PhD Studies Hurt Mental Health, but Less than Previously Feared*

Matti Keloharju

Aalto University School of Business, CEPR, and IFN

Samuli Knüpfer

Aalto University School of Business, BI Norwegian Business School, and IFN

Dagmar Müller

Swedish Public Employment Service

Joacim Tåg

IFN and Hanken School of Economics

December 2023

Abstract

We study the mental health of PhD students in Sweden using comprehensive administrative data on prescriptions, specialist care visits, hospitalizations, and causes of death. We find about 7% (5%) of PhD students receive medication or diagnosis for depression (anxiety) in a given year. These prevalence rates are less than one-third of the earlier reported survey-based estimates, and even after adjusting for difference in methodology, 43% (72%) of the rates in the literature. Nevertheless, PhD students still fare worse than their peers not pursuing graduate studies. Our difference-in-differences research design can attribute at least 80% of this health disadvantage to the time in the PhD program. This deterioration suggests doctoral studies causally affect mental health.

Keywords: PhD studies, mental health, depression, anxiety, suicide

* Corresponding author: Matti Keloharju, Aalto University School of Business, P.O. Box 21220, FI-00076 Aalto, Finland, tel. +358-40-353-8043, e-mail matti.keloharju@aalto.fi. This work was supported by the Academy of Finland (Grant 319316 to M.K.) and Marianne and Marcus Wallenberg foundation (Grant 2020.0049 to J.T.). We thank Teodor Duevski, Sina Ghavamabadi, Maija Löyskä, and Kimmo Niinimäki for excellent research assistance. The text was revised with assistance of ChatGPT. Declarations of interest: none.

1. Introduction

Many PhD students are overworked and overstressed, and their mental health is often thought to suffer from work stress (e.g., Forrester, 2021; Woolston, 2017). Satinsky et al.'s (2021) recent meta-analysis finds 24% (17%) of PhD students have clinically significant symptoms of depression (anxiety), and even suicidal ideation is not uncommon. Reports highlighting the prevalence of mental health issues, believed to be caused by the educational program and its environment, have led to calls for stronger policy responses (Council of Graduate Schools, 2021; Evans et al., 2018; Forrester, 2021; Nature 2019a; Nature 2019b; Woolston, 2017) to what Evans et al. (2018) label as "mental health crisis in graduate education."

Existing research on the prevalence of mental health issues of PhD students relies on cross-sectional surveys conducted on samples that are often small, heterogenous, and lack appropriate benchmarks. Previous work indicates that such survey methods tend to overstate the prevalence of mental health issues (Levis et al., 2020). This overestimation, coupled with a lack of longitudinal data and appropriate benchmarks, undermines accurate evaluation of both mental health status among PhD students and the causal impact of doctoral studies. These limitations in data quality not only hamper researchers' understanding of the mental health challenges among PhD students but also make it harder for policy makers to manage them.

We address these concerns by systematically analyzing medically validated indicators of the mental health of PhD students in Sweden. We use administrative data on prescriptions, specialist care visits, hospitalizations, and causes of death in the entire country over the 2005–15 period. We compare the prevalence of depression, anxiety, and suicide among three groups: PhD students, Master's graduates not pursuing a PhD (in Sweden, PhD programs generally require admitted students to have a completed Master's degree), and the general population. In a longitudinal analysis, we follow the mental health of four cohorts of PhD students and their peers in the nine years surrounding the entry into the PhD program (or graduation from the Master's program).

We find that 6.7% of PhD students receive treatment or a diagnosis for depression in a given year. After adjusting for methodological differences, we estimate that PhD students' depression rate is no more than 43% of the corresponding meta-estimate of 24% by Satinsky et al. (2021). We also find lower prevalence of anxiety and completed suicides than previous research suggests. For example, prior studies find that the prevalence of suicidal ideation in PhD students can exceed 10%, whereas our findings suggest these ideations almost never culminate in

completed suicides. In sum, these results offer hope graduate studies may be less harmful to mental health than previously feared.

Our benchmark groups allow us to gauge how PhD students differ from their peers and the population. We find the prevalence rate of depression among Master's graduates not pursuing a PhD education is 5.5%, or 1.2% less than that for PhD students. With this gap, the depression rate for PhD students aligns closely with the 7.0% rate seen in individuals aged 20–39 but remains lower than the 9.0% in the 18–70 age group. Anxiety follows a similar pattern, with PhD students showing a 0.8% higher prevalence compared to Master's graduates.

Two possible explanations may account for the elevated mental health problems in PhD students. One is self-selection: those with existing mental health conditions are more likely to enroll in PhD programs. The other is that PhD students develop mental health problems during their studies. Our data, which spans years before and after entering the program, enables us to discern between these explanations.

We employ difference-in-differences regressions to examine mental health outcomes. The model includes a treatment indicator for PhD student status, event time indicators surrounding PhD program entry, and their interactions. We estimate this regression using the Callaway and Sant'Anna (2021) method and condition on covariates using doubly robust inverse probability weighting by Sant'Anna and Zhao (2020). These covariates include gender, age at entry to PhD program or Master's graduation, parental mental health, and high-school GPA.

We find most of the disparity in mental health appears to emerge during the program, particularly in the later years. For example, the treatment effect of 1.0% for depression equals 81% of the total depression difference between PhD students and their peers. This inference relies on the assumption that the mental health trajectory of PhD students' peers, appropriately weighted by covariates, serves as a reasonable counterfactual for the PhD students. Our data show similar mental health trends for both groups before starting graduate studies, allowing us to attribute the effects to the program. Given the limitations of past research relying on cross-sectional surveys, these findings present a more credible case on the adverse causal effects of doctoral studies on mental health.

Our study relates to a large body of literature on the mental health effects of the work environment and the underlying stressors (see, e.g., Stansfeld and Candy, 2006; Harvey et al., 2017). Our work contributes to this literature by focusing on the prevalence and development of mental health issues among PhD students who face a unique set of challenges and pressures related to aspects such as resources, work methodologies, mentor relationships, and dynamics within the academic community (see e.g., Hyun et al., 2006; Pyhältö et al., 2012). Levecque et

al. (2017) suggest that the mental health issues of PhD students are linked to work and organizational contexts. More recent studies delve into these nuances of academic life, particularly mentoring and peer interactions (e.g., Broström 2019; Corsini et al., 2022; Wuestman et al., 2023).

This paper is organized as follows. Section 2 describes our data sources, defines the variables, and explains how we selected our sample. Section 3 compares our evidence on the prevalence of mental health issues among PhD students with prior evidence and shows how entering the doctoral program alters their mental health status compared with their peers. Section 4 concludes with policy recommendations.

2. Data

2.1. Data sources

The data combine information on individuals from two sources, which are linked together using masked social security numbers.

National Board of Health and Welfare. The health data come from the National Board of Health and Welfare, which maintains comprehensive records of hospital visits, open care offered by specialized doctors, prescriptions, and causes of death in Sweden. The hospital and specialized open care data include primary and secondary diagnoses along with the associated four-digit ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th revision) codes for each diagnosis. The prescription data include all prescriptions along with the associated ATC-code (Anatomical Therapeutic Chemical Classification System) with at least four digits. These ATC codes are further translated into diagnoses using established medical literature.

Statistics Sweden. Data for all variables are sourced from Statistics Sweden's registers. Specifically, we use the LISA database for employer information, immigration status, level and field of educational, registration for master's and PhD studies, and examination year. Other registers consulted include the Population Register for age and gender, the Multigenerational Register for biological parent data, and the Education Register for high school GPA. The LISA database encompasses the entire Swedish population aged 16 and above who are residents as of year-end. Compiled from multiple government authorities, this database spans 2001–15, serving as our base register. Additional variables are integrated from other registers. Our analysis focuses

on individuals aged 18 to 70 and excludes the few individuals with reused social security numbers.

2.2. Sample selection

Our sample consists of all individuals who started PhD studies or received a Master's degree at the age of 35 or less in Sweden in 2009–11. We exclude individuals who have unknown parents or unknown high school GPA, or that have immigrated to Sweden. We merge these data with records of prescriptions, specialist care visits, hospitalizations, and causes of death in Sweden in 2005–15 (the prescribed drug register begins in July 2005). Below we specify how we selected the sample components used in our analysis.

PhD students. The sample of PhD students contains all individual-year observations fulfilling the following criteria: started PhD studies in 2009–11 in a non-medical field at the age of 35 or lower, born in Sweden, and has a known high school GPA and known parents. In addition, the observation must be from 0–4 years after the start year of the PhD studies.

Peers. The sample of peers contains all individual-year observations fulfilling the following criteria: graduated with a Master's degree in 2009–11 from a non-medical field at the age of 35 or lower, did not start a PhD, born in Sweden, and has a known high school GPA and known parents. In addition, the observation must be 0–4 years after the graduation year.

Treatment group. The treatment group contains all individuals in the sample of PhD students who started their PhD studies in 2009–11. Each individual is followed for nine years in total: from four years before to four years after their start year of PhD studies. The restriction on the start year ensures the panel is balanced.

Control group. The control group contains all individuals in the sample of peers who graduated in 2009–11 and did not start a PhD degree. Each individual is followed for nine years in total: from four years before to four years after their graduation year. The restriction on the graduation year ensures the panel data is balanced.

Population. The sample of population includes all individual-year observations from 2005–15 when the age of the individual is between 18 and 70.

Age peers. Subsample of population as defined above, constrained to 20–39-year-olds.

2.3. Variables

Depression and anxiety. An individual is defined to have depression (anxiety) each year if she has a record of diagnosis or prescription assigned to depression (anxiety) during that year.

Depression (anxiety) diagnoses are indicated with ICD-10 codes F32–F33 (F40–F41) and with ATC codes N06A (N05B, N05C) (Schäfer et al., 2010; Fishman et al, 2003).

Suicide. An individual is defined to have committed suicide in year *t* if she has died in that year or the next and the cause-of-death database considers her cause of death to be suicide.

Parental mental health. Parents are defined to have prior depression (anxiety) if at least one parent has at least one record of diagnosis assigned to depression (anxiety) in 2001–04. Parents' diagnoses can either originate from the hospital or specialized open care data. If parents have no prior depression or anxiety, they are classified as healthy; otherwise, they are classified as non-healthy.

PhD status. PhD status is assigned to an individual-year observation conditional on the individual being registered as PhD student for the spring or fall semester during the year.

Start year of PhD studies. The start year of PhD studies is the first year an individual is recorded as having PhD status.

Graduation year. The graduation year is the first year an individual is recorded as having a Master's degree as the highest education attained. Following Statistics Sweden methodology, the highest education is measured at the end of spring semester each year.

Hard sciences and soft sciences. The field of study is coded based on the Swedish Educational Terminology (SUN). Based on previous literature, we use its first digit to divide it into two categories, hard and soft sciences (Biglan, 1973; Stoecker, 1993). Hard sciences include natural sciences, mathematics and computing; engineering and manufacturing; and agriculture, forestry, and veterinary medicine. Soft sciences represent all other fields and include teaching methods and teacher education; humanities and arts; social sciences, law, commerce, and administration; health care and nursing; social care (excluding medicine); services; and unknown. Medicine is considered as a separate field and, following existing literature (Satinsky et al., 2021), excluded from the analysis.

Field of study. The data contain information on the field of study during Master's studies but not during PhD studies. We derive the field of a PhD student based on the establishment she works at. To assign a field to an establishment, we consider all PhD students who worked in the establishment and graduated with a Master's degree in 2006–15. If the majority of the individuals in an establishment studied hard sciences in the year they graduated with a Master's degree, we assign the field of the establishment as hard sciences. Other establishments belong to soft sciences. We also separately categorize individuals as studying in the medical field. For graduated individuals not pursuing a PhD degree, the field of study is defined by the recorded field of study in their graduation year.

3. Results

3.1. Prevalence of mental health problems

When reporting the prevalence of anxiety and depression in Figure 1, we use the samples of population, age peers, treatment group (referred to as PhD students) and control group (referred to as "Other Master's graduates") defined above. The treatment (control) group observations preceding the start year of PhD studies (graduation year) are excluded. In the statistics of PhD students' subgroups, the students missing a classifying variable are dropped.

Figure 1 Panel A reports the prevalence of depression among PhD students, other Master's graduates, two population samples representing different age groups, and various PhD student subsamples. In a given year, 6.7% of the PhD students receive medication or diagnosis for depression, while the corresponding prevalence for other Master's graduates is 5.5%. Sample splits among the PhD student population suggest depression is more prevalent among women, students who were older at the time of enrolling in a PhD program and had lower GPA in high school, students of soft sciences, and students who have a parent with a prior depression or anxiety diagnosis. The highest prevalence of depression among the above sample splits, recorded for PhD students with parents struggling with mental health, is the same as that for the general population of 18–70-year-olds (9.0%). Table 1 reports full details.

Previous studies report a higher prevalence of depression than what we find in our sample of PhD students. Specifically, a meta-analysis suggests that 24% exhibit clinically significant symptoms of depression, which is 2.6 times higher than our findings (Satinsky et al., 2021). Given that these prior estimates are based on screening tools, they are not necessarily directly comparable with estimates obtained from medical data. To make our results more comparable, we present a derived estimate. The derived estimate of 10.4% corresponds to 43% of the meta-analysis estimate and falls notably below the meta-analysis estimate's 95% confidence interval's lower bound of 18%.

The derived estimate multiplies our base estimate of 6.7% with two factors. The first factor, 1.2, captures differences between the measures used in this study and the meta-analysis. It is obtained by dividing screened prevalence of clinically significant depression in a representative sample of 18–70-year-olds in the Swedish population (Johansson et al., 2013) with prevalence of medically validated depression in the population of the same age range in our data. The second factor, 1.29, eliminates the possible effects of age on treatment behavior, including the likelihood to seek treatment (Alonso et al., 2004; Moitra et al., 2022), the type of treatment received

(Forslund et al, 2020), and the healthcare setting utilized. The factor is computed as the ratio of medically validated depression among 18–70-year-olds and 20–39-year-olds in the Swedish population. The latter group has the same average age, 30, as PhD students and other Master's graduates in our sample, and its age range also reflects that of these populations.

Given realistic assumptions and bounds for unobserved parameters, the derived estimate cannot have any significant negative bias (see supplement for details). Rather, if anything, our derived estimate is inflated, making comparisons between it and the meta-analysis estimate conservative.

Figure 1 Panel B reports on the prevalence of anxiety. In a given year, 5.1% of the PhD students have anxiety, while the corresponding prevalence for Master's graduates (general population of 20–39-year-olds) is 4.2% (6.1%). The sample splits produce qualitatively similar results for anxiety and depression except for parental mental health that does not appear to generate differences in anxiety.

A meta-analysis estimates the prevalence of clinically significant symptoms of anxiety at 17%, i.e. at 3.4 times the prevalence in our sample. By using the same method as for depression to compute the derived estimate for PhD students' prevalence of anxiety, we arrive at $5.1\% \times 1.47 \times 1.64 = 12.2\%$. This number is 72% of the corresponding meta-analysis estimate and narrowly within its 95% confidence interval. As for depression, our estimate for the difference between the derived estimate and the meta-analysis estimate for anxiety is conservative (see supplement for details).

Past research finds prevalence of suicidal ideation among PhD students can exceed 10% (Satinsky et al., 2021). However, our results suggest these ideations are highly unlikely to lead to completed suicides: there are 18,796 individual-PhD year observations in our sample but only one suicide. This small propensity makes it challenging to compare the smaller PhD student group to the larger benchmark groups. Nevertheless, the propensity is only 48% of the corresponding propensity in the population of 20–39-year-olds.

3.2. Why do PhD students have worse mental health than their peers?

While our data show lower rates of depression and anxiety among PhD students compared to prior research, these students still fare worse than their peers. This disparity may stem from two factors. First, Master's graduates with mental health issues or a higher likelihood of

developing them might opt for PhD studies. Second, PhD students may enter the program in a similar state of mental health as their peers but deteriorate over time.[1]

Panels A and C of Table 1 detail the attributes of PhD students compared to their peers. Panel A shows these students are slightly older and more often male. Panel C reports the parents of PhD students exhibit lower rates of depression and anxiety, and as expected, PhD students have substantially higher high-school GPAs. These patterns offer no systematic evidence that would lead us to expect PhD students are more prone to mental health problems than their peers.

To understand the factors affecting the poorer mental health of PhD students, we employ longitudinal data that cover periods both before and after entering the program. Three notable patterns emerge in Figure 2, Panel A, which compares the prevalence of depression between PhD students around the start of their doctoral studies and other Master's graduates around their graduation. First, depression prevalence more than doubles during the seven years the subjects are followed, likely reflecting the effect of age on health (Baxter et al., 2014; Ferrari et al., 2013) and treatment behavior (Alonso et al., 2004; Forslund et al., 2020; Moitra et al., 2022). Second, PhD students have higher prevalence rates, suggesting potential differences in vulnerability to mental health problems among the two groups. Third, the prevalence rates increase more for PhD students upon entering the program than for their peers whereas the two groups show a similar pre-trend. Panel B shows anxiety follows similar patterns.

The decline in mental health following entry into the PhD program, along with the pre-existing differences, suggests that both factors may contribute to the inferior mental health of PhD students. We further investigate these contributing forces by using a difference-in-differences linear probability model. This model estimates annual prevalence of depression using a treatment indicator for PhD students, event time indicators that mark the first year in the PhD program or the year of Master's graduation, and their interactions. We estimate this regression using the recently introduced Callaway and Sant'Anna (2021) method and condition on covariates using doubly robust inverse probability weighting by Sant'Anna and Zhao (2020). The covariates include gender, age at entry to PhD program or Master's graduation, parental mental health, and high-school GPA. The benefit of using this estimator is that it circumvents recently highlighted inference problems arising from treatment effect heterogeneity in

[1] PhD students might also improve their mental health during the program compared to their peers. This health advantage could arise from higher intellectual rewards, greater independence, and other beneficial aspects of PhD studies. This explanation is not able to deliver the observed health disadvantage of PhD students unless accompanied by strong negative selection on mental health into the PhD program. As we detail below, the data only shows weak selection and no evidence of improvement in PhD students' mental health during the program.

difference-in-differences designs and can flexibly account for covariates (Baker et al. 2022; Roth et al., 2023).

Panel A in Table 2 presents the difference-in-differences estimates. In the pre-treatment years before starting their PhD, doctoral students have a 0.09% higher prevalence of depression whereas the pre-treatment estimate is –0.005% for anxiety. The *z*-values of 0.74 and –0.04 below the estimates show the differences are well below conventional thresholds for statistical significance. These small and insignificant estimates suggest that there are no meaningful differences in how mental health develops prior to treatment when we account for compositional differences between the treatment and control groups.

The coefficients for the post-treatment interactions help us understand the contribution of post-admission years to PhD students' poorer mental health. For added clarity, Panels C and D in Figure 2 illustrate these coefficients along with their 95% confidence intervals. A visual review confirms that PhD students and other Master's graduates move in parallel before the treatment year. For depression, all interactions are positive after entering the PhD program and become statistically significant in year three. In the fifth year, the point estimate for the interaction term suggests depression rates are 2.1% higher for PhD students (z=4.4). Anxiety shows a similar, although weaker pattern. The largest difference occurs in the fourth year, with an estimate of 1.4% (z=3.2).

These findings help to understand how much of the unconditional differences between PhD students and their peers are attributable to the PhD program. The average unconditional differences in Table 1, Panel A equal 1.2% and 0.8% for depression and anxiety. Table 2, Panel A, reports that the average post-treatment coefficients are 1.0% and 0.7% (z-values 3.0 and 2.3). Accordingly, the treatment effects amount to 81% of the depression difference and 88% of the anxiety difference. This calculation points to the PhD program as the main contributor to PhD students' poor mental health.

Another way for evaluating our estimates' magnitudes compares them to the 8.9% prevalence rate of depression among fifth-year PhD students. Using the fifth-year coefficient in Table 2, Panel A shows the years in the PhD program account for 24% (2.1%/8.9%) of the overall rate of depression whereas it is 15% (0.9%/5.8%) for anxiety. These findings suggest a significant portion of doctoral students' mental health issues stem from their experiences in the program.

Since most of the decline in PhD mental health occurs during the doctoral program, it is natural to ask which students are most affected. Table 2, Panel B reports separate treatment effects by the characteristics reported in Table 1, Panel B, emanating from regressions run

separately in each subsample. The panel reports the treatment effects and their differences in the subsamples along with *z*-values that assess statistical significance.

Across all the ten subsample comparisons, three differences are statistically significant. Both women and candidates in hard sciences display 1.3% higher treatment effect in anxiety than men and students in soft sciences.[2] Students with parents in poor mental health display 2.8% higher anxiety than students with healthy parents. Because the standard errors of the differences are about 0.6% in all subsample comparisons (except for the breakdown by parental mental health that creates unbalanced subsample sizes), all the other mean differences between subgroups fall below the 1.2% threshold for statistical significance. The largest insignificant differences obtain for depression: men and students with parents in poor health, in older ages, and in hard sciences have higher effects than others (0.5%, 0.9%, 0.7%, and 0.2%, respectively). All in all, these subsample comparisons point to gender, field of study, and parental mental health exposing students differentially to mental health problems while other characteristics appear to play a smaller role.

4. Summary and implications for research policy

This study finds a lower prevalence of depression, anxiety, and completed suicides among PhD students than previous research suggests. For example, our conservative estimate for the prevalence of depression among PhD students, adjusted for differences in methodology, is 43% of the corresponding meta-estimate in the literature. Suicides among PhD students are exceedingly rare. Nevertheless, compared to the benchmark group of Master's graduates not pursuing a PhD, the prevalence of medically validated depression and anxiety among PhD students are both about 20% higher. Our rich longitudinal data and a powerful difference-in-differences design allow us to estimate the causal effects of the PhD program. These treatment effects amount to at least 81% of the difference in mental health between PhD students and their peers.

These results inform research policy on the strength and types of appropriate responses. Our findings suggest graduate studies may be less harmful to the mental health of students than previously feared. At the same time, our data and research design allow us to present a more credible case than the earlier literature about the adverse causal effects of doctoral studies on

[2] While not detailed in the table, female PhD candidates in hard sciences exhibit a treatment effect for anxiety that is 1.8% greater compared to their male counterparts (*t*-value of 2.1). All the other gender differences by field of study are statistically insignificant.

mental health. This credibility gives further weight to calls for an appropriately designed policy response to students' mental health worries (Council of Graduate Schools, 2021; Evans et al., 2018; Forrester, 2021; Nature 2019a; Nature 2019b; Woolston, 2017).

Our research offers three policy recommendations. First, since most mental health deterioration in PhD students occurs during their study period, resources should be directed towards improving the actual program experience, rather than focusing on the selection of students resilient to academic pressures. Second, our findings indicate that women and candidates in hard sciences are more adversely affected by PhD studies, indicating a higher need for specialized support interventions for these groups. Third, the final years of the program, laden by significant stressors such as thesis completion and job searching are often when mental health problems surface. This observation indicates that these stages warrant particular attention and targeted interventions.

References

Alonso J., Angermeyer, M.C., Bernert, S., Bruffaerts, R., Brugha, T.S., Bryson, H., de Girolamo, G., de Graaf, R., Demyttenaere, K., Gasquet, I., 2004. Use of mental health services in Europe: Results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. Acta Psychiat. Scand. 109, 47–54.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, Washington, DC, ed. 5).

Auerbach, R.P., Alonso, J., Axinn, W.G., Cuijpers, P., Ebert, D.D., Green, J.G., Hwang, I., Kessler, R.C., Liu, H., Mortier, P., Nock, M.K., 2016. Mental disorders among college students in the World Health Organization World Mental Health Surveys. Psychol. Med. 46, 2955–2970.

Baker, A.C., Larcker, D.F., Wang, C.C.Y., 2022. How much should we trust staggered difference-in-differences estimates?, J. Financ. Econ. 144, 370–395.

Baxter, A.J., Scott, K.M., Ferrari, A.J., Norman, R.E., Vos, T., Whiteford, H.A., 2014. Challenging the myth of an "epidemic" of common mental disorders: trends in the global prevalence of anxiety and depression between 1990 and 2010. Depress. Anxiety 31, 506–516.

Biglan, A., 1973. The characteristics of subject matter in different academic areas. J. Appl. Psychol. 57, 195–203.

Blanco, C., Okuda, M., Wright, C., Hasin, D.S., Grant, B.F., Liu, S.M., Olfson, M., 2008. Mental health of college students and their non-college-attending peers: results from the National Epidemiologic Study on Alcohol and Related Conditions. Arch. Gen. Psychiatry 65, 1429–1437.

Broström, A., 2019. Academic breeding grounds: Home department conditions and early career performance of academic researchers. Res. Policy 48, 1647–1665.

Corsini, A., Pezzoni, M., Visentin, F., 2022. What makes a productive Ph. D. student? Res. Policy 51, 104561.

Council of Graduate Schools & the Jed Foundation. Supporting Graduate Student Mental Health and Well-being (CGS, 2021).

Evans, T.M., Lindsay, B., Beltran Gastelum, J., Todd Weiss, L., Vanderford, N.L., 2018. Evidence for a mental health crisis in graduate education. Nat. Biotechnol. 36, 282–284.

Ferrari, A.J., Charlson, F.J., Norman, R.E., Flaxman, A.D., Patten, S.B., Vos, T., Whiteford, H.A., 2013. The epidemiological modelling of major depressive disorder: Application for the Global Burden of Disease Study 2010. PLOS One 8, e6963.

Fishman, P., Goodman, M., Hornbrook, M., Meenan, R., Bachman, D., O'Keeffe Rosetti, M., 2003. Risk adjustment using automated ambulatory pharmacy data: The RxRisk model. Med. Care 41, 84–99.

Forrester, N., 2021, Mental health of graduate students sorely overlooked. Nature 595, 135–137.

Forslund, T., Kosidou, K., Wicks, S., Dalman, C., 2020. Trends in psychiatric diagnoses, medications and psychological therapies in a large Swedish region: A population-based study. BMC Psychiatry 20, 1–9.

GBD 2019 Mental Disorders Collaborators, 2022. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: A systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 9, 137–150.

Harvey, S.B., Modini, M., Joyce, S., Milligan-Saville, J.S., Tan, L., Mykletun, A., Bryant, R.A., Christensen, H., Mitchell, P.B., 2017. Can work make you mentally ill? A systematic meta-

review of work-related risk factors for common mental health problems. Occup. Environ. Med. 74, 301–310.

Henriksson, S., Asplund, R., Boëthius, G., Hällström, T., Isacsson, G., 2006. Infrequent use of antidepressants in depressed individuals (an interview and prescription database study in a defined Swedish population 2001–2002). Eur. Psychiatry 21, 355–360.

Hyun, J.K., Quinn, B.C., Madon, T., Lustig, S., 2006. Graduate student mental health: Needs assessment and utilization of counseling services. J. Coll. Stud. Dev. 47, 247–266.

Ibrahim, A.K., Kelly, S.J., Adams, C.E., Glazerbrook, C., 2013. A systematic review of studies of depression prevalence in university students. J. Psychiatr. Res. 47, 391–400.

Johansson, R., Carlbring, P., Heedman, Å., Paxling, B., Andersson, G., 2013. Depression, anxiety and their comorbidity in the Swedish general population: Point prevalence and the effect on health-related quality of life. PeerJ 1, e98.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., Monahan, P.O., Löwe, B., 2007. Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. Ann. Intern. Med. 146, 317–325.

Levecque, K., Anseel, F., De Beuckelaer, A., Van der Heyden, J., Gisle, L., 2017. Work organization and mental health problems in PhD students. Res. Policy 46, 868–879.

Levis, B., Benedetti, A., Ioannidis, J.P., Sun, Y., Negeri, Z., He, C., Wu, Y., Krishnan, A., Bhandari, P.M., Neupane, D., Imran, M., 2020. Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: Individual participant data meta-analysis. J. Clin. Epidemiol. 122, 115–128.

Manea, L., Gillbody, S., McMillan, D., 2012. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): A meta-analysis. CMAJ 184, E191–E196.

Moitra, M., Santomauro, D., Collins, P.Y., Vos, T., Whiteford, H., Saxena, S., Ferrari, A.J., 2022. The global gap in treatment coverage for major depressive disorder in 84 countries from 2000–2019: A systematic review and Bayesian meta-regression analysis. PLOS Med. 19, e1003901.

Moriarty, A.S., Gillbody, S., McMillan, D., Manea, L., 2015. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. Gen. Hosp. Psychiatry 6, 567–576.

Nature, 2019a. Academia's mental-health woes. Nature 569, 307.

Nature, 2019b. A cry for help. Nature 575, 257–258.

Negeri, Z.F., Levis, B., Sun, Y., He, C., Krishnan, A., Wu, Y., Bhandari, P.M., Neupane, D., Brehaut, E., Benedetti, A., Thombs, B.D., 2021. Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: updated systematic review and individual participant data meta-analysis. BMJ 375, n2183.

Plummer, F., Manea, L., Trepel, D., McMillan, D., 2016. Screening for anxiety disorders with the GAD-7 and GAD-2: A systematic review and diagnostic meta-analysis. Gen. Hosp. Psychiatry 39, 24–31.

Pyhältö, K., Toom, A., Stubb, J., Lonka, K., 2012. Challenges of becoming a scholar: A study of doctoral students' problems and well-being. International Scholarly Research Notices 2012, 934941.

Roth, J., Sant'Anna, P.H.C., Bilinski, A., Poe, J., 2023. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. J. Econom. 235, 2218–2144.

Satinsky, E.N., Kimura, T., Kiang, M.V., Abebe, R., Cunningham, S., Lee, H., Lin, X., Liu, C.H., Rudan, I., Sen, S., Tomlinson, M., 2021. Systematic review and meta-analysis of depression, anxiety, and suicidal ideation among Ph.D. students. Sci. Rep. 11, 14370.

Schäfer, I., von Leitner, E.C., Schön, G., Koller, D., Hansen, H., Kolonko, T., Kaduszkiewicz, H., Wegscheider, K., Glaeske, G., van den Bussche, H., 2010. Multimorbidity patterns in the elderly: A new approach of disease clustering identifies complex interrelations between chronic conditions. PLOS One 5, e15941.

Sihvo, S., Isometsä, E., Kiviruusu, O., Hämäläinen, J., Suvisaari, J., Perälä, J., Pirkola, S., Saarni, S., Lönnqvist, J., 2008. Antidepressant utilisation patterns and determinants of short-term and non-psychiatric use in the Finnish general adult population. J. Affect. Disord. 110, 94–105.

Stansfeld, S., Candy, B., 2006. Psychosocial work environment and mental health—a meta-analytic review. Scand. J. Work Environ. Health 32, 443–462.

Statistics Sweden, LISA, 2020. Det statistiska registrets framställning och kvalitet, Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier (LISA). https://www.scb.se/contentassets/f0bc88c852364b6ea5c1654a0cc90234/lisa-am9901_dokstar_2020.pdf, Accessed 15th July 2022.

Stoecker, J. L., 1993. The Biglan classification revisited. Res. High. Educ. 34, 451–464.

Thombs, B.D., Kwakkenbos, L., Levis, A.W., Beneditti, A., 2018. Addressing overestimation of the prevalence of depression based on self-report screening questionnaires (Appendix 2). CMAJ 190, E44–E49.

Woolston, C., 2017. Graduate survey: A love–hurt relationship. Nature 550, 549–552.

Wuestman, M., Wanzenböck, I., Frenken, K., 2023. Local peer communities and future academic success of Ph. D. candidates. Res. Policy 52, 104844.

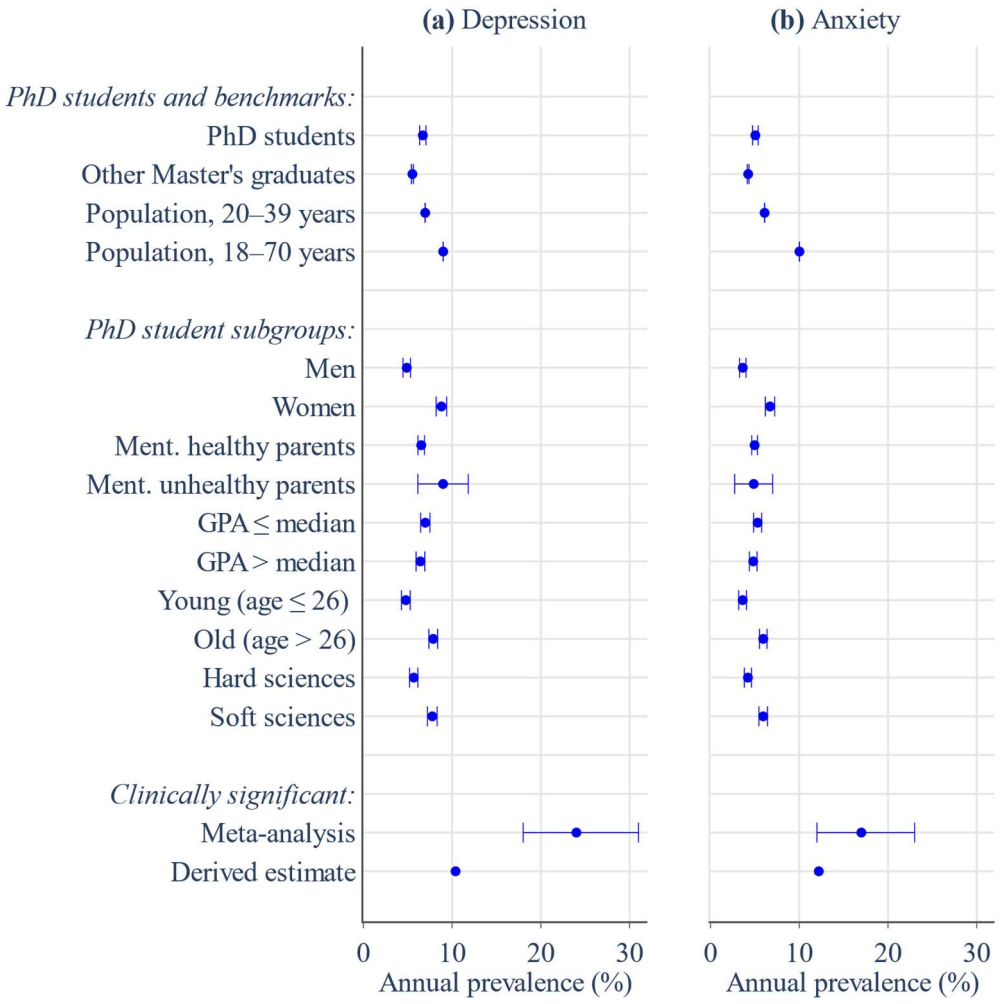

Figure 1. Annual prevalence of depression and anxiety among PhD students, their subgroups, and benchmarks. Age refers to cohort ages in the year of starting PhD studies. "Meta-analysis" refers to estimates from a meta-analysis on PhD students' mental health (Satinsky et al., 2021), whereas "Derived estimate" adjusts our base estimate for PhD students by multiplying it with two factors: the ratio of clinically significant screened depression and medically validated depression in a representative sample of the Swedish population of 18–70-year-olds (Johansson et al., 2013), and the ratio of medically validated depression among 18–70-year-olds and 20–39-year-olds in the Swedish population. Error bars represent 95% confidence intervals.

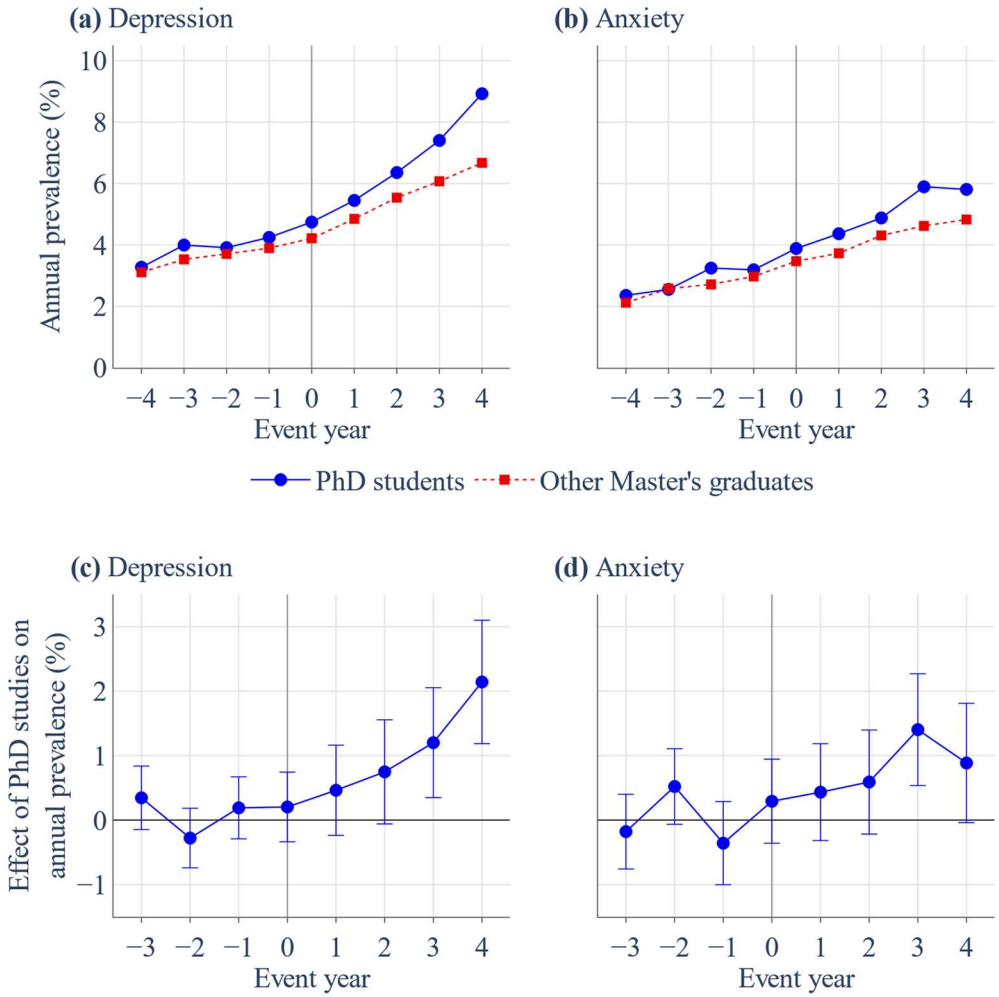

Figure 2. Development of mental health over time for PhD students and control group. Panels (a) and (b) report the annual prevalence of depression and anxiety for PhD students and a control group around the start of PhD studies or graduation with a Master's degree. Panels (c) and (d) report difference-in-differences regression estimates of depression and anxiety on PhD student status. The independent variables are the treatment indicator (being a PhD student), the event time indicators (for treated, t=0 equals first year in PhD program; for control, t=0 equals year of graduation with Master's degree), and their interactions. Event time t=–4 serves as the omitted category. The estimates are based on Callaway and Sant'Anna (2021) and condition on covariates using the doubly robust inverse probability weighting method by Sant'Anna and Zhao (2020). The covariates are gender, age, high school GPA, and indicators for parental anxiety and depression.

Table 1. Annual prevalence of depression, anxiety, and suicides (‰), and average characteristics, among PhD students, their subgroups, and benchmarks. Panel B reports on student age in the year of starting PhD studies and excludes statistics on suicide to preserve privacy. "Meta-analysis" in Panel C refers to estimates from a meta-analysis on PhD students' mental health (Satinsky et al., 2021), whereas "Derived estimate" adjusts our base estimate for PhD students by multiplying it with two factors: the ratio of clinically significant screened depression and medically validated depression in a representative sample of the Swedish population of 18–70-year-olds (Johansson et al., 2013), and the ratio of medically validated depression among 18–70-year-olds and 20–39-year-olds in the Swedish population. Panel D reports high-school GPA and parental mental health indicators (originating from the hospital or specialized open care data) available for subsamples. Standard errors are reported below the means. *N* refers to the number of individual-year observations.

Panel A: PhD students and benchmarks

	Depression (%)	Anxiety (%)	Suicide (‰)	Age	Female (%)	*N*
PhD students	6.69	5.06	0.53	30.04	46.29	18,796
	0.18	0.16	0.53	0.02	0.36	
Other Master's graduates	5.52	4.24	0.25	29.91	55.85	158,293
	0.06	0.05	0.13	0.01	0.12	
Population, 20–39 years	6.96	6.09	1.11	29.53	48.92	26,689,985
	0.005	0.005	0.02	0.001	0.01	
Population, 18–70 years	8.99	10.02	1.36	43.49	49.35	69,822,141
	0.003	0.004	0.01	0.002	0.01	

Panel B: PhD student subgroups

	Depression (%)	Anxiety (%)	Suicide (‰)	Age	Female (%)	*N*
Men	4.88	3.64	Not rep.	29.79	0.00	10,095
	0.215	0.186	Not rep.	0.033	0.00	
Women	8.79	6.71	Not rep.	30.32	100.00	8,701
	0.304	0.268	Not rep.	0.037	0.00	
Mentally healthy parents	6.52	4.97	Not rep.	29.97	46.14	17,537
	0.186	0.164	Not rep.	0.025	0.38	
Mentally unhealthy parents	8.97	4.87	Not rep.	29.95	52.56	390
	1.449	1.092	Not rep.	0.170	2.53	
GPA ≤ median	6.97	5.31	Not rep.	30.37	42.04	9,164
	0.266	0.234	Not rep.	0.035	0.52	
GPA > median	6.43	4.82	Not rep.	29.72	50.33	9,632
	0.250	0.218	Not rep.	0.035	0.51	
Age ≤ 26	4.78	3.62	Not rep.	27.10	41.35	7,158
	0.252	0.221	Not rep.	0.020	0.58	
Age > 26	7.87	5.95	Not rep.	31.85	49.33	11,638
	0.250	0.219	Not rep.	0.027	0.46	
Hard sciences	5.67	4.22	Not rep.	28.91	40.42	9,625
	0.236	0.205	Not rep.	0.028	0.50	
Soft sciences	7.76	5.94	Not rep.	31.22	52.46	9,171
	0.279	0.247	Not rep.	0.037	0.52	

Panel C: Clinically significant		
	Depression (%)	Anxiety (%)
Meta-analysis	24	17
	3	3
Derived estimate	10.38	12.20

Panel D: Additional characteristics available for subsamples					
	Parental mental health			High-school GPA	
	Depression (%)	Anxiety (%)	*N*	Normalized GPA	*N*
PhD students	1.73	0.73	17,927	1.00	18,796
	0.10	0.06		0.01	
Other Master's graduates	1.94	1.02	151,141	0.70	158,293
	0.04	0.03		0.002	
Population, 20–39 years	2.65	1.70	20,383,325	–0.02	18,791,794
	0.004	0.003		0.0002	
Population, 18–70 years	2.50	1.57	31,483,769	–0.01	34,608,011
	0.003	0.002		0.0002	

Table 2. This table reports difference-in-differences regression estimates of depression and anxiety on PhD student status. The independent variables are the treatment indicator (being a PhD student), the event time indicators (for treated, t=0 equals first year in PhD program; for control, t=0 equals year of graduation with Master's degree), and their interactions. Event time t=–4 serves as the omitted category. The estimates are based on Callaway and Sant'Anna (2021) and condition on covariates using the doubly robust inverse probability weighting method by Sant'Anna and Zhao (2020). The covariates are gender, age, high school GPA, and indicators for parental anxiety and depression. Panel A reports the results in the full sample whereas Panel B splits the sample by the characteristics reported in Table 1, Panel B except for parental mental health that generates a subsample too small for reliable inference.

Panel A: Difference-in-differences estimates in full sample				
Dependent variable	Depression		Anxiety	
	Coefficient	*z*-value	Coefficient	*z*-value
Average coefficients				
Pre-treatment	0.09	(0.74)	–0.005	(–0.04)
Post-treatment	0.95	(2.95)	0.72	(2.31)
Coefficients by event year				
–3	0.35	(1.38)	–0.18	(–0.60)
–2	–0.28	(–1.18)	0.52	(1.74)
–1	0.19	(0.78)	–0.36	(–1.09)
0	0.20	(0.74)	0.29	(0.88)
+1	0.46	(1.30)	0.43	(1.13)
+2	0.75	(1.82)	0.59	(1.44)
+3	1.20	(2.76)	1.40	(3.17)
+4	2.14	(4.39)	0.89	(1.88)

Panel B: Difference-in-differences estimates in subsamples				
Dependent variable	Depression		Anxiety	
	Coefficient	*z*-value	Coefficient	*z*-value
Men	1.20	(3.17)	0.13	(0.34)
Women	0.67	(1.23)	1.44	(2.80)
Difference	0.53	(0.81)	–1.31	(–2.06)
Mentally healthy parents	0.93	(2.87)	0.66	(2.08)
Mentally unhealthy parents	1.84	(0.67)	3.50	(2.47)
Difference	–0.91	(–0.33)	–2.84	(–1.96)
GPA ≤ median	0.96	(2.08)	0.73	(1.60)
GPA > median	0.94	(2.05)	0.72	(1.65)
Difference	0.01	(0.02)	0.01	(0.02)
Age ≤ 26	0.50	(1.05)	0.68	(1.62)
Age > 26	1.22	(2.80)	0.74	(1.70)
Difference	–0.72	(–1.12)	–0.06	(–0.09)
Hard sciences	1.07	(2.44)	1.39	(3.48)
Soft sciences	0.85	(1.77)	0.12	(0.24)
Difference	0.22	(0.33)	1.27	(2.02)

Supplementary Note for

PhD Studies Hurt Mental Health, but Less than Previously Feared

Matti Keloharju

Aalto University School of Business, CEPR, and IFN

Samuli Knüpfer

Aalto University School of Business, BI Norwegian Business School, and IFN

Dagmar Müller

Swedish Public Employment Service

Joacim Tåg

Hanken School of Economics and IFN

December 2023

This supplementary note defines the terms and provides the analysis needed to understand the statistical properties of an estimator that aims to make our results comparable with existing literature.

1. Definitions of terms

Medical validation. When needed, the term medical validation is used to emphasize the distinction between our definition of depression (anxiety) and screened or true depression (anxiety).

The other definitions listed below are in accordance with literature and are not specific to our paper.

Screening. Refers to the usage of questionnaires, such as PHQ-9 and GAD-7, for detecting potential cases of depression or anxiety. Individuals are given points based on their answers to the multiple-choice questions, and if these points exceed a set threshold, they are classified as having screened positive for depression (anxiety). While screening is commonly used to estimate the prevalence of mental health problems, it is known to be prone to overestimation (Levis et al., 2020).

Clinically significant symptoms of depression (anxiety). Used to emphasize the distinctions between screened positive case of depression (anxiety) and medically validated or true depression (anxiety).

True depression (anxiety). Defined based on the Fifth edition of Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013).

Sensitivity. Refers to the probability that a measure (such as screening questionnaire with some threshold) correctly identifies a true positive case.

Specificity. Refers to the probability that a measure (such as screening questionnaire with some threshold) correctly identifies a true negative case.

Yearly prevalence of depression (anxiety). Represents the proportion of population that has suffered from depression (anxiety) at any point within the measurement year.

Point prevalence of depression (anxiety). Represents the proportion of population that is suffering from depression (anxiety) at any given point of time.

2. Statistical analysis

We aim to transform our yearly prevalence estimate of medically validated depression (anxiety) to align with the point prevalence estimate of screened depression (anxiety) presented

in Satinsky et al.'s (2021) meta-analysis. To achieve this, we reference the findings from Johansson et al.'s (2013) study on a representative sample of the Swedish population aged 18–70. Johansson et al. (2013) and the studies covered by the meta-analysis are comparable in the sense that they tend to assess the point prevalence of depression (anxiety) using the same screening questionnaires (PHQ-9 and GAD-7, respectively). While the thresholds are the same for PHQ-9 (10 points), the threshold used for GAD-7 is lower in Johansson et al. (2013) than in the substudies covered by the meta-analysis (8 versus 10 points). We constrain our population sample to the same age range (18–70 years) and comparable years (2005–15) as Johansson et al. (2013), whose screening was conducted in fall 2009.

The derived estimate $\widehat{Screen}_{phd}$ is defined as

$$\widehat{Screen}_{phd} = Med_{phd} * \frac{\widehat{Screen}_{pop}}{Med_{agepeers}}, \qquad (E1)$$

where Med refers to the prevalence of medically validated cases, $\widehat{Screen}$ to the estimate of the prevalence of screened cases, phd to PhD students, pop to 18–70-year-olds, and $agepeers$ to 20–39-year-olds.

For any population P we have:

$$Med_P = (1 - I_P) * \left(1 - Spec_{M,P}\right) + I_P * Sens_{M,P} \qquad (E2)$$

$$Screen_P = (1 - I_P) * \left(1 - Spec_{S,P}\right) + I_P * Sens_{S,P}, \qquad (E3)$$

where I_P is the true prevalence of depression for population P, $Spec_{X,Y}$ and $Sens_{X,Y}$ refer to the specificity and sensitivity of method X applied on population Y, and M and S refer to medical validation and screening, respectively (Thombs et al., 2018).

As our data is comprehensive, we observe Med_{phd} and $Med_{agepeers}$. The only random variable in formula (E1), $\widehat{Screen}_{pop}$, is obtained from Johansson et al. (2013) and it is assumed to be unbiased. The expected value of the derived estimate (E1) is

$$E\left(\widehat{Screen}_{phd}\right) = Med_{phd} * \frac{Screen_{pop}}{Med_{agepeers}}. \qquad (E4)$$

The bias of the estimator is the expected difference between (E1) and (E3):

$$E\left(\widehat{Screen}_{phd} - Screen_{phd}\right) =$$
$$Med_{phd} * \frac{Screen_{pop}}{Med_{agepeers}} - \left(1 - I_{phd}\right) * \left(1 - Spec_{S,phd}\right) - I_{phd} * Sens_{S,phd}. \qquad (E5)$$

We analyze the bias in our derived estimate for depression based on three assumptions. Firstly, we assume that both specificity and sensitivity of medical validation are the same for PhD students and their age peers. While we found no studies that directly address this, existing research on college students and their age peers finds no difference in treatment seeking for anxiety or depression (Blanco et al., 2008), or in the reception of minimally adequate treatment for mental disorders (Auerbach et al., 2016). Secondly, we assume constant specificity and sensitivity of screening across all groups, an assumption also implicit in prior research comparing screened depression prevalence between students and the general population (Ibrahim et al., 2013; Satinsky et al. 2021). Thirdly, we assume the prevalence of depression in the 20–39-year age group is not higher than in the 18–70-year age range (Baxter et al., 2014; Ferrari et al., 2013).

If there were no false positive cases (i.e., the specificities equaled one), the formula for the bias would simplify significantly. Moreover, if the true prevalence of depression in the population additionally equaled that among age peers, the bias would be zero. However, without such assumptions, the ratio between screened and medically validated depression depends not only on the properties of the measures but also on the underlying true prevalence of depression. We analyze the effects of these additional complications on the bias by applying broad bounds for the unobserved parameters based on existing literature: 1–15% for the prevalence of depression among both 18–70-year-olds and 20–39-year-olds (Baxter et al., 2014; Ferrari et al., 2013; GBD 2019 Mental Disorders Collaborators, 2022), 1–35% for the prevalence of depression among PhD students (American Psychiatric Association, 2013), 9.1–12.5% for the screened prevalence of depression among 18–70-year-olds (Johansson et al., 2013). In terms of test accuracy, we apply 50–99% for both the specificity and sensitivity of screening (Moriarty et al., 2015; Negeri et al., 2021) and 89–99.9% and 10–90% for the specificity and sensitivity of medical validation for all groups, respectively.

The bounds for medical validation are based on two studies, which utilize Swedish and Finnish register data on prescriptions (Henriksson et al., 2006; Sihvo et al., 2008). The Swedish (Finnish) study on prescriptions measures depression with SCID (CIDI), that is, a standardized

semi-structured (fully structured) interview, both considered gold standards for measuring depression (Manea et al., 2012). The studies also account for antidepressant users with signs of alleviated depression possibly due to the medication (Henriksson et al., 2006; Sihvo et al., 2008). The same proportion (52%) of the individuals with either current or alleviated depression use antidepressants in both studies (Henriksson et al., 2006; Sihvo et al., 2008). In addition, these studies yield a similar estimate for the proportion of depressed among antidepressant users as a study utilizing more comprehensive Swedish diagnostic data (Forslund et al., 2020). We set the limits for sensitivity conservatively (10–90%). The two studies also allow us to calculate the proportion of population who are non-depressed and do not use medication (99% and 97%, respectively) (Henriksson et al., 2006; Sihvo et al., 2008), which represents the specificity of medical validation. The proportion of negative cases in our data (about one tenth in the general population) is an obvious lower bound for specificity. We set the bounds for specificity conservatively (89–99.9%).

We minimize the bias (i.e., maximize the negative bias) over the unobserved parameters given our observations of medically validated cases, assumptions, and bounds for the parameters. The minimization is performed using Python's (version 3.9.13) Scipy library (version 1.7.3), which employs a sequential least squares programming algorithm. The global minimum of the bias for the derived estimate is –0.15% for depression, suggesting it does not have any significant negative bias. Rather, if anything, our derived estimate is inflated, making comparisons between it and the meta-analysis estimate conservative.

We use the above bounds also for anxiety disorders. The sensitivity and specificity of screening are well within these bounds (Kroenke et al., 2007; Plummer et al., 2016), but we cannot confirm the same for medical validation. However, based on the research on depression, our bounds are highly conservative. In addition, we slightly modify two of our assumptions because the prevalence of anxiety is likely to increase in age (Baxter et al., 2014) and because Johansson et al. (2013) uses a lower GAD-7 threshold for screening than the comparable studies in the meta-analysis (8 versus 10 points). These two deviations from our original assumptions have opposite effects: allowing the prevalence of depression among age peers to exceed that among the population decreases the global minimum of bias, while the higher sensitivity and lower specificity of screening among the population (due to the lower threshold used) increases it.

Existing literature suggests the prevalence of anxiety decreases between ages 20 and 74 roughly by 2.5 percentage points (Baxter et al., 2014); we allow the difference between 20–39-year-olds and 18–70-year-olds to be as high as 5 percentage points. The sensitivity (specificity)

of screening with the lower threshold is assumed to be 9 (6) percentage points higher (lower) than that with the higher threshold (Kroenke et al., 2007). Given these assumptions, the global minimum of the bias for the derived estimate is 0.1%. This means our derived estimate is likely too high, making comparisons between it and the meta-analysis estimate conservative.