7 presents a set of specification checks on an index of all physical rather than mental health outcomes in our dataset (e.g., asthma, diabetes, hepatitis).[34] Consistent with intuition, the effects of the introduction of Facebook on physical health are significantly smaller than the effects on mental health across all specifications and, in our preferred specification with college rather than Facebook-expansion-group fixed effects, also statistically indistinguishable from zero. As an additional check, Figure A.6 displays the cumulative distribution of coefficients on the individual components of our indices of poor mental and poor physical health. As shown in the figure, the distribution of coefficients on the components of the index of poor mental health first-order stochastically dominates the distribution of coefficients on the components of the index of poor physical health. A Mann-Whitney U test rejects the hypothesis of equality of the two distributions at the 1% significance level.

Next, we show that the results on our index of poor mental health are not driven by any one outcome variable, any particular Facebook expansion group, or by how we define treatment status when the semester in which a student took the survey coincides with the semester in which Facebook was rolled out at her college. To address the first concern, we construct various versions of the index of poor mental health, each time excluding a different component from the index. Appendix Figure A.7 shows that our estimates are robust to separately dropping each individual component of the index of poor mental health. To address the second concern, we run our TWFE and length-of-exposure models on a restricted dataset that excludes from the analysis colleges belonging to each Facebook expansion group in turn. Appendix Table A.8 shows that the results remain fairly stable across the various restricted datasets.[35] Lastly, to address the third concern, Appendix Table A.9 shows that our results are qualitatively similar independently of whether we consider respondents who took the survey in the semester in which Facebook was rolled out at their colleges treated, untreated, or whether we assign them a treatment status of 0.5.[36]

[34] We recognize that the introduction of Facebook at a college could in principle affect physical health: it could do so directly (e.g., back pain from sitting in front of a computer) or indirectly as a result of compromised mental health. Nevertheless, we would have found it surprising had physical health been severely affected by the introduction of Facebook. The fact that it is not is in line with our prior.

[35] In fact, both in panel (a) and in panel (b), we fail to reject the hypothesis of equality of coefficients across the various restricted datasets at conventional significance levels.

[36] We note that imputing a treatment status for such participants, especially a treatment status of 0 or 1, might introduce substantial measurement error and weaken the magnitudes of the effects.

As another robustness check, we estimate a specification in which we interact the survey-wave fixed effects with college- or Facebook-expansion-group-level characteristics that are correlated with Facebook roll-out timing (baseline mental health, geographic region, and selectivity).[37] Appendix Table A.10 shows that our results are not meaningfully affected by the inclusion of these additional controls.

Our most powerful robustness check shows that we obtain qualitatively similar results using a specification that does not rely on the parallel trends assumption required by our baseline difference-in-differences model. In particular, for our baseline model to identify causal effects, we had to impose the assumption that, absent the introduction of Facebook, the mental health outcomes of students attending colleges in different Facebook expansion groups would have evolved along parallel trends. A version of the length-of-exposure specification—Equation (3)—that includes college×survey-wave fixed effects does not rely on such parallel trends assumption for identification.[38] Instead, in this specification, identification comes from comparing students within the same college–survey-wave, but who were exposed to Facebook for different lengths of time based on the year in which they entered college. The results are included in Table 2 and show that, even after the inclusion of college×survey-wave fixed effects, students exposed to Facebook for longer periods of time report being in worse mental health.

Finally, we show that our estimates are virtually unaffected when replacing the TWFE estimator from Equation (1) with the estimator suggested in De Chaisemartin and d'Haultfoeuille (2020). The DCDH estimator, which shuts down the 2×2 difference-in-differences comparisons between newly-treated and already-treated units, is designed to be consistent even in the presence of heterogeneous treatment effects across units and across time. Furthermore, the DCDH estimator lends itself to a set of intuitive placebo tests. Table A.11 shows that the DCDH estimate is virtually identical to our baseline TWFE estimates, and that all placebo estimates are statistically indistinguishable from zero.

Alternative Explanations One might worry that Facebook affected the stigma associated with mental illness and that our results might not reflect an increase in the prevalence of mental

[37]See Appendix Tables A.1 and A.2 for evidence that those characteristics are correlated with the timing of the Facebook roll-out.

[38]The college×survey-wave fixed effects would absorb all the college-level differences that would arise if, absent the Facebook introduction, colleges in different Facebook expansion groups were not on parallel mental health trends.

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 articles containing the word Facebook that appeared in college newspapers around the time of Facebook's expansion and checked whether any of them mentions 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 discussing mental health 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 Facebook introduction.[39] Consistent with the effects being driven by increased prevalence of mental illness rather than by stigma, columns (1)–(3) of Appendix Table A.12 show that the prevalence of missing answers was not significantly affected by the introduction of Facebook. Third, in Section 6, 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. If reduction in stigma was indeed the driving force behind our mental health results, it would be surprising not to find similar results on other stigmatized behaviors.

One could also worry that the introduction of Facebook affected the way individuals respond to mental health survey questions. For instance, the introduction of Facebook might have made mental health more salient, which in turn might have induced individuals to more easily remember instances in which they felt depressed. Binary outcomes, such as whether someone felt depressed at least once or whether somebody is on anti-depressants, are less likely to be affected by this concern. Column 4 of Appendix Table A.12 shows that our conclusions are qualitatively unaffected if we redefine all continuous variables as binary variables and only consider extensive-margin responses.

5.5 Downstream Implications of Poor Mental Health

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

One of the questions in the NCHA survey inquires as to whether various conditions af-

[39] Indeed, missing values are more common in the NCHA survey among more sensitive questions (Kays et al., 2012).

fected the students' academic performance.[40] The conditions related to mental health are: attention deficit disorder, depression/anxiety disorder/seasonal affect disorder, eating disorders, stress, and sleep difficulties.[41]

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 an equally-weighted index summarizing them is positive and significant. The effect size on the index is 0.067 standard deviation units. The largest effect is on the depression/anxiety-disorder measure. The number of students who reported that depression/anxiety-disorder affected their academic performance increased by three percentage points over a baseline of 13%. Overall, according to the students' reports, the negative effects on mental health caused by the introduction of Facebook at a college had detrimental downstream effects on academic performance.[42]

6 Mechanisms

Many of the channels through which social media can affect mental health were already noticeable in the very early days of Facebook, as evidenced by the concluding paragraph of a column published in Harvard's daily newspaper only 13 days after Facebook's launch in February 2004:

> "The thefacebook.com scene includes reams of carefully coiffed, immaculately manicured, evening-garbed Harvard students grinning eagerly on page after page as we present our own ideal image of selfhood to fellow browsers... there are plenty of other primal instincts evident at work [on thefacebook.com]: an element of wanting to belong, a dash of vanity and more than a little voyuerism probably go a long way in explaining most addictions (mine included). But most of all it's about performing—striking a pose, as Madonna might put it, and letting the world

[40] Of course, Facebook might also affect the students' academic performance due to channels other than mental health; however, the set of questions we are leveraging in this part of the analysis ask specifically about the extent to which issues related to mental health affected the students' academic performance.

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

[42] The NCHA dataset also includes a question inquiring about the students' cumulative GPA. The effects of the introduction of Facebook on cumulative GPA are small and noisy, likely because the answer options to the GPA question are rather coarse (A, B, C, D/F) and because cumulative GPA is a stock variable whose value might largely be determined before the introduction of Facebook at a college.

know why we're important individuals." (Lester, 2004)

Consistent with the insights from the column, recent scholarship identified two main channels whereby Facebook might directly affect mental health: unfavorable social comparisons (Appel et al., 2016) and disruptive internet use (Griffiths et al., 2014). Another, albeit 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.[43] To the extent that these social comparisons are unfavorable, they could be detrimental to users' self-esteem and mental health (Vogel et al., 2014).[44]

Theoretically, the set of individuals who might be negatively affected by social comparisons is unclear. A simple model of social comparisons would posit that individuals compare themselves to the median member of their group along some dimension of interest (e.g., popularity, wealth, or looks).[45] 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 would expect around half of social media users to benefit from social comparisons and about half to suffer from it. 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 et al., 2016; Chou and Edge, 2012). 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 suffer from social comparisons.

In this section, we present evidence showing: i) that sub-populations which, in virtue of their baseline characteristics, might be more likely to suffer from social comparisons exhibit

[43] 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 et al., 2008).

[44] We consider so-called "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.

[45] 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.

larger effects;[46] ii) the introduction of Facebook did not correct the students' misperceptions of 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 socio-economic 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. Therefore, consistent with the social comparison mechanism, the introduction of Facebook at a college has particularly detrimental effects on the mental health of students who might view themselves as comparing unfavorably to their peers.[47]

To test whether the introduction of Facebook affected the students' beliefs about their peers' social lives, we estimate the impact of the Facebook roll-out on all survey questions that elicit students' perceptions of their peers' drinking behaviors.[48] Specifically, we study the following three sets of beliefs: i) beliefs about the number of alcoholic drinks the typical student

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

[47] 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 heath. Appendix Figure A.8 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 affected.

[48] We focus on drinking behavior because alcohol is the most commonly consumed intoxicant among college students and because of the existence of studies showing that pictures of students drinking and positive references to alcohol were common on Facebook profiles at the time (Watson et al., 2006; Kolek and Saunders, 2008).

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. Table 3a finds a positive and significant effect on each of the three outcomes above and on an equally-weighted index summarizing the three outcomes. Furthermore, Table A.13 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.

Did Facebook affect beliefs about alcohol-consumption because it led students to actually drink more often, or did Facebook affect beliefs without a concurrent increase in drinking behaviors? Table 3b shows that the effects on actual usage are substantially smaller than the effect on perceptions, suggesting the effect on perceptions is not driven by a change in actual behavior.

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 bringing them more in line with reality. Table A.14 estimates the effect on the differences between actual alcohol usage and perceptions and shows that the introduction of Facebook at a college did not lead students to develop more accurate perceptions. For one of the outcomes, it even moved their beliefs further away from the truth. 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.[49]

Disruptive Internet Use The second direct channel whereby social media may negatively affect mental health is disruptive internet use (Griffiths et al., 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 et al., 2012; Meier et al., 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/video-

[49] 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).

games affected their academic performance.[50] 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/video-games as an issue and that it affected their academic performance. Appendix Table A.15 shows that the share of students experiencing internet/video-games as an issue increased by around 5%, though the effect is not statistically significant.

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 Facebook roll-out might have popularized illicit drug use.[51]

Appendix Tables A.16–A.18 present estimates of the effects of the Facebook roll-out 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, we find no strong effects on most outcomes related to relationships or drug use. Combined with the null results on drinking behaviors, we do not find much evidence that the introduction of Facebook at a college had a strong effect on various self-reported behaviors that could have a bearing on mental health.

7 Discussion

Implications for social media today Our estimates of the effects of social media on mental health rely on quasi-experimental variation in Facebook access among college students around the years 2004 to 2006. Such population and time window are directly relevant to the discussion about the severe worsening of mental health conditions among adolescents and young adults over the last two decades. 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 algo-

[50]Unfortunately, the question does not ask about the internet separately from video games.

[51]Indeed, some of the early Facebook groups made not-so-veiled references to drug use (Hirschland, 2006).

rithm; 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 in this paper does not allow us to identify the impact of these features of social media. For example, the introduction of pages, along with other changes, made news consumption on Facebook more common over the last decade than it was at inception. 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.

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; second, the technological changes undergone by Facebook and related platforms might have amplified rather than mitigated the effect of those mechanisms.

At their core, Facebook and similar platforms are online forums where individuals share information in a semi-public fashion.[52] Much of the information is about the individuals themselves: it includes 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 naturally invite social comparisons. To the extent that the effects of Facebook use 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 be-

[52]The sharing of most information on Facebook and related social media platforms like Instagram is semi-public in the sense of being directed to one's friends and followers rather than to a single individual.

cause 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.[53]

Estimates of Productivity Loss In order to translate our effects into a dollar-denominated measure, we perform a back-of-the-envelope calculation using our results on depression diagnoses. The point estimate of the effect of Facebook's introduction on the share of students who were diagnosed with depression in the last school year is 0.5 percentage points over a baseline of 4.7%. Goetzel et al. (2004) estimate the productivity loss from depression, including medical and absenteeism costs, at $348 per employee-year.[54] Assuming the effects of Facebook on U.S. employees are similar to the ones estimated in our study and ignoring length of exposure, we estimate that the productivity loss due to the effect of Facebook on depression amounts to over 202 million dollars per year.[55] We note that a full assessment of the mental health costs exacted by the introduction of social media would likely be much larger than the estimate above, because it would also include a dollar-denominated measure of the loss in well-being caused by worsened mental health conditions.

8 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 leverage the staggered introduction of Facebook across U.S. colleges to estimate the impact of social media on mental health. We find that the introduction of Facebook at a college led to a worsening of mental health symptoms and to an increase in take-up of depression-related services. The effects are particularly pronounced among individuals who,

[53]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.

[54]A more recent study from Denmark finds that depression results in a 34% earnings penalty (Biasi et al., 2019).

[55]We assume that among approximately 152 million employees, 76% use Facebook. The number of employees is based on the U.S. Bureau of Labor Statistics Total Nonfarm Payroll statistic for December 2019 (United States Department of Labor, 2021). Facebook usage is based on the Pew Research Center January 2019 Core Trends Survey (Pew, 2019)

based on immutable characteristics, are predicted to be more susceptible to mental illness. We also find that the detrimental effects on mental health have negative downstream consequences on the students' academic performance. Lastly, our exploration of mechanisms suggests the results are consistent with Facebook enhancing people's abilities to engage in unfavorable social comparisons.

The results presented in this study should be interpreted with caution for several reasons. First, as discussed, our estimates cannot speak directly to the effects of social media features—e.g., news pages—that were introduced after the time period considered in our study. Second, despite being the core component of most mental health diagnoses, self reports may still suffer from measurement error for reasons related to recall bias and lack of incentives. Finally, we note that our results apply 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.

We emphasize once again that this paper does not aim to estimate the overall welfare effects of social media; rather, it aims to shed light on a very important component of such welfare calculation, namely mental health. Clearly, social media might have positive effects on other outcomes affecting welfare. Indeed, the fact that individuals keep using social media despite the documented negative effects on subjective well-being and mental health suggests that social media platforms might have benefits that compensate for such costs. Ideally, future iterations of these platforms will be able to preserve the benefits while mitigating the mental health costs.

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 (2005). The american college health association national college health assessment (ACHA-NCHA), Spring 2003 reference group report. *Journal of American College Health 53*(5), 199.

ACHA (2006a). American college health association-national college health assessment (acha-ncha) spring 2004 reference group data report (abridged). *Journal of American College Health 54*(4), 201.

ACHA (2006b). American college health association national college health assessment (acha-ncha) spring 2005 reference group data report (abridged). *Journal of American College Health 55*(1), 5.

ACHA (2007). American college health association national college health assessment spring 2006 reference group data report (abridged). *Journal of American College Health 55*(4), 195.

ACHA (2008). American college health association-national college health assessment spring 2007 reference group data report (abridged). *Journal of American College Health 56*(5), 469–479.

ACHA (2009). American college health association-national college health assessment spring 2008 reference group data report (abridged): the american college health association. *Journal of American College Health 57*(5), 477–488.

Acquisti, A. and R. Gross (2006). Imagined communities: Awareness, information sharing, and privacy on the facebook. In *International workshop on privacy enhancing technologies*, pp. 36–58. Springer.

Allcott, H., L. Braghieri, S. Eichmeyer, and M. Gentzkow (2020). The welfare effects of social media. *American Economic Review 110*(3), 629–676.

Allcott, H., M. Gentzkow, and L. Song (2021). Digital addiction. *NBER Working Paper*.

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders: DSM-5.

Anderson, M. 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.

Ang, D. (2021). The effects of police violence on inner-city students. *The Quarterly Journal of Economics 136*(1), 115–168.

Appel, H., A. L. Gerlach, and J. Crusius (2016). The interplay between facebook use, social comparison, envy, and depression. *Current Opinion in Psychology 9*, 44–49.

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

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

Baranov, V., S. Bhalotra, P. Biroli, and J. Maselko (2020, March). Maternal depression, women's empowerment, and parental investment: Evidence from a randomized controlled trial. *American Economic Review 110*(3), 824–59.

Bekalu, M. A., R. 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, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika 93*(3), 491–507.

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

Biasi, B., M. S. Dahl, and P. Moser (2019). Career effects of mental health. *Available at SSRN 2544251*.

Blattman, C., J. C. Jamison, and M. Sheridan (2017, April). Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in liberia. *American Economic Review 107*(4), 1165–1206.

Borusyak, K., X. Jaravel, and J. Spiess (2021). Revisiting event study designs: Robust and efficient estimation.

Brügger, N. (2015). A brief history of facebook as a media text: The development of an empty structure. *First Monday*.

Bursztyn, L., G. Egorov, R. Enikolopov, and M. Petrova (2019). Social media and xenophobia: evidence from Russia.

Cabral, M., B. Kim, M. Rossin-Slater, M. Schnell, and H. Schwandt (2021). Trauma at school: The impacts of shootings on students' human capital and economic outcomes. Technical report, National Bureau of Economic Research.

Callaway, B. and P. H. Sant'Anna (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.

Cassidy, J. (2006). Me media: How hanging out on the internet became big business. *The New Yorker 82*(13), 50.

Chan, D. (2010). So why ask me? are self-report data really that bad? In *Statistical and methodological myths and urban legends*, pp. 329–356. Routledge.

Chou, H.-T. G. and N. Edge (2012). "they are happier and having better lives than i am": The impact of using facebook on perceptions of others' lives. *Cyberpsychology, Behavior, and Social Networking 15*(2), 117–121.

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

Dienlin, T., P. K. Masur, and S. 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.

Enikolopov, R., A. Makarin, and M. Petrova (2020). Social media and protest participation: Evidence from russia. *Econometrica 88*(4), 1479–1514.

Enikolopov, R., M. Petrova, and K. 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, L. and C. Molina (2020). Facebook causes protests. Technical report, The Latin American and Caribbean Economic Association-LACEA.

Fujiwara, T., K. Müller, and C. Schwarz (2020). The effect of social media on elections: Evidence from the united states. *Available at SSRN 3719998*.

Goetzel, R. Z., S. R. Long, R. J. Ozminkowski, K. Hawkins, S. Wang, and W. Lynch (2004). Health, absence, disability, and presenteeism cost estimates of certain physical and mental health conditions affecting us employers. *Journal of Occupational and Environmental Medicine 46*(4), 398–412.

Golberstein, E., G. Gonzales, and E. 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, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.

Griffiths, M. D., D. J. Kuss, and Z. Demetrovics (2014). Social networking addiction: An overview of preliminary findings. *Behavioral Addictions*, 119–141.

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

Hass, N. (2006). In your Facebook.com. *The New York Times 8*, 30–31.

Haushofer, J. and J. 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.

Hirschland, J. (2006). Busted on the facebook. *Columbia Spectator 136*(1).

Jacobs, A. Z., S. F. Way, J. Ugander, and A. Clauset (2015). Assembling thefacebook: Using heterogeneity to understand online social network assembly. In *Proceedings of the ACM Web Science Conference*, pp. 1–10.

James, S. L., D. Abate, K. H. Abate, S. M. Abay, C. Abbafati, N. Abbasi, H. Abbastabar, F. Abd-Allah, J. Abdela, A. 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, K., K. Gathercoal, and W. Buhrow (2012). Does survey format influence self-disclosure on sensitive question items? *Computers in Human Behavior 28*(1), 251–256.

Kelly, Y., A. Zilanawala, C. Booker, and A. Sacker (2018). Social media use and adolescent mental health: Findings from the uk millennium cohort study. *EClinicalMedicine 6*, 59–68.

Kessler, R. C., G. P. Amminger, S. Aguilar-Gaxiola, J. Alonso, S. Lee, and T. B. Ustun (2007). Age of onset of mental disorders: a review of recent literature. *Current opinion in psychiatry 20*(4), 359.

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

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

Kroenke, K. and R. L. Spitzer (2002). The PHQ-9: a new depression diagnostic and severity measure.

Lampe, C., N. B. Ellison, and C. Steinfield (2008). Changes in use and perception of facebook. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pp. 721–730.

Layard, R. (2017). The economics of mental health. *IZA World of Labor*.

Lester, A. E. (2004). Show your best face: Online social networks are a hop, click and jump from reality. *The Harvard Crimson*. https://www.thecrimson.com/article/2004/2/17/show-your-best-face-lets-talk/.

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

Lin, L. Y., J. E. Sidani, A. Shensa, A. Radovic, E. Miller, J. B. Colditz, B. L. Hoffman, L. M. Giles, and B. A. Primack (2016). Association between social media use and depression among us young adults. *Depression and anxiety 33*(4), 323–331.

Meier, A., L. Reinecke, and C. E. Meltzer (2016). "facebocrastination"? predictors of using facebook for procrastination and its effects on students' well-being. *Computers in Human Behavior 64*, 65–76.

Mosquera, R., M. Odunowo, T. McNamara, X. Guo, and R. Petrie (2020). The economic effects of facebook. *Experimental Economics 23*(2), 575–602.

Müller, K. and C. 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, S. and L. 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, V., A. J. Flisher, S. Hetrick, and P. McGorry (2007). Mental health of young people: a global public-health challenge. *The Lancet 369*(9569), 1302–1313.

Paul, J. A., H. M. Baker, and J. D. Cochran (2012). Effect of online social networking on student academic performance. *Computers in Human Behavior 28*(6), 2117–2127.

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

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

Pew (2019). Pew research center january 2019 core trends. https://www.pewresearch.org/internet/dataset/core-trends-survey/.

Plomin, R., C. M. A. Haworth, and O. S. P. Davis (2009). Common disorders are quantitative traits. *Nature Review Genetics 10*, 872–878.

Ridley, M., G. Rao, F. Schilbach, and V. Patel (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science 370*(6522).

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

Shapiro, B. (2021). Promoting wellness or waste? evidence from antidepressant advertising. *American Economic Journal: Microeconomics*.

Sheldon, P. (2008). Student favorite: Facebook and motives for its use. *Southwestern Mass Communication Journal 23*(2).

Stutzman, F. (2006). Student life on the Facebook. https://bit.ly/3keTneb.

Sun, L. and S. Abraham (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*.

Thompson, S. H. and E. 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, A. L., P. J. Mucha, and M. A. Porter (2012). Social structure of facebook networks. *Physica A: Statistical Mechanics and its Applications 391*(16), 4165–4180.

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

Twenge, J. M., A. B. Cooper, T. E. Joiner, M. E. Duffy, and S. 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.

United States Department of Labor (2021). All employees: Total nonfarm. https://fred.stlouisfed.org/series/PAYEMS (accessed July 2011).

Verduyn, P., N. Gugushvili, K. Massar, K. Taht, and E. Kross (2020). Social comparisons on social networking sites. *Current Opinion in Psychology 36*, 32–37.

Vogel, E. A., J. P. Rose, L. R. Roberts, and K. Eckles (2014). Social comparison, social media, and self-esteem. *Psychology of Popular Media Culture 3*(4), 206.

Watson, S. W., Z. Smith, and J. Driver (2006). Alcohol, Sex and Illegal Activities: An Analysis of Selected Facebook Central Photos in Fifty States. *Online Submission*.

We Are Social (2021). Digital 2021 april global statshot report.

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

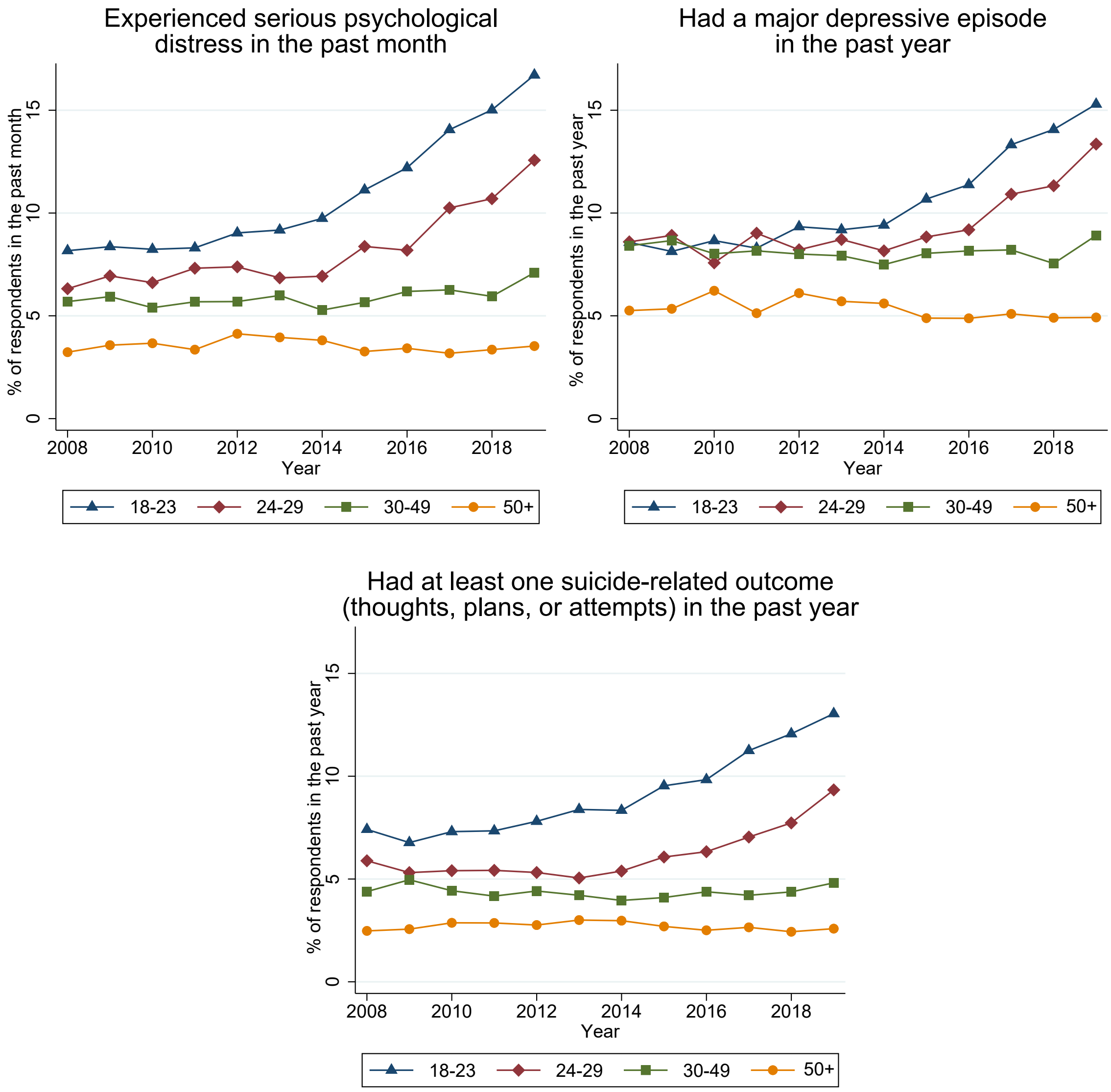
WHO (2018). Mental health: strengthening our response.

Wilson, R. E., S. D. Gosling, and L. T. Graham (2012). A review of facebook research in the social sciences. *Perspectives on psychological science 7*(3), 203–220.

Yang, L., Y. Zhao, Y. Wang, L. Liu, X. Zhang, B. Li, and R. Cui (2015). The effects of psychological stress on depression. *Current neuropharmacology 13*(4), 494–504.

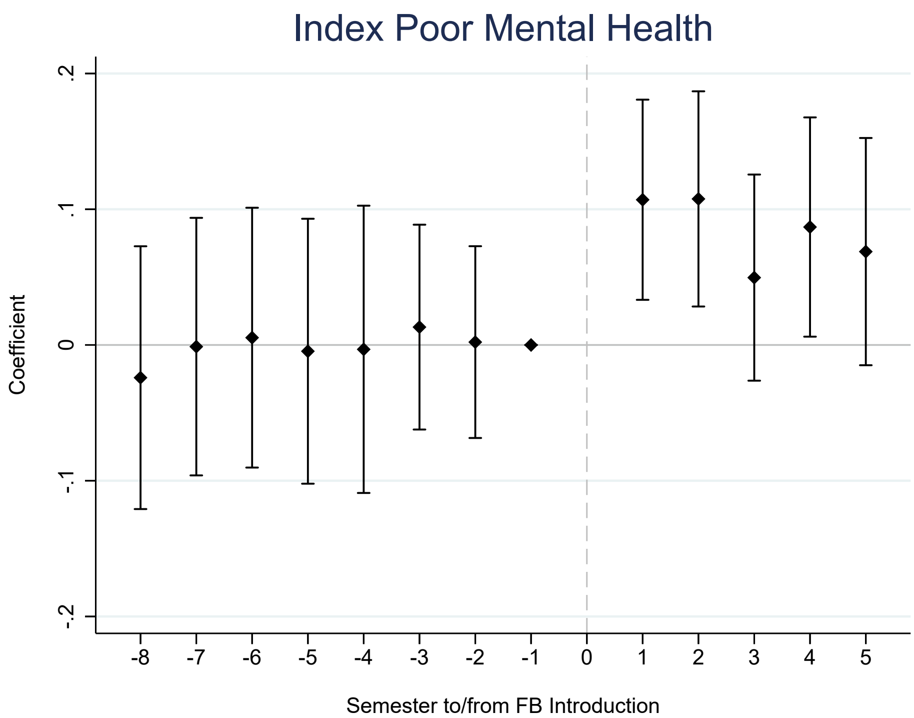
Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political effects of the internet and social media. *Annual Review of Economics 12*, 415–438.

Figure 1: Mental Health Trends in the United States, 2008–2019



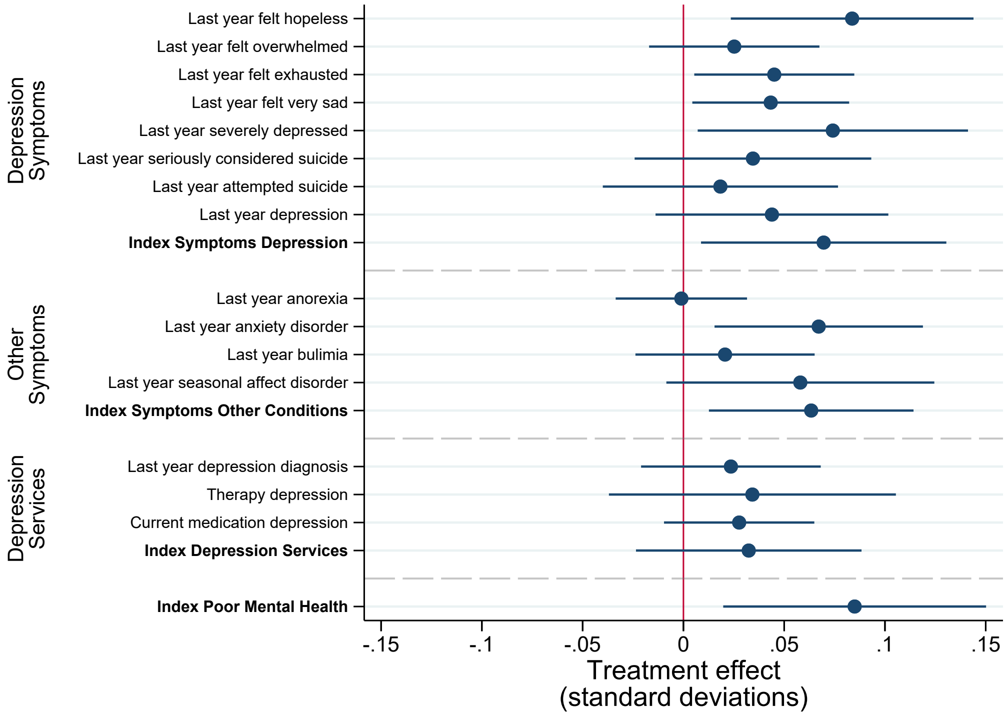
Notes: This figure displays the mental health trends in the United States by age group in the time period 2008–2019. The data come from the National Survey on Drug Use and Health. The data are not available for respondents younger than 18 or for years earlier than 2008. For the precise question formulations and variable definitions, see NSDUH (2019). For a more detailed analysis and discussion of these trends, see Twenge et al. (2019).

Figure 2: Effects Based on Distance to/from Facebook Introduction



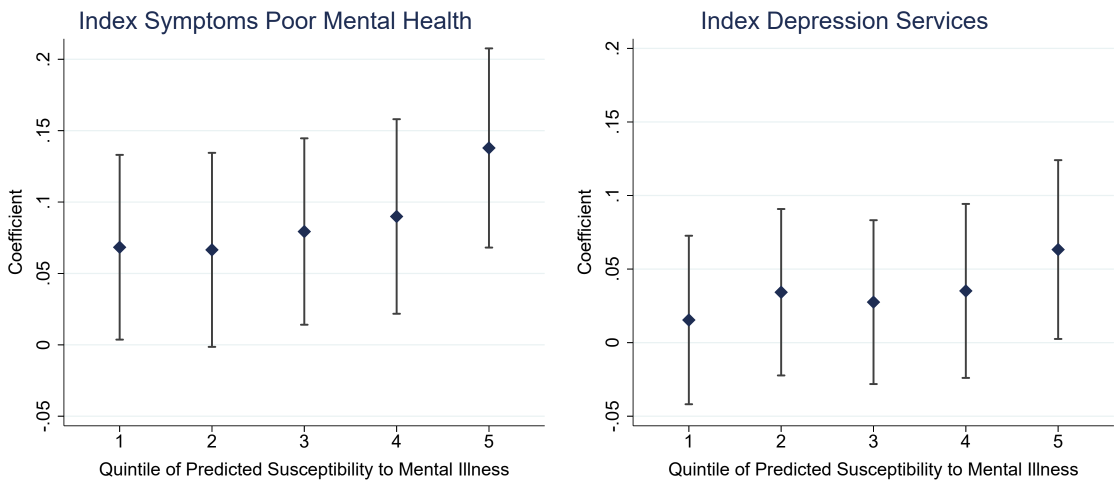
Notes: This figure presents an event-study-like plot; specifically, it displays coefficients β_k from Equation (2). The time spanned by the x-axis (8 semesters in the pre-period and 5 in the post-period) is the largest span of time for which we have data from all four Facebook expansion groups. The outcome variable is our overall index of poor mental health. The index is standardized so that, in the pre-period, it has a mean of zero and a standard deviation of one. The regression does not include controls. The coefficient on $t = 0$ is omitted because we exclude from the analysis respondents who took the survey in the semester in which Facebook was rolled out at their colleges. Given the dates in which Facebook was introduced at the various colleges in our dataset and the starting dates of the subsequent semesters, $t = 1$ corresponds to at least two months after Facebook's introduction at a college. The coefficient on $t = -1$ is normalized to zero. For a detailed description of the outcome and treatment variables, see Appendix Table A.19. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure 3: Effect of Facebook Introduction on Student Mental Health



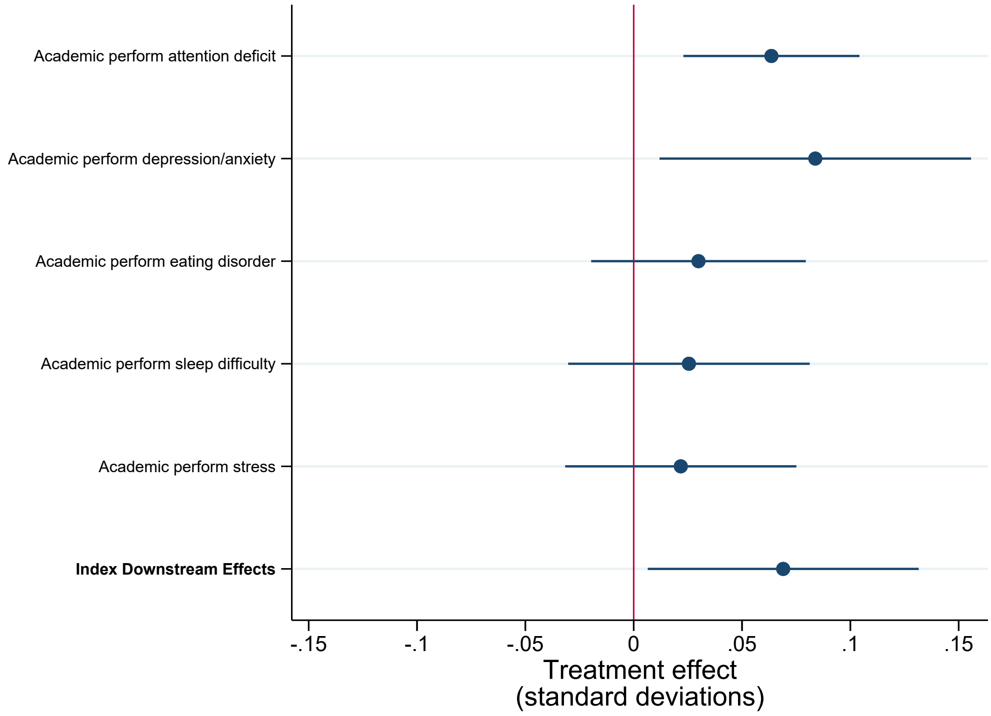
Notes: This figure explores the effects of the introduction of Facebook at a college on all our mental-health outcome variables and on the related indices. Specifically, it presents estimates of coefficient β from Equation (1) using our preferred specification, namely the one including survey-wave fixed 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. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.19. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure 4: Heterogeneous Effects by Predicted Susceptibility to Mental Illness



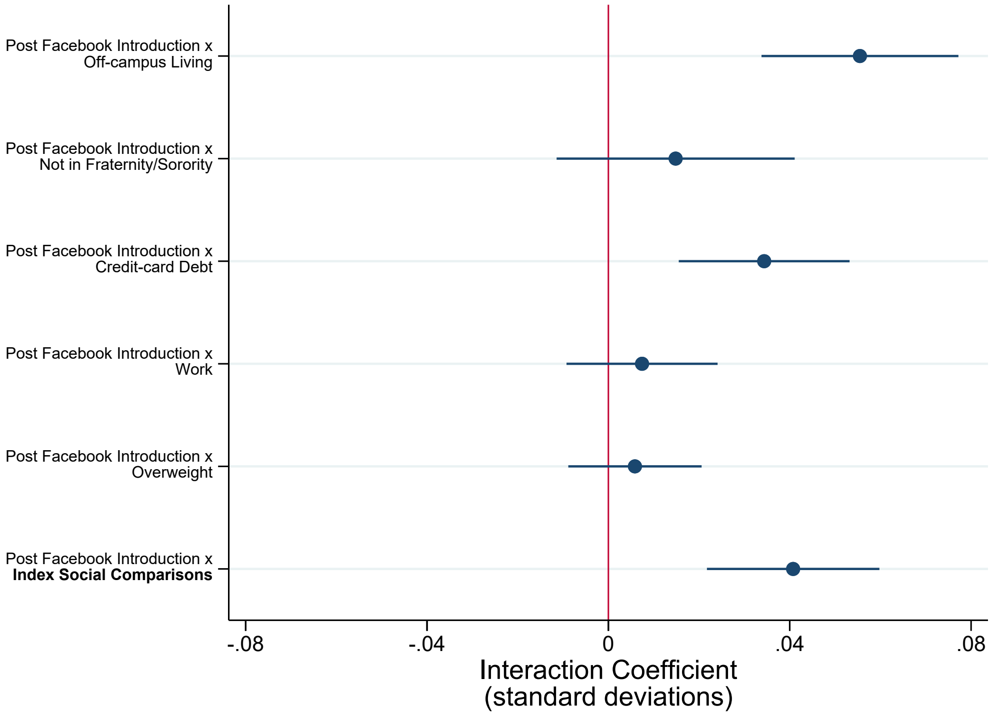
Notes: This figure explores the extent to which the effects of the introduction of Facebook at a college are heterogeneous depending on students' predicted susceptibility to mental illness. Specifically, it presents estimates from a version of Equation (1) in which our indicator for post Facebook introduction 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 left panel is our index of symptoms of poor mental health; the outcome variable in the right panel is our index of depression services. Both indices are standardized so that, in the pre-period, they have a mean of zero and a standard deviation of one. The displayed coefficients represent the linear combination of the baseline effect for the omitted category (lowest quintile) and the heterogeneity component. The estimates 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.19. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure 5: Downstream Effects on Academic Performance



Notes: This figure explores downstream effects of the introduction of Facebook on the students' academic performance. 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 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. 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, and control variables, see Appendix Table A.19. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure 6: Heterogeneous Effects as Evidence of Unfavorable Social Comparisons Mechanism



Notes: This figure explores the mechanisms behind the effects of Facebook on mental health. Specifically, 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 sub-populations of students are: students who live off-campus, students who do not belong to a fraternity or sorority, students who carry some credit card debt, students who work alongside studying, and students who are overweight according to body mass index (BMI). The estimates 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.19. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

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 effects 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.19. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table 2: Length-of-exposure Specification

	Index Poor Mental Health		Index Symptoms Poor Mental Health		Index Depression Services	
	(1)	(2)	(3)	(4)	(5)	(6)
Num. Treated Semesters	0.020***	0.024***	0.019***	0.022***	0.012***	0.019***
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)
Observations	315,155	315,155	316,256	316,256	332,011	332,011
Survey Wave FE	✓		✓		✓	
College FE	✓		✓		✓	
Controls	✓	✓	✓	✓	✓	✓
Survey Wave × College FE		✓		✓		✓

Notes: This table explores the effects of length of exposure to Facebook on student mental health. Specifically, it presents estimates of coefficient β from Equation (3). The outcome variables are the overall index of poor mental health (columns (1) and (2)), the index of symptoms of poor mental health (columns (3) and (4)), and the index of depression services (columns (5) and (6)). All indices are standardized so that, in the pre-period, they have a mean of zero and a standard deviation of one. Odd-numbered columns present estimates of Equation (3) including survey-wave fixed effects, college fixed effects and controls; even-numbered columns replace survey-wave fixed effects and college fixed effects with survey-wave × college fixed effects. 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. Cohorts of students who might have been exposed to Facebook in high school are excluded from the regression. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.19. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table 3: Effects on Alcohol Use and Perceptions

(a) Perceptions of typical alcohol use

	Typical drink count	Share used 30 days	Typical student used daily	Index std. dev.
	(1)	(2)	(3)	(4)
Post Facebook Introduction	0.154**	0.020***	0.043***	0.120***
	(0.072)	(0.004)	(0.011)	(0.030)
Baseline mean	5.71	0.70	0.38	0.00
Observations	375,025	370,390	378,503	380,886
Survey Wave FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
College FE	✓	✓	✓	✓

(b) Reported alcohol use

	Drink count	Used 30 days	Used daily	Index std. dev.
	(1)	(2)	(3)	(4)
Post Facebook Introduction	0.099	0.004	0.001	0.019
	(0.068)	(0.011)	(0.004)	(0.021)
Baseline mean	4.15	0.68	0.04	0.00
Observations	377,844	378,590	378,590	380,886
Survey Wave FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
College FE	✓	✓	✓	✓

Notes: This table explores the effects of the introduction of Facebook at a college on students' perceptions and behaviors related to alcohol use. Specifically, it presents estimates of coefficient β from Equation (1). Panel (a) presents results on perceptions; Panel (b) presents results on actual alcohol use. All columns are in original units, besides column (4) which is an index of the outcomes in columns (1) through (3). All indices are standardized so that, in the pre-period, they have a mean of zero and a standard deviation of one. All estimates 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, and control variables, see Appendix Table A.19. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Appendix For Online Publication

A Internal Validation of Symptoms Variables

The NCHA survey contains both questions about symptoms of depression and questions related to depression diagnoses. As a validation exercise, we study the relationship between exhibiting symptoms of depression and having ever been diagnosed with depression in our sample. We note that, in the NCHA dataset, it is impossible to distinguish individuals who, if evaluated by a mental healthcare professional, would not be diagnosed with depression from individuals who never visited a healthcare professional in the first place. In other words, the absence of a depression diagnosis might mean that the individual is not affected by depression or that the individual is affected by depression but never visited a mental healthcare professional. With this caveat in mind, we study how well our index of depression symptoms predicts ever having received a depression diagnosis.

As shown in Appendix Figure A.9, the index of symptoms of depression is highly predictive of ever having received a depression diagnosis. Specifically, for each ventile of our index of symptoms of depression, the figure plots the average index of symptoms of depression against the fraction of individuals who have ever received a depression diagnosis. The correlation coefficient between the two measures is 0.37.

As an additional validation exercise, Appendix Figure A.10 shows the Receiver Operating Characteristic (ROC) curve for a binary classifier constructed by running a logit model of ever having been diagnosed with depression on our index of depression symptoms. As shown in the figure, the binary classifier performs fairly well. For instance, it can achieve a true positive rate of 75% at the cost of a false positive rate of 30%. In other words, the classifier classifies as having received a depression diagnosis 75% of individuals who indeed have ever received a depression diagnosis and only 30% of individuals who have never received a depression diagnosis. As discussed before, some of the individuals who have never been diagnosed with depression might actually be affected by depression and might have simply never been evaluated by a healthcare professional. Therefore, the actual performance of the classifier is likely be even higher because some of the observations that are currently being counted as false positive might actually be true positives.[56]

[56] Similarly, some of the observations that are counted as false negatives might actually be true negatives. That is because our index of depression symptoms might classify individuals who received a depression diagnosis in the past but have since recovered as not being affected by depression. Such classification is counted as a false negative in the figure above, but it would be counted as a true negative in a world in which the variable being

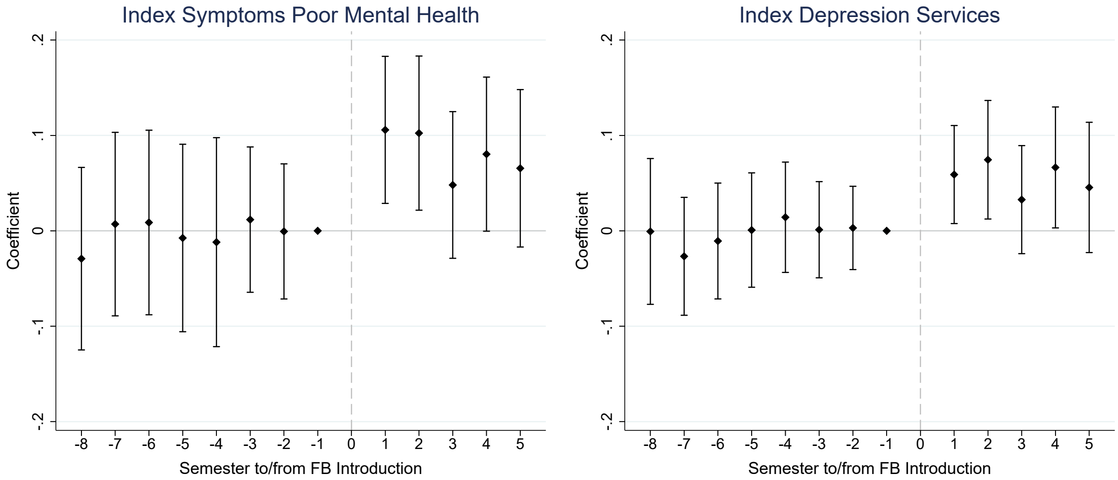
B Additional Tables and Figures

Figure A.1: Facebook Homepage as of June 2004

Notes: The figure shows a snaphsot of the homepage of thefacebook.com as of June 15th, 2004 recovered via the Wayback Machine. The colleges that, by that date, had been granted access to Facebook are listed on the home page.

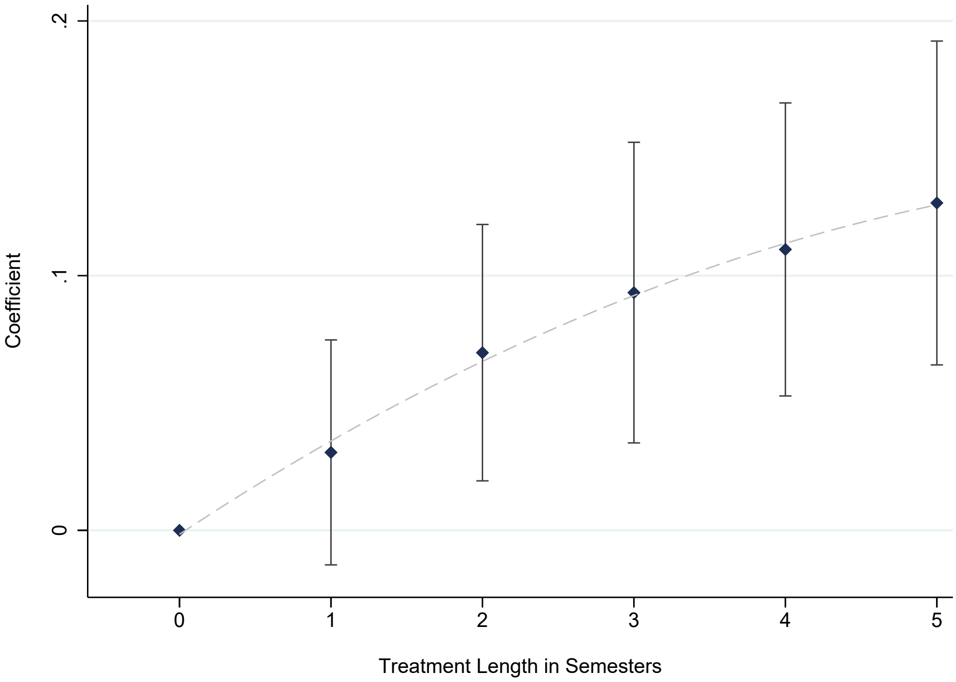
predicted is whether the student has an active depression diagnosis at the time in which she takes the survey. Unfortunately, the NCHA dataset does not contain such a variable.

Figure A.2: Effects Based on Distance to/from Facebook Introduction: Separating Symptoms of Mental Illness and Take-up of Mental Healthcare Services



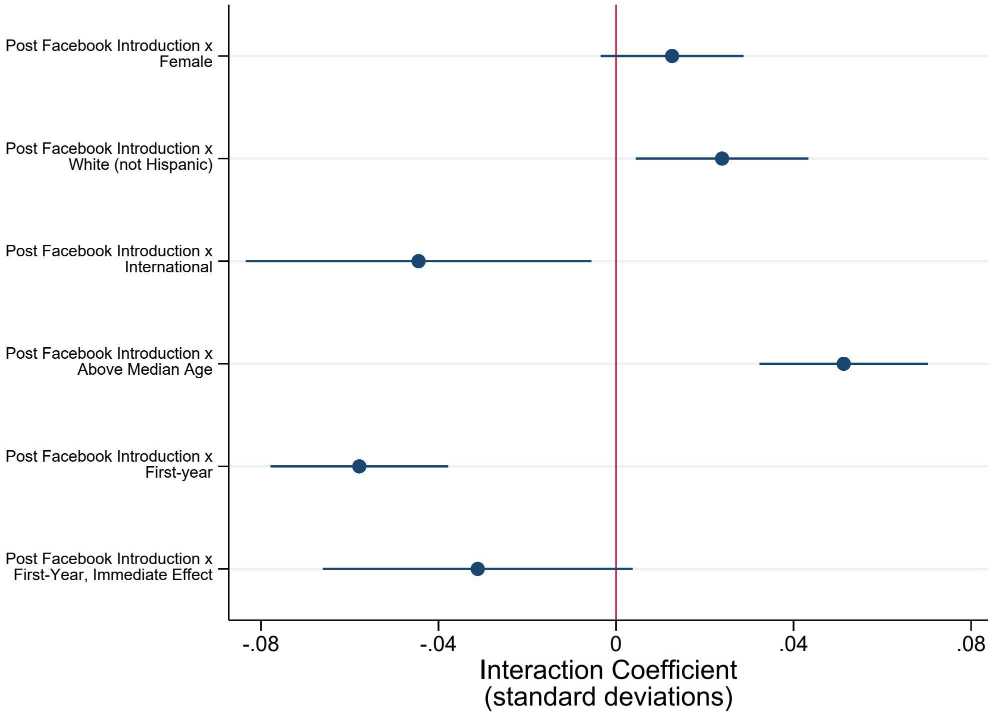
Notes: This figure presents two event-study-like plots, separating symptoms of poor mental health from the take-up of mental healthcare services. Specifically, each figure presents estimates of coefficients β_k from Equation (2). The time spanned by the x-axis (8 semesters in the pre-period and 5 in the post-period) is the largest span of time for which we have data from all four Facebook expansion groups. The outcome variable in the left panel is our index of symptoms of poor mental health; the outcome variable in the right panel is our index of depression services. Both indices are standardized so that, in the pre-period, they have a mean of zero and a standard deviation of one. The regressions do not include controls. For the same figure involving the overall index of poor mental health (which encompasses both symptoms and services), see Figure 2. The coefficient on $t = 0$ is omitted because we exclude from the analysis respondents who took the survey in the semester in which Facebook was rolled out at their colleges. Given the dates in which Facebook was introduced at the various colleges in our dataset and the starting dates of the subsequent semesters, $t = 1$ corresponds to at least two months after Facebook's introduction at a college. The coefficient on $t = -1$ is normalized to zero. For a detailed description of the outcome and treatment variables, see Appendix Table A.19. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure A.3: Effect on Mental Health by Length of Exposure to Facebook



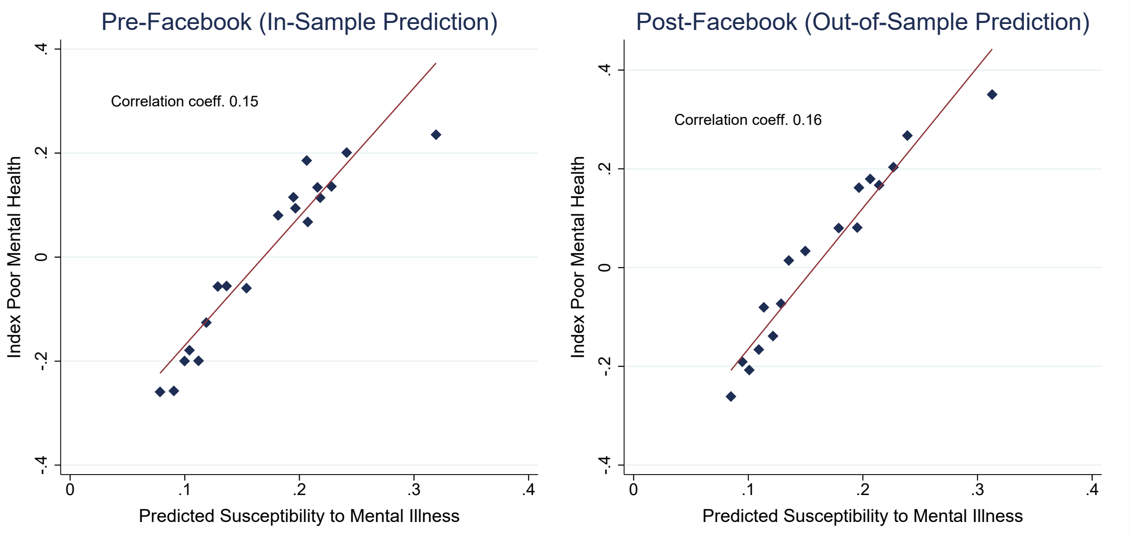
Notes: This figure explores the effects of length of exposure to Facebook on student mental health without imposing the assumption that length of exposure to Facebook enters the regression linearly. Specifically, it presents estimates from a version of Equation (3) in which "treatment length in semesters" enters as a set of indicators rather than linearly. The outcome variable is our index of poor mental health. 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. Cohorts of students who might have been exposed to Facebook in high school are excluded from the regression. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.19. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure A.4: Heterogeneous Effects



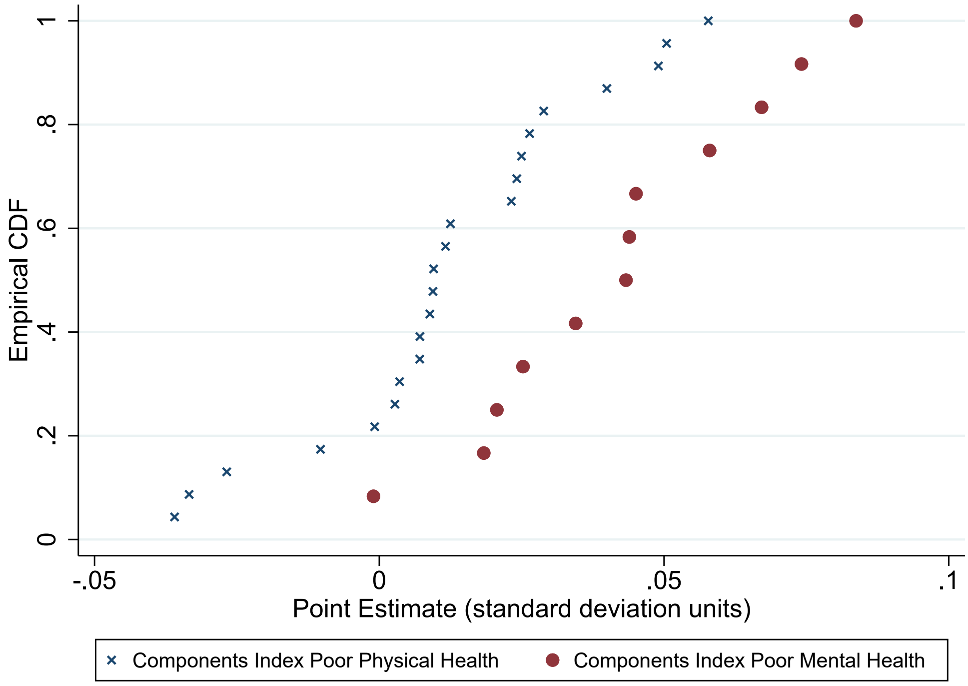
Notes: This figure explores whether the effects of the introduction of Facebook on student mental health are heterogeneous across a host of demographic characteristics. Specifically, it presents estimates from a version of Equation (1) in which our treatment indicator is interacted with various moderators. The outcome variable is our index of poor mental health. The index is standardized so that, in the pre-period, it has a mean of zero and a standard deviation of one. The moderators are indicators for: identifying as female, identifying as white (non Hispanic), being an international student, being above median age, and being a first-year student (freshman). In the last row, we restrict our sample to only include students who took the survey at most one semester after the introduction of Facebook at their college. 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, and control variables, see Appendix Table A.19. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure A.5: Relationship between LASSO-predicted Measure of Susceptibility to Mental Illness and Index of Poor Mental Health



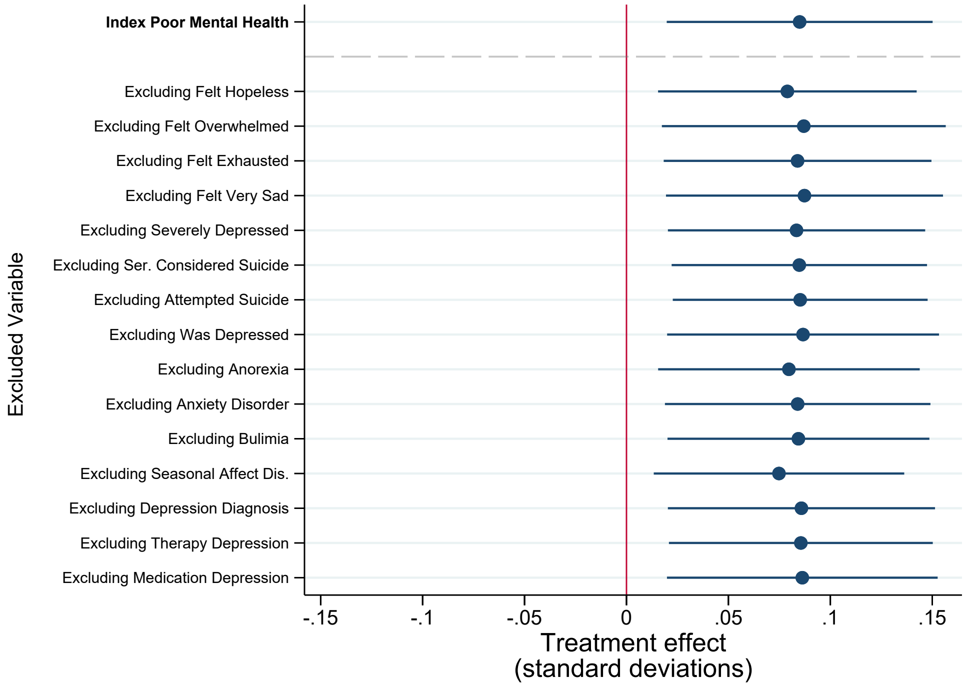
Notes: This figure explores the relationship between our LASSO-predicted measure of susceptibility to mental illness and our index of poor mental health. Specifically, for each ventile of our LASSO-predicted measure of susceptibility to mental illness, the figure plots the average predicted susceptibility to mental illness against the average index of poor mental health. The index is standardized so that, in the pre-period, it has a mean of zero and a standard deviation of one. See Section 5.2 for details about the LASSO procedure. The left panel presents data from the period before the Facebook introduction at a college; the right panel presents data from the period after the Facebook introduction at a college. Since the LASSO algorithm is trained on pre-period data, the left figure shows in-sample predictions, whereas the right figure shows out-of-sample predictions. The figure also displays correlation coefficients between the index of poor mental health and our LASSO-predicted measure of susceptibility to mental illness.

Figure A.6: Cumulative Distribution of Coefficients on Components of Index of Poor Mental Health and Index of Poor Physical Health



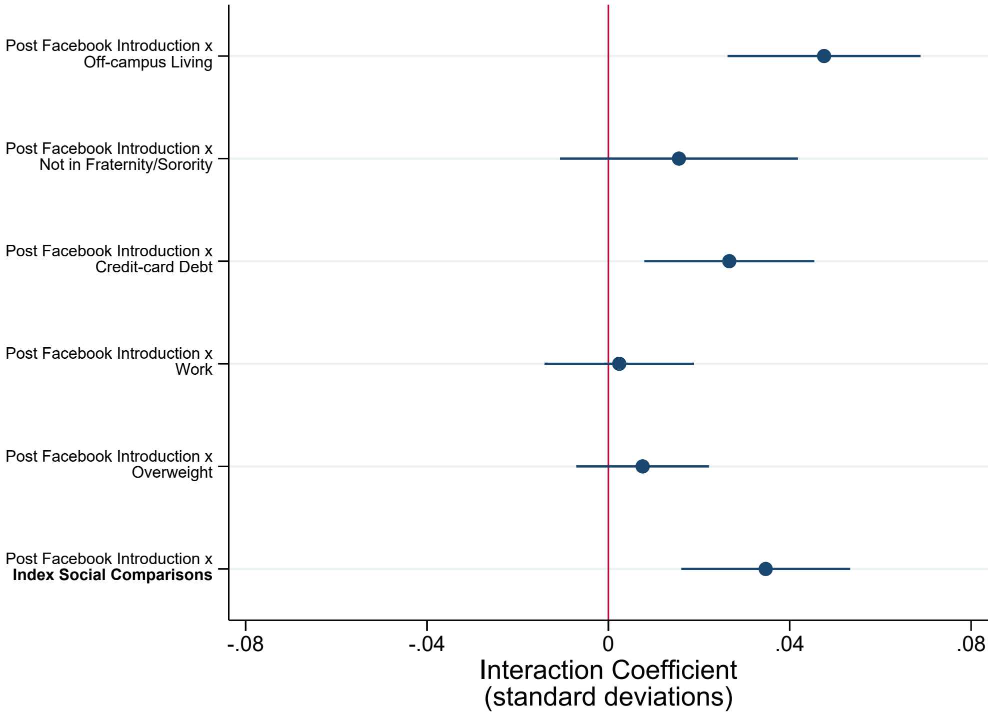
Notes: This figure displays cumulative distribution functions of the coefficients on the components of the indices of poor mental and poor physical health. The figure is constructed as follows: first, we computed estimates of coefficients β from Equation (1) for each component of the index of poor physical health and for each component of the index of poor mental health. Second, we constructed two cumulative distribution functions using the estimated coefficients: one for the components of the index of poor physical health, and one for the components of the index of poor mental health. All estimates are obtained using our preferred specification, namely the one including survey-wave fixed effects, college fixed effects, and controls. The outcomes are always 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. For a detailed description of the questions used in the construction of the two indices, as well as for a description of treatment and control variables, see Appendix Table A.19. Standard errors in parentheses are clustered at the college level.

Figure A.7: Robustness to Excluding Each Variable from the Index of Poor Mental Health



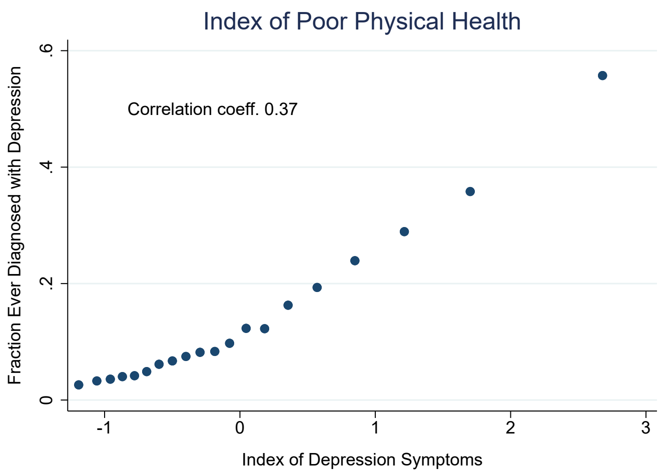
Notes: This figure explores the robustness of our baseline results to excluding each individual variable from the construction of the index of poor mental health. Specifically, it presents estimates of coefficient β from Equation (1). Each row excludes a different variable from the construction of the index. The index is always standardized so that, in the pre-period, it has a mean of zero and a standard deviation of one. All estimates 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, and control variables, see Appendix Table A.19. The bars represent 95% confidence intervals. Standard errors in parentheses are clustered at the college level.

Figure A.8: Heterogeneous Effects as Evidence of Unfavorable Social Comparisons Mechanism, Controlling for Predicted Susceptibility to Mental Illness



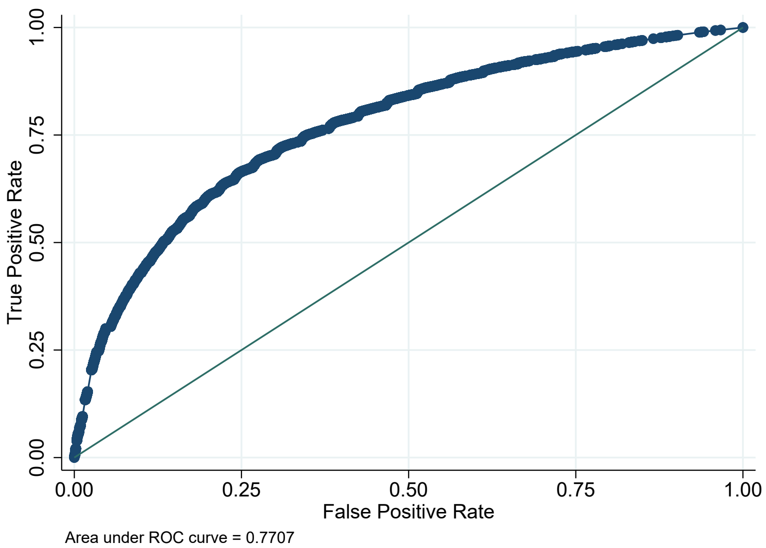
Notes: This figure presents a version of Figure 6 controlling for heterogeneity by the predicted susceptibility to mental illness. Specifically, it presents estimates from a version of Equation (1) in which our treatment indicator is interacted with a set of indicators for belonging to a certain sub-population of students and in which our treatment indicator is also interacted with our LASSO-predicted measure of susceptibility to mental illness. The outcome variable is our overall index of poor mental health. The sub-populations of students are: students who live off-campus, students who do not belong to a fraternity or sorority, students who carry some credit card debt, students who work alongside studying, and students who are overweight according to body mass index (BMI). The estimates 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.19. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure A.9: Relationship between Index of Symptoms of Depression and Ever Having Been Diagnosed with Depression



Notes: This figure explores the relationship between our index of symptoms of depression and ever having been diagnosed with depression. Specifically, for each ventile of our index of depression symptoms, the figure plots the fraction of individuals who have ever received a depression diagnosis. The index is standardized so that, in the pre-period, it has a mean of zero and a standard deviation of one. The figure also displays the correlation coefficient between the index of depression symptoms and the fraction of individuals ever diagnosed with depression.

Figure A.10: Performance of Binary Classifier based on Index of Symptoms of Depression



Notes: The figure presents the Receiver-Operating-Characteristic curves of the binary classifiers constructed by running a logit model of ever having been diagnosed with depression on our index of depression symptoms.

Table A.1: Summary Statistics by Facebook Expansion Group: IPEDS data

	(1) FB Expansion Group 1 (Spring 2004) mean	(2) FB Expansion Group 2 (Fall 2004) mean	(3) FB Expansion Group 3 (Spring 2005) mean	(4) FB Expansion Group 4 (Fall 2005) mean
Panel A. University Characteristics				
Four-year	1.00	0.99	0.98	0.84
Public	0.28	0.52	0.51	0.42
Private non-profit	0.72	0.48	0.49	0.56
Offers doctoral degrees	0.86	0.63	0.41	0.22
Offers graduate degrees	0.91	0.86	0.87	0.69
Offers medical degrees	0.62	0.20	0.05	0.02
Has tenure system	1.00	0.98	0.96	0.84
Land grant institution	0.14	0.15	0.02	0.03
Located in a city with >250k population (or suburb)	0.47	0.47	0.37	0.38
Located in a rural area	0.03	0.03	0.04	0.08
Huge (>20k students)	0.41	0.29	0.08	0.03
Large (10–20k students)	0.29	0.22	0.23	0.09
Medium-sized (5–10k students)	0.10	0.17	0.25	0.21
Small (<5k students)	0.19	0.32	0.44	0.66
Region: Midwest	0.14	0.18	0.23	0.25
Region: Northeast	0.45	0.35	0.32	0.24
Region: South	0.28	0.28	0.33	0.43
Region: West	0.14	0.19	0.12	0.08
Panel B. Undergraduate program characteristics				
Highly selective	0.93	0.61	0.30	0.07
Medium selective	0.07	0.35	0.59	0.53
Low selective	0.00	0.04	0.11	0.40
Large (>10k students)	0.69	0.47	0.22	0.04
Medium-size (3–10k students)	0.16	0.31	0.47	0.35
Small (<3k students)	0.16	0.22	0.31	0.61
Highly residential	0.66	0.43	0.41	0.39
Primarily residential	0.26	0.35	0.34	0.34
Primarily non-residential	0.09	0.23	0.24	0.26
Number of colleges	58	236	268	213
Number of colleges (NCHA subsample)	40	124	120	137

Notes: This table presents college-level summary statistics by Facebook expansion group. The data is obtained by merging our Facebook introduction dates dataset to data from the Integrated Postsecondary Education Data System (IPEDS). Selectivity tiers are defined as follows: colleges classified as being "highly selective" have average incoming student test scores in the first (top) quintile of all baccalaureate-granting institutions. Colleges classified as being "medium selective" have average incoming student test scores in the second and third quintile of all baccalaureate-granting institutions. The remaining colleges are classified as "low selective". We note that the summary statistics do not refer to the subset of colleges from the Facebook introduction dates dataset that appears in the NCHA dataset; they refer to the full set of 775 colleges from the Facebook introduction dates dataset. The rationale is that, for privacy reasons, the NCHA dataset was stripped of college identifiers and, therefore, cannot be matched to the IPEDS dataset. The second-to-last row of the table shows the distribution of colleges in the Facebook expansion dates dataset across Facebook expansion wave; the last row of the table shows the distribution of colleges in the NCHA dataset across Facebook expansion wave.

Table A.2: Summary Statistics by Facebook Expansion Group: NCHA data

	(1) FB Expansion Group 1 (Spring 2004) mean	(2) FB Expansion Group 2 (Fall 2004) mean	(3) FB Expansion Group 3 (Spring 2005) mean	(4) FB Expansion Group 4 (Fall 2005) mean
Panel A. Baseline Characteristics				
Female	0.65	0.63	0.63	0.61
White	0.70	0.80	0.82	0.77
Year in School	2.38	2.34	2.69	2.21
Weekly Work Hrs	1.99	2.30	2.51	2.63
Credit Card Debt	2.73	2.84	3.19	2.96
Panel B. Baseline Mental Health				
Index Poor Mental Health	0.06	-0.02	-0.02	-0.03
Index Symptoms Poor Mental Health	0.07	-0.02	-0.02	-0.03
Index Depression Services	-0.00	-0.03	-0.02	-0.01
Number of students	16514	41014	21964	16575

Notes: This table presents student-level summary statistics by Facebook expansion group. The data is obtained by averaging student-level characteristics from the NCHA dataset across colleges in different Facebook expansion groups. The averages are taken in the pre-period; i.e., up to and excluding year 2004. All indices 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 variables, see Appendix Table A.19.

Table A.3: Balance

Variable	(1) Pre FB introduction Mean/SE	(2) Post FB introduction Mean/SE	T-test P-value (1)-(2)
Age	20.84	20.68	0.87
	(0.11)	(0.07)	
Female	0.63	0.65	0.26
	(0.01)	(0.01)	
Year in School	2.44	2.48	0.64
	(0.05)	(0.02)	
White	0.80	0.78	0.17
	(0.01)	(0.01)	
International	0.03	0.03	0.78
	(0.00)	(0.00)	
Height (inches)	67.40	67.15	0.39
	(0.08)	(0.05)	
N	123235	254379	
Clusters	224	318	
F-test of joint significance (p-value)			0.86
F-test, number of observations			377614

Notes: This table presents a balance table on the following characteristics: age, gender (indicator for identifying as female), year in school, race (indicator for identifying as white), international status, and height in inches. For a detailed description of the variables, see Appendix Table A.19. The first column shows the mean value of the demographic characteristics in the pre-period; the second columns shows the mean value of those characteristics in the post-period. The *p*-values are calculated after residualizing each demographic characteristic on survey-wave fixed effects and college fixed effects. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A.4: Baseline Results: Individual Variables

	Treatment effect (original units)	Standard error (original units)	Treatment effect (SD units)	Standard error (SD units)	p-value	Sharpened FDR-adjusted q-value
Last year felt hopeless	0.16	0.06	0.08	0.03	0.01	0.09
Last year felt overwhelmed	0.05	0.04	0.03	0.02	0.24	0.29
Last year felt exhausted	0.09	0.04	0.05	0.02	0.03	0.09
Last year felt very sad	0.09	0.04	0.04	0.02	0.03	0.09
Last year severely depressed	0.13	0.06	0.07	0.03	0.03	0.09
Last year seriously considered suicide	0.03	0.02	0.03	0.03	0.25	0.29
Last year attempted suicide	0.01	0.01	0.02	0.03	0.54	0.37
Last year anorexia	-0.00	0.00	-0.00	0.02	0.95	0.59
Last year anxiety disorder	0.02	0.01	0.07	0.03	0.01	0.09
Last year bulimia	0.00	0.00	0.02	0.02	0.36	0.33
Last year depression	0.02	0.01	0.04	0.03	0.13	0.22
Last year seasonal affect disorder	0.01	0.01	0.06	0.03	0.09	0.17
Last year depression diagnosis	0.01	0.00	0.02	0.02	0.30	0.32
Therapy depression	0.01	0.01	0.03	0.04	0.34	0.33
Current medication depression	0.01	0.00	0.03	0.02	0.14	0.22

Notes: This table 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. Columns (1) and (2) present effects and standard errors on un-normalized outcomes. Columns (3) and (4) present effects and standard errors on normalized outcomes, where the normalization is such that the mean in the pre-period is zero and the standard deviation in the pre-period is one. Columns (5) and (6) present unadjusted p-values and sharpened False Discovery Rate-adjusted two-stage q-values, respectively. 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, and control variables, see Appendix Table A.19. Standard errors are clustered at the college level.

Table A.5: Length of Exposure to Facebook and Depression Services

	Last Year Depression Diagnosis	Therapy For Depression	Current Medication Depression
	(1)	(2)	(3)
Num. Treated Semesters	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)
Observations	332,292	332,271	332,216
Baseline mean	0.047	0.030	0.045
Controls	✓	✓	✓
Survey Wave × College FE	✓	✓	✓

Notes: This table explores the effects of length of exposure to Facebook on the take-up of depression-related services. Specifically, it presents estimates of coefficient β from Equation (3), including controls and survey-wave times college fixed effects. The outcome variables are the components of the index of depression services (in original units), namely whether a student was diagnosed with depression within the last year, whether a student was in therapy for depression in the last year, and whether a student was taking anti-depressants over the last year. 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. Cohorts of students who might have been exposed to Facebook in high school are excluded from the regression. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.19. Standard errors in parentheses are clustered at the college level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A.6: Placebo Check: Predicted Susceptibility to Mental Illness

	Predicted Susceptibility to Mental Illness			
	(1)	(2)	(3)	(4)
Post Facebook Introduction	0.139	-0.003	-0.006	-0.007
	(0.116)	(0.015)	(0.005)	(0.005)
Observations	380,886	380,886	380,886	380,886
Survey Wave FE	✓	✓	✓	✓
FB Expansion Group FE	✓	✓		
Controls		✓	✓	✓
College FE			✓	✓
FB Expansion Group Linear Time Trends				✓

Notes: This table presents a placebo check by exploring the effects of the introduction of Facebook at a college on the LASSO-predicted measure of susceptibility to mental illness. Specifically, it presents estimates of coefficient β from Equation (1) with our measure of predicted susceptibility to mental illness as the outcome variable. The outcome variable 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) 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.19. Standard errors in parentheses are clustered at the college level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A.7: Placebo Check: Index of Physical Health

	Index of Poor Physical Health			
	(1)	(2)	(3)	(4)
Post Facebook Introduction	0.064**	0.052**	0.032	0.030
	(0.027)	(0.021)	(0.032)	(0.032)
Observations	365,217	350,481	350,481	350,481
Survey Wave FE	✓	✓	✓	✓
FB Expansion Group FE	✓	✓		
Controls		✓	✓	✓
College FE			✓	✓
FB Expansion Group Linear Time Trends				✓
P-value coeff. physical health vs. coeff. mental health	0.043	0.008	0.055	0.056

Notes: This table presents a placebo check by exploring the effects of the introduction of Facebook at a college on student physical health. Specifically, it presents estimates of coefficient β from Equation (1) with our index of poor physical 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) replaces Facebook-expansion-group fixed effects with college fixed effects; column (4) includes linear-time trends estimated at the Facebook-expansion-group level. The last row of the table shows the p-value on a test of the null hypothesis that the coefficient on the index of poor physical health equals the coefficient on the index of poor mental health from Table 1. 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.19. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.8: Results Excluding each Facebook Expansion Group in Turn

(a) Baseline difference-in-differences specification

	Index of Poor Mental Health			
	(1) Excluding FB Expansion Group 1	(2) Excluding FB Expansion Group 2	(3) Excluding FB Expansion Group 3	(4) Excluding FB Expansion Group 4
Post Facebook Introduction	0.059	0.096***	0.094**	0.084*
	(0.040)	(0.034)	(0.038)	(0.044)
Observations	293,112	216,328	268,554	301,487
Survey Wave FE	✓	✓	✓	✓
College FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓

(b) Length-of-exposure specification

	Index of Poor Mental Health			
	(1) Excluding FB Expansion Group 1	(2) Excluding FB Expansion Group 2	(3) Excluding FB Expansion Group 3	(4) Excluding FB Expansion Group 4
Num. Treated Semesters	0.015***	0.017***	0.020***	0.023***
	(0.005)	(0.006)	(0.005)	(0.005)
Observations	253,501	194,853	233,266	263,851
Survey Wave FE	✓	✓	✓	✓
College FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓

Notes: This table explores the robustness of our baseline results to excluding colleges belonging to each Facebook expansion group in turn. Specifically, it presents estimates of coefficient β from Equation (1) (Panel (a)) and Equation (3) (Panel (b)). Each column excludes all observations from a particular Facebook expansion group in turn. The outcome variable is always the index of poor mental health. The index is standardized so that, in the pre-period, it has a mean of zero and a standard deviation of one. All estimates 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. In Panel (b), cohorts of students who might have been exposed to Facebook in high school are excluded from the regression. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.19. Standard errors in parentheses are clustered at the college level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A.9: Alternative Treatment Assignments for Individuals Taking the Survey in the Semester of the Introduction of Facebook at their College

	Index of Poor Mental Health			
	(1)	(2)	(3)	(4)
Post Facebook Introduction	0.085^{**}	0.043^{***}	0.071^{***}	0.041^{**}
	(0.033)	(0.016)	(0.025)	(0.020)
Observations	359,827	389,878	389,878	389,878
Survey Wave FE	✓	✓	✓	✓
College FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
Imputed Treatment Status	Missing	0	0.5	1

Notes: This table explores whether and how our results vary depending on alternative treatment assignments for respondents who took the survey in the semester in which Facebook was rolled out at their colleges. Since we have no information about whether such respondents took the NCHA survey before or after the introduction of Facebook at their colleges, we do not know whether they are treated or untreated by the time they take the survey. The outcome variable is our index of poor mental health. The index is standardized so that, in the pre-period, it has a mean of zero and a standard deviation of one. All estimates are obtained using our preferred specification, namely the one including survey-wave fixed effects, college fixed effects, and controls. Column (1) presents again our main results, obtained by excluding respondents who took the survey in the semester in which Facebook was rolled out at their colleges. Column (2) presents results assuming such respondents are untreated. Column (3) presents results assigning a treatment status of 0.5 (partially-treated) to those respondents. Column (4) presents results assuming such respondents are fully treated. 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, and control variables, see Appendix Table A.19. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.10: Robustness Check Controlling for College Characteristics Interacted with Survey Wave

	Index Poor Mental Health		
	(1)	(2)	(3)
Post Facebook Introduction	0.104***	0.071*	0.078*
	(0.032)	(0.041)	(0.043)
Observations	359,827	359,827	359,827
Survey Wave FE	✓	✓	✓
College FE	✓	✓	✓
Controls	✓	✓	✓
Survey-wave FE × College Baseline Mental Health	✓		
Survey-wave FE × College Region FE		✓	
Survey-wave FE × Expansion-Group Selectivity Factor			✓

Notes: This table presents a robustness check in which we interact survey-wave fixed effects with college- or Facebook-expansion-group-level characteristics that are correlated with Facebook roll-out timing. Column (1) controls for survey-wave fixed effects interacted with a variable that computes, at the college level, the pre-period average of the index of poor mental health. If a college does not appear in the pre-period, that college is assigned the average value of the variable across all colleges in the same Facebook expansion group that do appear in the pre-period. Column (2) controls for survey-wave fixed effects interacted with college region fixed effects (Northeast, Midwest, West, South). Finally, column (3) controls for survey-wave fixed effects interacted with a summary variable of selectivity computed at the Facebook expansion group level. The variable consists of the first factor predicted from a factor analysis of the following variables: whether the college is four-year, whether it is public, whether it offers doctoral, graduate, or medical degrees, whether it has a tenure system, whether it is a land grant college, and whether its undergraduate program is highly or medium selective. Notice we cannot construct a selectivity measure at the college level, because all college level variables other than geographic region were stripped away from the NCHA dataset for privacy reasons. The outcome variable is our index of poor mental health. The index is standardized so that, in the pre-period, it has a mean of zero and a standard deviation of one. All estimates are obtained using a specification that includes 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, and control variables, see Appendix Table A.19. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.11: Alternative Difference-in-differences Estimator

	Point Estimate	Standard Error	*p*-value
TWFE Estimate	0.085	0.033	0.010
DCDH Estimate	0.073	0.034	0.032
DCDH Placebo Estimate t-1	-0.002	0.037	0.956
DCDH Placebo Estimate t-2	-0.030	0.041	0.470
DCDH Placebo Estimate t-3	0.026	0.034	0.441
DCDH Placebo Estimate t-4	-0.002	0.030	0.952

Notes: This table presents robustness to using the alternative difference-in-differences estimator by De Chaisemartin and d'Haultfoeuille (2020). The outcome variable is always our index of poor mental health. The index is standardized so that, in the pre-period, it has a mean of zero and a standard deviation of one. The first row shows our baseline two-way-fixed-effect (TWFE) estimates, standard errors, and *p*-values. The estimates are obtained using our preferred specification, namely the version of Equation (1) that includes 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, and control variables, see Appendix Table A.19. The second row shows estimates, standard errors, and *p*-values obtained using the estimator suggested in De Chaisemartin and d'Haultfoeuille (2020). We refer to that estimator as DCDH. See De Chaisemartin and d'Haultfoeuille (2020) for a detailed description of how the estimator is constructed and why it is robust to treatment effects being heterogeneous across time and treated units. Rows 3 through 6 display the DCDH placebo estimates. The placebo estimates are constructed assuming that, for units whose treatment actually happens at t, treatment occurred at time $t-k$ for $k \in \{1,2,3,4\}$.

Table A.12: Addressing Alternative Explanations

	Index of Missing Values (1)	Any Missing Values (2)	Total Missing Values (3)	Extensive Margin (4)
Post Facebook Introduction	0.010	0.003	0.014	0.051*
	(0.049)	(0.008)	(0.067)	(0.031)
Baseline mean	0	0.07	0.27	0
Observations	380,886	380,886	380,886	380,886
Survey Wave FE	✓	✓	✓	✓
College FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓

Notes: This table addresses alternative explanations for our results, such as a reduction in the stigma associated with mental illness as a result of the introduction of Facebook or increased ability to recall instances of poor mental health as a result of the introduction of Facebook. Specifically, it presents estimates of coefficient β from Equation (1). In column (1), the outcome is an index of missing values. The index is constructed by summing the number of missing values among all variables composing our index of poor mental health, and then standardizing the final outcome. In column (2), the outcome is an indicator equal to one if a respondent did not answer at least one question composing the index of poor mental health, and equal to zero otherwise. In column (3), the outcome is the total number of questions composing the index of poor mental health left unanswered by a respondent. In column (4), the outcome is an index that takes into account only extensive margin responses. The index is constructed by: first, converting each variable in our index of poor mental health into a binary variable that takes value 0 if and only if the respondent reported not experiencing any issue related to that variable; second, averaging the newly constructed binary variables; third, standardizing the final outcome. All estimates 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, and control variables, see Appendix Table A.19. Standard errors in parentheses are clustered at the college level.* p<0.1, ** p<0.05, *** p<0.01.

Table A.13: Heterogeneous Effects on Perceptions of Alcohol Use

	Typical drink count (1)	Share used (2)	Typical student used daily (3)	Index std. dev. (4)
Post Facebook Introduction	0.121	0.020***	0.036***	0.105***
	(0.076)	(0.005)	(0.012)	(0.032)
Post Facebook Introduction ×	0.094**	0.001	0.020***	0.041***
Off-Campus Living	(0.038)	(0.002)	(0.006)	(0.015)
Observations	374,041	369,422	377,503	379,864

Notes: This figure explores whether the effects of the introduction of Facebook on perceptions of alcohol use are heterogeneous depending on whether the respondent lives off-campus. Specifically, it presents estimates from a version of Equation (1) in which our treatment indicator is interacted with living off-campus. The outcome variables are the perceived number drinks a typical student had the last time she partied, the perceived percent of students who used alcohol in the last 30 days, perceptions about whether a typical student in the school uses alcohol daily, and a standardized index of the three outcomes. All estimates 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, and control variables, see Appendix Table A.19. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.14: Effect on Differences Between Perceived and Reported Alcohol Use

	Difference drink count	Difference share used	Typical Student incorrect	Index std. dev.
	(1)	(2)	(3)	(4)
Post Facebook Introduction	-0.010	-0.003	0.066***	0.055
	(0.057)	(0.005)	(0.020)	(0.048)
Baseline mean	2.20	0.15	0.45	-0.00
Observations	375,025	370,390	377,869	380,886
Survey Wave FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
College FE	✓	✓	✓	✓

Notes: This figure explores the effects of the introduction of Facebook on the differences between perceptions of alcohol use and actual use. Specifically, it presents estimates of coefficient β from Equation (1). All columns are in original units, besides column (4) which is an index of the outcomes in columns (1) through (3). All indices are standardized so that, in the pre-period, they have a mean of zero and a standard deviation of one. All estimates 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, and control variables, see Appendix Table A.19. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.15: Effect on Outcomes related to Disruptive Internet Use

	Internet, computer games experienced	Internet, computer games academics
	(1)	(2)
Post Facebook Introduction	0.023	0.004
	(0.016)	(0.009)
Baseline mean	0.52	0.11
Observations	375,263	375,263
Survey Wave FE	✓	✓
Controls	✓	✓
College FE	✓	✓

Notes: This table explores the effects of the introduction of Facebook at a college on outcomes related to disruptive internet use. Specifically, it presents estimates of coefficient β from Equation (1). In column (1), the outcome is whether a student experienced the internet/video games as an issue; in column (2), the outcome is whether the issue affected the student's academic performance. The outcome variables are in original units. All estimates 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, and control variables, see Appendix Table A.19. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.16: Effect of the Introduction of Facebook on Assaults and Sexual Assaults

	Assault, fight last year	Sexual assault last year	Sexual threat last year	Abusive relationship last year
	(1)	(2)	(3)	(4)
Post Facebook Introduction	0.002	-0.006	0.001	0.005
	(0.008)	(0.008)	(0.004)	(0.006)
Baseline mean	0.15	0.15	0.04	0.15
Observations	380,809	380,803	379,916	379,539
Survey Wave FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
College FE	✓	✓	✓	✓

Notes: This table explores the effects of the introduction of Facebook at a college on assaults and sexual violence. 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 relate to various dimensions of physical and sexual violence. 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, and control variables, see Appendix Table A.19. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.17: Effect of the Introduction of Facebook on Relationships

	Straight	Single	Relationship difficulties experienced	Partners number
	(1)	(2)	(3)	(4)
Post Facebook Introduction	0.000	-0.005	0.015	0.053
	(0.005)	(0.009)	(0.014)	(0.032)
Baseline mean	0.95	0.58	0.46	1.40
Observations	376,505	377,078	375,278	376,118
Survey Wave FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
College FE	✓	✓	✓	✓

Notes: This table explores the effects of the introduction of Facebook at a college on relationships. 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 relate to various dimensions of romantic relationships or sexual orientation. 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, and control variables, see Appendix Table A.19. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.18: Effect of the Introduction of Facebook on Drug Use

	Cigarettes	Cigars	Smokeless tobacco	Marijuana	Cocaine	Amphetamines	Rohypnol	MDMA	Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Post Facebook Introduction	0.009	0.002	0.001	0.010	-0.000	-0.003	0.000	-0.001	0.006
	(0.009)	(0.004)	(0.006)	(0.007)	(0.003)	(0.003)	(0.001)	(0.001)	(0.005)
Survey Wave FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
College FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baseline mean	0.25	0.07	0.04	0.18	0.02	0.05	0.00	0.00	0.04
Observations	379,708	379,002	376,399	378,805	379,157	379,257	379,160	243,555	367,087

Notes: This table explores the effects of the introduction of Facebook at a college on drug use. 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 relate to the use of various drugs. 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, and control variables, see Appendix Table A.19. Standard errors in parentheses are clustered at the college level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A.19: Variables definitions, constructions, and associated NCHA survey questions

Variable	Description
Treatment Variables	
Post Facebook Introduction	Coding: 1 = Facebook was available at the respondent's college at the time she took the survey; 0 = Facebook was not available at the respondent's college at the time she took the survey; . = Impossible to determine whether Facebook was available at the respondent's college at the time she took the survey, because the semester in which the respondent took the survey coincides with the semester in which Facebook was introduced at her college.
Number of semesters exposure	Number of semesters that a student might have been exposed to Facebook given: i) the college the student goes to, ii) the survey wave the student participated in, and iii) the year in which the student started college.
Main Indices	
Index Poor Mental Health	The index is constructed as follows: i) we standardized all variables related to *symptoms of poor mental health* (see below) and all variables related to *depression services* (see below) so that they have a mean of 0 and a standard deviation of 1 in the pre-period; ii) we took an equally-weighted average of the standardized variables; iii) we re-standardized the equally-weighted average so that it has a mean of 0 and a standard deviation of 1 in the pre-period.
Index Symptoms Poor Mental Health	The index is constructed as follows: i) we standardized all variables related to *symptoms of poor mental health* (see below) so that they have a mean of 0 and a standard deviation of 1 in the pre-period; ii) we took an equally-weighted average of the standardized variables; iii) we re-standardized the equally-weighted average so that it has a mean of 0 and a standard deviation of 1 in the pre-period.
Index Depression Services	The index is constructed as follows: i) we standardized all variables related to *depression services* (see below) so that they have a mean of 0 and a standard deviation of 1 in the pre-period; ii) we took an equally-weighted average of the standardized variables; iii) we re-standardized the equally-weighted average so that it has a mean of 0 and a standard deviation of 1 in the pre-period.
Index Symptoms Depression	The index is constructed as follows: i) we standardized all variables related to *symptoms of depression* (see below) so that they have a mean of 0 and a standard deviation of 1 in the pre-period; ii) we took an equally-weighted average of the standardized variables; iii) we re-standardized the equally-weighted average so that it has a mean of 0 and a standard deviation of 1 in the pre-period.
Index Symptoms Other Conditions	The index is constructed as follows: i) we standardized all variables related to *symptoms of other conditions* (see below) so that they have a mean of 0 and a standard deviation of 1 in the pre-period; ii) we took an equally-weighted average of the standardized variables; iii) we re-standardized the equally-weighted average so that it has a mean of 0 and a standard deviation of 1 in the pre-period.
Index Downstream Effects	The index is constructed as follows: i) we standardized all variables related to *downstream effects of poor mental health* (see below) so that they have a mean of 0 and a standard deviation of 1 in the pre-period; ii) we took an equally-weighted average of the standardized variables; iii) we re-standardized the equally-weighted average so that it has a mean of 0 and a standard deviation of 1 in the pre-period.

Table A.19 (cont'd): Variable definition, construction, and associated NCHA survey questions

Variable	Description
Symptoms of Poor Mental Health	
Symptoms of Depression	
Last year felt hopeless	Question: "Within the last school year how many times have you: Felt things were hopeless"; Scale: 1 = never; 2 = 1-2 times; 3 = 3-4 times; 4 = 5-6 times; 5 = 7-8 times; 6 = 9-10 times; 7 = 11 or more times.
Last year felt overwhelmed	Question: "Within the last school year how many times have you: Felt overwhelmed by all you had to do"; Scale: same as above.
Last year felt exhausted	Question: "Within the last school year how many times have you: Felt exhausted (not from physical activity)"; Scale: same as above.
Last year felt very sad	Question: "Within the last school year how many times have you: Felt very sad"; Scale: same as above.
Last year severely depressed	Question: "Within the last school year how many times have you: Felt so depressed that it was difficult to function"; Scale: same as above.
Last year seriously considered suicide	Question: "Within the last school year how many times have you: Seriously considered attempting suicide"; Scale: same as above.
Last year attempted suicide	Question: "Within the last school year how many times have you: Attempted suicide"; Scale: same as above.
Last year depression	Question: "Within the last school year, have you had any of the following?: Depression"; Scale: 1 = yes; 0 = no.
Symptoms of Other Conditions	
Last year anorexia	Question: "Within the last school year, have you had any of the following?: Anorexia"; Scale: 1 = yes; 0 = no.
Last year anxiety disorder	Question: "Within the last school year, have you had any of the following?: Anxiety disorder"; Scale: 1 = yes; 0 = no.
Last year bulimia	Question: "Within the last school year, have you had any of the following?: Bulimia"; Scale: 1 = yes; 0 = no.
Last year seasonal affect disorder	Question: "Within the last school year, have you had any of the following?: Seasonal Affect Disorder"; Scale: 1 = yes; 0 = no.
Depression Services	
Last year depression diagnosis	Question: "Have you been diagnosed with depression within the last school year?"; Scale: 1 = yes; 0 = no. Coding: the question is asked only to individuals who answered affirmatively to a previous question asking whether they had ever been diagnosed with depression. We impute a value of 0 for all individuals who reported never having been diagnosed with depression and who, therefore, are not asked the question about being diagnosed with depression in the last school year. See Section 4.1 for a discussion about the imputation.
Therapy depression	Question: "Are you currently in therapy for depression?"; Scale: 1 = yes; 0 = no. Coding: the question is asked only to individuals who answered affirmatively to a previous question asking whether they had ever been diagnosed with depression. We impute a value of 0 for all individuals who reported never having been diagnosed with depression and who, therefore, are not asked the question about being in therapy for depression. See Section 4.1 for a discussion about the imputation.
Current medication depression	Question: "Are you currently taking medication for depression?"; Scale: 1 = yes; 0 = no. Coding: the question is asked only to individuals who answered affirmatively to a previous question asking whether they had ever been diagnosed with depression. We impute a value of 0 for all individuals who reported never having been diagnosed with depression and who, therefore, are not asked the question about being in taking medication for depression. See Section 4.1 for a discussion about the imputation.

Table A.19 (cont'd): Variable definition, construction, and associated NCHA survey questions

Variable	Description
Drinking Perceptions and Behaviors	
Perceptions	
Typical drink count	Question: "How many alcoholic drinks do you think the typical student at your school had the last time he/she partied/socialized?" Open numeric response. Coding: Winsorized at 9
Share used, 30 days	Question: "Within the last 30 days, what percent of students at your school used Alcohol? State your best estimate." Open numeric response.
Typical student used daily	Question: "Within the last 30 days, how often do you think the typical student at your school used alcohol (beer, wine, liquor)?" Coding: 1 = Used daily; 0 = {Never Used, One or more days}.
Perceptions Index	The index is constructed as follows: i) we standardized the three variables above so that they have a mean of 0 and a standard deviation of 1 in the pre-period. ii) we took an equally-weighted average of the standardized variables. iii) we re-standardized the equally-weighted average so that it has a mean of 0 and a standard deviation of 1 in the pre-period.
Usage	
Drink count	Question: "The last time you partied/socialized, how many alcoholic drinks did you have? State your best estimate." Open numeric response. Coding: Winsorized at 9
Used 30 days	Question: "Within the last 30 days, on how many days did you use alcohol (beer, wine, liquor)?" Coding: 1 = {1–2 days; 3–5 days; 6–9 days; 10–19 days; 20–29 days; All 30 days}; 0={Never used; Have used, but not in last 30 days}
Used daily	Question: "Within the last 30 days, on how many days did you use alcohol (beer, wine, liquor)?" Coding: 1 = {20–29 days; All 30 days}; 0={1–2 days; 3–5 days; 6–9 days; 10–19 days; Never used; Have used, but not in last 30 days}
Usage Index	The index is constructed as follows: i) we standardized the three variables above so that they have a mean of 0 and a standard deviation of 1 in the pre-period. ii) we took an equally-weighted average of the standardized variables. iii) we re-standardized the equally-weighted average so that it has a mean of 0 and a standard deviation of 1 in the pre-period.
Differences between perceptions and usage	
Difference drink count	Absolute value of the difference between the typical drink count variable and the average drink count in the same university and survey-wave. The average drink count variable is constructed using the drink count variable described above.
Difference share used	Absolute value of the difference between the share used, 30 days variable and the share of respondents in the same university and survey-wave who reported using alcohol at least once in the last 30 days. The share of respondents using alcohol at least once in the last 30 days is constructed using the used daily variable described above.
Typical student incorrect	Binary variable indicating whether the typical student used daily response does not equal the modal value of the used daily variable in the university and survey wave of the respondent.
Difference Index	The index is constructed as follows: i) we standardized the three variables above so that they have a mean of 0 and a standard deviation of 1 in the pre-period. ii) we took an equally-weighted average of the standardized variables. iii) we re-standardized the equally-weighted average so that it has a mean of 0 and a standard deviation of 1 in the pre-period.

Table A.19 (cont'd): Variable definition, construction, and associated NCHA survey questions

Variable	Description
Downstream Effects	
Academic perform attention deficit	Question: "Within the last school year, have any of the following affected your academic performance?: Attention Deficit Disorder"; Scale: 1 = {Received a lower grade on an exam or important project; Received a lower grade in the course; Received an incomplete or dropped the course.}; 0 = {This did not happen to me/not applicable; I have experienced this issue but my academics have not been affected}.
Academic perform depression	Question: "Within the last school year, have any of the following affected your academic performance?: Depression/Anxiety Disorder/Seasonal Affective Disorder"; Scale: same as above.
Academic perform eating disorder	Question: "Within the last school year, have any of the following affected your academic performance?: Eating disorder/problem"; Scale: same as above.
Academic perform sleep difficulty	Question: "Within the last school year, have any of the following affected your academic performance?: Sleep difficulty"; Scale: same as above.
Academic perform stress	Question: "Within the last school year, have any of the following affected your academic performance?: Stress"; Scale: same as above.
Social Comparisons Moderators	
Off-campus living	Question: "Where do you currently live?"; Coding: 1 = {Off-campus housing, Parent/guardian's home, Other}; 0 = {Campus residence hall, Fraternity or sorority house, Other university/college housing}.
Not in fraternity/sorority	Question: "Are you a member of a social fraternity or sorority?"; Scale: 1=yes; 0=no.
Credit-card debt	Question: "If you have a credit card(s) how much total credit card debt did you carry last month? That is, what was the total unpaid balance on all of your cards (that you are responsible for paying)?"; Coding: 1 if reported debt is at least $1; 0 otherwise.
Work	Question: "How many hours a week do you work for pay?"; Coding: 1 = more than 1 hour; 0 = 0 hours.
Overweight	Use recoded BMI (BMI $= kg/m^2$); Coding: 1 = if recoded BMI>25 (indicating overweight or obesity); 0 otherwise.
Index of Social Comparisons	Coding: Index sums the binary variables defined above. As an additional moderator to study heterogeneous treatment effects, we consider whether a respondent is above the median value of the index of social comparisons or below the median value.
Disruptive Internet Use	
Internet, computer games experienced	Question: "Within the last school year, have any of the following affected your academic performance? Internet use/computer games." Coding: 1 = {I have experienced this issue but my academics have not been affected; Received a lower grade on an exam or important project; Received a lower grade in the course; Received an incomplete or dropped the course.}; 0 = {This did not happen to me/not applicable}.
Internet, computer games academics	Question: "Within the last school year, have any of the following affected your academic performance? Internet use/computer games." Coding: 1 = {Received a lower grade on an exam or important project; Received a lower grade in the course; Received an incomplete or dropped the course.}; 0 = {This did not happen to me/not applicable; I have experienced this issue but my academics have not been affected}.

Table A.19 (cont'd): Variable definition, construction, and associated NCHA survey questions

Variable	Description
Other Behaviors and Perceptions	
Assaults and Sexual Assaults	
Assault, fight last year	Questions: "Within the last school year, were you: in a physical fight?", "Within the last school year, were you: physically assaulted?" Scale: yes, no. Coding: 1 = answering yes to either of the two questions; 0 = otherwise.
Sexual assault last year	Questions: "Within the last school year, have you experienced: sexual touching against your will?", "Within the last school year, have you experienced: attempted sexual penetration against your will?", "Within the last school year, have you experienced: sexual penetration against your will?" Scale: yes, no. Coding: 1 = answering yes to at least one of the three questions; 0 = otherwise.
Sexual threat last year	Question: "Within the last school year, have you experienced: verbal threats for sex against your will?" Scale: yes, no. Coding: 1 = yes, 0 = no.
Abusive relationship	Question: "Within the last school year, have you been in a relationship that was: sexually abusive?" Scale: yes, no. Coding: 1 = yes, 0 = no.
Relationships	
Straight	Question: "Which of the following best describes you?" Coding: 1 = {Heterosexual}; 0 = {Gay/Lesbian, Bisexual, Transgender, Unsure}.
Single	Question: "What is your current relationship status?" Coding: 1 = {Single}; 0 = {Married/domestic partner, Engaged or committed dating relationship, Separated, Divorced, Widowed}.
Relationship difficulties experienced	Question: "Within the last school year, have any of the following affected your academic performance? Relationship difficulty." Coding: 1 = {I have experienced this issue but my academics have not been affected; Received a lower grade on an exam or important project; Received a lower grade in the course; Received an incomplete or dropped the course.}; 0 = {This did not happen to me/not applicable}
Partners number	Question: "Within the last school year, with how many partners, if any, have you had sex (oral, vaginal, or anal)?" Open numeric response. Coding: Winsorized at 9.
Drug use	
Cigarettes	Question: "Within the last 30 days, on how many days did you use: cigarettes?" Scale: 1 = never used; 2 = have used, but not in last 30 days; 3 = 1-2 days; 4 = 3-5 days; 5 = 6-9 days; 6 = 10-19 days; 7 = 20-29 days; 8 = all 30 days.
Cigars	Question: "Within the last 30 days, on how many days did you use: cigars?" Scale: same as above.
Smokeless tobacco	Question: "Within the last 30 days, on how many days did you use: smokeless tobacco?" Scale: same as above.
Marijuana	Question: "Within the last 30 days, on how many days did you use: marijuana (pot, hash, hash oil)?" Scale: same as above.
Cocaine	Question: "Within the last 30 days, on how many days did you use: cocaine (crack, rock, freebase)?" Scale: same as above.
Amphetamines	Question: "Within the last 30 days, on how many days did you use: amphetamines (diet pills, speed, meth, crank)?" Scale: same as above.
Rohypnol	Question: "Within the last 30 days, on how many days did you use: rohypnol (roofies), GHB, or Liquid X (intentional use)?" Scale: same as above.
MDMA	Question: "Within the last 30 days, on how many days did you use: MDMA (Ecstasy, XTC, E, X, Adam)?" Scale: same as above.
Other	Question: "Within the last 30 days, on how many days did you use: other drugs?" Scale: same as above.

Table A.19 (cont'd): Variable definition, construction, and associated NCHA survey questions

Variable	Description
Physical Health	
Index poor physical health	The index is based on the following question: "Within the last school year, have you had any of the following?" The physical health conditions are: allergy, asthma, chronic fatigue, diabetes, endometriosis, genital herpes, genital warts, hepathites B or C, high blood pressure, high cholesterol, HIV, carpal tunnel, back pain, broken bones, bronchitis, chlamydia, ear infection, gonorrhea, mono, pelvic inflamation, sinus infection, strep, tuberculosis. The answer options are yes and no. The index is constructed as follows: i) we standardized all the variables above so that they have a mean of 0 and a standard deviation of 1 in the pre-period. ii) we took an equally-weighted average of the standardized variables. iii) we re-standardized the equally-weighted average so that it has a mean of 0 and a standard deviation of 1 in the pre-period.
Control variables	
Female	Question: "What is your sex?"; Coding: 1 = female; 0 = male
White	Question: "How do you usually describe yourself? (Mark all that apply)"; Coding: 1 if chose "White-not Hispanic (includes Middle Eastern)"; 0 otherwise.
Black	Question: "How do you usually describe yourself? (Mark all that apply)"; Coding: 1 if chose "Black-not Hispanic"; 0 otherwise.
Hispanic	Question: "How do you usually describe yourself? (Mark all that apply)"; Coding: 1 if chose "Hispanic or Latino"; 0 otherwise.
Asian	Question: "How do you usually describe yourself? (Mark all that apply)"; Coding: 1 if chose "Asian or Pacific Islander"; 0 otherwise.
Native American	Question: "How do you usually describe yourself? (Mark all that apply)"; Coding: 1 if chose "American Indian or Alaskan Native"; 0 otherwise.
Other race	Question: "How do you usually describe yourself? (Mark all that apply)"; Coding: 1 if chose "Other"; 0 otherwise.
International	Question: "Are you an international student?"; Scale: 1 = yes; 0 = no.
Age	Question: "How old are you?". Used in regression as separate indicators.
Year in school	Question: "Year in school"; Scale: 1 = 1st year undergraduate; 2 = 2nd year undergraduate; 3 = 3rd year undergraduate; 4 = 4th year undergraduate; 5 = 5th year or more undergraduate. Used in regression as separate indicators.
Region	Macro-region of a college: Northeast, Midwest, South, or West; used in regressions as four separate indicators.
Missing Values Variables	
Index of missing values	The index is constructed as follows: i) we considered all variables that comprise the index of poor mental health. ii) we assinged a value of 1 to a variable if the answer is missing and 0 otherwise. iii) we standardized the newly constructed variables so that they have a mean of 0 and a standard deviation of 1 in the pre-period. iv) we took an equally-weighted average of the standardized variables. v) we re-standardized the equally-weighted average so that it has a mean of 0 and a standard deviation of 1 in the pre-period.
Any missing values	1 = respondent left unanswered at least one question composing the index of poor mental health; 0 = respondent answered all of the questions that compose the index of poor mental health.
Total missing values	The number of questions composing the index of poor mental health that a respondent left unanswered.

Table A.19 (cont'd): Variable definition, construction, and associated NCHA survey questions

Variable	Description
Other variables	
Index extensive margin	The index is constructed as follows: i) we considered all variables that comprise the index of poor mental health. ii) we binarized the variables by assigning them a value of 1 whenever the original variable takes value greater than or equal to 1 and 0 otherwise. iii) we standardized the newly constructed variables so that they have a mean of 0 and a standard deviation of 1 in the pre-period. iv) we took an equally-weighted average of the standardized variables. v) we re-standardized the equally-weighted average so that it has a mean of 0 and a standard deviation of 1 in the pre-period.
Predicted susceptibility to mental illness	The variable is constructed as follows: i) we constructed an indicator that takes value one if and only if a student has ever been diagnosed with a mental health condition. ii) we considered a set of immutable individual-level characteristics (age, year in school, gender, race, an indicator for U.S. citizenship and height). iii) we generated all two-way interactions between the characteristics, and generated second- and third-order monomials of each characteristic. iv), we implemented a LASSO procedure in the pre-period to predict our indicator for ever having been diagnosed with a mental health condition using the immutable individual-level characteristics and functions thereof described above. v) we used the model selected by the Extended Bayesian Information Criterion (EBIC) to generate a prediction of our indicator for ever having been diagnosed with a mental health condition.
First-year	Question: "Year in school"; Coding: 1 if chose first year undergraduate; 0 otherwise.