Social Media and Mental Health*

Luca Braghieri, Ro'ee Levy, and Alexey Makarin

April 2022

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 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. We also find that, among students predicted to be most susceptible to mental illness, the introduction of Facebook led to increased utilization of mental healthcare services. Lastly, we find that, after the introduction of Facebook, students were more likely to report experiencing impairments to academic performance resulting from poor mental health. Additional evidence on mechanisms suggests that the results are due to Facebook fostering unfavorable social comparisons.

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

*Braghieri: Ludwig Maximilian University of Munich, CEPR, and CESifo. Email: luca.braghieri@econ.lmu.de. Levy: Tel Aviv University. Email: roeelevy@tauex.tau.ac.il. 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 the Editor, three anonymous referees, Levi Boxell, Davide Cantoni, Daniel Chen, Scott Cunningham, Georgy Egorov, Ruben Enikolopov, Amy Finkelstein, Matthew Gentzkow, Luigi Guiso, Martin Mattson, Jack Mountjoy, Samuel Norris, Petra Persson, Maria Petrova, Andrea Prat, Maya Rossin-Slater, Frank Schilbach, Sebastian Schweighofer-Kodritsch, Joseph Shapiro, Andrey Simonov, and seminar participants at EIEF, LMU, MIT, NNH Norwegian School of Economics, TSE, UPF, the Italian National Health Institute (Istituto Superiore di Sanità), the Southern Economic Association, and the Center for Rationality and Competition for helpful comments. Luca Braghieri gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119). Ro'ee Levy gratefully acknowledges financial support from the Foerder Institute for Economic Research. We thank Juan Carlos Cisneros, Valerio Sergio Castaldo, Gleb Kozlyakov, and Meruyert Tatkeyeva for excellent research assistance. Disclosures: Levy is an unpaid member of Facebook's 2020 Election Research Project.

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). For instance, the total number of individuals aged 18–23 who reported experiencing a major depressive episode in the past year increased by 83% between 2008 and 2018 (NSDUH, 2019). Similarly, over the same time period, suicides became more prevalent and are now the second leading cause of death for individuals 15–24 years old (National Center for Health Statistics, 2021). Although the ultimate causes of these trends are still largely unknown, scholars have hypothesized that the diffusion of social media might be an important contributing factor (Twenge et al., 2019). In fact, concerns about a potential negative effect of social media on mental health have become so prominent that the U.S. Senate held a committee hearing about the topic in late 2021 (Wells et al., 2021). Well-identified causal evidence, however, remains scarce.

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

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

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

Our analysis relies on a generalized difference-in-differences research design, where one of the dimensions of variation is the college a student attends, and the other dimension is whether the student took the survey before or after the introduction of Facebook at her college. Under a parallel trends assumption, the college by survey-wave variation generated by the sharp but staggered introduction of Facebook allows us to obtain causal estimates of the introduction of Facebook on student mental health. Our empirical strategy allows us to rule out various confounding factors: first, college-specific differences fixed in time (e.g., students at more academically demanding colleges may have worse baseline mental health than students at less demanding colleges); second, differences across time that affect all students in a similar way (e.g., certain macro-economic fluctuations); third, mental health trends affecting colleges in different Facebook expansion groups differentially, but smoothly (e.g., colleges where Facebook was rolled out earlier may be on different linear trends in terms of mental health than colleges where Facebook was rolled out later).[1] We also address recent econometric concerns with staggered difference-in-differences research designs by showing robustness to the use of a variety of alternative estimators.[2] Lastly, we complement the difference-in-differences strategy with a specification that exploits variation in length of exposure to Facebook across students within a college and survey wave, and that, therefore, does not rely on our baseline college-level parallel trends assumption for identification.

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

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

Our main finding is that the introduction of Facebook at a college had a negative effect on student mental health. Our index of poor mental health, which aggregates all the relevant mental health variables in the NCHA survey, increased by 0.085 standard deviation units as a result of the introduction of Facebook. As a point of comparison, this magnitude is around 22% of the effect of losing one's job on mental health, as reported in a meta-analysis by Paul and Moser (2009). We further benchmark our results against a clinically-validated depression scale: the PHQ-9 (Kroenke et al., 2001). The effect of the introduction of Facebook on our index of poor mental health is equivalent to a two-percentage-point increase in the share of students suffering from depression according to the PHQ-9 over a baseline of 25%. The negative effects on mental health are strongest for students who, based on immutable characteristics such as gender and age, are predicted to be most susceptible to mental illness. For those students, we also observe a significant increase in depression diagnoses, take-up of psychotherapy for depression, and use of anti-depressants. Finally, we find that, after the introduction of Facebook, students were more likely to report that their academic performance was negatively affected by conditions related to poor mental health.

What explains the negative effects of Facebook on mental health? The pattern of results is consistent with Facebook increasing students' ability to engage in unfavorable social comparisons. Two main pieces of evidence bear on this conclusion. First, we find that the effects are particularly pronounced for students who 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 of lower socioeconomic status, and students not belonging to fraternities/sororities. Second, we show that the introduction of Facebook directly affected the students' beliefs about their peers' social lives and behaviors, 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 are due to disruptive internet use. We also rule out several alternative mechanisms, such as reduced stigma about mental illness and direct effects on drug use, alcohol consumption, and sexual behaviors.

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

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

This paper contributes to the literature by providing the most comprehensive causal evidence to date on the effects of social media on mental health. The three closest papers to ours—Allcott et al. (2020, 2021), and Mosquera et al. (2020)—feature experiments that incentivize a randomly-selected subset of participants to reduce their social media use.[5] Those studies find negative effects of social media use on self-reported well-being, and Allcott 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 the experimental papers above. Specifically, our outcome variables include eleven questions related to depression—covering symptoms, diagnoses, take-up of psychotherapy, and use of anti-depressants—and various questions related to other mental health conditions, ranging from seasonal affect disorder to anorexia. By contrast, the three ex-

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

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

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

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

This paper also relates to the rapidly-growing literature in economics about the determinants and consequences of mental illness (Ridley et al., 2020). Research on the determinants of mental illness showed that unconditional cash transfers, in-utero exposure to the death of a maternal relative, unemployment shocks, and economic downturns can affect mental health (Paul and Moser, 2009; Haushofer and Shapiro, 2016; Persson and Rossin-Slater, 2018; Golberstein 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

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

[7]In the case of the experiments listed above, subjects in the treatment group are paid to reduce their social media use and are therefore not blind to treatment status. 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. Finally, social media experiments often screen participants who do not meet certain criteria and, therefore, employ rather selected samples. For instance, the main sample analyzed in Allcott et al. (2020) includes participants who reported using Facebook more than 15 minutes per day and were willing to accept $102 to deactivate their Facebook accounts for a month.

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, after the introduction of Facebook, students were more likely to report experiencing impairments to academic performance as a result of poor mental health.

Lastly, this paper contributes to an emerging literature exploiting the expansion of 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 the social media platform VK in Russia and of Facebook worldwide, respectively, to show that social media use increases protest participation. Bursztyn et al. (2019) and Müller and Schwarz (2020) exploit the expansion of VK and Twitter, respectively, and find that social media use increases the prevalence of hate crimes.[8] A unique feature of our setting is that it allows us to measure the effects of the sharp roll-out of the biggest social media platform in the world at a time in which very few close substitutes were available.

The remainder of the paper is organized as follows: Section 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.

2 Background

Mental Health Mental illnesses, such as depression, anxiety, bipolar disorder, and schizophrenia, are disturbingly common and can be highly debilitating. According to the Global Burden of Disease Study, around a billion people in the world suffered from mental disorders in 2017, with depression and anxiety-related disorders as the leading conditions (James et al., 2018). In the U.S., around 1 in 5 adults experiences some form of mental illness each year, and 1 in 20 experiences serious mental illness (NAMI, 2020). Mental health conditions can have severe

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

adverse effects, hampering people's ability to work, study, and be productive. According to the WHO's Global Burden of Disease, mental illness is the most burdensome disease category in terms of total disability-adjusted years for adults younger than 45, and depression is one of the most taxing conditions (WHO, 2008; Layard, 2017).

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

A Brief History of Facebook's Expansion and Initial Popularity Facebook was created at Harvard in February 2004 and was rolled out gradually to other colleges in the U.S. and abroad over the subsequent two-and-a-half years. The staggered nature of the roll-out was enforced by requiring users to be in possession of verified email addresses (e.g., addresses ending in @harvard.edu). The roll-out of Facebook across U.S. colleges was not random: as shown in the descriptive statistics section, more selective colleges were granted access to Facebook relatively earlier than less selective colleges. The staggered nature of the expansion is arguably due to three factors: first, scale constraints due to limited server capacity (Kirkpatrick, 2011); second, Facebook's willingness, at least at the outset, to maintain a flavor of exclusivity; third, Facebook's desire to strengthen network effects by keeping the fraction of users who likely knew each other offline artificially high (Aral, 2021).

Even in its infancy, Facebook was extremely popular. Upon being granted access to the platform, colleges witnessed rapid and very widespread adoption among students.[9] To get a sense of the early adoption rates among college students, we matched data provided by Facebook on the number of users at each of the first 100 colleges that were granted access to the platform with IPEDS (Integrated Postsecondary Education Data System) data on the number

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

of full-time undergraduate students at those colleges (U.S. Department of Education, 2005; Traud et al., 2012). Appendix Figure A.2 presents a histogram of the number of users per 100 undergraduate students at those colleges and shows that, in September 2005, there were on average 86 Facebook users for every 100 undergraduate students. This result is consistent with Facebook's statement that, across all the colleges with access to the platform as of September 2005, approximately 85% of students had a Facebook profile (Arrington, 2005).[10]

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

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 Facebook's expansion.

Facebook Expansion Dates Data The Facebook Expansion Dates dataset was assembled in two steps: for the first 100 colleges that received Facebook access, we rely on introduction dates collected and made public in previous studies (Traud 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 that allows users to visit old versions of those websites. Specifically, between the spring of 2004 and the spring of 2005, the front page of Facebook's website was regularly

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

updated to show the most recent set of colleges that had been given access to the platform.[11] As an example, Appendix Figure A.3 shows the front page of Facebook as of June 15th 2004, recovered via the Wayback Machine. As shown in the figure, Facebook was open to 34 colleges at that point in time.

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

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

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. We emphasize that reliance on self-reported symptoms is part of standard medical practice in the domain of mental health (Chan, 2010). Specifically, according to the official diagnostic man-

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

ual of the American Psychiatric Association (DSM-V), the diagnosis of many mental health disorders including depression relies almost exclusively on patients' self-reports of symptoms such as difficulty sleeping, anhedonia, fatigue, feelings of worthlessness and guilt, diminished ability to think or concentrate, and recurrent suicidal thoughts (APA, 2013). In fact, self-administered questionnaires inquiring about depression symptoms have been shown to predict medical diagnoses with accuracy rates up to 90% (Kroenke and Spitzer, 2002).[12]

The NCHA dataset includes the universe of responses to all NCHA survey waves administered between the spring of 2000 and the spring of 2008, the longest stretch of time around Facebook's early expansion in which the content of the survey remained constant.[13] Colleges included in the NCHA dataset administered the survey to randomly selected classrooms, randomly selected students, or all students. The average response rate across the survey waves for which we have such information is 37% (ACHA, 2009). In order to assuage concerns about the possibility that the introduction of Facebook affected the composition of students who participated in the survey, Appendix Tables A.3 and A.9 show that, along the demographic characteristics elicited in the NCHA survey, there are no meaningful compositional changes following the introduction of Facebook. Throughout our analysis, we limit our sample to full-time undergraduate students.[14]

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

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

[12]Section 4, Appendix B, and Appendix C discuss our symptom measures in detail and present an array of exercises to validate them.

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

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

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

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.[16] Appendix Table A.1 shows that colleges in earlier Facebook expansion groups are more selective in terms of test scores, larger, more likely to be on the East Coast, and have more residential undergraduate programs than colleges in later Facebook expansion groups. Panel A of Appendix Table A.2, which averages student-level variables available in the NCHA dataset across the different Facebook expansion groups, shows that colleges in earlier Facebook expansion groups enroll students from relatively more advantaged economic backgrounds. Lastly, Panel B of Appendix Table A.2 shows that students in colleges that received Facebook relatively earlier have worse baseline mental health outcomes than students attending colleges in later Facebook expansion groups.[17] The baseline differences across Facebook expansion groups may lead one to wonder about the plausibility of the parallel trends assumption in this setting; we address concerns related to parallel trends in Section 4.

Appendix Table A.1 also shows the number of colleges in the NCHA dataset that received

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

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

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

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.

4 Empirical Strategy

Construction of the Primary Outcome Variables In order to mitigate concerns about cherry-picking outcome variables, we consider all the questions in the NCHA survey that are related to mental health and that inquire about a respondent's recent past (e.g., "Within the last school year, how many times have you felt so depressed that it was difficult to function?").

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

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

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

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

Appendix Table A.30 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.

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

We validate the NCHA survey questions externally by conducting an original survey on more than 500 college students. Our survey contained both the questions from the NCHA survey that feature in the construction of our index of poor mental health and the questions from canonical depression and generalized anxiety disorder screeners—the PHQ-9 and GAD-7 respectively—known to be highly predictive of medical diagnoses (Kroenke et al., 2001;

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

Spitzer et al., 2006). Appendix Figures A.14-A.15 show that our index of poor mental health is strongly correlated with the PHQ-9 and GAD-7 scores (correlation coefficients of 0.66 and 0.61 respectively). The validation exercise is described in detail in Appendix C.

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

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

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

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

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

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 on student mental health.

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

One may worry about the plausibility of the parallel trends assumption in our setting—that is, one might worry that colleges belonging to different Facebook expansion groups might be on different mental health trends. We address this concern in four ways. First, we estimate a fully dynamic version of Equation (1) and check for potential pre-trends. Second, we explore the existence of pre-trends by estimating a fully dynamic version of the alternative estimators introduced in Borusyak et al. (2021), Callaway and Sant'Anna (2021), De Chaisemartin and d'Haultfoeuille (2020), and Sun and Abraham (2021). Third, to the extent that the trends are linear, we would be able to account for them in a robustness check that includes expansion-group-level linear time trends. Fourth, we present results using a specification that does not rely on our baseline college-level parallel trends assumption. In particular, we present results using a specification that includes college$\times$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 homogeneity in treatment effects (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; 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.

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

5 Results

5.1 Baseline Results

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

The 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 and that does not include linear time trends, is 0.085 standard deviation units. The effect above is estimated on the entire population of students taking the NCHA survey, which includes both students who did and who did not sign-up for a Facebook account after Facebook was made available at their college. Therefore, the point estimate captures both the direct effect of Facebook on students who joined the platform and the indirect effect of Facebook on students who did not join the platform, but whose peers did.[19]

In order to help build intuition about the magnitude of our baseline effects, we provide a few benchmarks. First, the magnitude of our baseline effect corresponds to approximately 84% of the difference in the index of poor mental health between students in our sample with and without credit card debt. Second, we benchmark the magnitude of our estimates against the effect of a sudden unemployment spell on mental health. Comparing our estimates to the most closely related ones in a meta-analysis by Paul and Moser (2009), we find that the impact of introducing Facebook at a college on mental health is around 22% of the effect of job loss.[20] Lastly, we benchmark our results against the canonical PHQ-9 and GAD-7 mental health scales. We use data from the validation survey mentioned in Section 4 and discussed in detail in Appendix C to determine how to weigh the variables contained in our index of

[19]Although we cannot separate these two channels in the absence of data on an individual's Facebook use, we note that it is unlikely that our results are primarily driven by students who did not have a Facebook account. As discussed in Section 2, the average penetration rate of Facebook at each college was around 85%. Therefore, an effect concentrated solely among students who did not join the platform would have to be implausibly large—approximately 0.57 standard deviations in our main specification—to be consistent with our baseline effect.

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

poor mental health in a way that best predicts an indicator for having depression according to the PHQ-9 and an indicator for having generalized anxiety disorder according to the GAD-7. Next, we apply these weights to the NCHA sample to predict whether a student taking the NCHA survey would be classified as having depression or generalized anxiety disorder according to the PHQ-9 and GAD-7. Appendix Table A.29 shows that the introduction of Facebook increased by two percentage points the fraction of students whom, according to our prediction, the PHQ-9 and GAD-7 would classify as having depression or generalized anxiety disorder. The two percentage point increase corresponds to a 9% increase over the pre-period mean of 25% for depression and a 12% increase over the pre-period mean of 16% for generalized anxiety disorder.[21]

Figure 1 presents results on our individual outcome variables and shows that most of the dimensions of mental health in our dataset were negatively affected by the introduction of Facebook.[22] For all but one of the mental health outcomes from Figure 1, the point estimates are positive, which indicates worsened mental health. The conditions that appear to be most affected are depression and anxiety-related disorders, while the point estimates on anorexia and bulimia are close to zero.[23] The effect on severe depression is similar in magnitude to the effect observed in Allcott et al. (2020) on whether a respondent felt depressed in the past month (0.07 vs. 0.09 standard deviations, respectively). Such similarity is striking, especially in light of the fact that the time period, target population, and empirical strategy in Allcott et al. (2020) are different from the ones in this paper.

The bottom section of Figure 1 also presents suggestive evidence that the introduction of Facebook at a college might have increased the extent to which students took up depression-related services. For all three items comprising the index of depression services—receiving an official depression diagnosis, going to therapy for depression, and taking anti-depressants—the point estimates are positive, though not significant at conventional levels.[24] Finding a

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

[22] 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.

[23] Similar patterns can be observed in Figure A.5 which is a version of Figure 1 with expansion-group-specific linear trends.

[24] 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%

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.[25] In Section 5.2 below, we show that students who, based on immutable baseline characteristics, are predicted to be most susceptible to mental illness—and therefore more likely to be on the margin of receiving a depression diagnosis—are indeed significantly more likely to take up depression-related services after the introduction of Facebook.

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

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

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. When estimating the model using OLS, we treat students who took the survey in the semester just before Facebook was rolled out at their college as the omitted category and compare them to students who took the NCHA survey in other semesters.

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

Figure 2 presents the event-study figures and shows that the estimates are consistent with the parallel trends assumption: independently of the estimator used, the coefficients on the semesters prior to the introduction of Facebook at a college are all close to zero and exhibit no

and 20% over the baseline mean, respectively.

[25]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.

discernible pre-trends.[26] Figure 2 also sheds light on the dynamics of treatment effects: all the recently-developed robust estimators show treatment effects that increase over time in the post-periods.[27] The increase in treatment effects over time could be explained by: i) higher adoption rates at a college over time; ii) higher intensity of usage at the individual level over time; iii) the effects becoming stronger as a function of length of exposure to the platform. Given the evidence presented in Section 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 tentatively lean in favor of the length-of-exposure explanation. We further study the effects of differential length of exposure to Facebook at the individual level in Section 5.3

5.2 Heterogeneity

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

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

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

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

acteristics and functions thereof described above.

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

Armed with our LASSO prediction, we can study how the introduction of Facebook at a college affected students across the mental-illness-susceptibility spectrum, and whether it induced students who are more likely to be on the margin of a depression diagnosis to seek out depression-related services such as psychotherapy. The upper left panel of Figure 3 presents the estimated effects on the index of poor mental health across quintiles of our LASSO-predicted measure of susceptibility to mental illness.[28] As shown in the figure, the effects of the introduction of Facebook on symptoms of poor mental health tend to be stronger for individuals with a higher baseline risk of developing mental illness.[29]

The upper right panel of Figure 3 shows that the introduction of Facebook on the take-up of depression-related services exhibits a similar pattern. We find weak positive effects throughout the distribution of predicted susceptibility to mental illness, though for most quintiles the point estimates are fairly small and not statistically significant. The effects become more pronounced for individuals in the top quintile; in particular, the point estimate on the top quintile is relatively large in magnitude (0.063 standard deviations) and four times as large as the point estimate on the bottom quintile. As indicated in column (2) of Table A.5, the difference between the coefficients for the top and the bottom quintiles is significant at the 1% level. These results suggest that, indeed, students who are predicted to be most susceptible to mental illness—and therefore more likely to seek mental healthcare due to a worsening in symptoms—are more likely to take up depression-related services such as psychotherapy for depression and anti-

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

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

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

[29] We note that, for predicting baseline susceptibility to mental illness, the stock variable of 'having ever been diagnosed' with a mental illness is arguably more relevant than the flow variable of having exhibited a certain symptom in the past year, because the former captures information covering a longer time span. Appendix Figure A.10 examines robustness of our results to an alternative measure of susceptibility to mental illness based on a LASSO regression predicting whether a respondent's index of poor mental health is in the top 10% of the pre-period sample. The results are qualitatively similar.

depressants as a result of the introduction of Facebook.

Other dimensions of heterogeneity. Appendix Figure A.6 estimates heterogeneous effects across several baseline characteristics. Consistent with surveys showing that women use social media more often and are more likely to report using Facebook for longer than they intend, we find suggestive evidence that the results are larger among women (Thompson and Lougheed, 2012; Lin et al., 2016).[30] We also find stronger effects on non-Hispanic whites, and a weaker effect on international students, younger students, and first-years.

5.3 Effects Based on Length of Exposure to Facebook

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

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

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

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

[30] Furthermore, baseline prevalence of depression is found to be higher among women, across different nations, cultures, and age groups (Nolen-Hoeksema and Hilt, 2008; Albert, 2015; Salk et al., 2017). Thus, the slightly stronger effects among women are also consistent with studies showing that women are more likely to be affected by certain mental illnesses.

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.[31]

Figure 4 displays the estimates of β_k and shows that the negative effects of the introduction of Facebook on mental health worsen the longer students are exposed to Facebook. Appendix Table A.6 presents the results in a regression framework where we assume that the effects grow linearly over time. The table shows that the number of treated semesters has a significant effect on our main index, on symptoms of poor mental health, and on the utilization of depression-related healthcare services.

Since the index of depression services only comprises binary variables that have a straightforward yes-no interpretation, we provide intuition for the magnitude of our results by presenting the effects on each component of the index of poor mental health services in original units. Specifically, Appendix Table A.7 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 Appendix A describes a battery of exercises that probe the robustness of our estimates. The exercises include various placebo tests on variables that should not be affected by the introduction of Facebook and modified versions of our main specifications that take into account a host of possible concerns related to: i) the construction of our index of poor mental health, ii) the construction of our treatment variable, iii) particular Facebook expansion groups driving the effects, iv) particular variables driving the effects, v) the parallel trends assumption, vi) the level at which standard errors are clustered. We highlight one of our most convincing robustness check, which consists of a variant of the length-of-exposure specification from Section 5.3 that includes college by survey-wave fixed-effects. Such specification, which delivers estimates consistent with the hypothesis that longer exposure to Facebook has

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

a negative impact on student mental health, does not rely on our baseline college-level parallel trends assumption for identification.

Stigma as an Alternative Explanation One might worry that Facebook affected the stigma associated with mental illness and that our results may not reflect an increase in the prevalence of mental illness per se but rather an increase in willingness to discuss it. To formally investigate the role of stigma, we adopt a three-pronged strategy. First, we collected all the college newspaper articles containing the word Facebook published around the time of Facebook's expansion and checked whether any of them mention stigma in relation to mental health. While we do find articles hinting at potential negative effects of Facebook on mental health, we do not find any articles mentioning stigma. Second, we study whether the fraction of missing answers to the mental health questions in the NCHA survey was affected by the introduction of Facebook. If Facebook made people more comfortable discussing mental illness, we would expect to observe fewer missing answers after the introduction of Facebook.[32] Consistent with the effects being driven by increased prevalence of mental illness rather than by stigma, Appendix Table A.17 shows that the prevalence of missing answers was not 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. Furthermore, we find no detectable effects of the introduction of Facebook on eating disorders, even though such conditions are often highly stigmatized (Puhl and Suh, 2015). If reduction in stigma was indeed the driving force behind our results, it would be surprising not to find similar effects on other stigmatized behaviors and conditions.

5.5 Downstream Implications of Poor Mental Health

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

One of the NCHA survey questions inquires as to whether various conditions affected the students' academic performance. The conditions related to mental health are: attention deficit disorder, depression/anxiety disorder/seasonal affect disorder, eating disorders, stress,

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

and sleep difficulties.[33] The main advantage of analyzing these questions is that they trace a pathway from the introduction of Facebook to perceptions of worsened academic performance via poor mental health. It is important to emphasize, however, that we do not directly measure effects on grades, and that we do not rule out potential positive effects of Facebook on students' academic performance due to channels unrelated to mental health, such as improved teamwork.[34]

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

6 Mechanisms

Recent scholarship identified two main channels whereby Facebook might directly affect mental health: unfavorable social comparisons (Appel et al., 2016) and disruptive internet use

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

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

(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.[35] Such social comparisons, if unfavorable, could be detrimental to users' self-esteem and mental health (Vogel et al., 2014).[36]

Theoretically, the set of individuals who might be negatively affected by social comparisons is unclear. A simple model of social comparisons might posit that individuals compare themselves to the median member of their group along some dimension of interest (e.g., popularity, wealth, or looks).[37] If social media users are sophisticated, they will be able to extract accurate information from social media platforms about their relative ranking vis-à-vis their peers along the dimension of interest. In that case, we might expect around half of social media users to benefit from social comparisons and about half to suffer from them. Conversely, if social media users are to some extent naive, they will fail to understand that the content that their peers post on social media is likely to be highly curated rather than representative (Appel et al., 2016). In that case, they will systematically underestimate their relative ranking vis-à-vis their peers and, as a result, more than half of them will perceive social comparisons on Facebook as unfavorable.

In this section, we present evidence showing that: i) sub-populations which, in virtue of their baseline characteristics, might be more likely to suffer from social comparisons exhibit larger effects;[38] ii) the introduction of Facebook did not correct the students' misperceptions about their peers' social lives and, in some cases, exacerbated them. The latter piece of evi-

[35] 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).

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

[37] 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.

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

dence is consistent with students exhibiting a degree of naivete in interpreting the information conveyed through social media.

Figure 6 shows that the introduction of Facebook at a college affected more severely the mental health of students who might be more likely to be affected by unfavorable social comparisons. The figure plots estimates of the coefficient on the interaction between our treatment indicator and various moderators in a regression with our index of poor mental health as the outcome variable. Specifically, we consider the following sub-populations of students: i) students who live off-campus and are therefore less likely to participate in on-campus social life; ii) students who have weaker offline social networks as measured by not belonging to a fraternity or sorority organization; iii) students who have lower socioeconomic status as measured by carrying credit card debt or working part-time alongside studying; and iv) students who are overweight. We generate an index of social comparisons based on the above variables and consider, as an additional moderator, an indicator that takes value one if a student is above the median value of the index of social comparisons. All of the point estimates are positive and we find a strong and statistically significant effect on the index, on students living off-campus, and on students with credit card debt. Consistent with the social comparison mechanism, the introduction of Facebook has particularly detrimental effects on the mental health of students who might view themselves as comparing unfavorably to their peers.[39]

To test whether the introduction of Facebook affected the students' beliefs about their peers' social lives, we estimate the impact of the roll-out of Facebook on all survey questions that elicit students' perceptions of their peers' drinking behaviors.[40] Specifically, we study the following three sets of beliefs: i) beliefs about the number of alcoholic drinks the typical student has at a party, ii) beliefs about the share of the student population who has had an alcoholic drink in the month before the survey, iii) beliefs about the share of the student population who drinks alcohol on a regular basis. Appendix Table A.18a finds a positive and significant effect on each of the three outcomes above and on an equally-weighted index summarizing the three outcomes. Furthermore, Appendix Table A.19 shows that the effects on perceptions

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

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

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.[41]

Did Facebook affect beliefs about alcohol consumption because it led students to actually drink more, or did Facebook affect beliefs without a concurrent increase in drinking behaviors? Appendix Table A.18b shows that the effects on self-reported alcohol usage are substantially smaller than the effects on perceptions, suggesting that the effects on perceptions are unlikely to be driven by a change in actual behavior.[42]

If peers' behaviors did not change, why did Facebook affect perceptions? One option is that baseline perceptions were incorrect and the additional information provided on Facebook corrected such misperceptions. An alternative explanation is that Facebook led students to update their beliefs, but without aligning them more closely to reality. Appendix Table A.20 shows that the introduction of Facebook at a college did not lead students to develop more accurate perceptions about their peers' drinking behaviors and, for one of the outcomes, significantly exacerbated misperceptions. Specifically, the table estimates the effects on the difference between a student's perception of the alcohol consumption of the typical student at her college and the actual typical consumption at the student's college calculated using self-reported alcohol usage in the student's college–survey-wave. The results are consistent with students failing to fully take into account the fact that the content they see on social media is a curated rather than representative portrayal of their peers' lives. Such naivete could lead to distorted beliefs and exacerbate the effects of social comparisons.[43]

[41] Appendix Table A.24 provides suggestive evidence that perceptions regarding other students' sexual behavior may have also been affected by the introduction of Facebook. Conversely, Appendix Table A.26 shows that perceptions regarding the usage of illicit substances did not change. Finding effects on the perceptions of alcohol consumption but not on the perceptions of drug consumption is consistent with the fact that drinking and positive references to alcohol were common on Facebook profiles at the time, whereas images of students using drugs were very rare (Watson et al., 2006; Kolek and Saunders, 2008; Morgan et al., 2010)

[42] If the introduction of Facebook decreased the stigma related to alcohol consumption, our results about alcohol usage could be biased (see also our discussion of stigma in the context of mental health in Section 5.4). Although we cannot rule out the possibility that changes in stigma due to the introduction of Facebook had an effect specifically on alcohol-related questions, such bias would, if anything, make our results even starker. Specifically, if the introduction of Facebook reduced the stigma around underage drinking, the actual effect on alcohol usage would be smaller than the effect we estimate. Thus, the gap between the changes in usage and changes in perceptions would be even larger than the effect we currently estimate.

[43] 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).

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 or computer games affected their academic performance. Students could answer that the issue affected their academic performance, that they experienced the issue but it did not affect their performance, and that they did not experience the issue. If, after the introduction of Facebook at their college, students found the internet more distracting and had a harder time focusing because of it, we would expect a larger number of students to answer that they experienced the internet or computer games as an issue and that it affected their academic performance. Appendix Table A.21 shows that the share of students experiencing internet or computer games as an issue increased by around 5%, but the effect is not statistically significant.

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

Appendix Tables A.22, A.23, and A.25 present estimates of the effects of the introduction of Facebook using Equation (1) on various offline behaviors measured in the survey that could plausibly affect mental health. Comfortingly, we do not find any effects on sexual assaults. Similarly, none of the outcomes related to relationships and drug use exhibit significant effects. Combined with the null results on drinking behaviors (Appendix Table A.18b), we do not find much evidence that the introduction of Facebook at a college had meaningful effects on various self-reported behaviors that could have a bearing on mental health.

7 Discussion

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

Over the last two decades, Facebook underwent a host of important changes. Such changes include: i) the introduction of a personalized feed where posts are ranked by an algorithm; ii) the growth of Facebook's user base from U.S. college students to almost three billion active users around the globe (Facebook, 2021); iii) video often replacing images and text; iv) increased usage of Facebook on mobile phones instead of computers; and v) the introduction of Facebook pages for brands, businesses, and organizations. The nature of the variation we are exploiting does not allow us to identify the impact of these features of social media. For instance, our estimates cannot shed light on whether the increased reliance on Facebook for news consumption has exacerbated or mitigated the effects of Facebook on mental health. Similarly, we cannot provide evidence as to whether years of experience with the platform mitigate or exacerbate the effects on mental health.[44]

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

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

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

make Facebook even more engaging and might exacerbate the effects on mental health.[45]

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 leveraged the staggered introduction of Facebook across U.S. colleges to estimate the impact of social media on mental health and found that the introduction of Facebook at a college had a negative effect on student mental health.

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 (2005–2009). American College Health Association-National College Health Assessment (ACHA-NCHA) Spring 2003–2009 Reference Group Data Report (Abridged). *Journal of American College Health 54*(4), 201.

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.

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

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*.

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.

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

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.

[45] 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.

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

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). 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). 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. *arXiv preprint arXiv:2108.12419*.

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

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

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The review of economics and statistics 90*(3), 414–427.

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.

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.

Fujiwara, T., K. Müller, and C. Schwarz (2021). The effect of social media on elections: Evidence from the united states. *NBER Working Paper 28849*.

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 225*(2), 254–277.

Griffiths, M. D., D. J. Kuss, and Z. Demetrovics (2014). Chapter 6 - social networking addiction: An overview of preliminary findings. In K. P. Rosenberg and L. C. Feder (Eds.), *Behavioral Addictions*, pp. 119–141. San Diego: Academic Press.

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

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

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.

Hedden, S., G. Blau, R. Lipari, L. Rubenstein, V. Forman-Hoffman, and C. Glasheen (2017). Comparison of physical health conditions among adolescents aged 12 to 17 with and without major depressive episode. *Bethesda, MD: Substance Abuse and Mental Health Services Agency*.

Jacobs, A. Z., S. F. Way, J. Ugander, and A. Clauset (2015). Assembling thefacebook: Using heterogeneity to understand online social network assembly. WebSci '15, New York, NY, USA. Association for Computing Machinery.

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.

Kroenke, K., R. L. Spitzer, and J. B. Williams (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine 16*(9), 606–613.

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*, CSCW '08, New York, NY, USA, pp. 721–730. Association for Computing Machinery.

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

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.

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

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.

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

Prince, M., V. Patel, S. Saxena, M. Maj, J. Maselko, M. R. Phillips, and A. Rahman (2007). No health without mental health. *The Lancet 370*(9590), 859–877.

Puhl, R. and Y. Suh (2015). Stigma and eating and weight disorders. *Current psychiatry reports 17*(3), 1–10.

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

Roodman, D., M. Ø. Nielsen, J. G. MacKinnon, and M. D. Webb (2019). Fast and wild: Bootstrap inference in stata using boottest. *The Stata Journal 19*(1), 4–60.

Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe (2022). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *arXiv preprint arXiv:2201.01194*.

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*.

Spitzer, R. L., K. Kroenke, J. B. Williams, and B. Löwe (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine 166*(10), 1092–1097.

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

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

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.

U.S. Department of Education, N. C. f. E. S. (2005). Integrated postsecondary. Fall enrollment: gender, attendance status, and level of student.

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.

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

Wells, G., J. Horwitz, and D. Seetharaman (2021). Facebook knows instagram is toxic for teen girls, company documents show. *Wall Street Journal*.

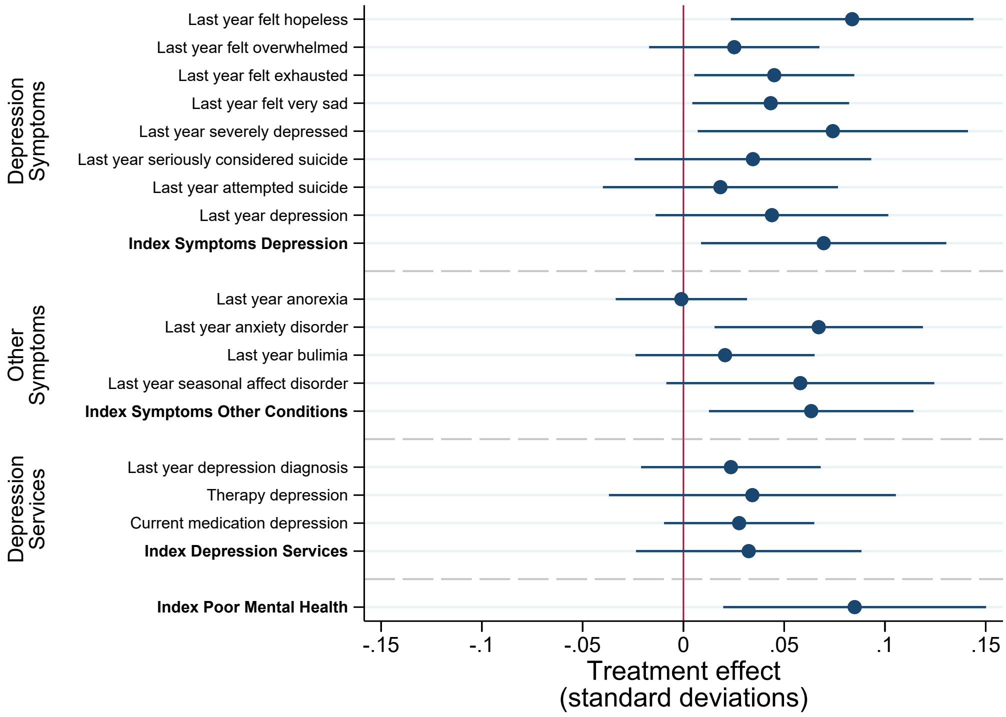
WHO (2008). *The Global Burden of Disease: 2004 Update*. World Health Organization.

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.

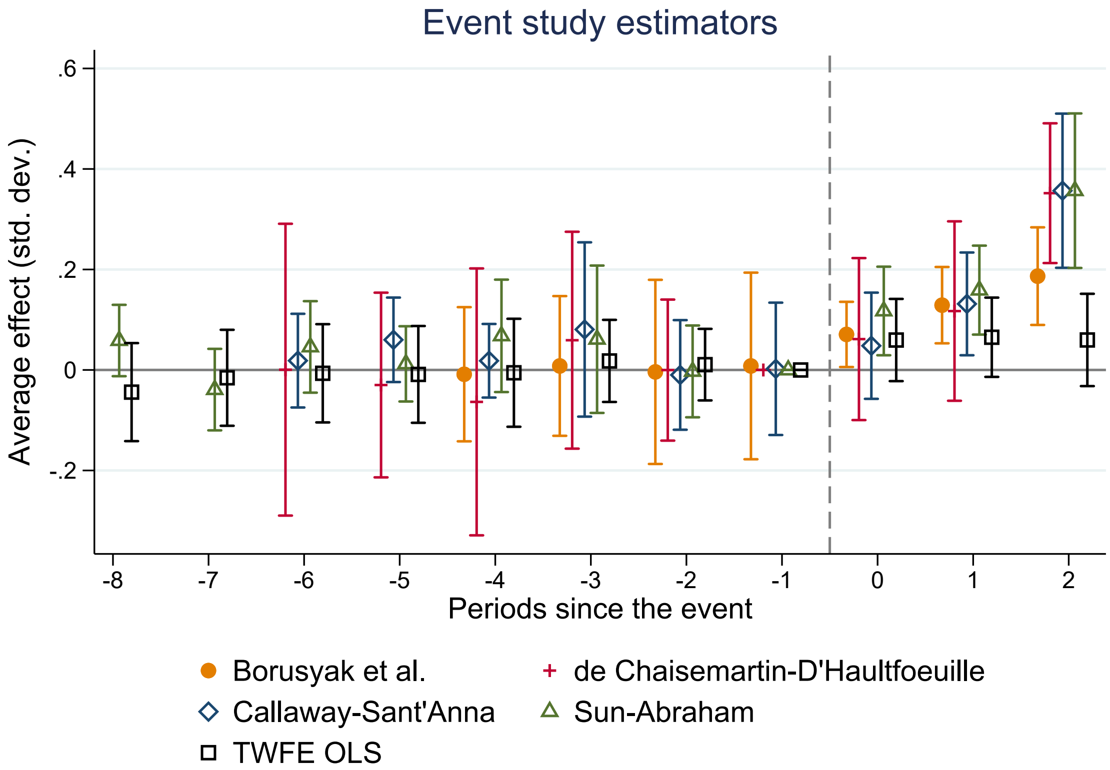
Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political effects of the internet and social media. *Annual Review of Economics 12*, 415–438.

Figure 1: Effects of the Introduction of Facebook on Student Mental Health



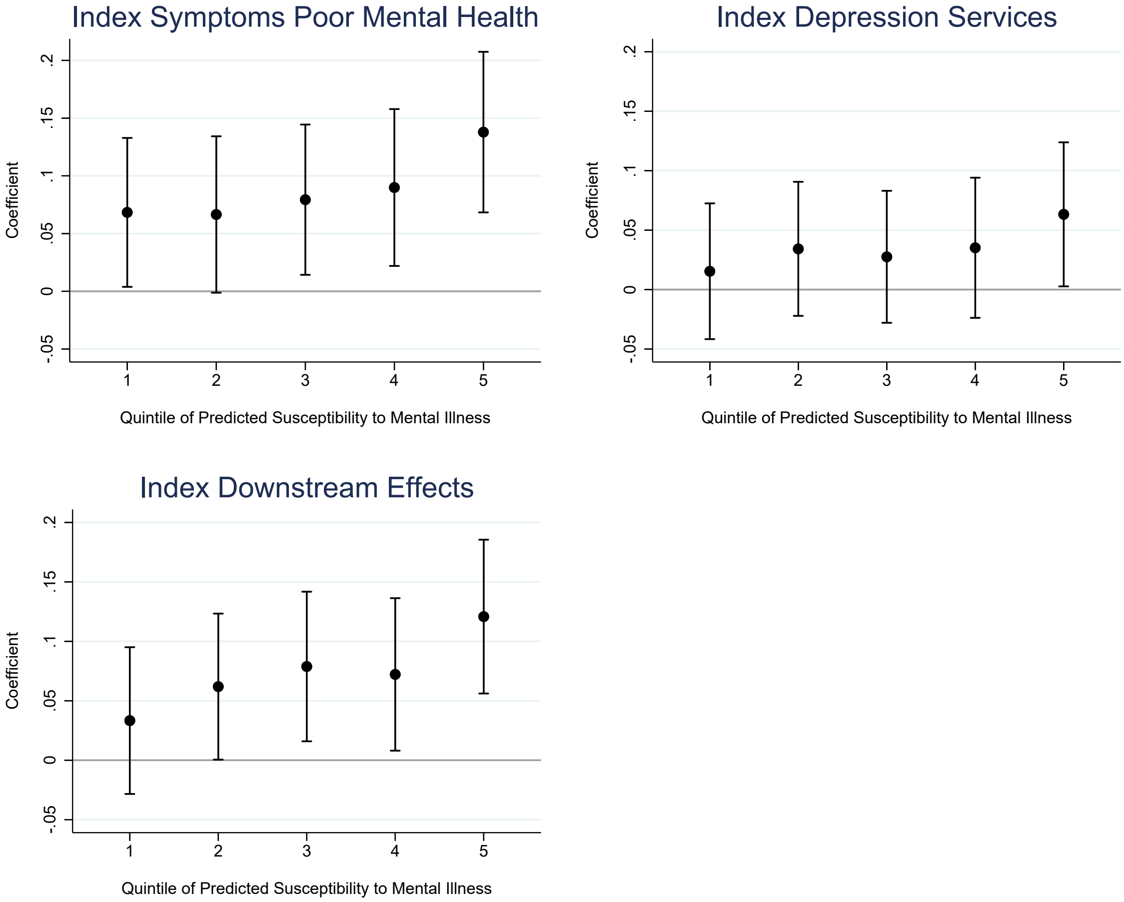
Notes: This figure explores the effects of the introduction of Facebook at a college on all our mental-health outcome variables and on the related indices. Specifically, it presents estimates of coefficient β from Equation (1) using our preferred specification, namely the one including survey-wave fixed effects, college fixed effects, and controls. The outcome variables are our overall index of poor mental health, the individual components of the index, and three sub-indices: the index of depression symptoms, the index of symptoms of other mental health conditions, and the index of depression services. All outcomes are standardized so that, in the pre-period, they have a mean of zero and a standard deviation of one. Our controls consist of: age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. The reason why the point estimate on an index might be relatively large compared to the point estimates on each of the components of the index is that averaging across the index components reduces noise and, as a consequence, might increase the effect size measured in standard deviation units. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.30. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure 2: Effects of Facebook on the Index of Poor Mental Health Based on Distance to/from Facebook Introduction



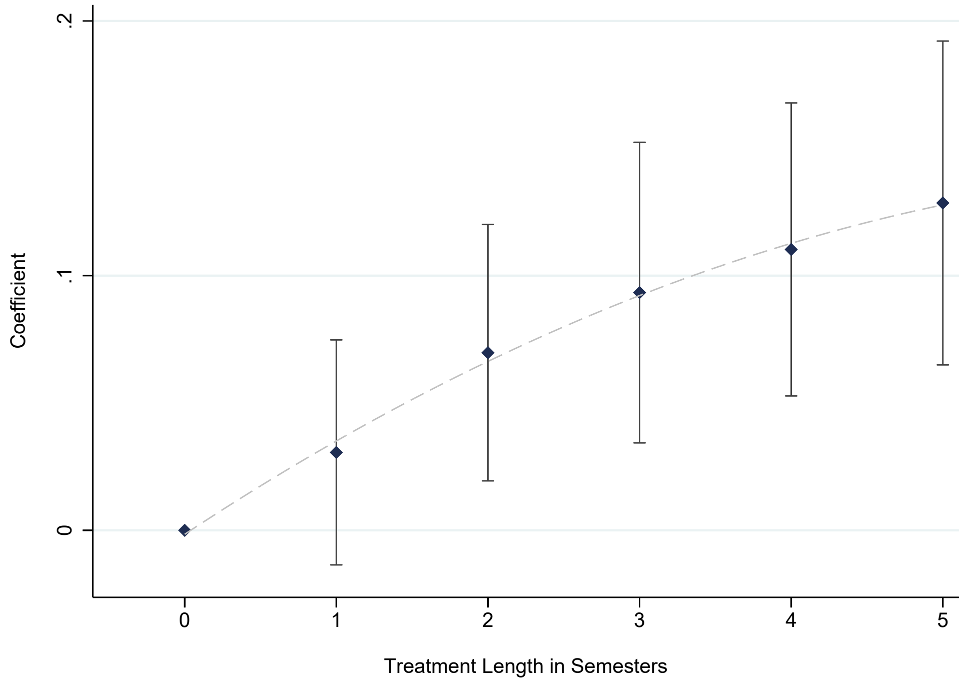
Notes: This figure overlays the event-study plots constructed using five different estimators: a dynamic version of the TWFE model—Equation (2)—estimated using OLS (in black with square markers), Sun and Abraham (2021) (in green with triangle markers), Callaway and Sant'Anna (2021) (in blue with diamond markers), De Chaisemartin and d'Haultfoeuille (2020) (in red with cross markers), and Borusyak et al. (2021) (in orange with circle markers). The outcome variable is our overall index of poor mental health. The time variable is the survey wave and the treatment group variable is given by the semester in which the college attended by the student was granted Facebook access. The figure displays only two post-periods because the estimation of additional post periods would require employing already-treated units as controls for newly-treated units. In the presence of heterogeneous dynamic treatment effects, such comparisons would bias the estimation and, therefore, they are shut down by all the newly-introduced robust estimators. As a result, the maximum number of post-periods that can be estimated robustly is two. For the Borusyak et al. (2021) estimator, we estimate four pre-periods since estimating more pre-periods dramatically increases the standard errors in the pre-period (Borusyak et al., 2021, p. 24). Similarly, for the estimator by De Chaisemartin and d'Haultfoeuille (2020), the maximum number of pre-periods that can be estimated in our panel is only five. In order to estimate the standard errors for the $t+2$ estimate, the De Chaisemartin and d'Haultfoeuille (2020) estimator includes controls for age and age squared. For appropriate estimation of the coefficients on $t=-8$ and $t=-7$ using the Sun and Abraham (2021) estimator, we include data from additional pre-periods, even though, in those pre-periods, we do not observe all four Facebook expansion groups (Sun and Abraham, 2021, p. 13). For a detailed description of the outcome and treatment variables, see Appendix Table A.30. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure 3: Heterogeneous Effects by Predicted Susceptibility to Mental Illness



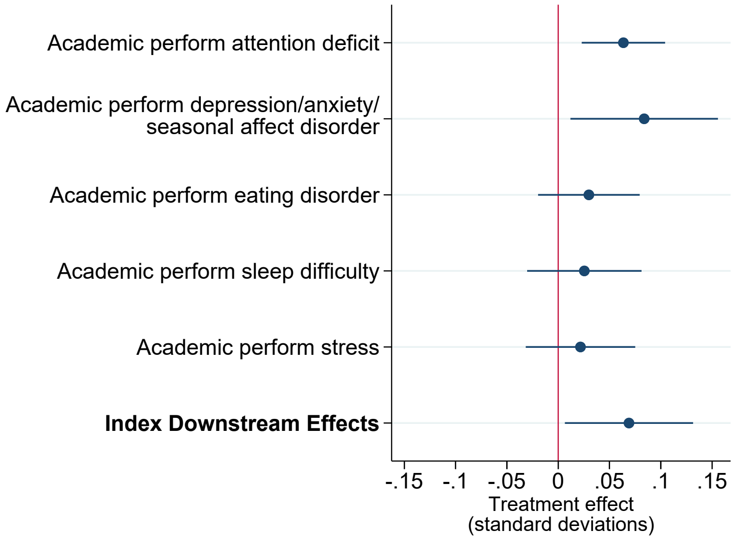
Notes: This figure explores the extent to which the effects of the introduction of Facebook at a college are heterogeneous depending on students' predicted susceptibility to mental illness. Specifically, it presents the estimates from equation (3) in which our indicator for post-Facebook introduction is interacted with a set of indicators for belonging to each quintile of a LASSO-predicted measure of susceptibility to mental illness. The outcome variable in the top-left panel is our index of symptoms of poor mental health; the outcome variable in the top-right panel is our index of depression services; the outcome variable in the bottom-left panel is our index measuring whether students reported that conditions related to poor mental health negatively affected their academic performance. All indices are standardized so that, in the pre-period, they have a mean of zero and a standard deviation of one. The estimates (also displayed in Table A.5) are obtained using our preferred specification, namely the one including survey-wave fixed effects, college fixed effects, and controls. Our controls consist of: age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. For a detailed description of the outcome, treatment, interaction, and control variables, see Appendix Table A.30. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure 4: Effect on Poor Mental Health by Length of Exposure to Facebook



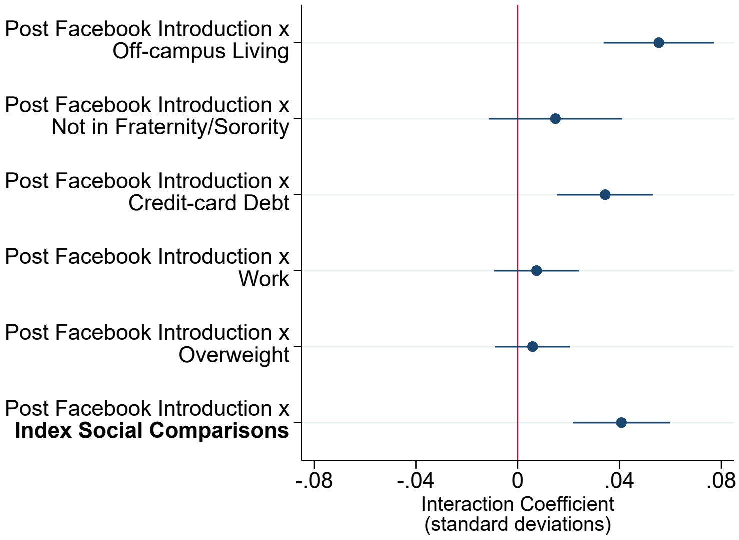
Notes: This figure explores the effects of length of exposure to Facebook on our index of poor mental health by presenting estimates of Equation (4). The index is standardized so that, in the pre-period, it has a mean of zero and a standard deviation of one. The dashed curve is the quadratic curve of best fit. Our controls consist of: age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. Students who entered college in 2006 might have been exposed to Facebook already in high-school, because, starting in September 2005, college students with Facebook access could invite high-school students to join the platform. Such students are excluded from the regression. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.30. 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. It presents estimates of coefficient β from Equation (1) using our preferred specification, including survey-wave fixed effects, college fixed effects, and controls. The outcome variables are answers to questions inquiring as to whether various mental health conditions affected the students' academic performance and our index of downstream effects. All outcomes are standardized so that, in the pre-period, they have a mean of zero and a standard deviation of one. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.30. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure 6: Heterogeneous Effects as Evidence of Unfavorable Social Comparisons



Notes: This figure explores the mechanisms behind the effects of Facebook on mental health. It presents estimates from a version of Equation (1) in which our treatment indicator is interacted with a set of indicators for belonging to a certain sub-population of students. The outcome variable is our overall index of poor mental health. The estimates are obtained using our preferred specification, namely the one including survey-wave fixed effects, college fixed effects, and controls. For a detailed description of the outcome, treatment, interaction, and control variables, see Appendix Table A.30. 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 effect of the introduction of Facebook at a college on student mental health. Specifically, it presents estimates of coefficient β from Equation (1) with our index of poor mental health as the outcome variable. The index is standardized so that, in the pre-period, it has a mean of zero and a standard deviation of one. Column (1) estimates Equation (1) without including controls; column (2) estimates Equation (1) including controls; column (3)—our preferred specification—replaces Facebook-expansion-group fixed effects with college fixed effects; column (4) includes linear-time trends estimated at the Facebook-expansion-group level. Our controls consist of: age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. Column (2) also includes indicators for geographic region of college (Northeast, Midwest, West, South); such indicators are omitted in columns (3) and (4) because they are collinear with the college fixed effects. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Appendix For Online Publication

A Robustness Checks

This section presents a battery of exercises that probe the robustness of our estimates.

First, as a placebo test, Table A.8 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 should not be affected by the introduction of Facebook at a college. In fact, if we did find an effect on this measure, we would worry that the introduction of Facebook affected the selection of students responding to the survey along dimensions that are predictive of mental illness. Comfortingly, the point estimates in Table A.8 are small and not significant. Table A.9 presents a similar robustness test using all available immutable individual-level characteristics as outcomes. Reassuringly, the point estimates are very small and only one out of the 13 estimates is statistically significant at the 10% level.

As an additional placebo test, Table A.10 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). 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. Figure A.8 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.[46]

Next, we show that the results on our index of poor mental health are not driven by the way in which we construct the index, by any one outcome variable, by any particular Facebook expansion group, or by how we define treatment status when the semester in which a student

[46] Although the effects on physical health are statistically smaller than the effects on mental health, the point estimates in Table A.10 are positive and may be considered non-negligible. Such effects might be due to noise, but they might also be capturing actual effects of the introduction of Facebook on physical health. There are two main reasons why Facebook might affect students' physical health. First, it could lead students to spend more time on their computers. Consistent with this narrative, the largest point estimate of Facebook's effect on physical health in Figure A.8 is for back pain. Second, Facebook might affect students' physical health indirectly as a result of its negative effect on mental health (Prince et al., 2007). Indeed, even respiratory diseases such as bronchitis have been linked to major depressive episodes (Hedden et al., 2017).

took the survey coincides with the semester in which Facebook was rolled out at her college. To address the first concern, we construct two additional indices: an index of poor mental health that includes observations for which some of the component variables are missing and an inverse-covariance weighted index that assigns a smaller weight to strongly correlated components (Anderson, 2008). Appendix Table A.11 shows that our results remain qualitatively similar using these alternative indices. To address the second concern, we construct various versions of the index of poor mental health, each time excluding a different component from the index. Appendix Figure A.9 shows that our estimates are robust to separately dropping each individual component of the index of poor mental health. To address the third concern, we run our TWFE and length-of-exposure models on a restricted dataset that excludes colleges belonging to each Facebook expansion group in turn. Appendix Table A.12 shows that the results remain fairly stable across the various restricted datasets.[47] Lastly, to address the fourth concern, Appendix Table A.13 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. Also, reassuringly, the coefficient on $t = 0$ in Figure 2 is in between the magnitudes of the coefficients on $t = -1$ and $t = 1$ for all estimators.

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).[48] Appendix Table A.14 shows that our results are not meaningfully affected by the inclusion of these additional controls, which allow for flexible differential trends based on expansion-group- and college-level features correlated with roll-out timing.

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 (4)—that includes college$\times$survey-wave fixed effects does not rely on this parallel

[47] In fact, in both panels, we fail to reject the hypothesis of equality of coefficients across the various restricted datasets at conventional significance levels.

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

trends assumption for identification.[49] 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 A.6 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.

Next, we show that our baseline estimates do not change substantially when we replace the TWFE estimator from Equation (1) with the estimators suggested in Borusyak et al. (2021), Callaway and Sant'Anna (2021), De Chaisemartin and d'Haultfoeuille (2020), and Sun and Abraham (2021). The latter set of estimators, which shut down the 2×2 difference-in-differences comparisons between newly-treated and already-treated units, are designed to be consistent even in the presence of heterogeneous treatment effects across across time and across treated units. Table A.15 shows that the estimates obtained using the robust estimators are qualitatively similar to our baseline TWFE estimates.

Finally, Table A.16 shows that the baseline results are robust to clustering standard errors at the Facebook-expansion-group level and at the expansion-group–by–survey-wave level. Since we only have four expansion groups, which is lower than the number of clusters necessary for asymptotics to work, we also include a p-value obtained using a wild bootstrap procedure that corrects for the few-clusters problem (Cameron et al., 2008; Roodman et al., 2019). The wild bootstrap p-value confirms the statistical significance of our baseline effect.

B 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

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

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.12, 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.13 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 correctly classifies as having received a depression diagnosis 75% of individuals who indeed have ever received a depression diagnosis and only incorrectly classifies 30% of individuals who have never received a depression diagnosis. As aforementioned, 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 to be even higher because some of the observations that are currently being counted as false positives might actually be true positives.[50]

C External Validation of Symptoms Variables

The mental health questions asked in the NCHA survey are non-standard; therefore, it is important to validate them against external benchmarks. In 1998–2000, the reliability of various NCHA survey questions was already validated against three external datasets: The CDC's National College Health Risk Behavior Survey (NCHRBS), the College Alcohol Study (CAS), and the National College Women's Sexual Victimization Survey (ACHA, 2009). In this section, we expand on the previous validation exercises by comparing the NCHA survey questions related to mental health to canonical depression and generalized anxiety disorder screeners: the Patient Health Questionnaire (PHQ-9) and the Generalized Anxiety Disorder (GAD-7) ques-

[50]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 predicted is whether the student has an active depression diagnosis at the time in which she takes the survey.

tionnaire.

The PHQ-9 is a widely-used depression screener that has been shown to be a "reliable and valid measure of depression severity" (Kroenke et al., 2001, page 606). The PHQ-9 asks nine questions about how often a person has been bothered by various problems (e.g., little interest or pleasure in doing things) over the past two weeks. For each question, a respondent receives a score from 0 ("not at all") to 3 ("nearly every day"). A respondent is classified as suffering from moderate or severe depression if their total PHQ-9 score equals 10 or above. When compared to medical diagnoses by trained psychiatrists, the PHQ-9 classification has been shown to have a sensitivity of 88% and a specificity of 88% for major depression (Kroenke et al., 2001).

The GAD-7 is a widely-used questionnaire screening for generalized anxiety disorder (Spitzer et al., 2006). The structure of the questions in the GAD-7 is similar to that in the PHQ-9, and a GAD-7 score of 10 and above indicates moderate or severe anxiety. When compared to medical diagnoses by trained psychiatrists, the GAD-7 classification has been shown to have a sensitivity of 89% and a specificity of 82% for generalized anxiety disorder (Spitzer et al., 2006).

In order to provide additional validation for the NCHA questions about mental health, we ran a survey on college students that included the NCHA questions, the PHQ-9 questions, and the GAD-7 questions. Specifically, we recruited full-time college students on Prolific to complete a survey on physical and mental health. The survey included basic demographics questions and the three modules of mental health questions, presented in a random order. Our original sample includes 523 respondents. We removed three duplicate respondents, one respondent who failed an attention check, one respondent who reported accidentally clicking the wrong response, and respondents in the bottom 2% of the survey time distribution (completed the survey in fewer than 110 seconds). Our final sample, therefore, includes 507 valid responses.

Appendix Table A.27 compares the sample we recruited to the NCHA sample analyzed in the paper. The samples have a similar share of international students and women (we intentionally attempted to recruit a sample that was balanced on sex), while the NCHA sample has a higher share of white respondents. The final three rows calculate the average response to all numeric and binary questions composing our index of poor mental health. The students who completed our survey on Prolific are more likely to suffer from poor mental health compared to the students in the NCHA sample. This may reflect the deterioration of mental health among young adults that occurred over the past two decades.

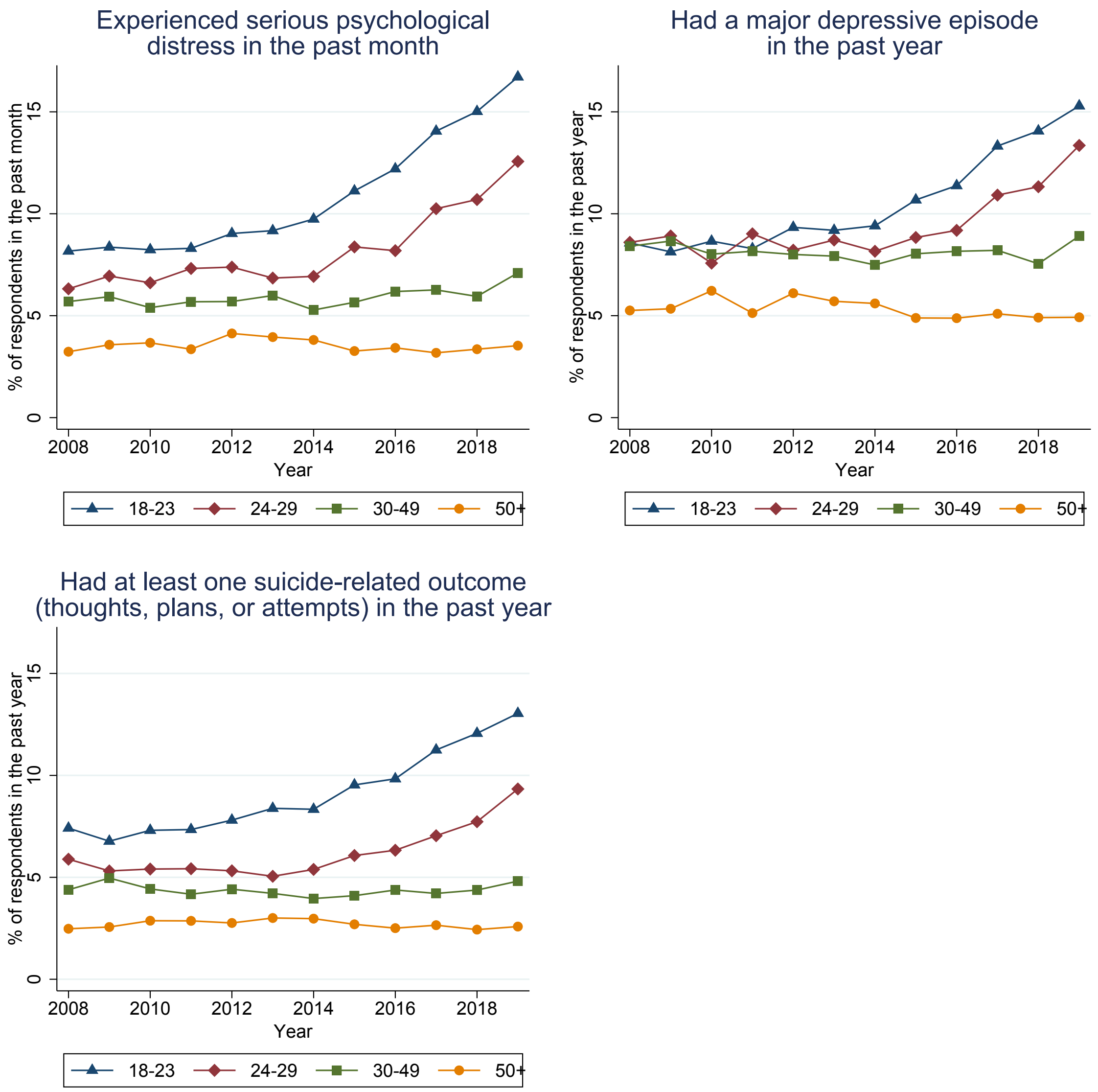
Appendix Figures A.14 and A.15 show that there is a strong correlation between our main index of poor mental health and the PHQ-9 and GAD-7 indices, respectively. The figures present binned scatter plots where each point shows the mean PHQ-9 or GAD-7 score for different ventiles of our index of poor mental health. The strong correlation suggests that if Facebook negatively affected our index of poor mental health, it also negatively affect the clinically-validated PHQ-9 and GAD-7 measures.

As discussed in Section 5, we can leverage our survey to get a better sense of the magnitude of our treatment effects. Specifically, using data from our survey, we can determine how to weigh the variables contained in our index of poor mental health in a way that best predicts an indicator for having depression according to the PHQ-9 (10≤PHQ-9) and an indicator for having generalized anxiety disorder according to the GAD-7 (10≤GAD-7). We calculate such weights using an OLS (linear probability model), a logistic regression, and a LASSO. The resulting weights are shown in Appendix Table A.28. Unsurprisingly, the symptom most predictive of depression is being severely depressed and the strongest predictor of anxiety is saying that one had anxiety disorder in the last year. Interestingly, taking medication for depression, conditional on the other coefficients, predicts that a respondent is less likely to suffer from depression.

Appendix Table A.29 shows that the introduction of Facebook increased by two percentage points the fraction of students whom, according to our prediction, the PHQ-9 and GAD-7 would classify as having depression or generalized anxiety disorder (the result is robust to the prediction methods used). Based on the OLS regressions, depression increased by 9% over a pre-period baseline mean of 25% and anxiety increased by 12% over a pre-period baseline mean of 16%.

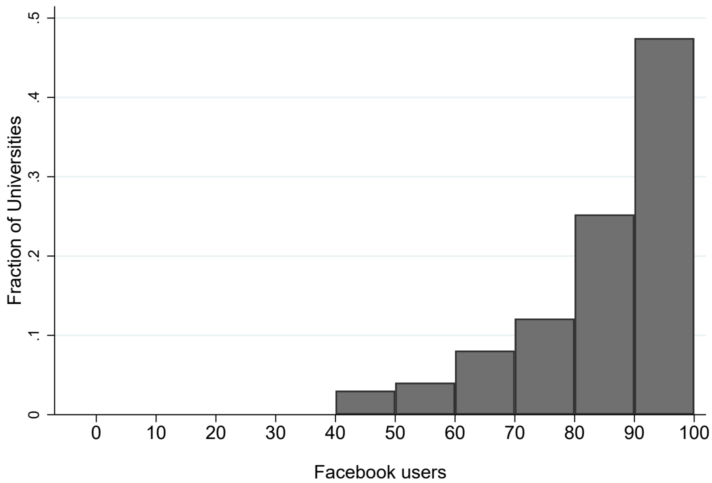
D Additional Tables and Figures

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



Notes: This figure displays mental health trends in the United States by age group in 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 A.2: Facebook Users Per 100 Undergraduate Students, September 2005

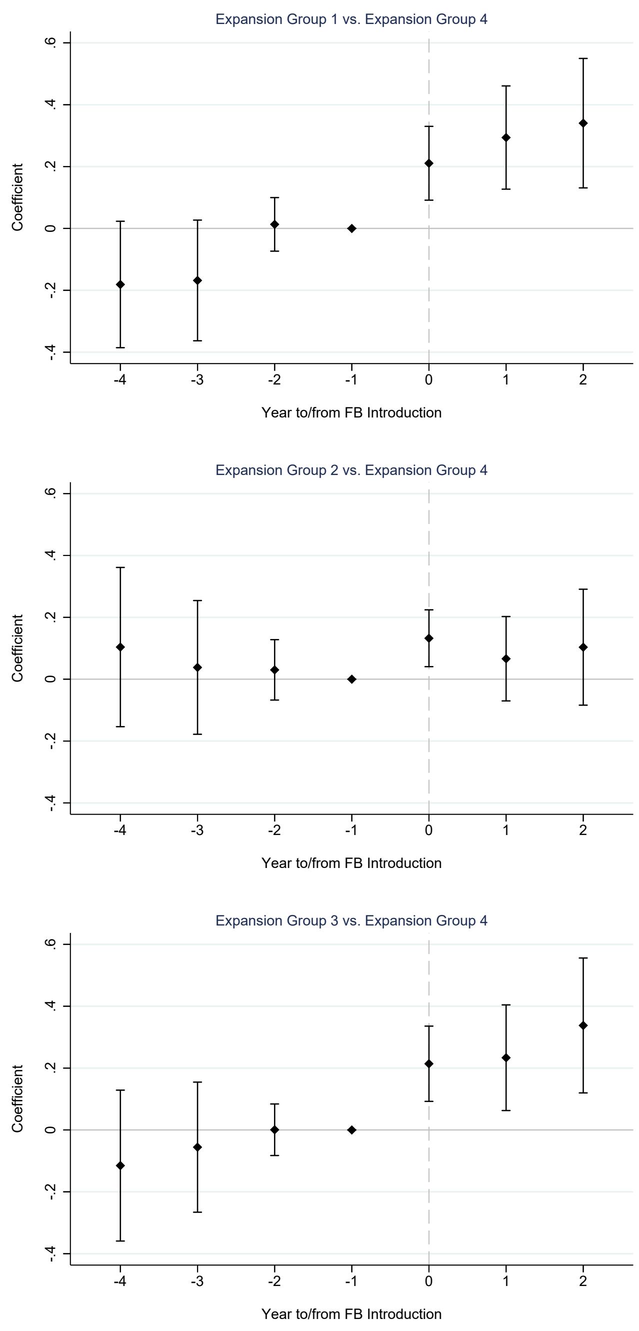


Notes: This figure presents a histogram of the number of Facebook users per 100 full-time undergraduate students in September 2005 for the first 100 universities that received access to the platform. The number of Facebook users is based on data provided by Facebook to Traud et al. (2012) and the number of full time students is based on IPEDS (U.S. Department of Education, 2005). We winsorize the number of users per 100 undergraduate students at 100. A value larger than 100 could occur, for instance, because, at early-adopting colleges, students who had graduated in the Spring of 2004 still had access to the platform in the fall of 2005. Tulane university is excluded since data on the number of full time students is not available for 2005.

Figure A.3: Facebook Homepage as of June 2004

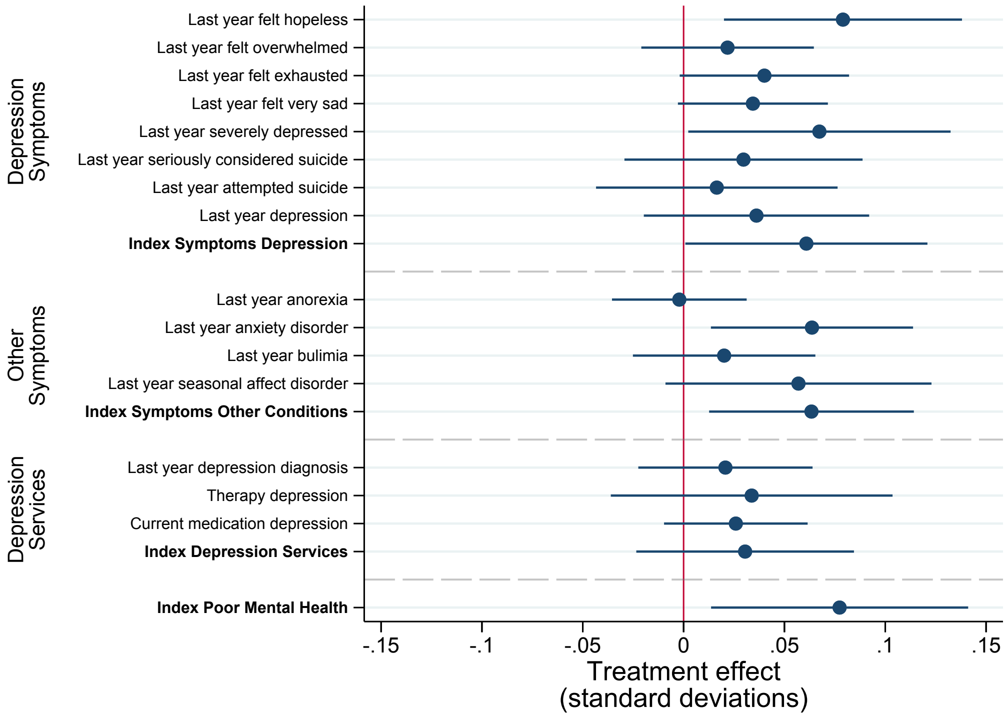
Notes: The figure shows a snaphsot of the homepage of thefacebook.com as of June 15th, 2004 recovered via the Wayback Machine. The colleges that, by that date, had been granted access to Facebook are listed on the home page.

Figure A.4: Event Studies Comparing each Expansion Group in Turn to Expansion Group 4



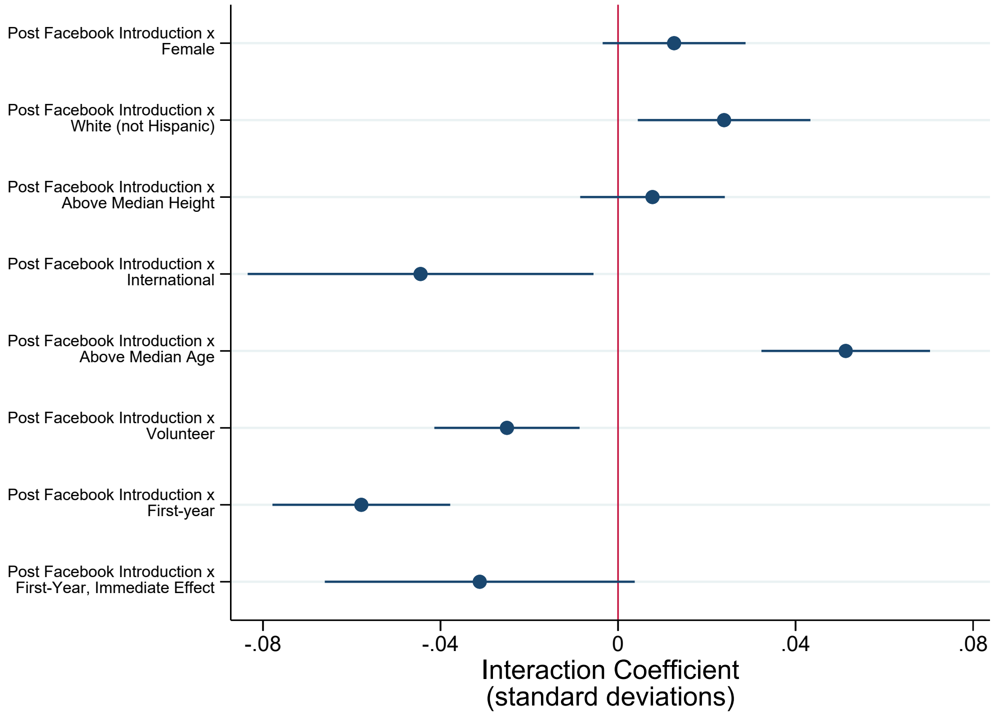
Notes: This figure presents three event study plots. Each plot isolates students attending colleges in one of the first three Facebook expansion groups (Spring 2004, Fall 2004, and Spring 2005) and compares them to students attending colleges in the last Facebook expansion group (post Spring 2005). The outcome variable is always 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 time variable is always the year in which the student participated in the survey and the treatment group variable is always given by the semester in which the college attended by the student was granted Facebook access. All three plots are based on a version of Equation (2) in which time is measured at the year rather than the semester level. We measure time at the year level because isolating each Facebook expansion group reduces the size of the samples used to construct each plot and increases noise. The regressions underlying the plots do not include controls. The coefficient on $t = -1$ corresponds to the omitted category and is normalized to zero. The coefficient on $t = 0$ corresponds to the semester when Facebook was introduced at the college, when it is impossible to determine if the student was treated, and the following semester. The time spanned by the x-axis (four years in the pre-period and three in the post-period) is the largest span of time for which we have data from all four Facebook expansion groups. For a detailed description of the outcome and treatment variables, see Appendix Table A.30. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure A.5: Effect of the Introduction of Facebook on Student Mental Health, with Expansion-Group-Specific Linear Trends



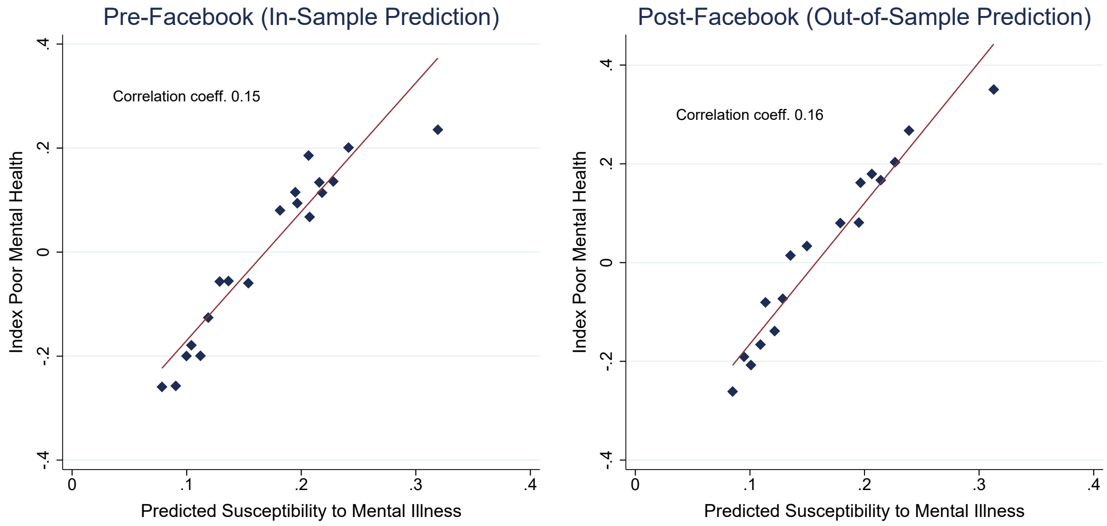
Notes: This figure explores the robustness of our baseline effects of the introduction of Facebook at a college on all our mental health outcome variables and on the related indices. Specifically, it presents estimates of coefficient β from Equation (1) using a specification that includes survey-wave fixed effects, college fixed effects, controls, and Facebook expansion-group-specific linear time trends. The outcome variables are our overall index of poor mental health, the individual components of the index, and three sub-indices: the index of depression symptoms, the index of symptoms of other mental health conditions, and the index of depression services. All outcomes are standardized so that, in the pre-period, they have a mean of zero and a standard deviation of one. Our controls consist of: age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. The reason why the point estimate on an index might be relatively large compared to the point estimates on each of the components of the index is that averaging across the index components reduces noise and, as a consequence, might increase the effect size measured in standard deviation units. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.30. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure A.6: Heterogeneous Effects



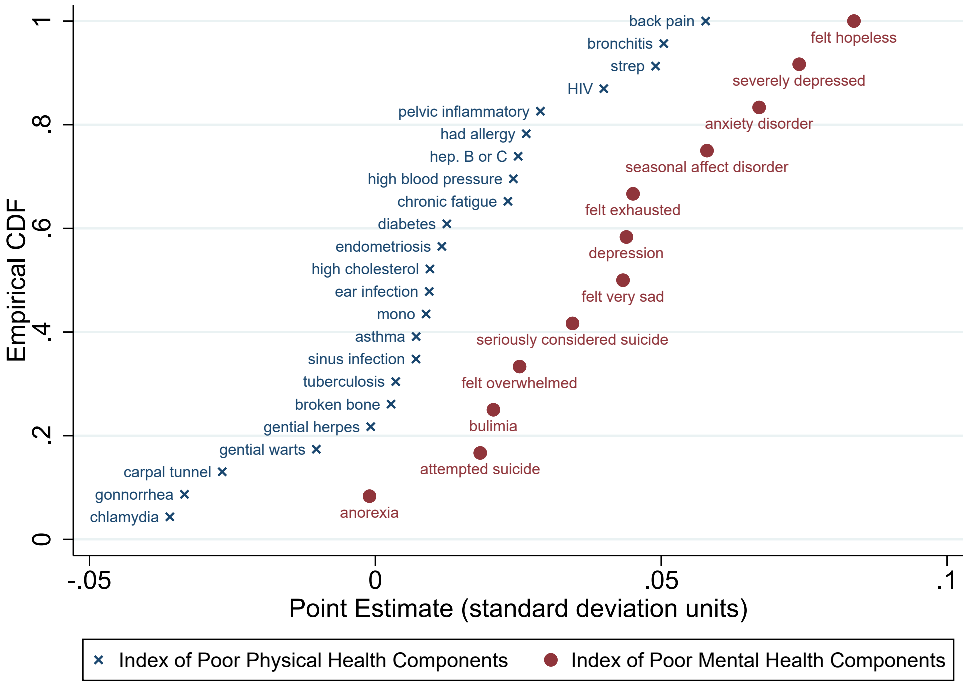
Notes: This figure explores whether the effects of the introduction of Facebook on student mental health are heterogeneous across a host of demographic characteristics. Specifically, it presents estimates from a version of Equation (1) in which our treatment indicator is interacted with various moderators. 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, volunteering, 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.30. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure A.7: Relationship between the LASSO-Predicted Measure of Susceptibility to Mental Illness and the Index of Poor Mental Health



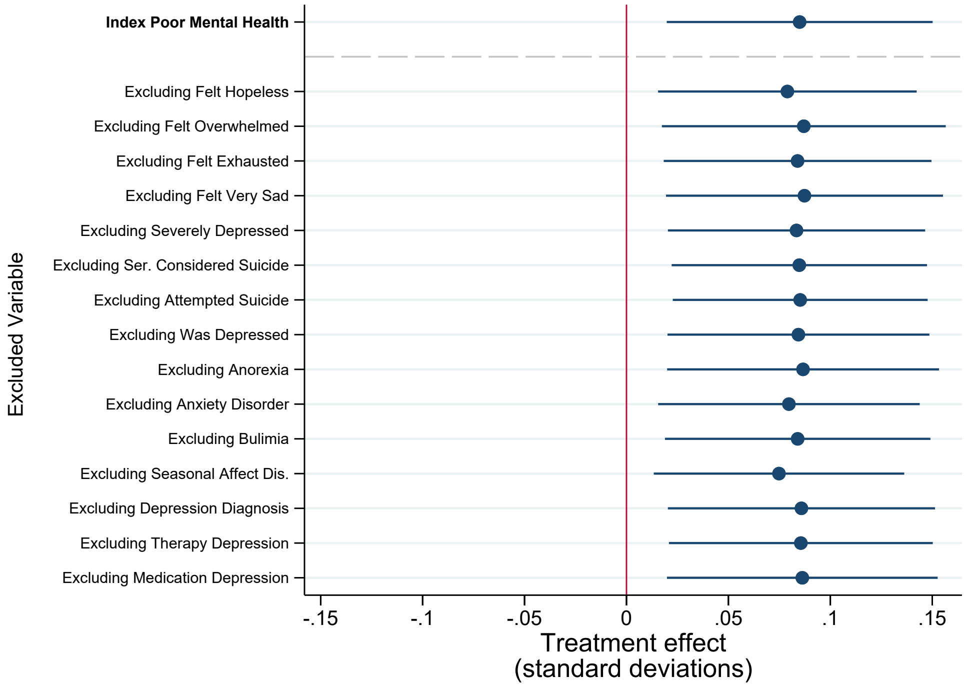
Notes: This figure explores the relationship between our LASSO-predicted measure of susceptibility to mental illness and our index of poor mental health. Specifically, for each ventile of our LASSO-predicted measure of susceptibility to mental illness, the figure plots the average predicted susceptibility 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 introduction of Facebook at a college; the right panel presents data from the period after the introduction of Facebook 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.8: Cumulative Distribution of Coefficients on Components of the Index of Poor Mental Health and the Index of Poor Physical Health



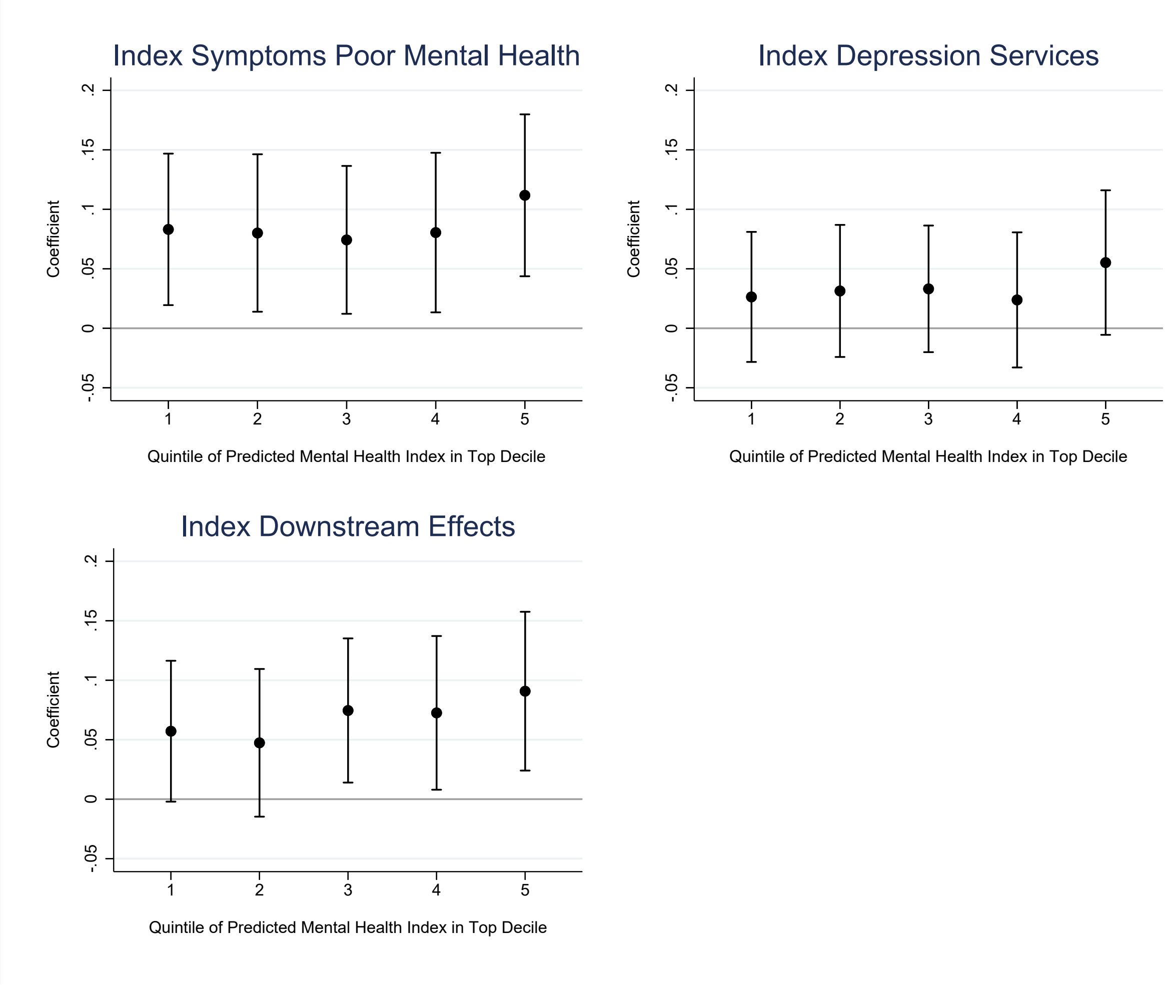
Notes: This figure displays cumulative distribution functions of the coefficients on the components of the indices of poor mental and poor physical health. The figure is constructed as follows: first, we computed estimates of coefficients β from Equation (1) for each component of the index of poor physical 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 index components, the treatment and the control variables, see Appendix Table A.30.

Figure A.9: Robustness to Excluding Each Variable from the Index of Poor Mental Health



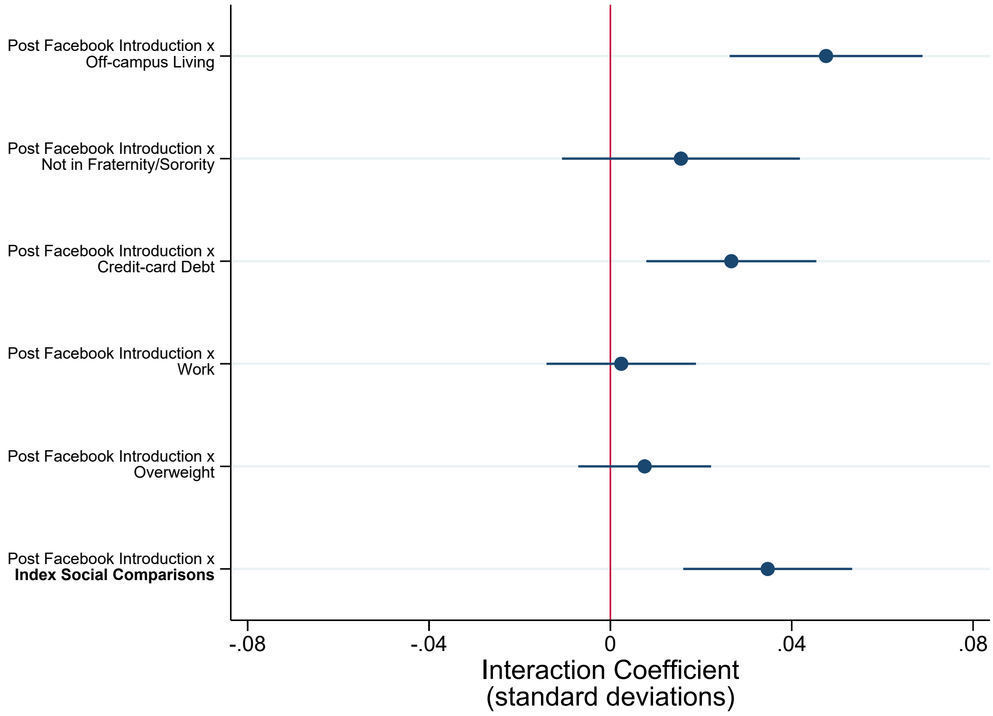
Notes: This figure explores the robustness of our baseline results to excluding each individual variable from the construction of the index of poor mental health. Specifically, it presents estimates of coefficient β from Equation (1). Each row excludes a different variable from the construction of the 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.30. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure A.10: Heterogeneous Effects by Predicted Susceptibility to Mental Illness, with Susceptibility Defined Using the Index of Poor Mental Health



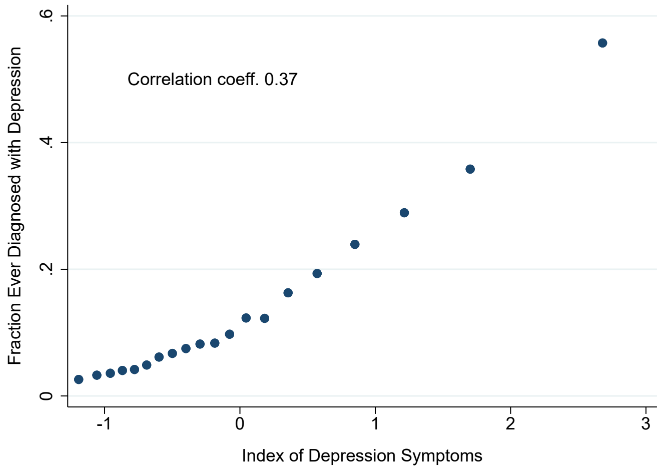
Notes: This figure explores the extent to which the effects of the introduction of Facebook at a college are heterogeneous depending on students' predicted susceptibility to mental illness. Specifically, it presents the estimates from equation (3) in which our indicator for post-Facebook introduction is interacted with a set of indicators for each quintile of a LASSO-predicted measure of susceptibility to mental illness. In this figure, susceptibility to mental illness is defined based on a LASSO predicting whether a respondent's index of poor mental health is among the top 10% of the pre-period sample. The outcome variable in the top-left panel is our index of symptoms of poor mental health; the outcome variable in the top-right panel is our index of depression services; the outcome variable in the bottom-left panel is our index of whether conditions related to poor mental health negatively affected a student's academic performance. All indices are standardized so that, in the pre-period, they have a mean of zero and a standard deviation of one. The estimates 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.30. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure A.11: Heterogeneous Effects as Evidence of Unfavorable Social Comparisons Mechanism, Controlling for Predicted Susceptibility to Mental Illness



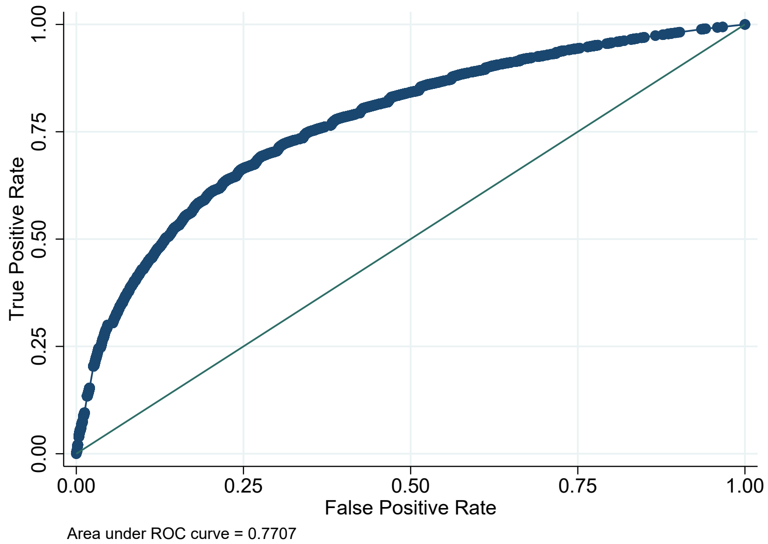
Notes: This figure presents a version of Figure 6 controlling for heterogeneity by the predicted susceptibility to mental illness. Specifically, it presents regressions similar to Equation (1) in which our treatment indicator is interacted with a set of indicators for belonging to a certain sub-population 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 the 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.30. The bars represent 95% confidence intervals. Standard errors are clustered at the college level.

Figure A.12: Relationship between the Index of Symptoms of Depression and Ever Having Been Diagnosed with Depression



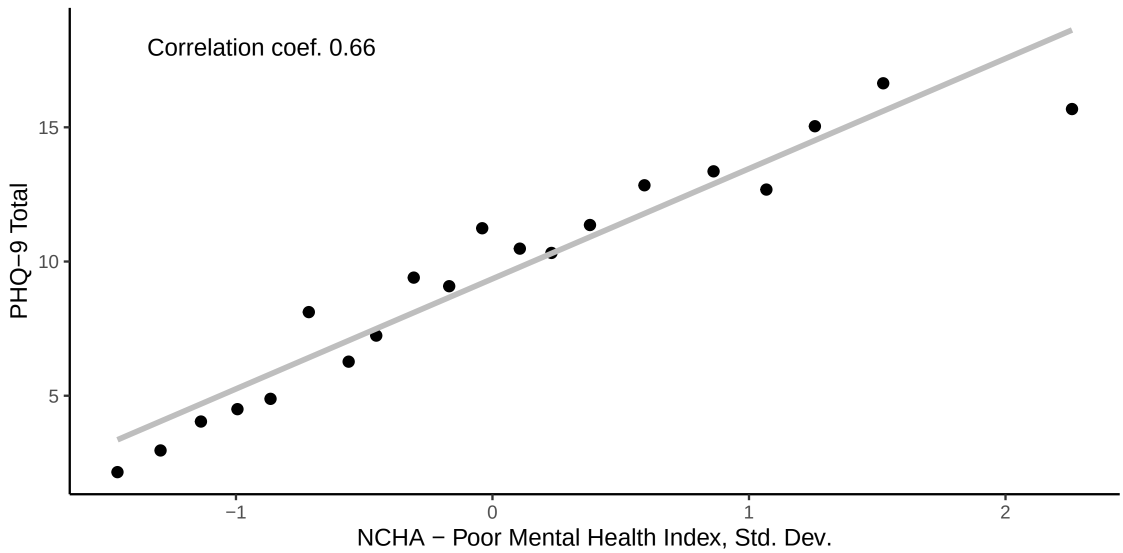
Notes: This figure explores the relationship between our index of symptoms of depression and ever having been diagnosed with depression. Specifically, for each ventile of our index of depression symptoms, the figure plots the fraction of individuals who have ever received a depression diagnosis. The 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.13: Performance of Binary Classifier based on Index of Symptoms of Depression



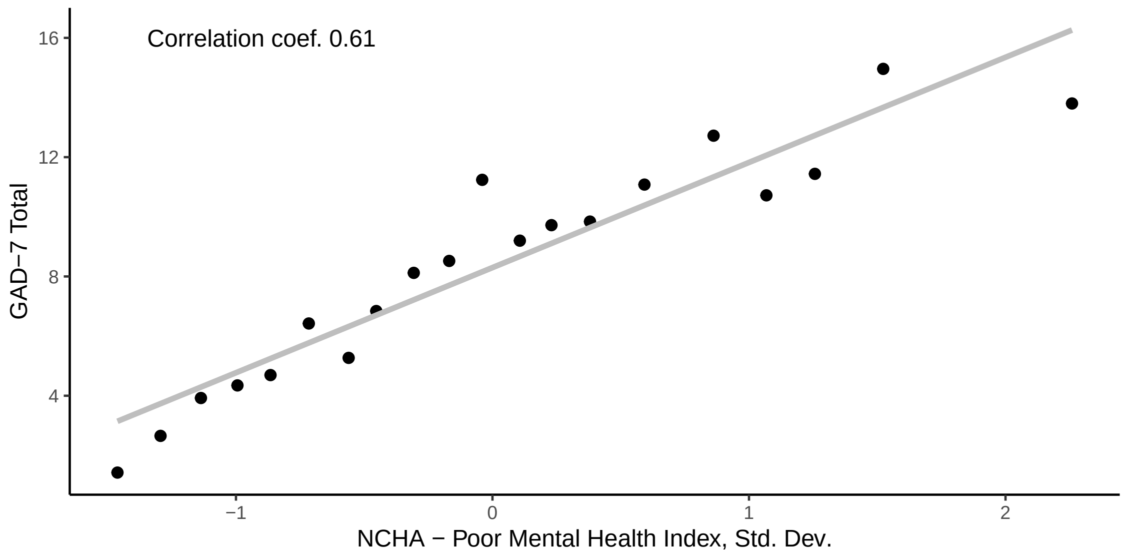
Notes: The figure presents the Receiver-Operating-Characteristic curves of the binary classifiers constructed by running a logit model of ever having been diagnosed with depression on our index of depression symptoms.

Figure A.14: Relationship Between the Index of Poor Mental Health and the PHQ-9



Notes: This figure explores the relationship between our index of poor mental health and PHQ-9 scores in a survey conducted among college students in 2022 and described in detail in Appendix C. Specifically, for each ventile of our index of poor mental health, the figure plots the mean PHQ-9 score. The figure also displays the correlation coefficient between the index of poor mental health and the PHQ-9 score.

Figure A.15: Relationship Between the Index of Poor Mental Health and the GAD-7



Notes: This figure explores the relationship between our index of poor mental health and GAD-7 scores in a survey conducted among college students in 2022 and described in detail in Appendix C. Specifically, for each ventile of the poor mental health index, the figure plots the mean GAD-7 score. The figure also displays the correlation coefficient between the index of poor mental health and the GAD-7 score.

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				
Top incoming test scores	0.93	0.61	0.30	0.07
Medium incoming test scores	0.07	0.35	0.59	0.53
Low incoming test scores	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 with data from the Integrated Postsecondary Education Data System (IPEDS). Colleges are classified as having "top incoming test scores" if their incoming student test scores are in the first (top) quintile of all baccalaureate-granting institutions. Colleges classified as having "medium incoming test scores" have average incoming student test scores in the second and third quintile of all baccalaureate-granting institutions. The remaining colleges are classified as "low incoming test scores." 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 waves; the last row of the table shows the distribution of colleges in the NCHA dataset across Facebook expansion waves.

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
Off-campus Living	0.40	0.47	0.57	0.61
In Fraternity/Sorority	0.14	0.10	0.09	0.09
Work for Pay	0.51	0.57	0.63	0.64
Have Credit Card Debt	0.26	0.29	0.35	0.32
Overweight	0.22	0.28	0.32	0.33
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
Observations	16441	40743	21819	16449

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 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.30.

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.30. 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.30. Standard errors are clustered at the college level.

Table A.5: Heterogeneous Effects by Predicted Susceptibility to Mental Illness

	Index Symptoms Poor Mental Health (1)	Index Depression Services (2)	Index Downstream Effects (3)
Post Facebook Introduction ×	0.068**	0.015	0.033
× 1st Quintile in Mental Illness Susceptibility	(0.033)	(0.029)	(0.031)
Post Facebook Introduction ×	0.067*	0.034	0.062**
× 2nd Quintile in Mental Illness Susceptibility	(0.035)	(0.029)	(0.031)
P-value for coeff. diff. with 1st quintile	*0.878*	*0.039*	*0.019*
Post Facebook Introduction ×	0.079**	0.028	0.079**
× 3rd Quintile in Mental Illness Susceptibility	(0.033)	(0.028)	(0.032)
P-value for coeff. diff. with 1st quintile	*0.358*	*0.228*	*0.000*
Post Facebook Introduction ×	0.089***	0.035	0.072**
× 4th Quintile in Mental Illness Susceptibility	(0.035)	(0.030)	(0.033)
P-value for coeff. diff. with 1st quintile	*0.066*	*0.103*	*0.002*
Post Facebook Introduction ×	0.138***	0.063**	0.121***
× 5th Quintile in Mental Illness Susceptibility	(0.035)	(0.031)	(0.033)
P-value for coeff. diff. with 1st quintile	*0.000*	*0.002*	*0.000*
Observations	361,045	378,456	368,344
College FE	✓	✓	✓
Surve Wave FE	✓	✓	✓
Controls	✓	✓	✓

Notes: This table explores the extent to which the effects of the introduction of Facebook at a college are heterogeneous depending on students' predicted susceptibility to mental illness. Specifically, it presents the estimates from equation (3) in which our indicator for post-Facebook introduction is interacted with a set of indicators for belonging to each quintile of a LASSO-predicted measure of susceptibility to mental illness. The outcome variable in column (1) is our index of symptoms of poor mental health; the outcome variable in column (2) is our index of depression services; the outcome variable in column (3) is our index of whether conditions related to poor mental health affected a student's academic performance. All indices are standardized so that, in the pre-period, they have a mean of zero and a standard deviation of one. We compare the coefficient for the first quintile to other quintiles using a Wald test. 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.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.6: 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. It presents estimates of coefficient β from an equation similar to Equation (4) where we assume that the number of treated semester has a linear effect on mental health and includes survey-wave by college fixed effects: $Y_{icgt} = \beta \times \text{FB}_{gt} \times [t - \max\{\tau_i, \tau_c\}] + \mathbf{X}_i \cdot \gamma + \lambda_{ct} + \varepsilon_{icgt}$. 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 (4) 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. See Footnote 31 for details. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.7: 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. It presents estimates of coefficient β from an equation similar to similar to Equation (4) where we assume that the number of treated semester has a linear effect on mental health and includes survey-wave by college fixed effects: $Y_{icgt} = \beta \times \text{FB}_{gt} \times [t - \max\{\tau_i, \tau_c\}] + \mathbf{X}_i \cdot \gamma + \lambda_{ct} + \varepsilon_{icgt}$. 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. See Footnote 31 for details. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.8: Placebo Check: Predicted Susceptibility to Mental Illness

	Predicted Susceptibility to Mental Illness				
	(1)	(2)	(3)	(4)	(5)
Post Facebook Introduction	0.139	-0.027	-0.003	-0.006	-0.007
	(0.116)	(0.031)	(0.015)	(0.005)	(0.005)
Observations	380,886	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 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) replaces Facebook-expansion-group fixed effects with college fixed effects; column (3) adds controls to the specification in column (1); column (4) replaces Facebook-expansion-group fixed effects with college fixed effects in the specification in column (3); column (5) 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 (3) also includes indicators for geographic region of college (Northeast, Midwest, West, South); such indicators are omitted in columns (2), (4), and (5) because they are collinear with the college fixed effects. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.9: Placebo Check: Demographics

	Age	Female	Year 1	Year 2	Year 3	Height (inches)
Post Facebook Introduction	0.034	-0.017	0.012	-0.009	0.005	-0.026
	(0.081)	(0.015)	(0.034)	(0.034)	(0.019)	(0.053)
Observations	380,886	380,886	380,886	380,886	380,886	380,162
Survey Wave FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
College FE	✓	✓	✓	✓	✓	✓

	White	Black	Hispanic	Asian	Indian	Other Race	International
Post Facebook Introduction	0.013	-0.004	-0.005	-0.008	0.001	0.003*	0.001
	(0.010)	(0.006)	(0.005)	(0.006)	(0.002)	(0.001)	(0.003)
Observations	380,886	380,886	380,886	380,886	380,886	380,886	380,886
Survey Wave FE	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓
College FE	✓	✓	✓	✓	✓	✓	✓

Notes: This table presents a placebo check exploring the effects of the introduction of Facebook at a college on student demographics. Specifically, it presents estimates of coefficient β from Equation (1) with all immutable individual-level characteristics included in the survey as outcome variables. 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. We do not control for covariates related to the outcome variable (e.g., we do not control for race indicators when the outcome variable is White). For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.10: 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 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.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.11: Alternative Index Construction Methods

	Equally-weighted index	Include obs. with missing values	Inverse-covariance index (Anderson 2008)
	(1)	(2)	(3)
Post Facebook Introduction	0.085**	0.073**	0.069**
	(0.033)	(0.031)	(0.030)
Baseline mean	-0.00	0.00	-0.00
Observations	359,827	380,036	359,827
Survey Wave FE	✓	✓	✓
Controls	✓	✓	✓
College FE	✓	✓	✓

Notes: This table explores the robustness of our results to different ways of constructing our index of poor mental health. Column (1) presents our baseline results, which rely on the index construction method described in Section 4. Column (2) presents results on a version of the index that includes observations for which some of the index components are missing and calculates the average value among all non-missing components. Column (3) presents results on an inverse-covariance weighted index (Anderson, 2008). 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.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.12: 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 (4) (Panel (b)). Each column excludes all observations from a particular Facebook expansion group. 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. See Footnote 31 for details. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.13: 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.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.14: 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 the test scores of income undergraduate students is high or medium. Note that 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.30. Standard errors in parentheses are clustered at the college level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A.15: Alternative Difference-in-differences Estimators

	Point Estimate	Standard Error	Lower Bound 95% Confidence Interval	Upper Bound 95% Confidence Interval
Borusyak-Jaravel-Spiess	0.107	0.030	0.048	0.166
Callaway-Sant'Anna	0.113	0.046	0.023	0.203
DeChaisemartin-D'Haultfeuille	0.075	0.073	-0.069	0.218
Sun-Abraham	0.164	0.042	0.081	0.247

Notes: This table presents robustness of our baseline estimate to using the alternative difference-in-differences estimators introduced in Borusyak et al. (2021), Callaway and Sant'Anna (2021), De Chaisemartin and d'Haultfoeuille (2020), and Sun and Abraham (2021). 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 time variable is the survey wave the student participated in and the treatment group variable is given by the semester in which the college attended by the student was granted Facebook access. The regressions underlying the table do not include controls. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.30. See Borusyak et al. (2021), Callaway and Sant'Anna (2021), De Chaisemartin and d'Haultfoeuille (2020), and Sun and Abraham (2021) for a detailed description of how the estimators are constructed and why they are robust to treatment effects heterogeneity across time and treated units.

Table A.16: Baseline Results with Alternative Clustering Methods

	Main regression (1)	Cluster by group (2)	Cluster by group*wave (3)
Post Facebook Introduction	0.085**	0.085***	0.085***
	(0.033)	(0.012)	(0.021)
Wild Bootstrap p-value		*0.015*	
Observations	359,827	359,827	359,827
Survey Wave FE	✓	✓	✓
Controls	✓	✓	✓
College FE	✓	✓	✓
Clusters	420	4	67

Notes: This table explores whether alternative methods of clustering standard errors impact our baseline results. Column (1) displays the baseline specification with the standard errors clustered at the college level; as such, it is identical to column (3) in Table 1. Column (2) presents the estimates with the standard errors clustered at the Facebook expansion group level. Since there are few expansion groups in the data, we also report a wild bootstrap *p*-value which corrects for the few-clusters problem (Cameron et al., 2008; Roodman et al., 2019). Finally, column (3) presents the estimates with the standard errors clustered at the Facebook expansion-group by survey-wave 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. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.30. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A.17: Missing Values

	Any Missing Values	Total Missing Values	Index of Missing Values
	(1)	(2)	(3)
Post Facebook Introduction	0.003	0.014	0.010
	(0.008)	(0.067)	(0.049)
Baseline mean	0.07	0.27	-0.00
Observations	380,886	380,886	380,886
Survey Wave FE	✓	✓	✓
Controls	✓	✓	✓
College FE	✓	✓	✓

Notes: This table addresses the potential reduction in the stigma associated with mental illness as a result of the introduction of Facebook. Specifically, it presents estimates of coefficient β from Equation (1) with three different ways of aggregating missing responses. In Column (1), 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 (2), the outcome is the total number of questions composing the index of poor mental health left unanswered by a respondent. In Column (3) the number of unanswered questions is standardized using means and standard deviations from the pre-period. 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.30. Standard errors in parentheses are clustered at the college level.* p<0.1, ** p<0.05, *** p<0.01.

Table A.18: 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 self-reported 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 self-reported 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.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.19: 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 x Off-Campus Living	0.094**	0.001	0.020***	0.041***
	(0.038)	(0.002)	(0.006)	(0.015)
Baseline mean	5.71	0.70	0.38	0.00
Observations	374,041	369,422	377,503	379,864
Survey Wave FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
College FE	✓	✓	✓	✓

Notes: This table 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 of drinks a typical student had the last time she partied, winsorized at nine, 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.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.20: Effects 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 table explores the effects of the introduction of Facebook on the difference between perceptions of alcohol use and self-reported use. Specifically, it presents estimates of coefficient β from Equation (1). We compare each respondent's perceptions to actual self-reported usage in the respondent's college and survey wave. Column (1) considers the absolute value of the difference between the respondent's perception of the typical drink count in her college and the 'actual' average number of drinks that students in the respondent's college and survey wave reported consuming (all drink counts are winsorized at nine). Column (2) considers the absolute value of the difference between the respondent's perception of the share of students drinking at her college and the actual share of students in the respondent's college and survey-wave who self-reported drinking at least once in the past 30 days. Column (3) is an indicator variable that equals one if the respondent's perception of whether the typical student at her college drinks daily differs from the behavior of the 'actual' typical student in the respondent's college and survey wave. We consider the typical student at a college a daily drinker if the modal response within a given college and survey-wave is using alcohol in at least 20 days out of the last 30 days. All columns are in original units, besides column (4) which is an index of the outcomes in columns (1) through (3). 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. For a detailed description of the outcome, treatment, and control variables, see Appendix Table A.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.21: Effects 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/computer 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.30. Standard errors in parentheses are clustered at the college level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A.22: Effects of the Introduction of Facebook on Assaults and Sexual Violence

	Assault, fight last year (1)	Sexual assault last year (2)	Sexual threat last year (3)	Abusive relationship last year (4)	Assault index (5)
Post Facebook Introduction	0.002	-0.006	0.001	0.005	0.000
	(0.008)	(0.008)	(0.004)	(0.006)	(0.025)
Baseline mean	0.15	0.15	0.04	0.15	-0.00
Observations	380,809	380,803	379,916	379,539	378,915
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. The first four columns are binary outcomes, and Column (5) is an index based on Columns (1)-(4). 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.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.23: Effects of the Introduction of Facebook on Relationships

	Straight (1)	Single (2)	Experienced relationship difficulties (3)	Partners number (4)	Relationship index (5)
Post Facebook Introduction	0.000	-0.005	0.015	0.053	0.024
	(0.005)	(0.009)	(0.014)	(0.032)	(0.024)
Baseline mean	0.95	0.58	0.46	1.40	-0.00
Observations	376,505	377,078	375,278	376,118	364,425
Survey Wave FE	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓
College FE	✓	✓	✓	✓	✓

Notes: This table explores the effects of the introduction of Facebook at a college on outcomes related to 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. Columns (1)–(3) are binary variables for whether the respondents are straight, single, and self-reported experiencing relationship difficulties. Column (4) is the number of sexual partners in the past year, winsorized at nine. Column (5) is an index, based on columns (1)–(4). 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.30. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.24: Effects on Perceptions Related to Sexual Behavior

	Num Partners	Oral Sex	Vaginal Intercourse	Anal Sex	Sexual Perceptions Index
	(1)	(2)	(3)	(4)	(5)
Post Facebook Introduction	0.011	0.027	0.024	0.035**	0.035*
	(0.018)	(0.021)	(0.021)	(0.018)	(0.020)
Baseline mean	0.00	0.00	-0.00	-0.00	-0.00
Observations	369,810	366,130	365,884	364,528	361,108
Survey Wave FE	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓
College FE	✓	✓	✓	✓	✓

Notes: This table explores the effects of the introduction of Facebook at a college on students' perceptions related to sexual behavior. Specifically, it presents estimates of coefficient β from Equation (1). Column (1) estimates the effect on the perceived number of sexual partners a typical student had sex with, winsorized at nine. Column (2)-(4) estimates the effect on the number of times a typical student is perceived to have engaged in sexual intercourse. Column (5) is an equally weighted index based on columns (1)-(4). All outcomes 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. Standard errors in parentheses are clustered at the college level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A.25: Effect of the Introduction of Facebook on Substance Use

	Cigarettes (1)	Cigars (2)	Smokeless tobacco (3)	Marijuana (4)	Cocaine (5)	Amphetamines (6)	Rohypnol (7)	MDMA (8)	Other (9)	Index (10)
Post Facebook Introduction	0.009	0.002	0.001	0.010	-0.000	-0.003	0.000	-0.001	0.006	0.016
	(0.009)	(0.004)	(0.006)	(0.007)	(0.003)	(0.003)	(0.001)	(0.001)	(0.005)	(0.026)
Baseline mean	0.25	0.07	0.04	0.18	0.02	0.05	0.00	0.00	0.04	0.00
Observations	379,708	379,002	376,399	378,805	379,157	379,257	379,160	243,555	367,087	380,540
Survey Wave FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
College FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table explores the effects of the introduction of Facebook at a college on substance 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. Columns (1)–(9) are binary variables indicating whether the respondent used the drug within the last 30 days. Column (10) is an index based on the standardized average of the other columns. Since many answers are missing for one specific question, we take the average for all non-missing questions. 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.30. Standard errors in parentheses are clustered at the college level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A.26: Effects on Perceptions of Using Illicit Substances

	Cigarettes (1)	Cigars (2)	Smokeless tobacco (3)	Marijuana (4)	Cocaine (5)	Amphetamines (6)	Rohypnol (7)	MDMA (8)	Other (9)	Index (10)
Post Facebook Introduction	0.018	0.001	-0.007	0.015	0.004	-0.015	-0.002	-0.005	0.009	0.026
	(0.012)	(0.013)	(0.011)	(0.009)	(0.014)	(0.013)	(0.009)	(0.011)	(0.021)	(0.029)
Baseline mean	0.93	0.60	0.60	0.84	0.38	0.53	0.32	0.37	0.50	0.00
Observations	378,668	377,846	377,077	377,750	376,988	375,614	374,885	242,380	361,088	379,329
Survey Wave FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
College FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table explores the effects of the introduction of Facebook at a college on students' perceptions related to substance use. Specifically, it presents estimates of coefficient β from Equation (1). Columns (1)–(9) are binary variables estimating the effect on whether respondents think that a typical student used the drug in the past 30 days. Column (10) is an index based on the standardized average of the other columns. Since many answers are missing for one specific question, we take the average for all non-missing questions. 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. Standard errors in parentheses are clustered at the college level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A.27: Comparison of the Authors' 2022 Mental Health Survey and the NCHA Sample

Variable	Authors' 2022 Survey	NCHA Survey
International	0.030	0.033
Female	0.690	0.642
White	0.643	0.780
How many times had symptom (1-7)	3.582	2.713
Had symptom/disorder (0-1)	0.232	0.082
Took up mental health service (0-1)	0.132	0.044

Notes: This table presents descriptive statistics for the NCHA respondents and participants in the survey we conducted for external validation (described in Appendix C). 'How many times had symptom' refers to the questions in the poor mental health index asking participants how many times they [felt hopeless, felt overwhelmed, felt exhausted, felt very sad, felt severely depressed, considered suicide, or attempted suicide] in the last year. 'Had symptom/disorder' refers to questions in our index asking participants whether they [had depression, had anorexia, had anxiety disorder, had bulimia, had seasonal affect disorder]. 'Took up mental health service' refers to questions asking participants whether they were diagnosed with depression in the last year, are currently in therapy for depression, or are currently taking medication for depression.

Table A.28: Coefficients predicting 10≤PHQ-9 or 10≤GAD-7

	10≤PHQ-9 Coefficients			10≤GAD-7 Coefficients		
Variable	OLS	Logit	LASSO	OLS	Logit	LASSO
Intercept	-0.228	-4.601	-0.199	-0.224	-4.741	-0.756
Depression Symptoms						
Last year felt hopeless	0.040	0.206	0.406	0.016	0.088	0.177
Last year felt overwhelmed	0.012	0.103	0.128	0.006	0.080	0.095
Last year felt exhausted	0.025	0.185	0.294	0.020	0.179	0.298
Last year felt very sad	-0.007	-0.046	0.000	0.001	0.032	0.057
Last year severely depressed	0.088	0.461	0.920	0.074	0.349	0.686
Last year seriously considered suicide	0.016	0.205	0.175	0.018	0.105	0.092
Last year attempted suicide	0.034	0.325	0.000	0.040	0.296	0.009
Last year depression	-0.033	-0.140	0.000	-0.064	-0.288	0.000
Other Symptoms						
Last year anorexia	-0.020	-0.067	0.000	-0.019	-0.200	0.000
Last year anxiety disorder	0.054	0.299	0.074	0.276	1.530	0.618
Last year bulimia	-0.051	-0.306	0.000	0.031	0.170	0.000
Last year seasonal affect disorder	0.031	0.187	0.000	-0.022	-0.131	0.000
Depression Services						
Last year depression diagnosis	0.116	0.881	0.140	0.006	0.072	0.000
Therapy depression	0.080	0.558	0.046	0.032	0.182	0.000
Current medication depression	-0.158	-1.125	-0.191	-0.193	-1.174	-0.313

Notes: This table presents the coefficients predicting a PHQ-9 score of at least 10 (moderate or severe depression) and a GAD-7 score of at least 10 (moderate or severe anxiety). In columns (2) and (5), the coefficients are created by regressing the binary outcomes on the components of our index of poor mental health using a linear probability model. Columns (3) and (6) are based on a logistic regression, and in columns (4) and (7) the binary outcomes are predicted using a LASSO regression (the coefficients in columns 3-4, 6-7 are in log-odds units). The regressions are based on data from the mental health survey conducted for external validation and described in Appendix C.

Table A.29: Effects on Predicted Depression and Anxiety

	10 ≤ PHQ-9			10 ≤ GAD-7		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Logit	LASSO	OLS	Logit	Lasso
Post Facebook Introduction	0.023**	0.022**	0.022**	0.019**	0.017**	0.022***
	(0.009)	(0.009)	(0.009)	(0.007)	(0.007)	(0.008)
Predicted baseline mean	0.25	0.25	0.42	0.16	0.17	0.34
Observations	359,827	359,827	359,827	359,827	359,827	359,827
Survey Wave FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
College FE	✓	✓	✓	✓	✓	✓

Notes: This table presents the effects of the introduction of Facebook on predicted moderate or severe depression, based on the PHQ-9 index (10≤PHQ-9), and predicted moderate or severe anxiety, based on the GAD-7 index (10≤GAD-7). The coefficients used to predict 10≤PHQ-9 and 10≤GAD-7 are described in Table A.28. In columns (1) and (4), the coefficients are created using a linear probability model, in columns (2) and (5), they are created using a logistic regression, and in columns (3) and (6) they are created using a LASSO regression. After creating measures for predicted depression and anxiety using the PHQ-9 and GAD-7 indices, respectively, we estimate the effects of the introduction of Facebook on these measures. Specifically, the table presents estimates of coefficient β from Equation (1). 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 more details see Appendix C. Standard errors in parentheses are clustered at the college level. * p<0.1, ** p<0.05, *** p<0.01.

Table A.30: 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.30 (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.30 (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 = no; 0 = yes.
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 = at least one hour; 0 = 0 hours.
Overweight	Use recoded BMI (BMI $= kg/m^2$); Coding: 1 = if recoded BMI>25 (indiciating 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.30 (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 college 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 college 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 same college 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.30 (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 assalted?" 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 last year	Question: "Within the last school year, have you been in a relationship that was: sexually abusive?" Scale: yes, no. Coding: 1 = yes, 0 = no.
Assault index	The index is constructed aggregating the four variables above following the same procedure as the Index Poor Mental Health.
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}.
Experienced relationship difficulties	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.
Relationship index	The index is constructed aggregating the five variables above following the same procedure as the Index Poor Mental Health.
Sexual Behavior Perceptions	
Num Partners	Question: "Within the last school year, with how many partners do you think the typical student at your school has had sex (oral, vaginal, or anal)?" Open numeric response. Coding: Winsorized at 9. The variable is standardized so that it has a mean of 0 and a standard deviation of 1 in the pre-period.
Oral Sex	Question: "How many times within the last 30 days do you think the typical student at your school has had: Oral sex?" Coding: 1 = {Never}; 2 = {Not in last 30 days}; 3 = {1–2 times}; 4 = {3–4 times}; 5 = {5–6 times}; 6 = {7–8 times}; 7 = {9–10 times}; 8 = {11 or more times}. The variable is standardized so that it has a mean of 0 and a standard deviation of 1 in the pre-period.
Vaginal Intercourse	Question: "How many times within the last 30 days do you think the typical student at your school has had: Vaginal Intercourse?" Coding: same as above.
Anal Sex	Question: "How many times within the last 30 days do you think the typical student at your school has had: Anal Intercourse?" Coding: same as above.
Sexual Perceptions Index	The index is constructed aggregating the four variables above following the same procedure as the Index Poor Mental Health.

Table A.30 (cont'd): Variable definition, construction, and associated NCHA survey questions

Variable	Description
Drug use	
Cigarettes	Question: "Within the last 30 days, on how many days did you use: cigarettes?" Scale: 0 = {never used; have used, but not in last 30 days}; 1 = {1-2 days; 3-5 days; 6-9 days; 10-19 days; 20-29 days; 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.
Perceptions variables	Within the last 30 days, how often do you think the typical student at your school used: State your best estimate: [Drug]. Scale: Coding: 1 = {Used daily, One or more days}; 0 = Never Used.
Index	The index of drug use (perceptions of drug use) is constructed aggregating the nine variables above (or perceptions related to their average use in college) following the same procedure as the Index Poor Mental Health not discarding observations when one of the nine variables above is missing.
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.

Table A.30 (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.
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 the questions composing 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.
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 calculate the total number of question that a respondent left unanswered; iii) we stadnardized the total so it has a mean of 0 and a standard deviation of 1 in the pre-period.
Other variables	
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.
Height	Question: "What is your height in feet and inches?"
Volunteer	Question: "How many hours a week do you volunteer?"; Coding: 1 = at least one hour; 0 = 0 hours.
First-year	Question: "Year in school"; Coding: 1 if chose first year undergraduate; 0 otherwise.