The Impact of the COVID-19 Pandemic on PhD Education: An Empirical Analysis of a Top Comprehensive European University*

Maxime François† and Kristof de Witte‡

Abstract:

While there is growing evidence on the impact of the COVID-19 pandemic on students in compulsory education, its effects on PhD education remain largely unknown. This paper aims to examine the impact of the pandemic on various aspects of PhD education (ISCED level 8), including time to graduation, diploma attainment, dropout rates, and new PhD student enrolment. The study utilizes a panel dataset from 2010 to 2022, consisting of 17,476 PhD students and 2,042 supervisors from a top 50 ranked European university. To account for both observed and unobserved supervisor characteristics, a fixed effects model is employed. The results of the study indicate a 12% decrease in diploma attainment in 2020, which was followed by an increase in 2021. Additionally, the average time to graduation was extended by 2 months in 2021 due to the pandemic. There was also a 25% reduction in scholarship discontinuations, although the numbers returned to previous levels later on. We argue that the closure of the university allowed professors to allocate more time for research, leading to an increase in successful funding applications. As a result, new PhD student enrolment did not significantly change in 2020 but rebounded in 2021 with a notable 22.5% increase. Further analysis reveals that the findings are primarily driven by the field of Biomedical Sciences. Graduates in this field experienced an average delay of approximately 6 months compared to students in other research areas. Interestingly, non-EU students tended to graduate earlier than EU students, with a reduction of 4 to 10 months in their time to graduation. No significant differences were found between genders.

Keywords: COVID-19; PhD education; University closures; Academic development; Research productivity; ISCED level 8.

JEL-classification : I21, I23.

Declarations of interest: None.

---

* We would like to thank the KU Leuven data center for providing the data and the Flemish Research Foundation (FWO) for funding (project grant G0D2722N). For their useful insights and comments regarding the paper or the data, we also would like to thank Giorgio Di Pietro, Klara Gijsbers, Marco Montanari and Sylke Schnepf.

† **Maxime François**: Methodology, software, formal analysis, validation, investigation, writing – original draft, visualization.
Leuven Economics of Education Research (LEER), KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium, E-mail: maxime.francois@kuleuven.be.

‡ **Kristof De Witte**: Conceptualization, resources, data curation, writing- Review & editing, supervision. Leuven Economics of Education Research (LEER), KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium, E-mail: kristof.dewitte@kuleuven.be; United Nations University - Maastricht Economic and Social Research Institute on Innovation and Technology (UNU-MERIT), Boschstraat 24, 6211 AX Maastricht, the Netherlands, E-mail: k.dewitte@maastrichtuniversity.nl.

1 Introduction

The closure of schools and universities worldwide as a preventive measure against the spread of the coronavirus has had a profound impact on the global education sector, significantly disrupting the learning process for millions of students. Considering the crucial role education plays in shaping wages, job prospects, and overall prosperity (Chetty et al., 2014; Currie and Thomas, 2001; Hanushek and Woessmann, 2020), there is a substantial interest in studying the effects of the COVID-19 pandemic on educational outcomes. Recent meta-analyses have examined the overall impact of the COVID-19 pandemic on compulsory education levels, providing valuable insights. For instance, De Witte and François (2023) observed that school closures are associated with learning deficits of approximately 0.11 standard deviation (SD) among the Member States of the European Union. On a global scale, Betthäuser et al. (2023) and Patrinos et al. (2022) have taken a comprehensive approach to assess the effects of COVID-19 on compulsory education, estimating average learning deficits ranging from 0.14 to 0.17 SD, respectively. Nevertheless, the evidence reveals substantial variations across different populations and countries, particularly concerning socio-economic status. Regrettably, as we shift our focus from compulsory education, the availability of research significantly diminishes, leaving a considerable knowledge gap regarding the effects of the pandemic on higher education. While studies have examined the impact on higher education and university students, there has been limited attention given to PhD students, and most of these studies have taken a qualitative approach.

Myers et al. (2020) discovered that research time experienced an average decrease of 24.4%, with a more substantial decline of 30-40% observed among biomedical scientists. Furthermore, Cui et al. (2022) noted a 13% reduction in productive time, as measured by the number of publications, specifically among women. This suggests that the pandemic has exacerbated pre-existing gender disparities within academia, particularly due to university closures. Considering the impact on senior researchers, it is plausible that the effects of the pandemic may be even more pronounced among young, unprepared scientists, including PhD students, leading to potential consequences for their academic trajectory and career prospects.

The present study aims to fill the existing research gap by offering a comprehensive analysis of the impact of COVID-19 on educational achievements among doctoral scholars. Our analysis contributes to the literature in three significant ways. First, it provides a quantitative assessment of the consequences of university closures on PhD graduates. This evaluation encompasses four crucial indicators of PhD progress, namely, dropout rates, diploma attainment, time to graduation, and the number of starting PhDs. Second, recognizing the diverse requirements for onsite research, such as laboratory work, this paper explores the heterogeneity among three distinct research fields: Biomedical Sciences, Humanities and Social Sciences, and Sciences, Engineering, and Technology. This analysis

aims to uncover variations in the impact of COVID-19 across these fields. Third, extending beyond the scope of the pandemic, our study delves into specific factors that influence the doctoral process. These factors include the student's birthplace (EU versus non-EU), gender, and the presence of a co-supervisor. To the best of our knowledge, this paper is the first to conduct a quantitative analysis of the drivers of success and failure in doctoral programs while considering the impact of the pandemic on PhD education.

The analysis focusses at a Times Higher Education (THE) top-50 ranked large comprehensive European university, namely the University of Leuven, Belgium. Using a sample of 2,042 supervisors and 17,476 PhD students graduated between 2010 and 2021, we apply a supervisor fixed effects model at PhD student and at supervisor level. As such, next to time varying variables (e.g. amount of students in a research group, supervisor's years of experience, etc.), we take into account observed (e.g. gender[1], school[2], etc.) and unobserved (e.g. support provided[3], salary, etc.) supervisor heterogeneity.

The remainder of the paper unfolds as follows. Section 2 provides a review of prior research findings and section 3 discusses the settings. Section 4 presents the data used in our analysis before section 5 details our empirical specification. The main results and heterogeneity analyses are reported and discussed in sections 6 and 7. Finally, some robustness checks are conducted in section 8 before the paper concludes.

2 Literature review and hypotheses

The impact of university closures on PhD progress has been a topic of interest since the beginning of the COVID-19 pandemic. While quantitative studies on the subject are limited, there is qualitative evidence that certain fields, such as medical and biomedical sciences, have been heavily impacted due to their reliance on university laboratories (Aydemir and Ulusu, 2020; Börgeson et al., 2021). Meanwhile, Communication Sciences and Disorders (CSD) have also seen a negative influence on PhD progress, as noted by Covington and Jordan (2022). However, the general conclusions regarding the pandemic's impact on PhD students are mixed. For example, Guest et al. (2021) conducted a qualitative analysis among 12 PhD students in the Education faculty of the University of Ottawa and found conflicting results, with 40% of respondents seeing the pandemic as a positive experience and 20% viewing it as neutral. Despite this, the authors also observed an increase in anxiety and uncertainty among PhD students. On the other hand, Covington and Jordan (2022) found that the majority of 160 PhD students in CSD experienced negative impacts and increased anxiety, along with reduced support

[1] More about the gender influence in academia can be found in Corbera et al. (2020), Cui et al. (2022), Mata et al. (2022), Minello (2020), Myers et al. (2020) and Rodrigues et al. (2021).
[2] More about the difference across schools can be found in Aydemir and Ulusu (2020) and Börgeson et al. (2021).
[3] See, e.g., Börgeson et al. (2021), Dohaney et al. (2020) and Rodrigues et al. (2021).

from their supervisor. In contrast, Börgeson et al. (2021) studied 254 PhD students in biomedical sciences and found that 185 students reported increased support from their mentor, which was associated with improved well-being and better supervision. Dohaney et al. (2020) also explored the factors that could mitigate the impact of a crisis on mental health, concluding that strong community, peer support, and leadership are crucial mitigators during pandemic times on learning and teaching. The authors also highlight the need for students to have interactions and meetings, possibly face-to-face, with their supervisor. Rodrigues et al. (2021) went even further, emphasizing the need for academic institutions to provide more conciliation, such as deadline adjustments and supervisor support, for PhD and master students.

Beyond PhD education, a significant body of the literature has focused on research productivity during the pandemic and found mixed results. On the one hand, Cui et al. (2022) estimated a difference-in-differences analysis on the number of publications written by nearly 77,000 authors from 25 countries. Their data covers the first wave of the pandemic, i.e. from December 2019 until May 2020. They found that research productivity increased by 35% during the first wave of the pandemic. Their study also found no significant decrease in productivity for PhD students when compared to other academic ranks, such as assistant or full professor. On the other hand, Myers et al. (2020) reached opposing conclusions. Based on a quantitative analysis of 4,535 academic surveys from the EU and the USA, the authors observed that the average work hours declined from 61.4 hours to 54.4. Additionally, the results of a LASSO regression indicated a 24.4% decrease in the average total hours dedicated to research, although there was significant variability in this outcome. The greatest decline in research time was found in the field of biomedical sciences, with a drop ranging from 30% to 40%. Conversely, fields that require less equipment such as statistics or social sciences saw only an average drop of 10%. The findings of Myers et al. (2020) are supported by other studies on productivity outside the academic realm. For instance, Bloom et al. (2020) evaluated the Total Productivity Factor (TPF) of 2,900 companies in the United Kingdom and found a 5% decrease in TPF in 2020 and a decrease in working hours in the private sector. In an early interview, Nicholas Bloom attributed the decrease in productivity to the difficulties encountered in establishing an appropriate work environment at home. Many workers were forced to work with children present, which greatly impacted their ability to concentrate (Gorlick, 2020).

Still, when focusing on the medium to long-run, Bloom et al. (2020) paint a relatively positive picture in their forecast of the lasting impacts of the pandemic. They assert that the "medium-run impact of COVID-19 implies only a small negative effect on productivity" (Bloom et al., 2020, p.4). This view is also echoed by Wigginton et al. (2020) who highlight that educational institutions in the USA and EU were already planning for the re-opening of their facilities and the resumption of academic work during the first wave of the pandemic. This proactive approach, they argue, would enhance the resilience of universities and mitigate the lasting impact of the pandemic in the medium to long term.

Individual characteristics such as gender, for instance, might also be a mechanism underlying the COVID-19's influence. This is not new as a gender gap was present in the academic world prior to the health crisis, as supported by numerous studies that have documented gender inequality (Corbera et al., 2020; Mata et al., 2022; Minello, 2020; Rodrigues et al., 2021). These studies attribute the differences between men and women to the fact that women typically take on a greater share of childcare and household responsibilities, even among highly educated couples. Following the university closure, this phenomenon seems to have been exacerbated. Indeed, despite their contrasting results on the impact of the pandemic on research productivity, both Cui et al. (2022) and Myers et al. (2020) agreed that the pandemic widened the gender gap in research. Cui et al. (2022) observed a decrease of 13% in women's research productivity, even as overall productivity rose by 35%. Myers et al. (2020) found that women's time fully dedicated to research declined by 4.2% more than men's.

Based on the previous findings, it is now possible to anticipate the effects of the university closure on PhD students. Following the decrease in productivity due to the lockdown, as a first hypothesis, we expect a decline in the number of graduated PhD during the pandemic years. Second, if fewer students complete their program, one could think that either more students dropped out or that their graduation had to be postponed. In the latter case, we would observe an extended time to graduation. However, according to the forecasts made by Bloom et al. (2020), these negative consequences of the pandemic should not persist for an extended period. As a third hypothesis, we anticipate a notable resurgence following the pandemic, such as a rise in the total number of new PhDs. Fourth, based on the greater impact of the university closure on biomedical research fields, the previously formulated hypotheses are expected to be more pronounced among biomedical PhD students (Aydemir and Ulusu, 2020; Börgeson et al., 2021). Fifth, the role of strong support from a supervisor as well as resilience factors such as community support and avoiding social isolation were identified as critical to avoiding mental health problems. Therefore, as a final hypothesis, we expect that the presence of a co-supervisor and supportive peer effects result in a more successful graduation probability for students.

3 Setting

This section provides insights about the typical trajectory of a PhD curriculum at the examined university. Next, it describes how the examined university faced the COVID-19 crisis among the staff and the students.

3.1 The PhD track

Graduate students have the option to pursue a PhD in one of three primary research fields: "Humanities and Social Sciences" (HSS), "Biomedical Sciences" (BS), or "Science, Engineering and Technology" (SET). The HSS doctoral school is composed of multiple subjects divided into schools, including economics, arts, law, philosophy, theology, and others. The BS doctoral school encompasses various departments such as biology, cellular research, pharmaceuticals, and more. Finally, the SET field is also organized into departments, such as mathematics, computer science, architecture, and chemistry. In the remainder of the paper, we will focus on the research field rather than the specific school or department.

In the PhD curriculum, irrespective of the research field, students have two options to initiate their doctoral journey: applying for an open position or proposing their own research idea. If a student chooses the latter, they are responsible for finding a supervisor who can support them in securing funding. The specific requirements to begin a PhD may vary depending on the circumstances under which the student applies. However, as a general principle, students are typically expected to have graduated with at least cum laude honours.[4] In certain departments, PhD students who have only completed a master's degree in their field may be required to fulfil additional prerequisites before commencing their doctoral studies. These prerequisites could include the completion of a research master's program or a pre-doctoral program. The duration of these programs typically ranges from 1 to 2 years, providing students with the necessary background and skills for their PhD research.

Upon acceptance, the student is typically committed to a 4-year, full-time research program, which is the standard length of time required to complete a doctoral program. During this period, the student is responsible for conducting and presenting their research, attending seminars and conferences, and may also be required to complete evaluated coursework. In certain cases, these additional courses may be incorporated into a research or pre-doctoral program that can be pursued alongside the PhD studies. It is worth noting that PhD students may also have educational obligations, such as providing tutoring, supervising master's theses, and invigilating exams. It is important to mention that, at the evaluated university, professors of any rank can serve as supervisors for PhD students. This means that an assistant professor, for instance, can supervise a doctoral student just as a full professor can. The only restriction is that postdoctoral researchers are not eligible to be supervisors of doctoral students.

In addition to meeting several milestones throughout the 4-year program, the PhD student is required to write a doctoral thesis. It might be either paper based or a comprehensive explanation of the research conducted during the program and its significance to the scientific community. After writing the thesis, the PhD candidate must defend his/her work in front of an examination committee, priorly

[4] Graduating with *cum laude* or with distinction means being graduated with a weighted grade point average of at least 14 out of 20.

constituted by the supervisory team. It is worth noting that the KU Leuven requires to PhD student to have at least one peer-reviewed publication to be allowed to defend the thesis. The defence is a formal discussion of maximum three hours, aiming to explain to the staff and external jury members, experts but also non-experts, the content of the work and the scope of it in, and contribution to, the research area (in the strict and less strict sense). During the defense, the PhD candidate has to demonstrate that he/she effectively masters the content through active discussion and answering of questions.

The PhD student receives a scholarship, corresponding to a monthly fee free of any charge that was around 2 000€ net at the time of the pandemic. Note that a PhD student does not pay any personal taxes but is subject to the social security contribution.[5] In some cases, the scholarship is funded by research agencies, and not directly by the university. Consequently, part of the supervisor's job is to elaborate research projects and to participate in competitive research bids, that ultimately fund new PhD students. Furthermore, approximately 25% of the PhD students are not financially supported and are required to secure their own funding. These individuals often engage in their research while simultaneously holding another job or employment. They are commonly referred to as 'part-time' PhD students, as they balance their doctoral studies alongside their other professional commitments.

In the event that a student is unable to complete their doctoral program within the standard 4-year timeframe, the student can continue the research without a scholarship or the supervisor can hire the student as a (more expensive) research or teaching assistant. As a result, these individuals may require more or less time beyond the standard 4-year program length to obtain their doctoral degree.

3.2 The research university dealing with the COVID-19 health crisis

To contain the transmission of the pandemic during the 2019-2020 academic year, the university administration took the decision to shift to fully online teaching starting from March 11. This measure remained in effect until the end of the term, resulting in a total closure of the university for approximately 9 weeks. Consequently, the administration announced a suspension of all non-essential on-campus research activities for PhD students and also cancelled any new research projects that necessitated the use of on-campus facilities. Despite the interruption of non-essential on-campus research activities, ongoing studies involving cell structures or lab animals, which were difficult to interrupt, were allowed to continue. In addition, labs working on COVID-19 research were also permitted to continue their investigations. However, the impact of these measures was primarily felt by biomedical sciences researchers working with labs and regular manipulations, while researchers in other fields were able to continue their research remotely, albeit with the difficulties that it entailed.

[5] The contribution to social security covers 13% of the gross wage. In the end, if a PhD student earns monthly a net wage of 2 000€, the gross wage is therefore around 2 298€.

During the examination period, PhD students were required to participate in some on-campus activities, such as invigilating or taking exams. However, aside from these activities, the majority of staff and researchers continued to work from home.

During the summer of 2020, the situation regarding exams was a hybrid one, with approximately 86% of exams being conducted on-campus as the pace of the pandemic had slowed down. This allowed for more staff members and PhD students to gradually return to campus, albeit with strict adherence to safety protocols. It is important to note that the precise number of individuals present on campus at any given time is not known, requiring us to make inferences. Although staff members were not explicitly forbidden from using their offices, they were strongly encouraged to work from home whenever feasible, prioritizing remote work arrangements.

The following academic year, 2020-2021, began in a similar mixed situation with lecture halls being occupied at a capacity ranging from 20% to 50%. On campus research was allowed, but there were restrictions on a maximum number of researchers per room. Consequently, many PhD students continued working from home in the first term. Academic activities, such as doctoral defences, were allowed to be conducted on-campus, albeit with a limited number of attendees in a single room, and individuals were advised to work from home if experiencing any symptoms of illness. The expansion of on-campus activities was largely motivated by the aim to offer the optimal educational experience for new students. However, social gatherings were strictly regulated, with a limited number of individuals allowed.

The aforementioned circumstances were not sustained, as all campuses were again closed from November 3, 2020 onwards. This persisted until March 5, 2021, after which the situation gradually improved and returned to normal by the end of the academic year. Overall, on top of the 9 weeks in 2020, the university was closed for a total of 14.5 weeks in 2021.

The series of transitions from complete closure to partial reopening posed significant challenges for biomedical scientists and their PhD students, who often required several months to initiate new research activities. Conversely, researchers in other fields were able to conduct their work either on-campus or remotely from home. As a result, these researchers experienced comparatively fewer disruptions than their counterparts in the biomedical sciences.

4 Methodology

To provide a thorough understanding of the effects of the COVID-19 pandemic on PhD education, we conduct an analysis that examines the impact of the pandemic using data at the level of the supervisor and the student. Clustering data at the supervisor level allows us for capturing common (un)observed effects that may be missed when looking at individual data points at the student level (e.g. supervisors

have a common approach to advising their PhD students). Additionally, it can reduce the potential for bias due to unobserved heterogeneity among students. On the other hand, looking at individual data points at the student level allows for the examination of the heterogeneity of impacts that students may have faced during the pandemic. This approach enables the identification of individual-level factors that may have affected PhD progress and success during the pandemic, such as student demographics. Examining individual-level data also provides a robustness check as the analysis at student level has more variation than an analysis with pooled data at supervisor level.

For our analysis, we begin by examining the impact of the COVID-19 pandemic on PhD education using data aggregated at the supervisor level. Specifically, we investigate the following outcome variables: the number of PhD candidates who successfully obtained their diploma in a given year, the average time to graduation in a given year, the annual number of PhD candidates who dropped out of the program, and the annual number of new PhD candidates supervised by the professor. As variables of interest, we focus on two dummy variables, '*COVID_2020*' and '*COVID_2021'*. The first variable equals 1 for observations in the year 2020 only, while the latter is set to 1 for observations in 2021.[6] The OLS model takes the following form:

$$(1) \qquad y_{i,t} = \alpha + \beta_1 COVID\_2020_{i,t} + \beta_2 COVID\_2021_{i,t} + \lambda Trend_t + \gamma GroupSize_{i,t} + \theta X_{i,t} + \delta_i + \epsilon_{i,t}$$

where y_{it} denotes the outcome variables depending on the specification by year *t* and supervisor *i*. Relatively to the 2010-2019 time period, the effect of the COVID-19 is captured by β_1 and β_2. The former focuses on the modifications that occurred immediately in the aftermath of the initial wave of university closures, while the latter examines the impacts one year later. The '*Trend'* coefficient captures the variations over time, such as the increasing seniority of the supervisor or the increased professionalism of doctoral schools, for instance. The '*GroupSize'* coefficient takes into account the number of students associated with supervisor *i* in a specific year *t*. Time varying controls (e.g., the share of women, share of non-EU students or share of students with a with more than one supervisor) are captured by the variable $X_{i,t}$. The supervisor fixed effects, δ_i, capture observed (e.g. gender, school, etc.) and unobserved (e.g. support provided, supervisory style, etc.) heterogeneity at supervisor level. Finally, we cluster the standard errors $\epsilon_{i,t}$ at supervisor level.

In addition, we undertake a second set of analyses using data at the level of PhD students. At this level, the variables indicating completion or interruption of the program are binary, given that each observation represents a single student. One of these variables is assigned a value of one if the PhD student graduated (rather than dropped out), and a value of zero if the PhD program is still in progress,

[6] Note that we only observe the year when the PhD was achieved or started, not the specific dates.

or if the student dropped out (rather than graduated). In our panel dataset, the outcome variables are 0 until the year of the event (graduation or dropout). The regression model takes the following form :

$$(2) \qquad y_{j,t} = \alpha + \beta_1 COVID\_2020_{j,t} + \beta_2 COVID\_2021_{j,t} + \lambda Trend_t + \gamma GroupSize_{j,t} + \theta X_j + \delta_i + \epsilon_{i,t}$$

where $y_{j,t}$ denotes the outcome variable (either the dummy either for graduating or dropping out from the PhD programme for student *j* in year *t;* or the total number of months the student *j* took to achieve the PhD in year *t* = T). The variables *'COVID_2020'* and *'COVID_2021'* are dummy variables that take on a value of one if the year of completion or dropping out occurred during the pandemic period, specifically in 2020 or 2021. Additionally, we control for student's characteristics such as the European Union (EU) citizenship, the gender and the number of supervisors guiding the student. The latter controls are captured by X_j, while λ and γ captures variations over time and the research group size effect. Using supervisor fixed effects, δ_i , we control for (un)observed time invariant information at the level of the supervisor. The residuals $\epsilon_{i,t}$ are standard errors clustered at the supervisors level.

5 Data

Panel A of Table 1 describes information at the supervisor level and aggregates the data from Panel B, which details the PhD student level. The dataset used in this study comprises all 2,042 supervisors who were active during the period spanning 2010 to 2021. The supervisors included in the study have been observed since either 2010 or the year of their appointment at the university, until either 2021 or the year in which they retired or left the university.

Our sample contains data on both the progress of doctoral students and their individual characteristics. As observed in Panel A, the average supervisor has every two years one graduated student (value 0.51), and about every year and three months a new PhD students starts under his/her supervision (value 0.80). There is a significant variation behind this mean, as there are also supervisors who have up to 11 graduates per year, and 16 new PhD students starting under their supervision. Others have none (see robustness section). About every 6 to 7 years, a supervisor experiences that a student leaves the PhD program, although there are also supervisors with 5 students who left in a given year. In other words, on average, 12% of all doctoral students graduate each year against 3.5% who drop out. The large variation in the data is also present from the research group size, which has on average 4 PhD students but a similar standard deviation. About a quarter of the supervisors are female. The average supervisor has 11.5 years of seniority. Further, Panel A indicates an average PhD length of approximately 59 months, while full-time PhD students are typically enrolled for a duration of 48

months. Furthermore, Panel A demonstrates that 21% of graduated PhDs had a more than one supervisor.

The sample includes information from 73 departments. The three main research fields have nearly equal frequencies. The 'Humanities and Social Sciences' (HSS) field accounts for 34.65% of the observations, followed by 'Biomedical Sciences' (BS) with 33.65% and 'Science, Engineering and Technology' (SET) with 31.44%. Additionally, the year in which the supervisor was hired is also known. Any student who graduated, started, or dropped out their PhD with a supervisor who was hired after these events are removed from the observations, as these situations may occur when a professor comes from another university or was a postdoc before becoming a full professor. The individual level data, presented in Panel B, indicates that about a third of the PhD students are non-EU students, and 44% are female. Comparing Panel A and Panel B indicates that the average age difference between a PhD student and their supervisor is about 20 years. Moreover, the underlying data reveal that the SET field has a higher concentration of students per supervisor, around 6.2 students against 3.1 and 3.6 respectively for the BS and HSS fields. Consequently, at student level the field of HSS holds for 28.32% of the observations, BS for 25.51%, and SET for 46.01%.

It should be noted that, by its nature, the data is censored as some PhD students are still being enrolled in their doctoral program at the time of the data collection. Additionally, we assume that if a student stops the PhD program, they cannot complete it later on, and vice versa. The individual level data are expanded for each student from PhD's starting year until their year of graduation. As a result, the dataset covers observations from 1995 (i.e., the oldest observed PhD starting year) until 2021, while the data aggregated at supervisor level covers a 2010 to 2021 period. The sample is unbalanced since we miss some observations for the research group size before 2011 and before 2010 for any other variable.[7]

Table 1. Descriptive statistics

	Obs	Mean	SD	Min	Max
Panel A. Sample at supervisor level					
Diploma attainment	16 651	0.51	0.90	0	11
Started PhD	16 651	0.80	1.15	0	16
Dropped out PhD	16 651	0.15	0.42	0	5
PhD time to graduation	5 609	58.88	17.38	0	236
Share of Diploma attainment with a co-supervisor	16 651	0.21	0.39	0	1
Research group size	14 609	4.25	4.29	0	69
Sex supervisor (female = 1)	16 651	0.26	0.44	0	1
Supervisor's birth year	16 651	1967	9.40	1944	1991
Seniority	16 651	11.51	8.58	0	42
Panel B. Sample at student level					

[7] Including only PhD students who started from 2010 onwards results in a dataset that is less comparable at both aggregation levels.

Non-EU student	17 476	0.27	0.45	0	1
Women	17 476	0.44	0.5	0	1
Student's birth year	17 476	1986.5	7.73	1923	2000

In a preliminary exploration of the sample, Figures A1-A2 in Appendix A illustrate the number of supervisors and students over the years.[8] The figures present a parallel tendency, as it can be seen that the number of PhD students has been steadily increasing, with a minor dip in 2021, the year after the first pandemic wave. Similarly, the total number of (co-)supervisors has been constantly increasing, with a minor dip in 2019. Given the common trends, the latter suggest a rather constant ratio between the number of PhD students managed by supervisors. As observed in the data, the average ratio of PhD students by supervisor moved slightly from 4.80 in 2011 to 4.40 in 2021.

According to Figure A3, the annual trend in the overall count of PhD graduates exhibited a reduction in the first year of the pandemic, followed by a resurgence in 2021. Remarkably, the decline observed in 2020 was also evident one year earlier in 2019. Moreover, Figure A4 depicts a comparable pattern, with a decrease in the number of PhD dropouts in 2020 and a substantial escalation in 2021. The latter graph seems much more volatile compared to the others, particularly in comparison to the constant growth of started PhDs since 2014, as shown in Figure A5. Despite a slight dip in 2020, there has been a significant increase in 2021 in the number of started PhDs. On the other hand, the average length of PhD education (for graduated students), as depicted in Figure A6, remains relatively stable over time, hovering around 59 months. However, a decrease in length was observed in 2020, followed by an increase to over 60 months in 2021.

Regardless of the variations within each variable, a common feature can be highlighted: in 2021, the likelihood of observing a student who had graduated, started, or dropped out their PhD was higher than in previous years. This suggests that, during or post-pandemic, there has been a growing number of students completing or starting their PhD, but also a higher number of students choosing to discontinue their studies. These overall increases in 2021 could be linked to the steady rise in the number of students, which mechanically results in potentially larger values for each variable. Furthermore, the higher number of graduated and discontinued PhDs could be the cause of the decrease in student numbers in 2021. Therefore, the following section will evaluate whether or not this conclusion remains true through the light of our econometric framework that accounts for a trend and (un)observed heterogeneity.

Finally, Figure A7 presents by research field the average time to graduation a given year. Figure A7 shows that graduates in HSS doctoral programmes took with an average of 60.27 months more time to graduate than BS students, who take on average 59.60 months and SET students who graduate after 57 months. In line with our hypothesis about the larger pandemic influence among BS, the greatest

[8] Given the censoring at the beginning of the dataset, we dropped 2010 observations.

increase in time to completion was observed in the BS field in 2021, although the average length is lower than in HSS. This pattern might be explained by the larger amount of part-time PhDs among the HSS research field.

6 Results

Table 2 presents the impact of the COVID-19 pandemic on time to graduation, diploma attainment, dropout and new enrolments in the PhD program, using both the aggregated data at supervisors level (left panel) and the individual level data (right panel).[9] First, consider the aggregated data at supervisor level. The results suggest that, relatively to other periods, the average time to successfully defend a PhD decreased insignificantly by 5 weeks (coefficient of -1.30) in 2020, i.e. the year of the university closures.[10] However, compared to before, students graduating in 2021 took 2.40 months longer (i.e. an increase of 4.1% to the average of 58.88 months) to complete the PhD. Over time, the 'trend' coefficient suggests that time to graduation increases by about 3 weeks per year. The number of PhD students per supervisor (as a proxy for the size of the research group) reveals the presence of economies of scale, as demonstrated by the significant reported negative coefficient (-0.36) of the group size. This indicates that, controlled for supervisor fixed effects, if the number of students supervised by the professor increases by 1 person, the PhD length decreases by 1.5 weeks. The coefficient for 'Non-EU' is statistically significant and negative, indicating that a larger share of non-EU students in the research group reduces the length of PhD studies. According to the findings, the composition of a research group exclusively comprising non-EU students yields a significant decrease of over 11 weeks in the duration of a doctoral program, corresponding to a 4.5% reduction compared to the mean duration. Both the share of female PhD students and the share of students with a co-supervisor does not significantly alter the average time to graduation.[11]

The outcomes of diploma attainment (column 2) and dropped out (column 3) PhDs exhibit similar trends with regards to the impact of the pandemic. Compared to the other periods, the number of PhDs graduated/dropped out decreased in 2020, with a reduction of -0.06 graduations and -0.04 dropped out PhDs per supervisor. Relatively to the average, this translates to a decrease of nearly 12% for graduated PhDs and almost 25% for dropped out PhDs.[12] However, in 2021, this pandemic influence was not

[9] Table D1 in appendix presents the same results, without supervisor fixed effects.
[10] As underlying mechanism, the decline in time to graduation in 2020 might be explained by the part-time PhD students, who combine their education with a position in companies and industry. Due to the general restrictions during the pandemic, this group of students might have made more progress in their research, and hence graduated quicker. Although we do not observe whether a student takes the program full or part time, we explore this mechanism more in detail in Section 8.
[11] Note that a Hausman test favored the use of Random Effect regression model in place of the Fixed Effect model.
[12] These figures are derived from dividing the coefficient values by the average number of completed (0.51) or dropped out (0.15) PhDs by supervisor. The low average values are a result of a large number of supervisors who did not have any completed or dropped out PhDs during the observation period.

reversed, as graduation and dropout were still lower in 2021 compared to earlier periods. Even though they are negative, the pandemic's coefficients in 2021 are close to zero and not statistically significant. In other words, the health crisis significantly dropped the graduations in 2020 but this effect did not last one year later. As an underlying mechanism, during the university closures dropping out of the PhD program was less attractive as the outside option became less attractive (due to the restrictions in the general economy). Hence, PhD students might have experienced more opportunity to re-focus on their research.

The time trends in columns 2 and 3 indicate an upward trend, which suggests that the number of graduated and discontinued PhDs increases with time, in line with the growing number of PhD students at the university. Additionally, the number of graduated and discontinued PhDs increases respectively by 26% and 29% as the group size increases by one student. Due to multicollinearity, we could not include time varying control variables such as 'share of non-EU students', 'share of females' and 'share of students with co-supervisor'.

Compared to the other periods, the number of students starting their PhD education (column 4) did not significantly change during the pandemic. However, one year later, we observe a significant increase in the number of new PhD students per supervisor. In particular, we observe an increase of 0.18 new students per supervisor, or an increase of 22.5%. As underlying mechanism, the increase in starting PhD students in 2021 might be explained by the extra time university professors received during the university closures, and that they used to write proposals for research funding on which new PhD students were later hired (in line with Wigginton et al., 2020; Myers et al., 2020).

The significant economies of scale within the research group indicate that the size of the group is positively correlated with the number of students they hire. In this light, Figure A8 in Appendix indicates that before 15 years of experience the average number of new students by supervisor increases steadily before rapidly dropping as the supervisor gets older.

Next, consider the right hand panel of Table 2 that presents results estimated at student level. The estimated coefficients of the pandemic for the time to graduation (column 5) are very similar to those estimated at supervisors level, suggesting robust findings. Due to the increased statistical power of the analysis, we observe more significant findings in the estimated coefficients. Overall, we observe that during the year of the pandemic related university closures, the individual likelihood of obtaining a PhD degree reduced significantly (-1.2%) just as the likelihood of leaving the curriculum. In addition, the graduating students finished their PhD education sooner than before. One year later, in 2021, the graduating students took longer to obtain their degree and the individual risk of dropout decreased. In 2021, the probability of obtaining the degree was not significantly different to before. It should be noted that, due to the student level perspective, we cannot estimate a model specification with 'started PhD' as outcome variable.

Table 2. Main results

	Data aggregated at supervisor level				*Data at individual level*		
	Time to graduation	Diploma attainment	Dropped out PhD	Started PhD	Time to graduation	Diploma attainment	Dropped out PhD
COVID 2020	-1.30	-0.06 ***	-0.04***	-0.02	-1.55*	-0.012**	-0.011***
	(0.93)	(0.025)	(0.014)	(0.03)	(0.80)	(0.005)	(0.003)
COVID 2021	2.41**	-0.005	-0.005	0.18***	1.99**	-0.001	-0.005
	(1.07)	(0.033)	(0.017)	(0.04)	(0.95)	(0.006)	(0.003)
Time trend	0.68***	0.03***	0.006***	-0.025***	0.65***	0.01***	0.002***
	(0.11)	(0.003)	(0.001)	(0.003)	(0.10)	(0.001)	(0.00)
Research group size	-0.36***	0.13***	0.044***	0.20***	-0.30***	0.0005	0.001*
	(0.11)	(0.005)	(0.003)	(0.01)	(0.12)	(0.00)	(0.00)
Constant	51.42***	-0.41***	-0.09***	0.26***	50.17***	-0.01	0.03***
	(1.79)	(0.03)	(0.02)	(0.05)	(2.04)	(0.01)	(0.006)
Non-EU	-2.69***				-3.79***	-0.01***	-0.004**
	(0.80)				(0.55)	(0.003)	(0.002)
Women	-0.84				-0.37	-0.002	0.005***
	(0.77)				(0.53)	(0.002)	(0.002)
Share of supervision	1.12				1.09***	0.001	-0.02***
	(0.72)				(0.54)	(0.002)	(0.001)
Supervisor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	5 239	14 609	14 609	14 609	7 945	62 148	62 148
R2-adjusted	0.04	0.12	0.04	0.16	0.03	0.01	0.00

Note*:* The supervisor fixed effects control for observed and unobserved heterogeneity at the level of the supervisor. Standard errors are clustered at supervisor level. In the data aggregated at supervisor level, 'non-EU' denotes the share of non-EU students supervised by the professor, 'women' the share of females supervised by the professor, 'share of supervision' stands for the share of students that are guided by more than one supervisor. In the estimations at individual level, these variables correspond to individual characteristics of the PhD student. As a result, 'share of supervision' captures the number of supervisors guiding the student.

The student level analysis provides interesting information on the control variables. Controlling for observed and unobserved supervisor characteristics, we observe that being born outside the EU reduces the duration of PhD education significantly by nearly 4 months, which holds for a decline of 6.4% compared to the average. Non-EU PhD students have a lower likelihood to dropout and graduate more, although the later coefficient is not significant. Being female and having a co-supervisor does not significantly influence the time to graduation. However, females dropout more often while having a co-

supervisor correlates with lower graduation and lower dropout. The time trend follows a similar pattern as the data aggregated at supervisor level. In larger research groups, students finish their PhD sooner, but also dropout more frequently.

7 Heterogeneity analysis

This section explores the heterogeneity in the data by focussing on the research field, gender, being born outside the EU, and additional supervision by multiple supervisors or co-supervisors. Although the results in Table 3 focus on time to graduation as outcome variable using the student level data, alternative heterogeneity analyses for other outcomes and analyses at aggregated data at supervisor level are presented in Appendix B.

First, consider the heterogeneity among research fields. The results indicate that the findings from section 6 are mainly driven by the field of Biomedical Sciences. Although the results are insignificant for the HSS and SET, students in the field of BS who graduated in 2021 took about 6 months longer than earlier cohorts. The intensive use of specialized laboratories in BS, might be an underlying mechanism for these differences. As these were closed, (PhD) research is more at risk of delay than in other research fields. Moreover, the constant indicates that, on average, BS PhD students take 5 years to graduate, whereas HHS and SET PhD students graduate on average after a little more than four years.

Second, if we focus on EU and non-EU students separately (columns 4 and 5), we observe similar estimates for both groups. The estimates for the 2020 and 2021 time dummies are insignificant, except for the EU students who seem to have used a longer time to graduation in 2021. Nevertheless, Appendices B2 and B3 indicate a lower probability of graduation and a raise of the dropout likelihood following the COVID-19 years for non-EU individuals only. This suggests that the effect on mental wellbeing is more outspoken for non-EU students. For example, due to the longer ban on intercontinental travelling and the less strong social network, non-EU individuals might be more anxious and depressed, which might have an influence on their PhD progress. Moreover, the peer effects[13] indicate a negative correlation with peers from non-EU countries. This is in line with already raised tendency of non-EU students to graduate sooner than others.

Third, from columns 6 and 7 we learn the change since the pandemic was very similar in 2020 and 2021 among males and females. However, the 2021 estimate is larger and significant for males only, indicating that males graduating in 2021 had a significant longer time to graduation than before.

[13] Both variables are between 0-1 and are constructed such that they capture the total number of non-EU individuals minus one, divided by the total number of PhD students by year and supervisor. The reason for subtracting one from the number of non-EU is that we are looking for the effect of the peer. In other words, for one individual born outside the EU, he/she cannot be taken into account in the number of non-EU peers, which implies to decrease the total number of non-EU students by one to compute the peer effect.

Appendices B2 and B3 indicate also a more severe impact of the pandemic among men, which goes against the gender gap evidence previously raised. As underlying mechanism, PhD women of our sample are most likely not to be mother during the PhD, which may explain the absence of gender gap. [14]

Finally, the heterogeneity analysis for PhD students with more than one supervisor and those with only one supervisor. Findings indicate that the prolonged effect on time to graduation in 2021 is mainly driven by the students with more than one supervisor. Additional findings in Appendix B2 indicate that students without one supervisor have a lower probability of graduation, both in 2020 and 2021. However, the drop out probability in Table B3 presents opposite findings with smaller dropout risk for student with one supervisor and no change for individuals guided by more than one supervisor.[15]

[14] The average age of women who graduated in our sample is 31.3 years, while the age of mothers at birth is 31 years old, on average in 2021 in Belgium (Statbel, 2021). This suggests that female PhD students are most likely not to be mothers (or only at the end of their program), implying less effect of the gender gap (Mata et al., 2022).

[15] An hypothesis test significantly rejected the null of no difference between the two coefficients (*pvalue* ≈ 0.00).

Table 3. Heterogeneity analyses in the time to graduation – Data at individual level

	Biomedical Sciences	Humanities and Social Sciences	Sciences, Engineering and Technology	EU students only	Non-EU students only	Women only	Men only	Student with one supervisor only	Student with more than one supervisor only
COVID 2020	-2.07	-1.16	-1.48	-1.63	-1.82	-1.22	-1.92*	-0.90	-1.52
	(1.44)	(2.19)	(0.95)	(0.98)	(1.72)	(1.20)	(1.16)	(1.45)	(1.08)
COVID 2021	6.04***	1.77	0.27	2.07*	1.94	1.55	2.66**	2.98*	1.96
	(2.08)	(2.27)	(1.08)	(1.17)	(1.84)	(1.42)	(1.28)	(1.61)	(1.34)
Research group size	-0.65**	-0.40	-0.22*	-0.20	-0.27	-0.18	-0.26*	-0.29***	-0.22
	(0.28)	(0.32)	(0.12)	(0.14)	(0.17)	(0.14)	(0.16)	(0.11)	(0.17)
Constant	60.41***	48.48***	46.41***	52.32***	46.66***	54.53***	52.88***	49.55***	51.03***
	(3.54)	(4.79)	(2.44)	(2.51)	(4.08)	(2.72)	(3.02)	(2.65)	(2.89)
Peer effect among EU				-0.81					
				(1.87)					
Peer effect among non-EU					-9.53***				
					(3.31)				
Peer effect among women						-5.52**			
						(2.26)			
Peer effect among men							-7.89***		
							(2.34)		
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Supervisor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	2 026	2 177	3 732	5 822	2 123	3 414	4 531	3 209	4 736
R2-adjusted	0.05	0.02	0.04	0.02	0.04	0.02	0.05	0.03	0.03

Note: "Additional controls" captures the variables X_j of equation (2), namely, dummies for the gender or being born with the EU and the number of supervisors.

8 Robustness checks

Five additional analyses are conducted to investigate the robustness of the results. First, we drop 5% of the outlying observations, corresponding to individuals with a PhD length of 95 months or more. These long PhD studies might be due to individuals who combine their PhD with other tasks (e.g. part-time students). The results in Table C1 (Appendix C) indicate nearly identical results compared with those from Section 6.1, except for the COVID-19 variables in columns (1) and (5). The pandemic influence on the time to graduation is much smaller in 2020 as the length decreased non-significantly by around 1 week while Table 2 reported nearly 6 weeks of delay, significantly under the students perspective. This marked decrease in delay is indicative of an underlying problem that warrants further investigation. In section 6.1, the decrease in the time to graduation, as evidenced by Table 2, may be linked to the increased amount of time spent on curriculum by long-term doctoral candidates during the lockdown period. Removing this subset of individuals from the sample would eliminate this phenomenon, thereby indicating a less substantial reduction in the time to graduation in 2020.[16] In column (6), the COVID-19 estimates of the year 2021 is greater than in Table 2 implying a smaller recovery in the second year of pandemic.

Second, to ensure that our findings are not driven by research groups composed of one student, which could shift estimators when this individual graduates, drops out, or starts the curriculum, we focus only on research groups with at least two PhD students. Consequently, we remove more than 30.5% of the supervisors' sample, and 6% of the students in the student sample. The results in Table C2 show the robustness of our earlier findings.

As a third robustness test, we replace the zero values for the outcome variables by missing observations when a supervisor has no student in charge in a given year. The results in Table C3 signal the robustness of the findings.

To further investigate the significancy of our main analysis, we implemented the Romano-Wolf (RW) multiple hypothesis correction[17] on our variables of interest. From Table C5, the closeness of the adjusted p-values with our main model certifies the robustness and significance of our findings.

Finally, we recode the 'COVID 2020' dummy to be 1 in 2020 and 2021, such that the 'COVID 2021' dummy presents an indicator for the resiliency after the pandemic. The results, presented in Table C4, suggest that the findings remain largely unchanged. The estimates indicate that since the pandemic, the time to graduation was stable, but increased in 2021 with about 3.72 months. Moreover, although

[16] As a further support to this statement, note that as we increase the limit of the PhD length (e.g. a maximum of 70 months) the coefficient of the first pandemic year loses in significance and gets closer to zero.

[17] The RW multiple hypothesis correction controls for the familywise error rate (i.e. the probability of rejecting at least one hypothesis *s* in a set of hypothesis *S*) by considering the dependence among the tested p-values. This is achieved by the use of a bootstrap resampling method in order to gain statistical power and account for the potential correlation across each test statistics for a given level of significancy (α). See Clarke, Romano and Wolf (2020) for more insights.

less students attained a PhD degree since the pandemic, but also dropout decreased. This resulted in the longer time to graduation for the students who graduated already. In 2021, we observe that more students obtained the degree, and also more students left the PhD education. Finally, we observe a sharp increase in the number of starting PhD students.

9 Conclusion

The aim of this study was to assess the effects of the COVID-19 pandemic on the advancement of PhD students' education. Additionally, we examined how individual characteristics, such as gender, nationality, and the presence of support from a co-supervisor or peers, influenced these effects. To achieve this objective, we employed a methodology based on supervisor fixed effect regressions, utilizing data from a large comprehensive European university ranked in the top 50.

We observe that, compared to previous years, the overall progress of PhD students slowed down during the first university closure in 2020, before picking up again in 2021. This is reflected by a 12% decline in the number of PhD graduates and a 1.2% lower probability of graduation in 2020, followed by a resurgence in 2021. As a result of the deferral of graduations from 2020 to 2021, the PhD students who graduated in 2021 took 2-months longer to graduate than before. We did not observe a similar pattern in 2020. Additionally, we observe a decline of 25% of the number of PhD candidates leaving PhD education in 2020, corresponding to a reduction of 1.1% in the probability to discontinue their scholarships, before rebounding in 2021. We argued that the university closure of 2020 gave more time to the professors to prepare research proposals, which led to a 22.5% increase in the number of new PhD students in 2021. Overall, the post-crisis situation appears to have returned to normal and is following pre-crisis trends. Yet, a back-of-the-envelope calculation indicates that the cost of the pandemic incurred by the university varies between 2,400,000€ and 3,412,000€.[18] Our heterogeneity analysis reveals that the effects were mainly driven by PhD students in the field of biomedical sciences. For these students, the time to graduation increased by 4 to 6 months, owing to the deferral of completions from 2020 (-5%) to 2021 (+1%) caused by laboratory closures.

Despite the use of supervisor fixed effects, we cannot claim causal relationships. However, it should be noted that most unobserved heterogeneity is taken into account by our econometric framework and the supervisor fixed effects. Moreover, there is no selection made to construct our sample since it

[18] The lower bound was computed by multiplying the gross wage of a PhD student (2 298€) times the number of student who graduated in 2021 from the biomedical doctoral school in more than 48 months (180) times the 6 months increase : 6*2 298€*174 ≈ 2 400 000€. We focus on the BS students since it is the only significant coefficient from Table 3. The upper bound comes from the multiplication of a PhD student's gross wage times the number of student who graduated in 2021 in more than 48 months (691) times the 2 months increase : 2.2*2 298€*675≈ 3 412 530€.

covers every PhD student in the university since 2010. Finally, as the COVID-19 health crisis was, in essence, unexpected, our findings are not prone to reversed causality.

Our findings align with previous qualitative research, particularly in regard to the impact of the health crisis and the influence of peer effects (Aydemir and Ulusu, 2020; Bloom et al., 2020; Börgeson et al., 2021 Dohaney et al., 2020; Myers et al., 2020). However, we observe that students born outside the EU are likely to finish their program 4 up to more than 10 months sooner compared to other students during non-pandemic times, which is a new pattern. During the COVID-19 pandemic, it appears that non-EU students have experienced greater challenges compared to their EU counterparts. Contrary to previous literature, our analysis did not uncover evidence of an increasing gender gap as a result of the health crisis (Corbera et al., 2020; Cui et al., 2022; Mata et al., 2022; Minello, 2020; Rodrigues et al., 2021).

This paper presents several avenues for further research. First, in this study, we gauge the progress of PhD students by examining their time to graduation, as well as the number of completed, discontinued, and initiated PhDs. While we acknowledge that PhD progress is not solely indicated by these factors, future research could explore the number and quality of publications as an additional proxy. Second, despite shedding light on the gender gap among PhD students, further research should examine this more in depth. The literature predicted an added-value of having a co-supervisor on top of the main supervisor. Our findings revealed mixed evidence of the co-supervisor influence since it seems to slows down the PhD progress, it does not reduce the dropout probability in 2021 but it helps to improve the graduation probability. Still, nothing is known about the mental health of the student during the lockdown and the potential positive link with the support of a (co-) supervisor suggested by Börgeson et al. (2021) and Dohaney et al. (2020). Therefore, further research could evaluate the link between the presence of a co-supervisor and mental health in times of crisis. Finally, in contrast with full-time students, our investigations indicate that the research productivity of “long term” or “parttime” PhD students increased during the first wave of the pandemic in 2020. Nevertheless, we did not manage to test and confirm this pattern, which could be also a subject for further studies.

Appendix A – Graphical analysis of the sample

Figure A1. Total number of supervisors by year

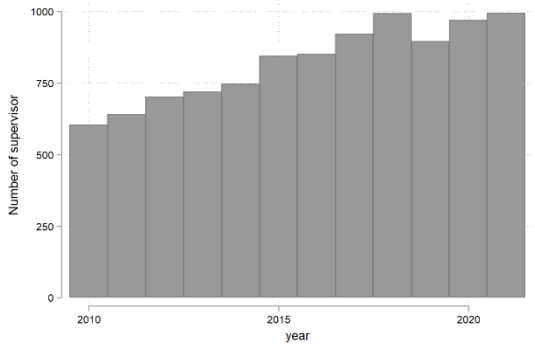


Figure A2. Number of students by year

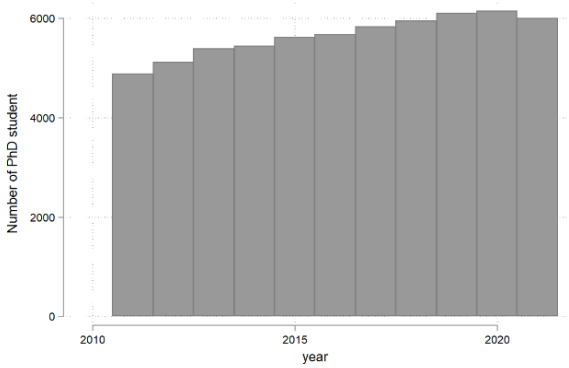


Figure A3. Graduated PhDs by year

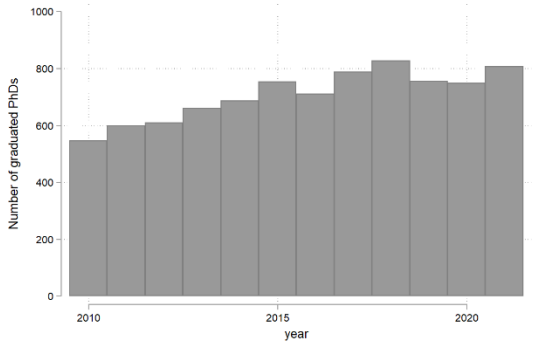


Figure A4. Dropped out PhDs by year

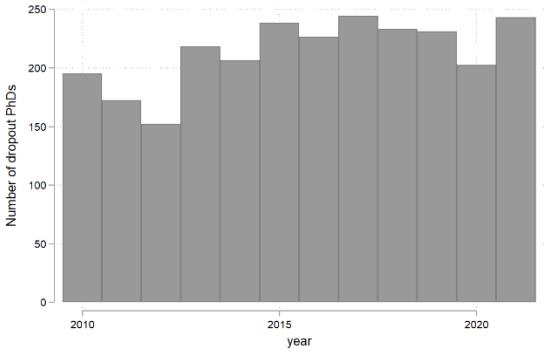


Figure A5. Newly started PhDs by year

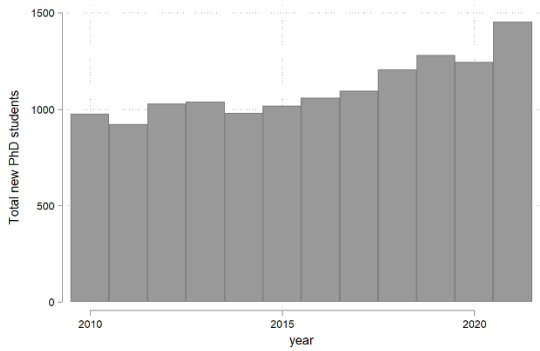


Figure A6. Average time to graduation

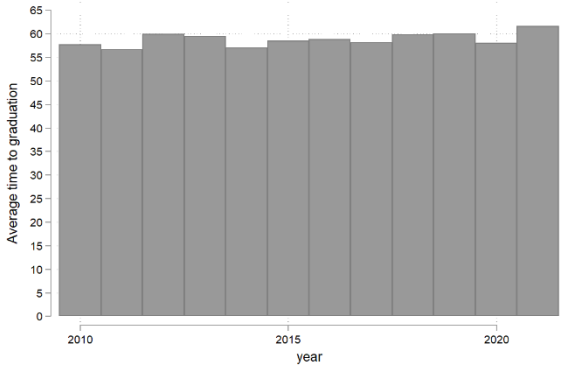


Figure A7. Average time to graduation by research field

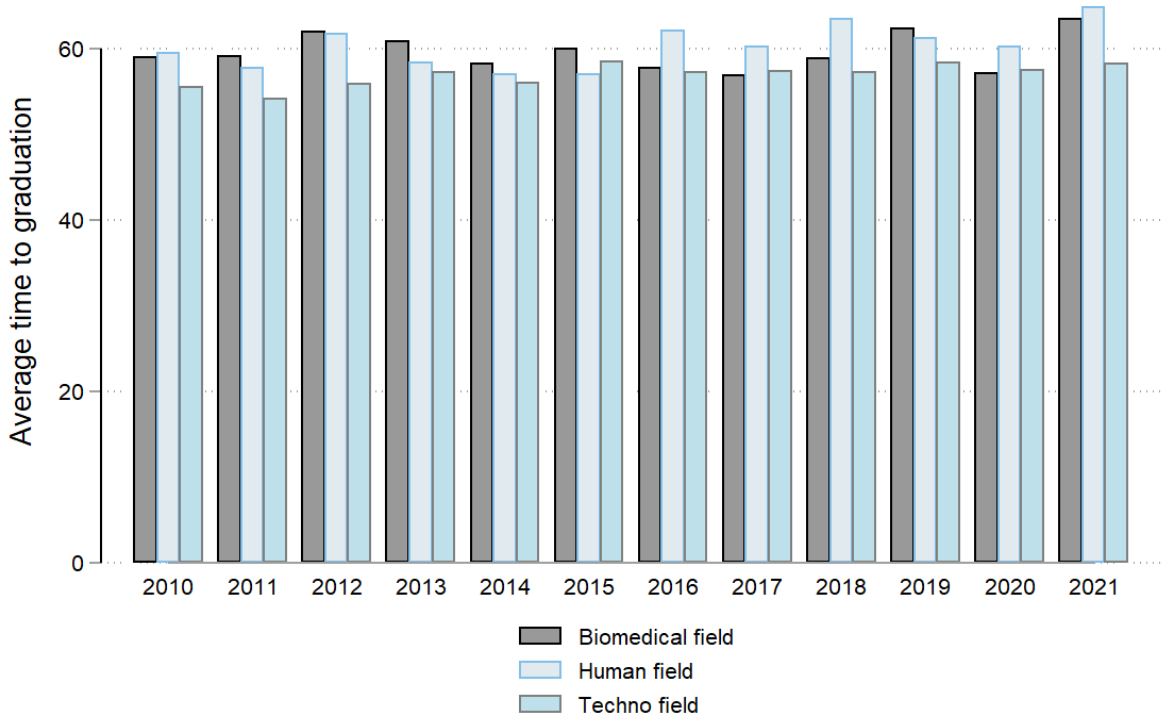


Figure A8. Yearly average number of initiated PhDs by seniority of supervisor

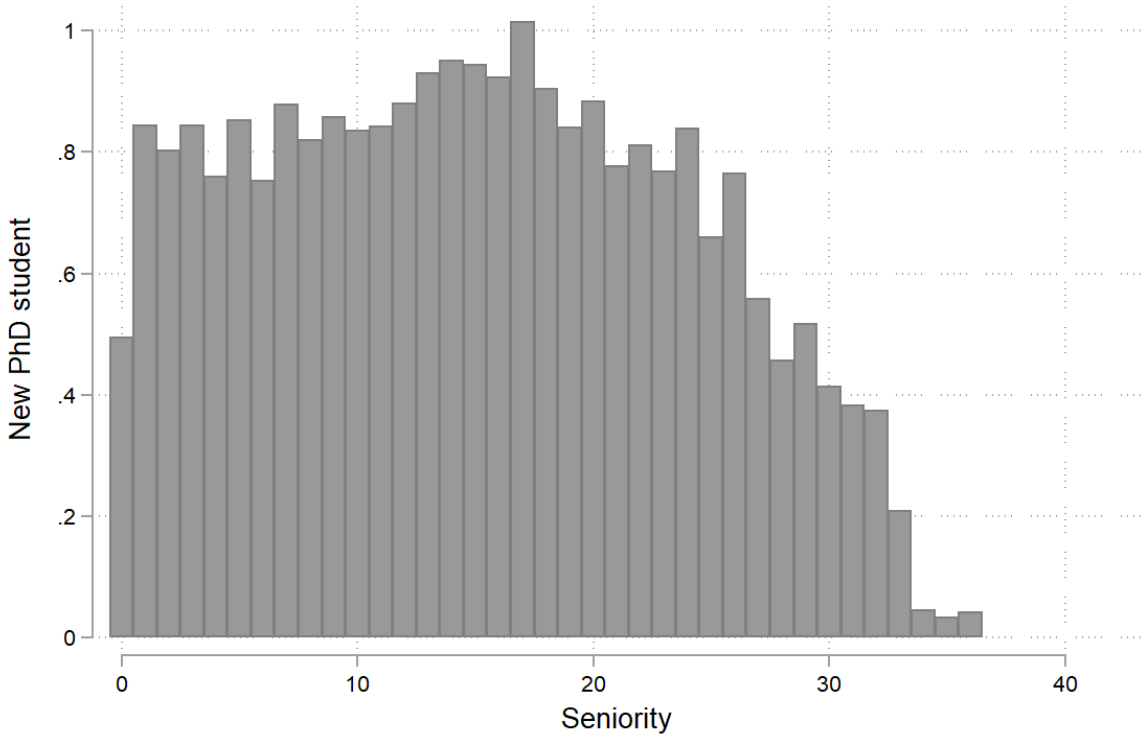


Figure A9. Share of graduated PhDs by EU citizenship

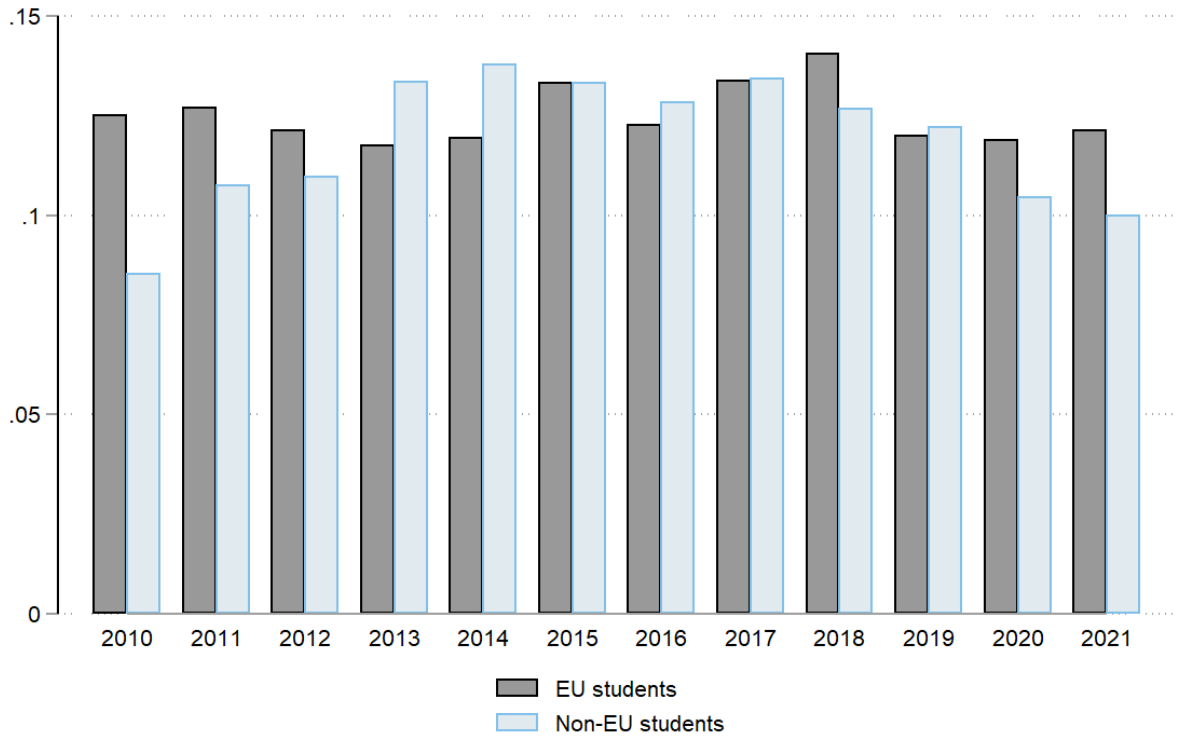


Figure A10. Share of discontinued PhDs by EU citizenship

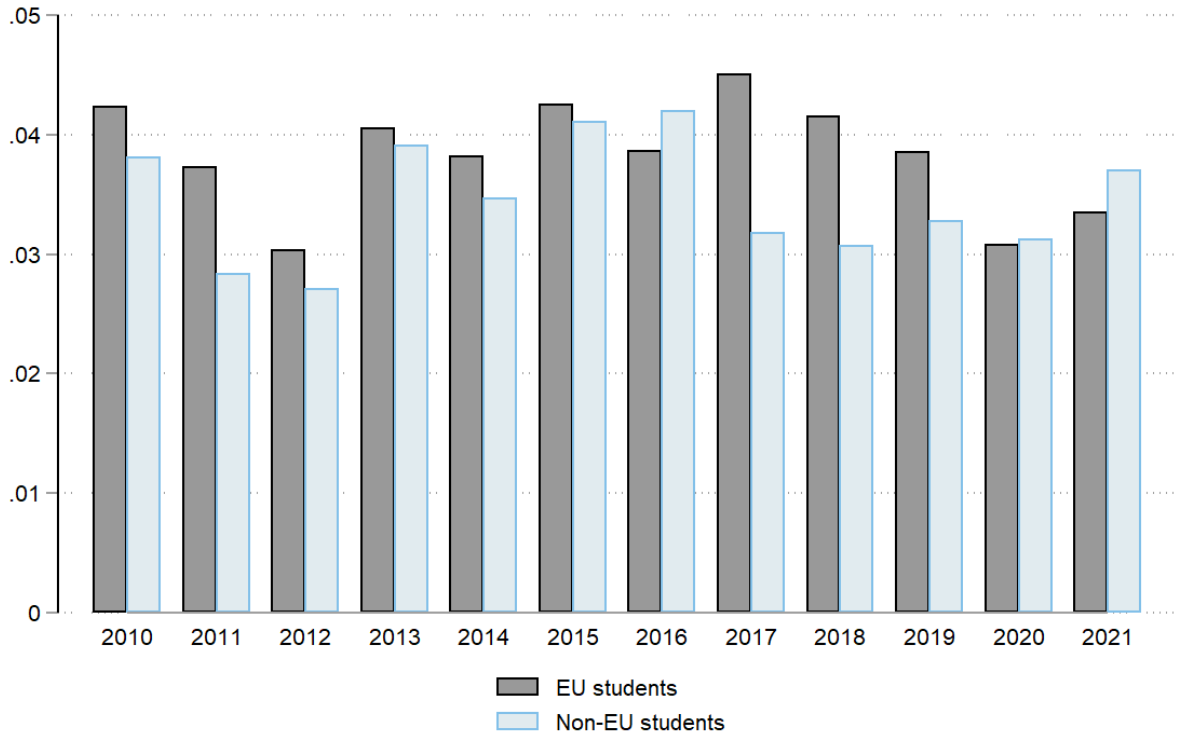


Figure A11. Share of graduated PhDs by the number of supervisors

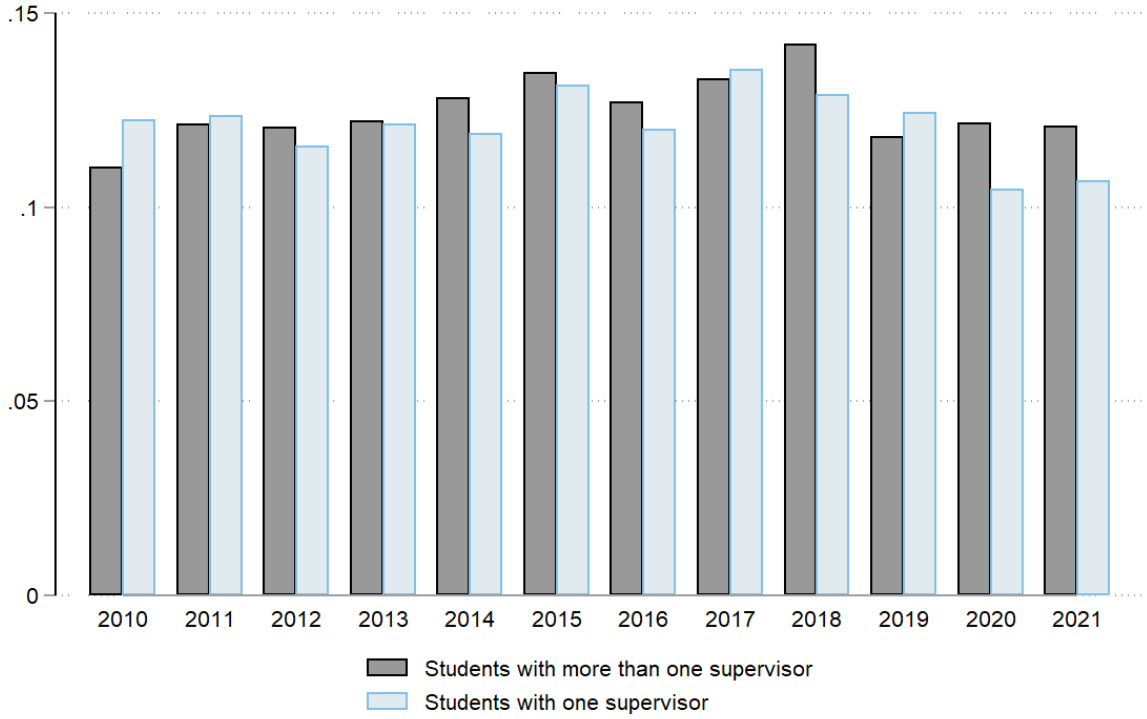
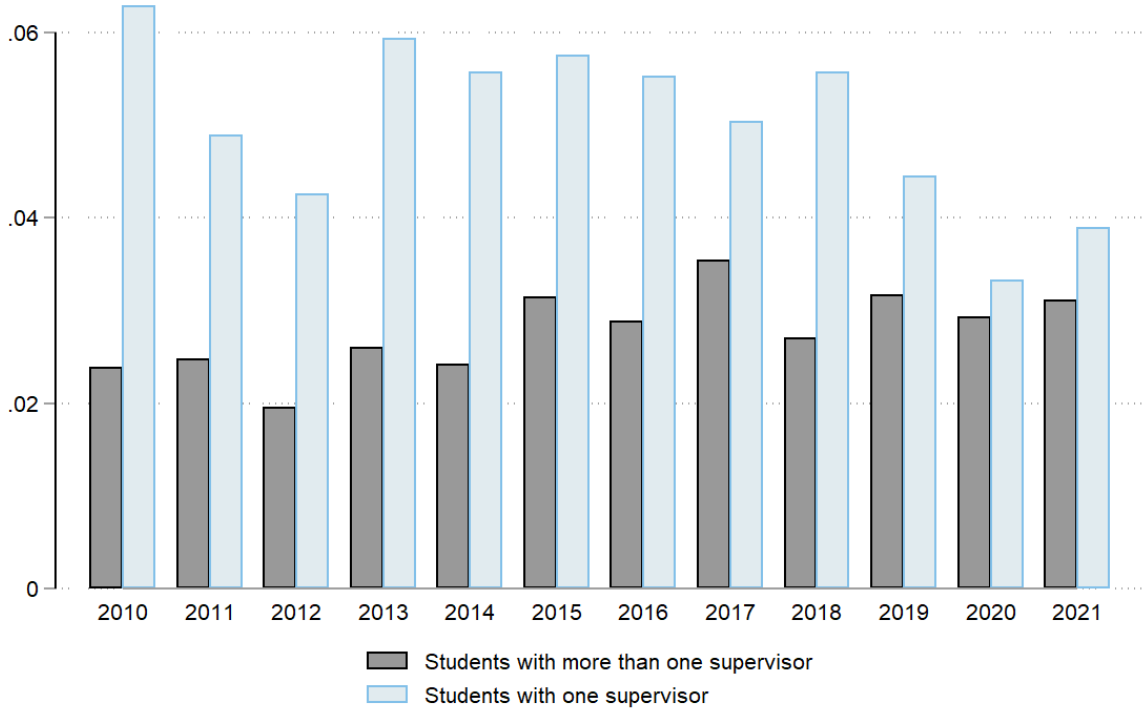


Figure A12. Share of discontinued PhDs by having a co-supervisor



Appendix B

Table B1. Heterogeneity analyses in time to graduation - Data aggregated at supervisor level

	Biomedical Sciences	Humanities and Social Sciences	Sciences, Engineering and Technology	Peer effects
COVID 2020	-2.35	-1.68	-0.55	-1.14
	(1.57)	(2.35)	(1.14)	(0.93)
COVID 2021	5.85***	1.25	1.10	2.57**
	(2.11)	(2.39)	(1.29)	(1.07)
Time trend	0.33*	1.00***	0.67***	0.66***
	(0.20)	(0.27)	(0.15)	(0.11)
Research group size	-0.52**	-0.52*	-0.25*	-0.19
	(0.25)	(0.31)	(0.13)	(0.12)
Constant	57.31***	48.86***	48.50***	52.59***
	(3.18)	(4.64)	(2.30)	(2.22)
Peer effect among EU				-1.18
				(2.62)
Peer effect among non-EU				-10.12***
				(3.86)
Peer effect among Men				-2.18
				(2.82)
Peer effect among women				-3.51
				(2.65)
Additional controls	Yes	Yes	Yes	No
Supervisor fixed effects	Yes	Yes	Yes	Yes
N	1 502	1 610	2 117	5 239
R2-adjusted	0.07	0.03	0.04	0.03

Note*:* The variable '*COVID_2020*' equals 1 for observations in the year 2020 only, while '*COVID_2021*' is set to 1 for observations in 2021. The 'Trend' coefficient captures the variations over time. The '*Research groupSize'* coefficient takes into account the number of students associated with supervisor *i* in a specific year *t*. The peer effect variables are between 0 and 1. Their coefficients capture the influence of the peers on an individual from the same group. Finally, "Additional controls" captures the variables X_i of equation (1), namely, the share of women or non-EU students and students with more than one supervisor.

Table B2. Heterogeneity sources among the graduation probability - Data at individual level

	Non-EU students only	EU students only	Women only	Men only	Student with one supervisor only	Student with more than one supervisor only
COVID 2020	-0.025**	-0.01	-0.01	-0.02**	-0.02***	-0.005
	(0.01)	(0.007)	(0.008)	(0.007)	(0.008)	(0.01)
COVID 2021	-0.028**	0.008	0.005	-0.01	-0.01	0.005
	(0.01)	(0.008)	(0.01)	(0.008)	(0.01)	(0.01)
Time trend	0.015***	0.01***	0.01***	0.01***	0.01***	0.01***
	(0.001)	(0.001)	(0.016)	(0.001)	(0.001)	(0.001)
Research group	-0.00	0.00	0.001	0.00	0.002*	-0.001
size	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-0.09***	-0.02**	-0.04***	-0.03**	-0.05***	-0.04**
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Non-EU			-0.03***	-0.004	-0.01*	-0.01***
			(0.005)	(0.003)	(0.004)	(0.004)
Women	-0.01**	0.001			-0.006	0.002
	(0.005)	(0.003)			(0.004)	(0.004)
Sharie of supervision	0.005***	0.001	0.003	-0.00		
	(0.003)	(0.002)	(0.003)	(0.003)		
Supervisor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	17 107	45 041	26 890	35 258	25 961	36 187
R2-adjusted	0.01	0.01	0.01	0.01	0.01	0.01

Note*:* The variable '*COVID_2020*' equals 1 for observations in the year 2020 only, while '*COVID_2021*' is set to 1 for observations in 2021. The 'Trend' coefficient captures the variations over time. The '*Research groupSize'* coefficient takes into account the number of students associated with supervisor *i* in a specific year *t*. The controls '*Non-EU'* and '*Women'* are dummies equal to 1 if the student was born outside the EU and if she is a women. '*Share of supervision*' indicates the number of supervisor by student.

Table B3. Heterogeneity sources among the dropout probability - Data at individual level

	Non-EU students only	EU students only	Women only	Men only	Student with one supervisor only	Student with more than one supervisor only
COVID 2020	-0.004	-0.01***	-0.01***	-0.01**	-0.02***	-0.003
	(0.005)	(0.003)	(0.005)	(0.004)	(0.005)	(0.003)
COVID 2021	0.004	-0.01**	-0.007	-0.004	-0.02***	0.003
	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)
Time trend	0.002***	0.003***	0.003***	0.002***	0.003***	0.002***
	(0.001)	(0.001)	(0.001)	(0.007)	(0.001)	(0.001)
Research group size	0.00	0.00	0.00	0.00	0.001**	-0.00
	(0.00)	(0.00)	(0.001)	(0.004)	(0.00)	(0.00)
Constant	0.03**	0.03***	0.04***	0.002***	-0.004	-0.005
	(0.01)	(0.006)	(0.00)	(0.008)	(0.01)	(0.006)
Non-EU			-0.01***	-0.00	-0.006*	-0.003
			(0.003)	(0.003)	(0.003)	(0.002)
Women	-0.003	0.01***			0.004***	0.005
	(0.004)	(0.002)			(0.003)	(0.002)
Share of supervision	-0.02***	-0.02***	-0.02***	-0.01***		
	(0.003)	(0.000)	(0.002)	(0.001)		
Supervisor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	17 107	45 041	26 890	35 258	25 961	36 187
R2-adjusted	0.003	0.004	0.005	0.003	0.001	0.002

Note*:* The variable '*COVID_2020*' equals 1 for observations in the year 2020 only, while '*COVID_2021*' is set to 1 for observations in 2021. The '*Trend'* coefficient captures the variations over time. The '*Research groupSize'* coefficient takes into account the number of students associated with supervisor *i* in a specific year *t*. The controls '*Non-EU'* and '*Women'* are dummies equal to 1 if the student was born outside the EU and if she is a women. '*Share of supervision*' indicates the number of supervisor by student.

Appendix C

Table C1. Main results - Top 5% deleted

	Data aggregated at supervisor level				*Data at individual level*		
	Time to graduation	Diploma attainment	Dropped out PhD	Started PhD	Time to graduation	Diploma attainment	Dropped out PhD
COVID 2020	-0.24	-0.07 ***	-0.03**	-0.06**	-0.58	-0.014**	-0.01***
	(0.70)	(0.02)	(0.013)	(0.03)	(0.62)	(0.006)	(0.003)
COVID 2021	1.94**	-0.03	-0.008	0.12***	1.42**	-0.007	-0.005
	(0.75)	(0.03)	(0.016)	(0.04)	(0.71)	(0.007)	(0.003)
Time trend	0.40***	0.02***	0.001	-0.02***	0.44***	0.01***	0.001***
	(0.08)	(0.003)	(0.001)	(0.003)	(0.07)	(0.00)	(0.00)
Research group size	-0.20**	0.14***	0.05***	0.21***	-0.14*	0.00	0.001**
	(0.08)	(0.005)	(0.002)	(0.006)	(0.08)	(0.001)	(0.00)
Constant	50.30***	-0.32***	-0.05***	0.20***	49.87***	0.001	0.03***
	(1.36)	(0.03)	(0.02)	(0.04)	(1.45)	(0.01)	(0.006)
Non-EU	-1.59***				-2.70***	-0.01***	-0.003*
	(0.55)				(0.43)	(0.003)	(0.002)
Women	0.88*				0.58	-0.003	0.006***
	(0.51)				(0.37)	(0.003)	(0.002)
Share of supervision	1.78***				1.94***	0.01**	-0.03***
	(0.54)				(0.43)	(0.003)	(0.002)
Supervisor fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	5 073	14 468	14 468	14 468	7 636	56 212	56 212
R2-adjusted	0.03	0.13	0.05	0.17	0.03	0.006	0,004

Note*:* The supervisor fixed effects control for observed and unobserved heterogeneity at the level of the supervisor. Standard errors are clustered at supervisor level. In the data aggregated at supervisor level, '*non-EU*' denotes the share of non-EU students supervised by the professor, '*women*' the share of females supervised by the professor, '*share of supervision*' stands for the share of students that are guided by more than one supervisor. In the estimations at individual level, these variables correspond to individual characteristics of the PhD student. As a result, '*share of supervision*' captures the number of supervisors guiding the student.

Table C2. Main results – Focus on research groups with at least two students

	Data aggregated at supervisor level				*Data at individual level*		
	Time to graduation	Diploma attainment	Dropped out PhD	Started PhD	Time to graduation	Diploma attainment	Dropped out PhD
COVID 2020	-0.82	-0.11***	-0.04**	-0.04	-1.18	-0.02***	-0.01***
	(0.97)	(0.03)	(0.02)	(0.04)	(0.83)	(0.005)	(0.003)
COVID 2021	2.43**	-0.03	-0.005	0.12**	2.02**	-0.01	-0.009**
	(1.11)	(0.04)	(0.02)	(0.05)	(1.00)	(0.006)	(0.003)
Time trend	0.70***	0.04***	0.005***	-0.03***	0.59***	0.01***	0.003***
	(0.12)	(0.003)	(0.002)	(0.004)	(0.10)	(0.001)	(0.00)
Research group size	-0.34***	0.14***	0.04***	0.19***	-0.25**	0.002**	0.001
	(0.11)	(0.006)	(0.003)	(0.006)	(0.12)	(0.001)	(0.00)
Constant	51.22***	-0.52	-0.10***	0.35***	50.20***	-0.01	0.02***
	(1.93)	(0.05)	(0.03)	(0.06)	(2.11)	(0.01)	(0.006)
Non-EU	-2.72***				-3.83***	-0.01***	-0.005**
	(0.83)				(0.56)	(0.003)	(0.002)
Women	-0.48				0.01	-0.002	0.004**
	(0.80)				(0.55)	(0.002)	(0.002)
Co-supervisor	1.06				1.27***	0.01***	-0.03***
	(0.73)				(0.35)	(0.001)	(0.002)
Supervisor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	4 586	10 293	10 293	10 293	6 955	57 407	57 407
R2-adjusted	0.04	0,13	0.04	0.16	0.03	0.005	0.004

Note*:* The supervisor fixed effects control for observed and unobserved heterogeneity at the level of the supervisor. Standard errors are clustered at supervisor level. In the data aggregated at supervisor level, '*non-EU*' denotes the share of non-EU students supervised by the professor, '*women*' the share of females supervised by the professor, '*share of supervision*' stands for the share of students that are guided by more than one supervisor. In the estimations at individual level, these variables correspond to individual characteristics of the PhD student. As a result, '*share of supervision*' captures the number of supervisors guiding the student.

Table C3. Main results - Missing observations

	Data aggregated at supervisor level		
	Diploma attainment	Dropped out PhD	Started PhD
COVID 2020	-0.07***	-0.04***	-0.03
	(0.025)	(0.014)	(0.03)
COVID 2021	-0.016	-0.006	0.18***
	(0.03)	(0.017)	(0.04)
Time trend	0.04***	0.006***	-0.03***
	(0.003)	(0.002)	(0.004)
Research group size	0.14***	0.045***	0.19***
	(0.005)	(0.003)	(0.01)
Constant	-0.50***	-0.10***	0.38***
	(0.04)	(0.02)	(0.05)
Supervisor fixed effects	Yes	Yes	Yes
N	13 638	13 638	13 638
R2-adjusted	0.13	0.04	0.15

Note*:* This table shows our main analysis using data aggregated at supervisor level for the number of graduated, started and dropped out PhDs on a new sample. In this case, if a supervisor has no student in charge in a given year, we replace the zero values for the graduate, dropped out and started variables by missing observations. The variable '*COVID_2020*' equals 1 for observations in the year 2020 only, while '*COVID_2021*' is set to 1 for observations in 2021. The '*Trend'* coefficient captures the variations over time. The '*Research groupSize'* coefficient takes into account the number of students associated with supervisor *i* in a specific year *t*.

Table C4. Main results - Resiliency analysis

	Data aggregated at supervisor level			
	Time to graduation	Diploma attainment	Dropped out PhD	Started PhD
COVID 2020	-1.30	-0.06 ***	-0.04***	-0.02
	(0.93)	(0.025)	(0.014)	(0.03)
COVID 2021	3.72***	0.06*	0.03*	0.20***
	(1.08)	(0.033)	(0.019)	(0.04)
Time trend	0.68***	0.03***	0.006***	-0.025***
	(0.12)	(0.003)	(0.001)	(0.003)
Research group size	-0.36***	0.13***	0.044***	0.20***
	(0.11)	(0.005)	(0.002)	(0.006)
Constant	51.42***	-0.41***	-0.09***	0.26***
	(1.84)	(0.03)	(0.02)	(0.05)
Non-EU	-2.68***			
	(0.80)			
Women	-0.84			
	(0.77)			
Share of supervision	1.12			
	(0.72)			
Supervisor fixed effects	Yes	Yes	Yes	Yes
N	5 239	14 609	14 609	14 609
R2-adjusted	0.03	0.12	0.04	0.16

Note*:* This table shows our main analysis using data aggregated at supervisor level for the number of graduated, started, dropped out PhDs and time to graduation, but exploring the resiliency of the students. To do so, we recode the 'COVID 2020' dummy to be 1 in 2020 and 2021, such that the 'COVID 2021' dummy presents an indicator for the resiliency after the pandemic. The '*Trend'* coefficient captures the variations over time. The '*Research groupSize'* coefficient takes into account the number of students associated with supervisor *i* in a specific year *t*. The controls '*Non-EU'* and '*Women'* are dummies equal to 1 if the student was born outside the EU and if she is a women. '*Share of supervision*' indicates the number of supervisor by student.

Table C5. Romano-Wolf Multiple Hypothesis

	Data aggregated at supervisor level			*Data at individual level*		
	Model p-values	Resample p-values	Romano-Wolf p-values	Model p-values	Resample p-values	Romano-Wolf p-values
Time to graduation						
COVID 2020	0.16	0.05	0.20	0.05	0.01	0.02
COVID 2021	0.02	0.003	0.018	0.04	0.003	0.01
Diploma attainment						
COVID 2020	0.008	0.002	0.007	0.03	0.001	0.01
COVID 2021	0.87	0.83	0.88	0.83	0.74	0.74
Dropped out PhD						
COVID 2020	0.005	0.001	0.001	0.0004	0.001	0.001
COVID 2021	0.76	0.64	0.88	0.11	0.03	0.05
Started PhD						
COVID 2020	0.50	0.35	0.73			
COVID 2021	0.00	0.001	0.001			

Note: Column 1 of the above table presents the p-values of the variable of interests estimated in our baseline model from Table 3. Column 2 presents the same p-values, but as average of a 1000 replications bootstrap. Finally, column 3 shows again the same p-values, but corrected by to the Romano-Wolf algorithm.

Appendix D

Table D1. Main results without supervisor's fixed effect

	Data aggregated at supervisor level				*Data at individual level*		
	Time to graduation	Diploma attainment	Dropped out PhD	Started PhD	Time to graduation	Diploma attainment	Dropped out PhD
COVID 2020	-0.55	-0.03	-0.02**	0.05*	-0.60	-0.02***	-0.008*
	(0.73)	(0.02)	(0.01)	(0.02)	(0.68)	(0.005)	(0.003)
COVID 2021	3.22***	0.06*	0.04**	0.24***	2.15***	0.004	0.001
	(0.96)	(0.02)	(0.02)	(0.04)	(0.77)	(0.005)	(0.003)
Time trend	0.55***	0.01***	0.002***	-0.02***	0.45***	0.004***	0.001***
	(0.04)	(0.001)	(0.000)	(0.001)	(0.04)	(0.000)	(0.00)
Research group size	-0.16***	0.12***	0.03***	0.18***	-0.14***	0.001	-0.001***
	(0.05)	(0.003)	(0.002)	(0.003)	(0.04)	(0.00)	(0.00)
Constant	51.28***	-0.16***	-0.007	0.29***	52.60***	0.07***	0.05***
	(0.81)	(0.01)	(0.007)	(0.01)	(0.75)	(0.004)	(0.003)
Non-EU	-2.40***				-3.93***	-0.005*	-0.006***
	(0.69)				(0.58)	(0.003)	(0.002)
Women	0.29				0.14	0.000	0.004**
	(0.55)				(0.48)	(0.002)	(0.002)
Share of supervision	1.15*				1.44***	0.01***	-0.02***
	(0.61)				(0.52)	(0.002)	(0.002)
Supervisor fixed effects	No	No	No	No	No	No	No
N	5 239	14 609	14 609	14 609	6 955	57 407	57 407
R2-adjusted	0.06	0.38	0.12	0.46	0.05	0.01	0,005

Note*:* The variable *'COVID_2020'* equals 1 for observations in the year 2020 only, while '*COVID_2021'* is set to 1 for observations in 2021. The '*Trend'* coefficient captures the variations over time. The '*Research groupSize'* coefficient takes into account the number of students associated with supervisor i in a specific year t. In the data aggregated at supervisor level, '*non-EU'* denotes the share of non-EU students supervised by the professor, '*women'* the share of females supervised by the professor, '*share of supervision*' stands for the share of students that are guided by more than one supervisor. In the estimations at individual level, these variables correspond to individual characteristics of the PhD student. As a result, '*share of supervision*' captures the number of supervisors guiding the student.

References

Aydemir, D., & Ulusu, N. N. (2020). Commentary: Challenges for PhD students during COVID-19 pandemic: Turning crisis into an opportunity. *Biochemistry and Molecular Biology Education*, 48(5), 428-429.

Betthäuser, B. A., Bach-Mortensen, A. M., & Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour*, 1-11.

Bloom, N., Bunn, P., Mizen, P., Smietanka, P., & Thwaites, G. (2020). The impact of Covid-19 on productivity (No. w28233). National Bureau of Economic Research.

Börgeson, E., Sotak, M., Kraft, J., Bagunu, G., Biörserud, C., & Lange, S. (2021). Challenges in PhD education due to COVID-19-disrupted supervision or business as usual: a cross-sectional survey of Swedish biomedical sciences graduate students. *BMC Medical Education*, 21(1), 1-11.

Chetty, Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers, [part] II: teacher value-added and student outcomes in adulthood. *The American Economic Review*, 104(9), 2633–2679.

Clarke, D., Romano, J. P., & Wolf, M. (2020). The Romano–Wolf multiple-hypothesis correction in Stata. *The Stata Journal*, *20*(4), 812-843.

Corbera, E., Anguelovski, I., Honey-Rosés, J., & Ruiz-Mallén, I. (2020). Academia in the Time of COVID-19: Towards an Ethics of Care. Planning Theory & Practice, 21(2), 191-199.

Covington, N. V., & Jordan, L. M. (2022). Impact of the COVID-19 Pandemic on PhD Students in Communication Sciences and Disorders. *Perspectives of the ASHA Special Interest Groups*, 7(2), 512-522.

Cui, R., Ding, H., & Zhu, F. (2022). Gender inequality in research productivity during the COVID-19 pandemic. *Manufacturing & Service Operations Management*, 24(2), 707-726.

Currie, J. & Thomas, D. (2001). Early test scores, school quality and SES: Long-run effects on wage and employment outcomes, *Research in Labor Economics*, 20, 103–132.

De Witte, K. and François, M. (2022). 'Learning deficits due to the COVID-19 analysis – A review of two years literature', EENEE Ad hoc report no. 04/2022.

Dohaney, J., de Róiste, M., Salmon, R. A., & Sutherland, K. (2020). Benefits, barriers, and incentives for improved resilience to disruption in university teaching. *International Journal of Disaster Risk Reduction*, 50, 101691.

Guest, E., McGinnis, S., Cao, X., Lee-Krueger, R., Mahjoub, G., McKie, K., ... & Varghese, D. (2021). Being a PhD student in the age of COVID-19. *Canadian Journal for New Scholars in*

Education/Revue canadienne des jeunes chercheures et chercheurs en éducation, 12(1), 30-38.

Hanushek, E. A., & Woessmann, L. (2020). The economic impacts of learning losses. OECD Education Working Papers, 225. Pp. 24.

Mata, J. A. L. D., & Samimi, S. A. B. (2022). Academic mothers during and post-COVID-19 pandemic. *Revista Brasileira de Saúde Ocupacional*, 47.

Minello, A. (2020). The pandemic and the female academic. Nature, 17(1).

Myers, K. R., Tham, W. Y., Yin, Y., Cohodes, N., Thursby, J. G., Thursby, M. C., ... & Wang, D. (2020). Quantifying the immediate effects of the COVID-19 pandemic on scientists.

Patrinos, H. A., Vegas, E., & Carter-Rau, R. (2022). An Analysis of COVID-19 Student Learning Loss. Policy Research Working Paper;10033. World Bank, Washington, DC.

Rodrigues, M., Silva, R., & Franco, M. (2021). Teaching and Researching in the Context of COVID-19: An Empirical Study in Higher Education. *Sustainability*, 13(16), 8718.

Statbel (2021). Une mère a en moyenne 31 ans à la naissance de son enfant. Retrieved January 27, 2023.

Wigginton, N. S., Cunningham, R. M., Katz, R. H., Lidstrom, M. E., Moler, K. A., Wirtz, D., & Zuber, M. T. (2020). Moving academic research forward during COVID-19. *Science*, 368(6496), 1190-1192.