Chapter 3: The effects of household cancer diagnosis on investment in higher education

3.1 Introduction

Education is classically understood to be an investment made by an individual in order to raise his productivity in the market, thereby justifying a demand for increased wages. While the notion that differences in human capital explains variation in earnings across individuals and jobs goes back to at least Mincer (1958),[26] major extensions by Becker (1962) and Ben-Porath (1967) laid the groundwork for countless theoretical additions, empirical investigations, and field journals in economics. Today Becker's Human Capital Model (HCM), which compares the stream of expected costs to the stream of expected benefits, is still at the heart of neoclassical economic models of education decisions.

This work is made possible in part due to a grant from the NBER and the Alfred P. Sloan Foundation. I also thank the Ohio Department of Health, Ohio Department of Higher Education, the Ohio Department of Job and Family Services, and CHRR at the Ohio State University for its generosity and support in allowing me to use its data for this analysis.

[26] Indeed, Toutkoushian and Paulsen (2016) and others trace the relationship between wages and differences in education to Adam Smith, Irving Fisher, and others.

Extensions of the HCM treat education investment decisions period by period, and allow a student to update his information set about himself or the labor market to dynamically compare costs and benefits (Aina et al., 2021). For instance, Manski (1992) finds that the graduation rate for students of affluent families significantly exceeds the rate of students from less wealthy families, which may be driven by additional costs that less affluent students undertake, for example by taking on a part-time job, or by increased costs through interest on student loans (Aina et al., 2021). On the other hand, a student making a decision to invest in education during a recession might expect that staying out of the labor force through higher education might allow the recession to pass to an economic boom upon graduation, thereby increasing expected earnings and thus increasing the likelihood of education attainment (e.g. Hampf et al. (2020), Ghignoni (2017)).

This essay uses a similar conceptual understanding as the cited example above. It considers a shock to the health of a student's household member (for example, a parent) as a cost that might affect a student's decision to remain enrolled in higher education. Past literature has shown, and I document here, that the effects of cancer diagnosis can be devastating to a household's financial well-being. For example, Gilligan et al. (2018) find that after two years, over 40% of patients over the age of 50 with a newly-diagnosed malignancy had already depleted their entire life savings.

The added health costs to the family may imply that financial assistance that would have otherwise been used towards paying for college is redirected to paying for medical treatment, thereby increasing the direct cost of education for the student. Indeed, a 2023 report from Sallie Mae found that in Academic Year 2022-2023, the mean cost of attending a 4-year public college was about $26.3 thousand. Of that $26,000, approximately $15,000 was paid for by parents or relatives (55%), and the ratio of dollars spent from parental income/savings to parental borrowing was about 5. By comparison, the sum of funding from student savings *and* borrowing was less than half of the amount financed through parental savings alone (Ipsos, 2023). This suggests that students are in large part depending on their parents for financial assistance in paying for college.

I use these motivating facts to hypothesize that the health shock increases the cost of obtaining education for the student, and therefore reduces the net benefits of college. As a result, I expect that students at the margin will re-evaluate their investment decision, and opt out of continuing their education. While the increased cost to the student is conceptually clear, the channel whereby the cost is increased is potentially ambiguous. In particular, the added time and concern (i.e. emotional costs) that the student spends with a household member may imply that additional effort is required to achieve the same level of benefit from higher education, thereby reducing the stream of expected economic benefits. In both channels, students at the margin may choose to reduce investment in education. I offer a more complete analysis of the Becker HCM in section 3.2.

While there is a very large literature that aims to empirically deduce the factors associated with college dropout, there is little empirical research that causally identifies the effect of an increase in cost on the investment in higher education in the United States.[27] In one study that directly queries students at an institution where direct

[27] While their analysis is not about the US, Melguizo et al. (2011) provide a useful state of the US literature and concerns about the interpretation of past results.

costs are nearly zero (Berea College), Stinebrickner and Stinebrickner (2008) show that credit constraints (as measured by demand for additional borrowing), particularly amongst students from low-income families, positively contributes to student dropout rates (i.e. decreased enrollment).[28] To my knowledge, mine is the first essay that uses a direct financial shock to the household, namely the diagnosis of cancer, to estimate the effect of increased cost on college enrollment. I assess this relationship empirically with a event study difference-in-difference model, which allows me to compare changes in the investment in higher education that results from a quasi-random shock to the household's finances, as well as a shock to the student's cost function, expanded on further below.

This essay proceeds as follows: First, I outline the mechanism whereby a medical shock to a household member alters the decision to invest in education, using advancements in the literature to Becker's HCM. Next, I outline specifics of the data, specifically as it relates to observing and classifying a medical shock to the household. I then describe my empirical strategy and accompanying results. Finally, I discuss the results as they relate to the conceptual model that I initially describe.

3.2 Conceptual model of investment in higher education

Consider a student who is currently enrolled in higher education in period t. The logic of the Human Capital Model specifies that this student will remain enrolled in higher

[28] In particular, dropout rates for students who were constrained were nearly two-times larger than students who were not credit constrained. However, the authors note that credit constraints were not the reason for dropout for 70% of the students who did not continue their education. Because Berea College is largely tuition free, one conclusion that could be drawn from this work is that finances are not a major factor in dropout determination.

education in $t + 1$ if and only if the net present value of expected benefit of going to college exceeds the net present value of expected costs of going to college. These costs include direct costs and indirect costs. Following Stinebrickner and Stinebrickner (2008), the decision to drop out in period t, $D(t)$, compares the expected utility of college, V_c, with the expected utility of dropping out of college (i.e. entering the LF immediately), V_d, evaluated with the information set at time t.

$$D(t) = 1 \text{ iff } E_t[V_c] - E_t[V_d] < 0 \tag{3.1}$$

A student currently enrolled in higher education, hence, by definition, is in a state where $E_t[V_c] > E_t[V_d]$. One can decompose $E_t[V_c]$ into a comparison of two components which allows for an equivalent and more intuitive interpretation, as outlined in Aina et al. (2021). That is, a student compares the expected utility of being in school with the expected costs of being in school. Utility is determined by the financial component of remaining in college – that is, does the state in which the student graduates college financially exceed the state in which the student disenrolls from college? – and a non-monetary component – that is, do the non-monetary benefits of college $[B_{N,M_t}]$ (e.g. social network, ecc.) outweigh the non-monetary costs of college $[C_{N,M_t}]$ (e.g. stress, ecc.), the latter ultimately being a function of effort. How much effort would it take for a student to make it through college, and does this non-monetary cost exceed the non-monetary benefits of college? He ultimately remains enrolled if the net benefit is positive.

Two additional observations may be conceptually important. First, the relationship between costs and benefits assumes that a student can accurately assess the net present value of the stream of future earnings. This assumption has often been rebutted in the empirical literature (Smith and Powell, 1990; Jerrim, 2015; Rouse, 2004; Betts, 1996), but is a critical component in the Becker HCM because it alters the parameters of the decision.

Second, is that the non-monetary costs may exceed the non-monetary benefits of college to such a degree that it even eclipses a positive financial benefit of college. For instance, a student might rationally unenroll from college, even if attending college has positive expected financial benefits, if the net non-monetary cost to the student of attending college while a parent is ill is so great that he is willing to forgo the financial benefit derived from higher education.

$$U(NPV_t, B_{NM_t}) > C_{NM}(e_t) \tag{3.2}$$

The first component of the utility function is merely an accounting exercise, but decomposing the Net Present Value, NPV_t, into the discounted streams of yearly earnings of a college graduate, Y_c, yearly earnings of a college dropout, Y_d, and direct monetary costs, C_M yields important intuition. Ultimately, one might first expect health shocks to manifest in the decision to invest in college by flipping the sign from the NPV from positive (i.e. it makes financial sense to attend college) to negative (i.e. the added cost has made it such that attending college is no longer financially beneficial).

$$NPV_t = \left[\sum_{j=x+1}^{L} \frac{Y_{c_t}^j}{(1+r)^j} - \sum_{j=1}^{x} \frac{C_{M_t}^j}{(1+r)^j} \right] - \sum_{j=1}^{L} \frac{Y_{d_t}^j}{(1+r)^j} \tag{3.3}$$

We can see from Equation (3.3) that an increase in the direct monetary cost of education thus decreases the NPV and, assuming that U is increasing in NPV, decreases utility. As a result, those students on the margin of Equation (3.2) may choose to dropout. The decrease in NPV from changes in cost due to a health shock from a household member may be the result of decreased financial family financial assistance, which increases the financial responsibility, C_M, of the student directly, and from an increase in the interest rate, r. This change in r may occur if the student previously receiving financial assistance from a parent instead must incur debt to finance higher education. Considering the first channel, we may expect to see changes in enrollment rates for those students whose family member is diagnosed with cancer, because the net benefit of higher education has decreased as cost increase.

A second channel whereby one might also expect a change in enrollment is through the right-hand side of the inequality in Equation (3.2). Due to a household member being sick, the student may reallocate time otherwise spent studying to aid the loved one. As a result, an increase in effort, e_t, for the increasing function of non-monetary costs C_{NM}, should result in students at the margin allocating time differently while enrolled. While this result is harder to observe practically than a change in the financial stream, this channel is also quite intuitive. For two comparable students, one with a sick parent and the second without, the amount of time and concern that the first student expends outside of his curricular studies is almost certainly greater than for the student without a sick parent. Given the time constraint on the students' time, at the margin the inference is that students with sick parents will have a greater likelihood of reducing investment in higher education than students without sick parents. Outcomes that might be related to the second channel include decreasing the number of courses or credits enrolled (or completed),[29] or reallocating time to less time-intensive courses.

In both channels identified and explained above, the expected behavior changes reenforce each other, leading me to predict *a priori* that a health shock to a household member, with accompanying financial distress, will result unambiguously in decreased enrollment rates. The empirical analysis that follows evaluates this prediction.

3.3 Data and Sample Construction

3.3.1 Data sources

Higher education information system

The source of data for enrollment and academic performance of students is the Higher Education Information System (HEI) from the Ohio Department of Higher Education.[30] At its most granular level, it provides course level enrollment and outcomes data for the universe of students enrolled in Ohio public colleges between academic years 2015 and 2020. It also provides some demographic information on the students that I take advantage of to define the sample, discussed more below. In particular, it provides information on the year of high school graduation and age. Using a secured hashing process, I am able to match the students in HEI to their credit data, which allows me to observe household members residing at the same address, as discussed more below.

[29] In certain cases, increasing or decreasing the number of credit hours below/above a threshold may have economic implications too by changing the cost for the student. I abstract away from that, here, and focus on the time aspect of changed course enrollment which is more directly related to the second channel than the first.

[30] https://highered.ohio.gov/data-reports/hei-system

Measures of student outcomes

Using the Ohio Department of Higher Education HEI, I construct measures of academic investment and performance to try to both understand the extensive and intensive margins of added costs, via household cancer diagnosis. In particular, I code several dependent variables, namely enrollment, course count, GPA, enrollment in a pass/fail course, a course ending in "Fail", "Incomplete", or "Withdraw", and the percentage of "Easy" courses in which the student enrolled. I aggregate data to the academic year.

To be considered enrolled, a student must have at least one enrolled course for two of the three semesters in an academic year. Once the student enters university, he remains in the sample until he graduated, which is noted in the HEI data, at which point he leaves the sample. Whereas the enrollment variable is unconditional, the additional outcomes are conditional on enrollment status. Year GPA is constructed in the conventional way for courses where credit is earned during a given academic year. A student who enrolls for a pass/fail course during the course of an academic year is coded as "1" for academic years where he is enrolled in any courses, and "0" when an academic year is completed with no pass/fail courses. Similarly, a student whose course is completed with the outcome "Incomplete", "Fail", or "Withdraw" is coded as "1" for academic years where he is enrolled in any courses, and "0" when courses are completed but none has the outcome of "Incomplete", "Fail", or "Withdraw". To construct the measure of the percent of easy courses, I used the HEI course data to identify courses at each campus of each university or college where the mean GPA was above the 75%-ile of courses offered at that campus. This measure is aggregated across courses for the academic year for each student. The intuition behind the selection of these variables is to try to capture the effect of the emotional burden described earlier, which may prompt dropout due to non-monetary cost. A student for whom a household cancer diagnosis becomes burdensome may remain in the college but suffer adverse course outcomes or enroll in easier courses.

Ohio cancer incidence surveillance system

For this research, I obtained access to the Ohio Cancer Incidence Surveillance System (OCISS) for 2015-2022, which is the state cancer registry for the state of Ohio, collected by the Ohio Department of Health. By Ohio law, all cancer diagnosis and treatments are required to be submitted to the OCISS, and thus this registry captures the universe of cancer diagnoses in Ohio. In addition to date of diagnosis, this registry provides information on the type of cancer, grade, laterality, site of tumor, and a few basic demographic characteristics of the patient. It provides an indicator for the mortality of the patient. Through a secure and anonymous hashing process that is congruent to the hashing process used to connect the HEI data to the other administrative datasets in this essay, the OCISS is linked to the Experian credit panel, and ultimately to the student via the student's household as defined in Experian and described above.

As shown in Figure 3.1, the number of cancer diagnoses is relatively consistent over time between 2015-2022, with a slight upward trend. An exception to this trend is the sharp decline in the first quarter of 2020 which coincided to COVID restrictions.

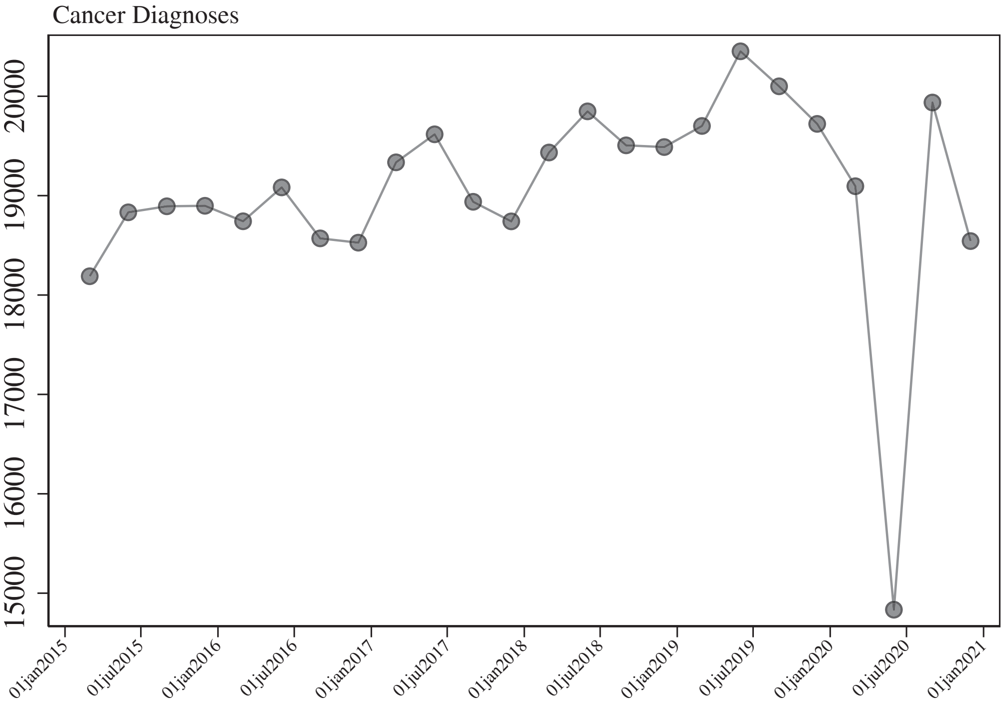

Figure 3.1: Ohio cancer diagnoses

Note: Plotted is the number of cancer diagnoses by quarter, as reported in the Ohio Cancer Incidence Surveillance System (OCISS).

	All [N=10,298]	Fatality [N=1,406]
1	Breast (25%)	Lung (24%)
2	Prostate (15%)	Pancreas (9%)
3	Skin (13%)	Misc. / Unclassified (7%)
4	Thyroid (5%)	Leukemia (6%)
5	Kidney/Pelvis (4%)	Breast (6%)

Table 3.1: Top 5 types of cancer affecting student households

Finally, using ICD-O-2 codes from the SEER database,[31] we can see that the distribution of cancer type for the household members of the student sample, shown in Table 3.1, is fairly consistent with the American population at large.[32]

Experian consumer credit data

I integrate consumer credit records from the Ohio Consumer Credit Panel (CCP). This dataset is comprised of the universe of consumer credit records from the state of Ohio quarterly from Q4 2015 to Q4 2021, approximately 8.8 million individuals or about 95% of the Ohio adult population.[33] It allows me to see a wide array of important financial characteristics at the quarterly level such as debt levels, measures of financial delinquency, public filings like bankruptcies, and credit score. Additionally, because I have the universe of consumers in Ohio, and because the dataset includes a household identifier that groups individuals together by the address listed in their credit records, I am also able to identify household members of the students. It, furthermore, allows

[31] `https://seer.cancer.gov/siterecode/icdo2_d01272003/`

[32] `https://seer.cancer.gov/statfacts/html/common.html`

[33] Prior studies estimate that approximately 11 percent of adults in the U.S. do not have a credit file (Brevoort et al., 2016). However, coverage in credit data has expanded over the past few years. Further, Ohio has a very small immigrant population and thus fewer people who have not yet established a credit file compared to states like California, Texas, Florida, and New York.

for the construction of household variables. I take advantage of both in later portions of this chapter.

Employment data

Though not the central focus of this study, I also integrate employment data about students and household members, which is provided by the Ohio Department of Job and Family Services (ODJFS). This employment data from ODJFS is quarterly data on all employed individuals in the state of Ohio, and includes quarterly wages and weeks employed. It is compiled as part of the Ohio Longitudinal Data Archive (OLDA), which is a project of the Ohio Education Research Center (oerc.osu.edu) and provides researchers with centralized access to administrative data. The OLDA is managed by The Ohio State University's Center for Human Resource Research (chrr.osu.edu) in collaboration with Ohio's state workforce and education agencies (ohioanalytics.gov), with those agencies providing oversight and funding.[34]

3.3.2 Sample Construction

Student sample

To construct the sample of individuals used in this analysis, I begin with the universe of college students enrolled in Ohio public institutions of higher education between the academic years of 2015 and 2020. I then apply sample restrictions to further refine the sample, as summarized in Table 3.2. First, I limit the sample to individuals who graduated from an Ohio high school. This restriction is necessary, as the cancer registry data that I have access to is collected by the Ohio Department of Health on Ohio diagnoses, and thus I cannot observe cancer diagnosis of household members

[34] For information on OLDA sponsors, see http://chrr.osu.edu/projects/ohio-longitudinal-data-archive .

for out-of-state students. Second, I limit the study population to undergraduate students. While the colloquial terms "Freshman", "Sophomore", "Junior", and "Senior" are employed in the dataset, they do not necessarily apply in the immediately-obvious way that matches the intuition of these labels. For example, an individual who took college credit in high school may not be considered a "Freshman" in his first year of university. Similarly, it is not uncommon for an individual continuously enrolled in a college to remain in a particular matriculation category for more than a year. As a result, I create cohorts of individuals based on their year of high school graduation corresponding to a semester of enrollment in a four-year college. Third, while 99.96% of students in the remaining are between 17 and 22 years old when they graduate high school, I further limit to the 99.9%-ile of age at high school graduation, which allows age to be up to 29-years old. High school graduation year is provided in the HEI dataset. Lastly, I limit the cohorts to those who begin in academic years 2015-2020, which are the years for which I have higher education data. I additionally make a few smaller limitations to the data, which are captured in a single row of Table 3.2. These exclusions include removing students who do not match to a conventional household or did appear in credit data in sufficient proximity to beginning university to conservatively attribute a household (details on both, below); students whose households are exceptionally large (more than ten individuals); and students who have a household member that was diagnosed with cancer before they began university.[35] Importantly, another feature that is captured in this step is that I exclude students that are from baseline households where the number of individuals of parent [32-67] or grandparent [68-100] age was 0 or more than 6. Because the baseline is typically in the quarter

[35] One limitation at this point is that the cancer registry data that I have only dates back until 2015, and so my visibility into which households are affected is imperfect.

	N of Students
0. Universe of students	1,247,567
1. Graduated from Ohio HS	948,786
2. Enrolled in undergraduate	915,652
3. Aged 17-29	521,188
4. Member of 2015-2020 cohort	223,923
5. Additional restrictions	188,957
6. Students without cancer diagnosis	188,468
Maximum Analytic Sample	183,902

Table 3.2: Sample restrictions

before beginning college, only about 6.5% of the remaining students are in atypical households. I exclude and note students who, themselves, are diagnosed with cancer during this period of study. Finally, I only consider students whose household members are affected by cancer in the years for which academic data is available, 2015-2020. The result of these exclusion criteria give a maximum analytic sample of 183,902 students, of whom about 10,000 are from households affected by cancer between 2015-2020, and just over 1,400 are from households affected by cancer that results in death.

Defining baseline household

Because the root of the analysis depends on identifying cancer diagnosis in a household member of the student, ideally I would be able to follow the family members of a student to track if one is diagnosed with cancer. Because this household data does not exist in either the Ohio administrative data, nor the Ohio cancer registry data that I have access to, I follow (Dettling and Hsu, 2018; Brown et al., 2012) by using consumer credit records (here, the Ohio CCP) to define relationships between household members.

I begin with the sample of students defined in 3.3.2, and locate their credit records in the credit panel. Using an anonymous household identifier, I select all the household members who reside at the same address as the student. The Ohio CCP includes nearly the entire universe of adults in Ohio, and so, when the student is available in credit data, I select the other individuals listed at this address as the baseline household members. Ideally, this is the last calendar year quarter of his first year of college. In reality, I am able to identify nearly 75% of students in credit data in their true baseline period.

One technical caveat to this general principle is that I only have access to credit data from 2015 to 2021, and so individuals who begin college in academic 2015 (i.e. calendar year 2014) use a quarter from 2015 instead of 2014. As I show below in Table 3.3, given the stickiness of households over time, this is unlikely to introduce much measurement error. Indeed, four-quarters after a quarter, 80.7% of the household remains intact, and nearly 60% of households are perfectly intact.

In this analysis, when credit data is unavailable for students in their true baseline periods, about 53% of the time, a credit record exists within 4-quarters of the true baseline. About 77% of the time, a credit record exists within 2-years of the true baseline. If a credit record does not exist for a student within 2-years of beginning university (before or after), he is excluded from the analysis. When there are multiple quarters which are not the true baseline period, I select the closest quarter to the baseline period, giving a preference to quarters before beginning university.

	Percent of Perfect Match	Percent of HH Match
Baseline	1.000	1.000
+1 Quarter	0.844	0.939
+2 Quarters	0.747	0.898
+3 Quarters	0.662	0.857
+4 Quarters	0.571	0.807
Total	0.772	0.902

Table 3.3: Household stability in credit data

Note: Households refers to the location where individuals report their credit, and does not necessarily mean that individuals live there. For example, a college student likely lives on campus, but reports his address to creditors as his "home address". Column (1) refers to the households that are exactly the same in quarter $baseline + q$. For example, 85% of the households after 1 quarter are exactly the same as households in the baseline quarter. Column (2) refers to the percent of the household in quarter $baseline + q$ that is the same as in the baseline quarter. For example, if there are 5 people in a household in the baseline quarter, after 4 quarters, 4 of the 5 household members (80%) are still reported at the same address.

One concern that one might have about selecting the student's household in the quarter that he began college is that he might have switched his address to a dormitory, and hence I am not capturing his family members. I noted above that one limitation on the sample that I made is to remove the students who have 0 parent or grandparent age individuals in their baseline household. This minimizes the risk that we are getting students in dorm or living with friends. Table 3.4 also reassures us that I am capturing an appropriate household. We can see the median baseline household is about three adult household members, with two members of the household being between 32 and 67 (parent age), and 0 members being between 68 and 100 (grandparent age) (Guzzo and Graham, 2022). I also provide some summary financial characteristics about the household in Table 3.4 to give a sense of the financial condition of the households at baseline. In particular, we can see that at baseline, the median student household appears to be in a relatively strong financial position, with low rates of delinquency, charge off, and bankruptcy, low levels of collections, and strong credit scores. We can also see that households that will eventually be treated are qualitatively similar to households in aggregate. In households where cancer results in fatality, I note that the household tends be slightly larger, with a higher likelihood of having a grandparent. Additionally, while certain repayment history metrics are comparable to non-treated households, there are several observable differences in baseline levels of credit score, balance in collections, and mortgage debt. This suggests that the household may be less financially advantaged than households where cancer diagnosis does not result in fatality and than untreated households.

	All				Ever Affected				Cancer Results in Fatality			
	Mean	Median	SD	N	Mean	Median	SD	N	Mean	Median	SD	N
Household Size	3.7	3.0	1.1	188468	4.0	4.0	1.2	11904	4.3	4.0	1.4	1421
Age of Youngest Non-Student	36.5	39.0	13.7	188468	36.4	38.0	14.6	11904	37.4	38.0	15.9	1421
Age of Oldest Non-Student	55.8	52.0	13.9	188468	60.1	55.0	14.2	11904	67.2	65.0	15.1	1421
Age Range of Non-Students	19.4	19.0	18.7	188468	23.6	27.0	19.4	11904	29.7	30.0	20.7	1421
Number of Parent Age	1.9	2.0	0.6	188468	2.0	2.0	0.6	11904	1.9	2.0	0.8	1421
Number of Grandparent Age	0.2	0.0	0.5	188468	0.3	0.0	0.7	11904	0.7	0.0	0.8	1421
Total Household Debt (K)	140.3	110.0	144.8	188468	146.0	116.7	149.4	11904	119.2	81.9	130.2	1421
Household Mtg Debt (K)	98.0	72.2	124.1	188468	102.6	77.1	128.3	11904	79.9	47.5	109.6	1421
Household CC Debt (K)	11.1	5.2	16.6	188468	12.0	5.9	17.3	11904	11.1	4.5	18.8	1421
Household Student Loan Debt (K)	24.5	7.1	43.7	188468	25.1	7.8	44.1	11904	24.3	7.4	44.0	1421
Total Collections	980.7	0.0	3421.6	188468	998.8	0.0	3659.8	11904	1789.8	0.0	5671.4	1421
Total Medical Collections	351.2	0.0	1783.2	188468	373.2	0.0	2216.0	11904	836.1	0.0	4591.7	1421
Maximum Credit Score in HH	741.9	764.0	87.4	188446	754.0	783.0	81.7	11904	737.5	756.0	85.9	1421
Charge off	0.0	0.0	0.2	188468	0.0	0.0	0.2	11904	0.0	0.0	0.2	1421
Discharged Bankruptcy	0.0	0.0	0.0	188468	0.0	0.0	0.0	11904	0.0	0.0	0.1	1421
60+ Day Delinquency Rate	0.1	0.0	0.3	188468	0.1	0.0	0.3	11904	0.1	0.0	0.3	1421

Table 3.4: Household characteristics of analytic sample of students at baseline

3.4 Empirical motivation

A significant amount of research has already shown that the impact of cancer to a household's financial health is significant. Shankaran et al. (2022) estimated that an individual was 1.71-times more likely to have an adverse financial event than those without cancer, and were 1.28-times more likely to have a past-due credit payment. Perhaps even more striking, Gilligan et al. (2018) estimated that after 2-years, nearly 43% of newly-diagnosed patients of malignant skin cancers had depleted their entire financial assets.

Before examining the effects of household cancer diagnosis on students' higher education outcomes, I first visualize trends of household financial outcomes centered around cancer diagnosis. To the extent that households commonly finance the higher education of children Ipsos (2023), understanding the relationship between cancer and financial distress is an important foundational step.

Figure 3.2 shows levels of household debt, collections (medical and total), credit score, charge off, and 60+ day delinquency relative to one quarter prior to cancer diagnosis. I plot trends for all cancer diagnoses and for the subset of cancer diagnoses where the patient ultimately dies. While there are striking characteristics for the full set of cancer diagnoses, the subset of cancers that resulted in fatality are especially alarming. We can see that while there is not much change to total household debt in the full sample, there is relatively immediate increase in mortgage and credit card debt relative to the pre-diagnosis trend, suggesting that households might be using debt to finance the medical burden of cancer treatment (Gupta et al., 2018). This is especially severe in households where cancer results in fatality, where there is a net increase.

Additionally, we observe an increase in medical and total collections in the three years after diagnosis. In terms of repayment behavior, we note an immediate increase in the likelihood of having a charged off debt or a delinquency at diagnosis, which remains elevated for at least four years after diagnosis. Perhaps most stark is the change in the household's maximum credit score around diagnosis. In the full sample, the growth appears to be linear pre-diagnosis, and remains stagnant for 3-4 years after diagnosis before it appears to begin to increase again. In the sample with fatal cancer, we observe a pre-diagnosis decline in credit score, but an immediately and notable drop in credit score even from the pre-diagnosis trends. This could be important for households who will potentially need to seek credit in the market, either to finance the disease management or higher education for students in the household. Finally, relating especially to this analysis, we observe a significant increase in household student loan debt in the years after cancer diagnosis. While it is unclear if this is directly related to the question studied in this analysis, its presence offers empirical support for the idea that cancer diagnosis may have inter-generational effects.

While Figure 3.2 only uses the household's prior behavior as a counterfactual, it provides suggestive evidence that the household is indeed under financial distress following cancer diagnosis. More specifically, it appears that there is little pre-diagnosis anticipation in terms of the financial distress indicators, other than linear trends that are likely associated with age. A second salient feature of Figure 3.2 is that cancers that result in fatality appear to be significantly more burdensome to the household than cancers that do not. This is not necessarily surprising, considering that two of

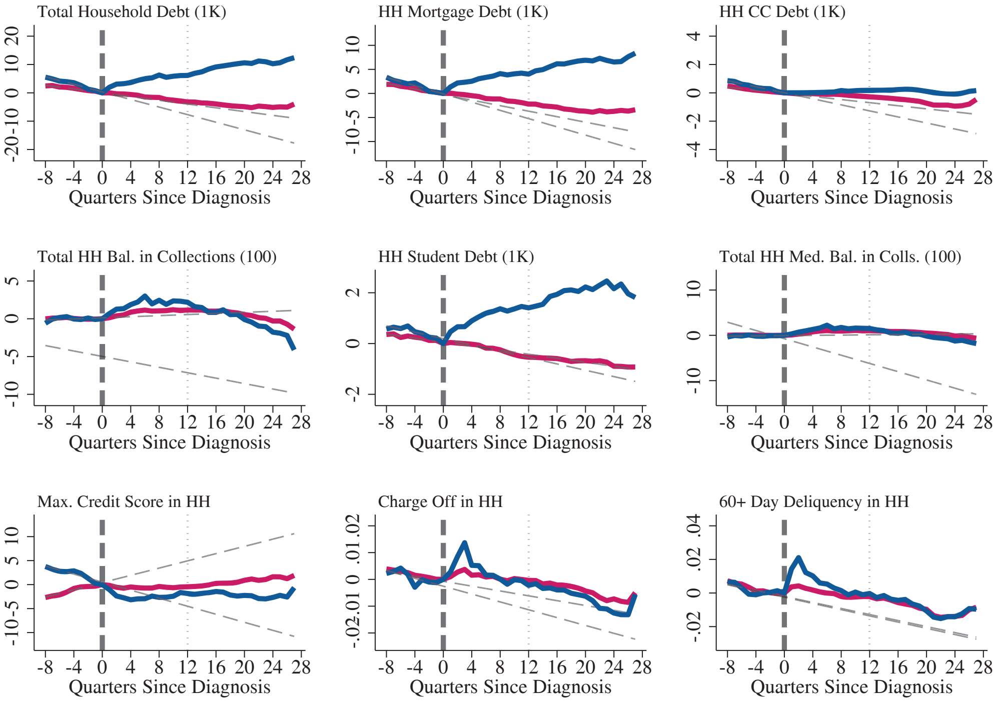

Figure 3.2: Household financial distress of cancer patient

Note: Measures are shown relative to *t-1*, where *t* is the quarter of cancer diagnosis. The cyclamen line represents all cancer types. The blue line is cancer diagnoses that result in fatality. Line of best fit is generated with data from eight proceeding quarters.

the three most common cancers in the registry are breast and prostate, which have 5-year survival rates of 91%[36] and 97%[37], respectively.

3.5 Empirical specification

I implement a stacked event study difference in differences model to estimate the impact of cancer diagnosis on a series of higher education outcomes. Though this style of model has been implemented before (Cengiz et al., 2019), I describe the model in greater detail below. Mathematically, I model as

$$Y_{i,n,s} = \alpha + \sum_{k=T_0}^{-2} \beta_k diag_{i,n,s} + \sum_{k=0}^{T_T} \beta_k diag_{i,n,s} + \theta_i + \sigma_s + \nu_n + \epsilon_{i,n,s}$$

Given the steep decline in likelihood of being enrolled in each subsequent year after beginning college, the primary time unit is a normalized year measure, n, which equates colloquially to “Freshman”, “Sophomore”, “Junior”, “Senior”, “Super Senior”, and “Super-Super Senior”, for student i in stack s. I include a fixed effect for normalized time, which captures typical changes in enrollment behavior over the span of college tenure. I also include a fixed effect for the individual, θ, which captures any unobservable time-invariant characteristics of the individual. Finally, given the modelling strategy that allows comparison students to be in multiple stacks, I include a stack fixed effect, σ, which isolates the proper comparison of treated students to other students who begin in the same year at the same college. I explain the model in detail in Appendix E.

[36] https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html

[37] https://www.cancer.org/cancer/types/prostate-cancer/detection-diagnosis-staging/survival-rates.html

In the event study difference in differences model, event time dummies are captured in β_k, where the parallel trends assumption suggests that $\beta_{k<0}$ is statistically zero, and the treatment effect relative to pre-treatment is captured in $\beta_{k\geq 0}$. The stacking strategy allows me to make the proper comparisons between individuals, and also solves the main critique of the two-way fixed effects estimator that has come under scrutiny lately (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; De Chaisemartin and d'Haultfoeuille, 2020), because I only compare students who are treated to students who will never be treated, and never to students who are already treated. It is worth noting, however, that because household cancer diagnosis is a relatively rare occurrence in this population, only a small share of the population will ever be treated, and so it is unlikely that this stacked strategy and a standard two-way fixed effects strategy will yield significantly different results since the relative share of "bad" comparisons (that is, comparisons of newly-treated to already-treated observations) is low.

One important note about the interpretation of the set of coefficients β_k: the intuitive interpretation of β_k is the difference in outcome, Y, relative to the baseline period $t-1$. For example, if β_0 is -0.1, then the interpretation is that in the first year after household cancer diagnosis, the decrease in enrollment is 10%. The event study difference in differences design allows this interpretation to have some inferential meaning, i.e. relative to a counterfactual group. However, one important point here is that due to the stark trends of some outcomes over the course of college *per se*, as is apparent below, it is more proper to understand β_k as the difference in outcome Y relative to the baseline period, conditional on the time fixed effect. In this case, conditional on the year of schooling. This is a subtle difference in interpretation, but can be helpful in understanding the coefficients in light of pronounced trends in dropout across college tenure.

3.5.1 Heterogeneity

As discussed earlier, Y is one of several dependent variables that aims to capture enrollment and outcomes that would indicate increased non-monetary cost or burden. I also assess heterogeneity in the treatment by comparing the effects across advantaged and disadvantaged students. To code this, I create four comparisons.

First, I consider heterogeneity by household wealth advantage, as measured by wage earnings. In particular, I suggest that there could be differences in academic outcomes for students whose families have greater wage earnings compared to students who come from relatively disadvantaged households. This is a direct implication of the theoretical model — students whose families are likely to be less financially burdened by cancer diagnosis are presumably less likely to have dramatic changes to their accounting costs. Empirically, I define compare individuals from households where the maximum wage earnings are in the bottom 25 percentile of household members to those in the top 75 percent (i.e. not the bottom 25 percentile).

Second, I consider heterogeneity by flexibility in household labor supply. The intuition behind this dimension of heterogeneity is that households that do not have two (or more) workers in the labor force may be ones that are financially constrained, and hence, similarly, accounting costs could be exacerbated. Conversely, a household that already has two individuals in the labor force can also be viewed as having less labor flexibility, and hence a limited ability to respond (Fisher et al., 2019). Non-monetary (caregiving) costs could be elevated. This dimension is explored empirically. I define disadvantaged in this dimension in the data as being from a household where no more than one individual is in the labor force.

Third, I consider differences in outcomes by students who have relatively high levels of student loans. Empirically, I define this as having student loans in the top 25 percentile of students. Importantly, this groups students who have zero student loans due to financial assistance, scholarships, or grants with young adults who have zero student loans due to a financially supportive household. While this is a counter intuitive grouping, the idea here is that it focuses on the group of students for whom the financial effects of the cancer diagnosis might be most severe.

The fourth dimension of heterogeneity that I exploit is geographic. While the travel (and hence financial and time) costs of a young adult who attends college far away is greater than for a local student, another important difference between these two groups of individuals is the extent to which caregiving may be feasible. I define the disadvantaged group in this context as those who live at or above the 75 percentile of distance from home to college.[38] Empirically this equates to about 100 miles (160 km), or roughly the distance from Cincinnati to Columbus. As noted, the majority of caregiving is done locally, and so we might expect that the demand to assist in offering care for this group of young adults is reduced compared to the group who is local.

[38] I use the Haversine formula to calculate the distance between the geographic coordinates of the centroid of the home zip code and college campus zip code of each young adult.

3.6 Results

3.6.1 Enrollment and graduation

The primary specification in this essay is an analysis of enrollment and graduation rates. The theoretical model suggests that a shock to the household that is sufficiently burdensome should decrease enrollment rates and graduation rates. Figure 3.3 depicts trends over college tenure in enrollment rates and in graduation rates in three groups. The black line depicts the trend for the majority of students whose household is never directly affected by cancer. The blue line depicts the trend for the group of students whose household is affected by cancer and the cancer results in mortality. And the cyclamen line depicts the majority of cancer diagnoses where the household member affected with cancer has not died. An important note about the trends is that the figure conceals event time because it aggregates students along their college tenure. For example, students who are affected later in college are still represented by the blue or cyclamen line even if cancer diagnosis has not yet occurred. We can see that the likelihood of being enrolled for at least two semesters in any particular year of college is actually greater for cancers that do not result in fatality compared to the group of students who never experience a household cancer diagnosis. By contrast, we see that for households where cancer ultimately results in mortality, the likelihood of enrollment is lower at all points. This characteristic is true for graduation rates as well.

I dive more deeply into graduation by modeling a series of cross-sectional logistic regressions to try to understand the difference in graduation rate in the cancer groups relative to the comparison group. Panel (b) of Figure 3.4 depicts these regressions,

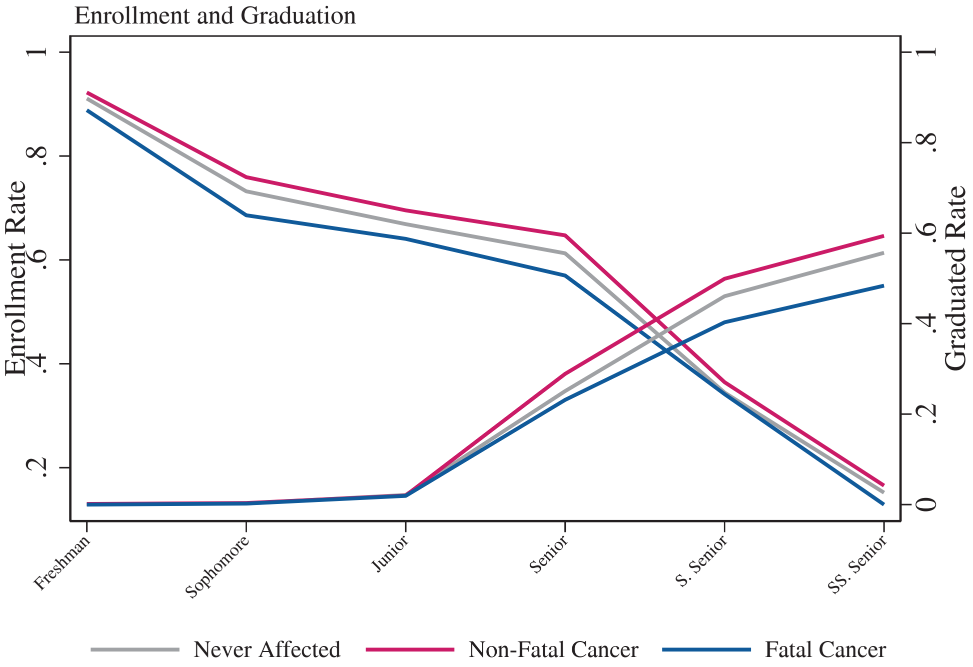

Figure 3.3: Trends of enrollment and graduate rates

Note: Plotted are the enrollment and graduation rates for students never affected by household cancer (black), affected by a diagnosis that does not result in fatality (cyclamen), and those with a fatal result (blue). Outcomes are plotted over college tenure. Enrollment is censored after graduation. Graduation is cumulative.

while Panel (a) models the same research question with a cox proportional hazard model.[39] The results of the two models generally tell us the same thing, namely that students of households with severe cancer are significantly less likely to graduate. The results from the logistic regression suggest a statistically-significant and consistent odds ratio of about 0.79 for years 4, 5, and 6, suggesting that individuals affected by severe types of cancer are about 80% as likely to graduate as students whose households are not affected by cancer. By contrast, students with households affected by cancers that do not result in mortality are (statistically) no more or less likely to graduate than the comparison group. These results, however, are only cross-sectional and the interpretation of the coefficients, while suggestive, could be biased by unobserved characteristics of the student.

Turning to enrollment, Figure 3.5 makes use of the stacked difference in differences dataset to visualize the mean enrollment rate by each of the three groups in event time. We can see that even though enrollment does decline for students affected by fatal and non-fatal cancer diagnoses, the decline in the counterfactual, comparison group is congruent. Indeed, the decline for the non-fatal cancer diagnosis tracks nearly identically with the comparison group. Unsurprisingly, then, the results in panel (b) reveal a precisely estimated null difference in enrollment for the entire sample of students from households affected by cancer. The lack of precisely estimated pre-trends in both groups suggests parallel trends pre-treatment, even though enrollment in Figure 3.3 was always below the comparison group in every year of college. Observationally, the trends in Figure 3.3 appear parallel, and panel (b) of Figure 3.5 confirms this. By

[39] In these models, controls for gender, cohort [i.e. which academic year the student began university], and campus were included.

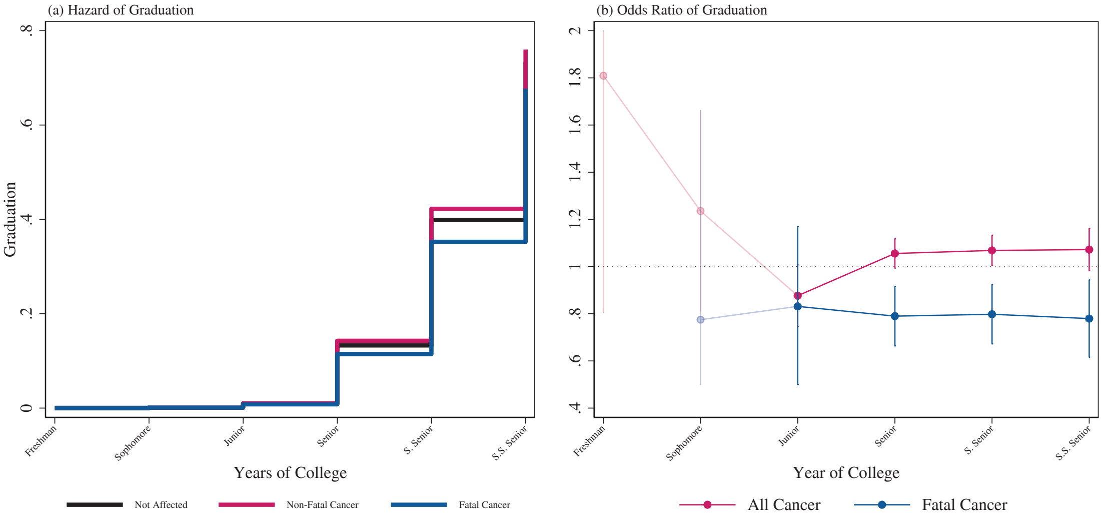

Figure 3.4: Cox and logit model results - graduation

Note: This figures plots descriptive evidence that graduation rates are lower for students whose households are affected by cancer. The left panel plots the results of a cox proportional hazard model, where graduation is the failure event. The results graphed do not include controls for academic cohort nor campus nor gender, though results are nearly equivalent when included. The right panel plots the results of logistic regressions of graduation at each year of college, from year one ("freshman") to year six ("super-super senior"). The odds ratios of each model are plotted in this panel with their 95% confidence intervals, and can be interpreted as the odds of graduating for each plotted group relative to the odds of graduating in the untreated group. These models include controls for academic cohort, campus, and gender.

contrast, panel (a) shows an acceleration in the decline in enrollment in event time zero and one before stabilizing in year two for students from households affected with severe cancer. Indeed, we see in panel (b) a precisely estimated point estimate of about a 2.6% decline in enrollment in the year of diagnosis, and an imprecisely estimated decline in enrollment in the first year after diagnosis of about 1.7%, relative to the baseline enrollment rate. These results suggest that for the majority of students affected with household cancer, their probability of enrollment is no different than students without the health shock, or rather that the effect of a health shock is not strong enough to observe differences in enrollment. For a small group of individuals with a severe health shock, there is some empirical evidence of an immediate effect that statistically attenuates after the first year. While we do not observe much difference in enrollment, it appears that the cumulative decrease in enrollment (at least the point estimate) ultimately results in a decreased likelihood of graduation at four years and beyond.

3.6.2 Additional education outcomes

Theoretically the decision to remain enrolled is a function of monetary and non-monetary costs of college. While we do not observe a difference in enrollment for students from cancer-affected households in aggregate, aside from from the first year for students of households with severe cancer, I examine a few additional education-related outcomes to assess the extent to which cancer may act as an emotional strain on the student, and affect his performance in school.

For enrolled students, Figure 3.6 shows that the trends in the number of enrolled courses for those with cancer and those without cancer are very similar. The results

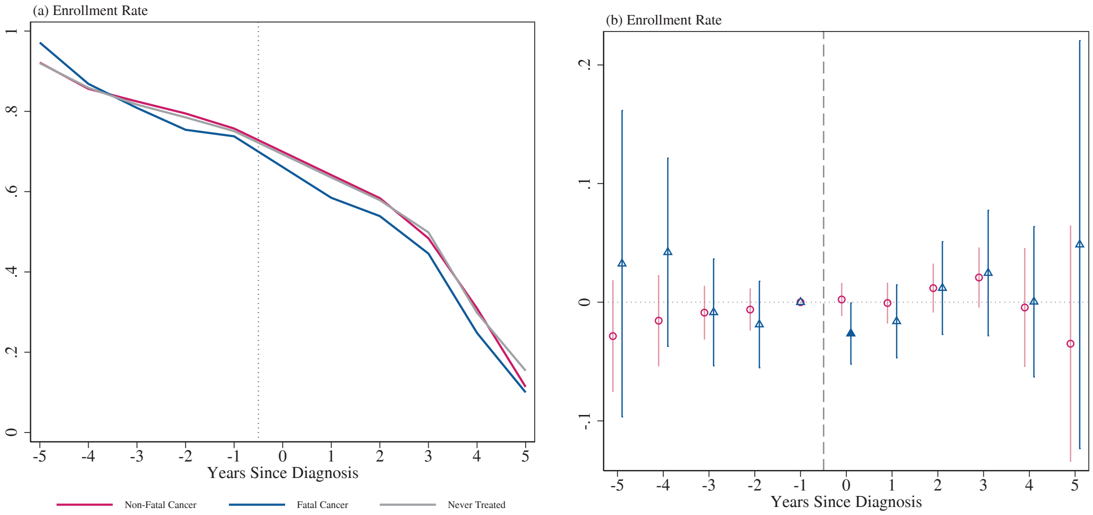

Figure 3.5: Means in event time and difference in differences results - enrollment

Note: The left panel of this figure displays the trends in enrollment, centered in event time, for students affected with a non-fatal household cancer diagnosis (cyclamen), and for students affected with a fatal household cancer diagnosis (blue). The mean of enrollment in the counterfactual group, i.e. students who were are never affected by household cancer diagnosis, is plotted in black. Event time for this group is created by the stacking process outlined in the Empirical Specifications section. In the right panel, the results of the event study difference in differences are displayed. The results come from a stacked event study difference in differences, where stacks are created for each cohort of student, 2015-2020, at each public college in Ohio.

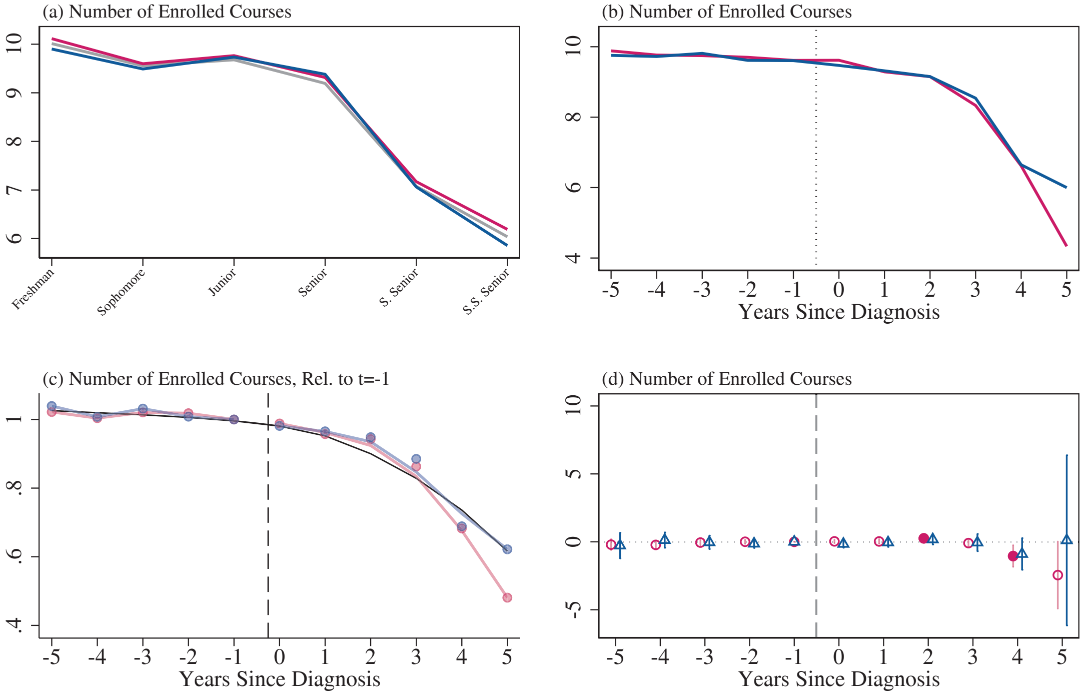

Figure 3.6: Difference in differences results - number of enrolled courses

Note: This figure presents four insights into understanding the relationship of cancer diagnosis and course enrollment for the college student. The number of courses enrolled is defined as the count of the number of courses in which a student enrolled in an academic year, regardless of whether the student completed the course or not. The measure allows for students to be enrolled at another Ohio public college. In panel (a), the count for three groups of students are plotted over college tenure, agnostic to cancer timing. The grey line depicts the group of students who will never have a household member that is diagnosed with cancer. The cyclamen and blue lines represent the groups of students who will experience non-fatal and fatal household cancer diagnoses, respectively. In panel (b), trends are depicted in event time, and are limited to the group of students who experience non-fatal household cancer diagnosis (cyclamen), or fatal household cancer diagnosis (blue). Panel (c) plots the ratio of count at each period in event time relative to t-1. The black line is the local regression for the non-treated counterfactual. The cyclamen line in panel (c) represents all cancer diagnoses, and the blue line represents only fatal cancer diagnoses. The cyclamen and blue lines are the local linear regression for the respectively-colored points. Panel (d) is the event study difference in differences model, as outlined in the Empirical Specification section. The results come from a stacked event study difference in differences, where stacks are created for each cohort of student, 2015-2020, at each public college in Ohio.

of the model suggest little difference in the first two years immediately after diagnosis in aggregate and for those with severe cancer. There is suggestive evidence that in the third year of cancer diagnosis that students may enroll in slightly more courses, though this effect dissipates immediately and reverses in the fourth year (i.e. isolated in students diagnosed early in college who have not yet graduated in the fourth or fifth year of college).

We similarly see no immediate effect on GPA (Figure 3.7), the percent of easy courses taken (Figure 3.11), or in the number of courses failed, left incomplete, or withdrawn (Figure 3.10) in the immediate years after diagnosis. Similar to course enrollment, there is evidence of increased GPAs, and decreased count of courses failed, left incomplete, or withdrawn for the subset of students diagnosed early in their college tenure who have not yet graduated by the fourth, fifth, or six years.

The two small pieces of evidence that point to any sign of distress that is manifesting in education outcomes for students are shown in the bottom right panels of Figures 3.8 and 3.9. In particular, we can see that summer enrollment increases marginally in the first year after diagnosis for students whose household member is affected by cancer. While summer term is sometimes considered to be the end of the calendar year, technically the way that it is classified is as the first semester of the year, and so an increase in the summer enrollment rate in the second year would actually correspond to the first summer after diagnosis. One possible significance of this is that students are potentially stretching their enrollment over three terms instead of two, relative to students whose household members are not affected by cancer.

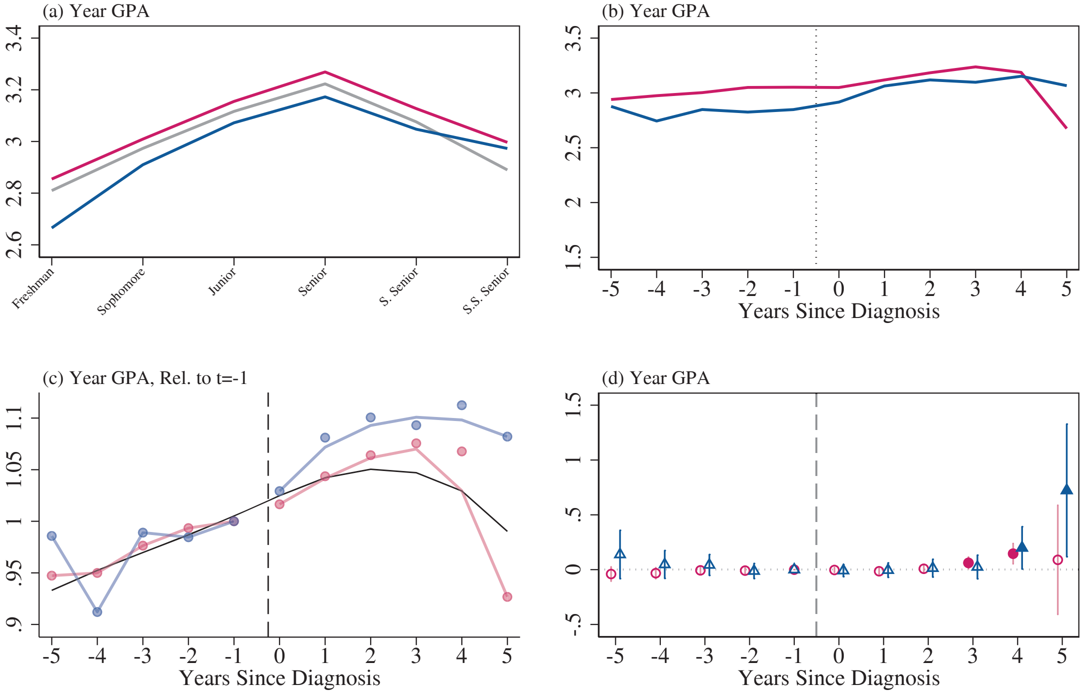

Figure 3.7: Difference in differences results - GPA of academic year

Note: This figure presents four insights into understanding the relationship of cancer diagnosis and academic achievement for the college student. The GPA of an academic year is defined as the ratio of GPA points earned in a year to GPA hours earned in a year. In the instance where students do not complete courses enrolled during an academic year, there is no GPA calculated. The measure allows for students to be enrolled at another Ohio public college. In panel (a), the GPA for three groups of students are plotted over college tenure, agnostic to cancer timing. The grey line depicts the group of students who will never have a household member that is diagnosed with cancer. The cyclamen and blue lines represent the groups of students who will experience non-fatal and fatal household cancer diagnoses, respectively. In panel (b), trends are depicted in event time, and are limited to the group of students who experience non-fatal household cancer diagnosis (cyclamen), or fatal household cancer diagnosis (blue). Panel (c) plots the ratio of GPA at each period in event time relative to t-1. The black line is the local regression for the non-treated counterfactual. The cyclamen line in panel (c) represents all cancer diagnoses, and the blue line represents only fatal cancer diagnoses. The cyclamen and blue lines are the local linear regression for the respectively-colored points. Panel (d) is the event study difference in differences model, as outlined in the Empirical Specification section. The results come from a stacked event study difference in differences, where stacks are created for each cohort of student, 2015-2020, at each public college in Ohio.

Secondly, Figure 3.9 shows a general increase in the number of pass/fail courses that a student from a cancer household enrolls for relative to non-affected students. While this is statistically insignificant in the first three years, and only significant in the fourth year, the general trend of the point estimates suggests a trend towards increased enrollment in courses with easier outcomes. Panels (a), (b), and (c) of Figure 3.9 also support this trend visually.

Taken together, these results present a counter-intuitive result in the relationship between household cancer diagnosis and student academic achievement. Despite mounting evidence that cancer is financially disastrous and extremely burdensome for households, I find little evidence that these time and financial burdens are reflected in student outcomes. Aside from an immediate 3% decrease in the enrollment rate of students from a household with severe cancer, I find no difference in the enrollment rates of affected students relative to non-affected students, even after implementing strategies to compare students from the same college in the same tenure of college. Sadly, it appears that when a cancer diagnosis in the household of a student results in dropout, that the student would have otherwise dropped out anyway. Moreover, I find little evidence that those who do remain enrolled are much different in other academic outcomes than their peers. I find some evidence of a small increase in summer enrollment in the first summer after diagnosis, and some suggestive (though largely statistically insignificant) evidence of an increase in enrollment in pass/fail courses. When paired with no change in the number of courses enrolled each academic year, it suggests that students might be spreading out course load or chronic stress load.

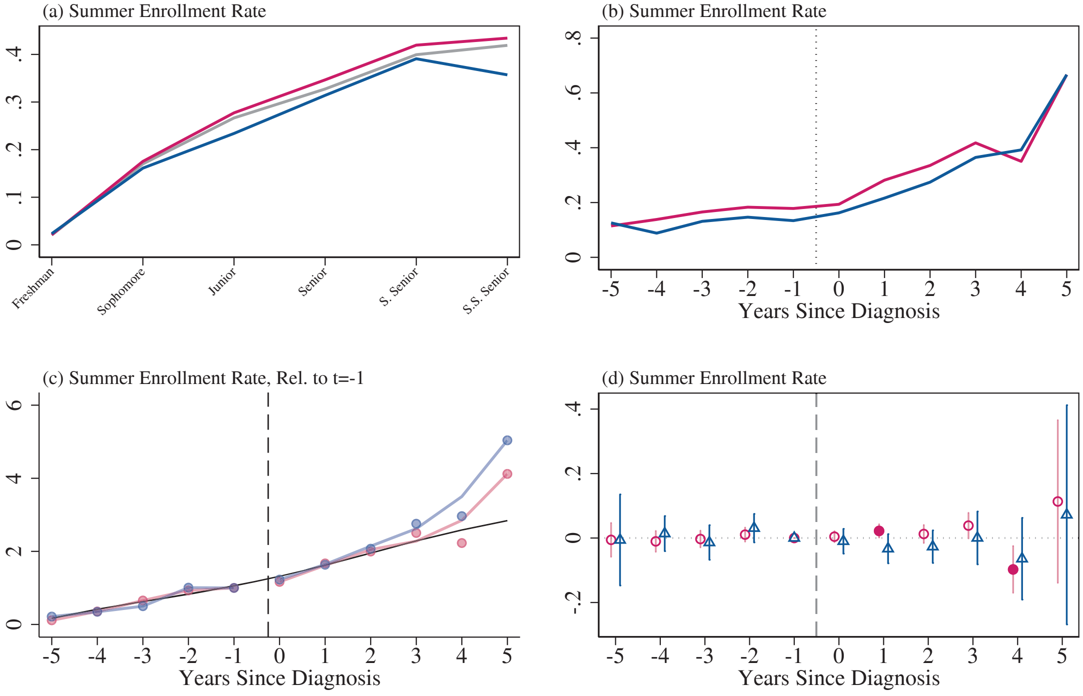

Figure 3.8: Difference in differences results - enrollment in summer term

Note: This figure presents four insights into understanding the relationship of cancer diagnosis and summer enrollment for the college student. Summer enrollment defined as an indicator for having enrolled in a course in summer term during an academic year, regardless of whether the student completed the course or not. The measure allows for students to be enrolled at another Ohio public college. In panel (a), the count for three groups of students are plotted over college tenure, agnostic to cancer timing. The grey line depicts the group of students who will never have a household member that is diagnosed with cancer. The cyclamen and blue lines represent the groups of students who will experience non-fatal and fatal household cancer diagnoses, respectively. In panel (b), trends are depicted in event time, and are limited to the group of students who experience non-fatal household cancer diagnosis (cyclamen), or fatal household cancer diagnosis (blue). Panel (c) plots the ratio of count at each period in event time relative to t-1. The black line is the local regression for the non-treated counterfactual. The cyclamen line in panel (c) represents all cancer diagnoses, and the blue line represents only fatal cancer diagnoses. The cyclamen and blue lines are the local linear regression for the respectively-colored points. Panel (d) is the event study difference in differences model, as outlined in the Empirical Specification section. The results come from a stacked event study difference in differences, where stacks are created for each cohort of student, 2015-2020, at each public college in Ohio.

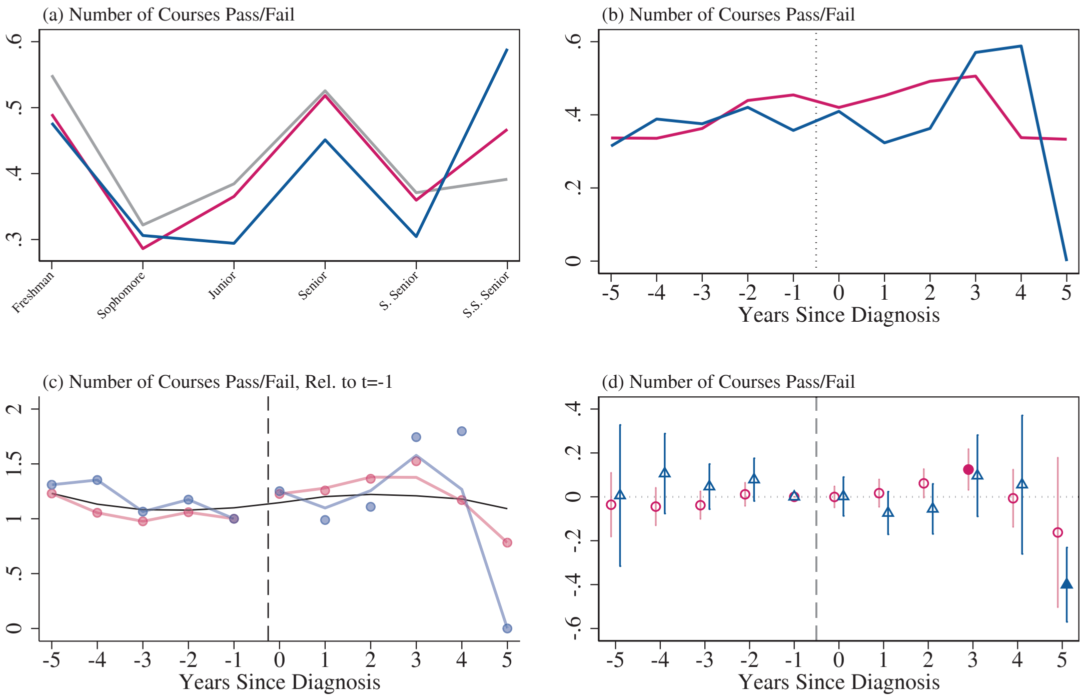

Figure 3.9: Difference in differences results - number of pass/fail courses enrolled

Note: This figure presents four insights into understanding the relationship of cancer diagnosis and enrollment in pass/fail courses for the college student. The number of courses enrolled in as pass or fail is defined as the count of courses enrolled during an academic year where the course type is pass or fail, regardless of whether the student completed the course or not. The student may have passed or failed the course. The measure allows for students to be enrolled at another Ohio public college. In panel (a), the count for three groups of students are plotted over college tenure, agnostic to cancer timing. The grey line depicts the group of students who will never have a household member that is diagnosed with cancer. The cyclamen and blue lines represent the groups of students who will experience non-fatal and fatal household cancer diagnoses, respectively. In panel (b), trends are depicted in event time, and are limited to the group of students who experience non-fatal household cancer diagnosis (cyclamen), or fatal household cancer diagnosis (blue). Panel (c) plots the ratio of count at each period in event time relative to t-1. The black line is the local regression for the non-treated counterfactual. The cyclamen line in panel (c) represents all cancer diagnoses, and the blue line represents only fatal cancer diagnoses. The cyclamen and blue lines are the local linear regression for the respectively-colored points. Panel (d) is the event study difference in differences model, as outlined in the Empirical Specification section. The results come from a stacked event study difference in differences, where stacks are created for each cohort of student, 2015-2020, at each public college in Ohio.

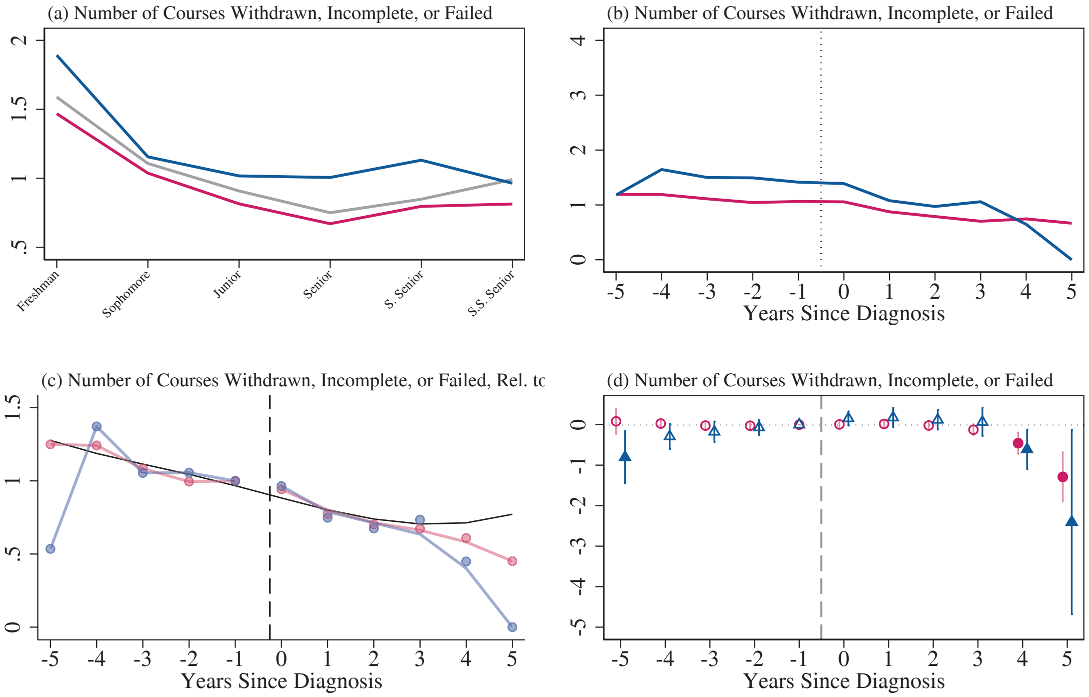

Figure 3.10: Difference in differences results - number of courses finished with fail, incomplete, or withdraw

Note: This figure presents four insights into understanding the relationship of cancer diagnosis and non-satisfactory course completion for the college student. The measure is defined as the count of courses in which the outcome is either fail, incomplete, or withdraw. Naturally, the student must have completed the course to be counted in this measure. The measure allows for students to be enrolled at another Ohio public college. In panel (a), the count for three groups of students are plotted over college tenure, agnostic to cancer timing. The grey line depicts the group of students who will never have a household member that is diagnosed with cancer. The cyclamen and blue lines represent the groups of students who will experience non-fatal and fatal household cancer diagnoses, respectively. In panel (b), trends are depicted in event time, and are limited to the group of students who experience non-fatal household cancer diagnosis (cyclamen), or fatal household cancer diagnosis (blue). Panel (c) plots the ratio of count at each period in event time relative to t-1. The black line is the local regression for the non-treated counterfactual. The cyclamen line in panel (c) represents all cancer diagnoses, and the blue line represents only fatal cancer diagnoses. The cyclamen and blue lines are the local linear regression for the respectively-colored points. Panel (d) is the event study difference in differences model, as outlined in the Empirical Specification section. The results come from a stacked event study difference in differences, where stacks are created for each cohort of student, 2015-2020, at each public college in Ohio.

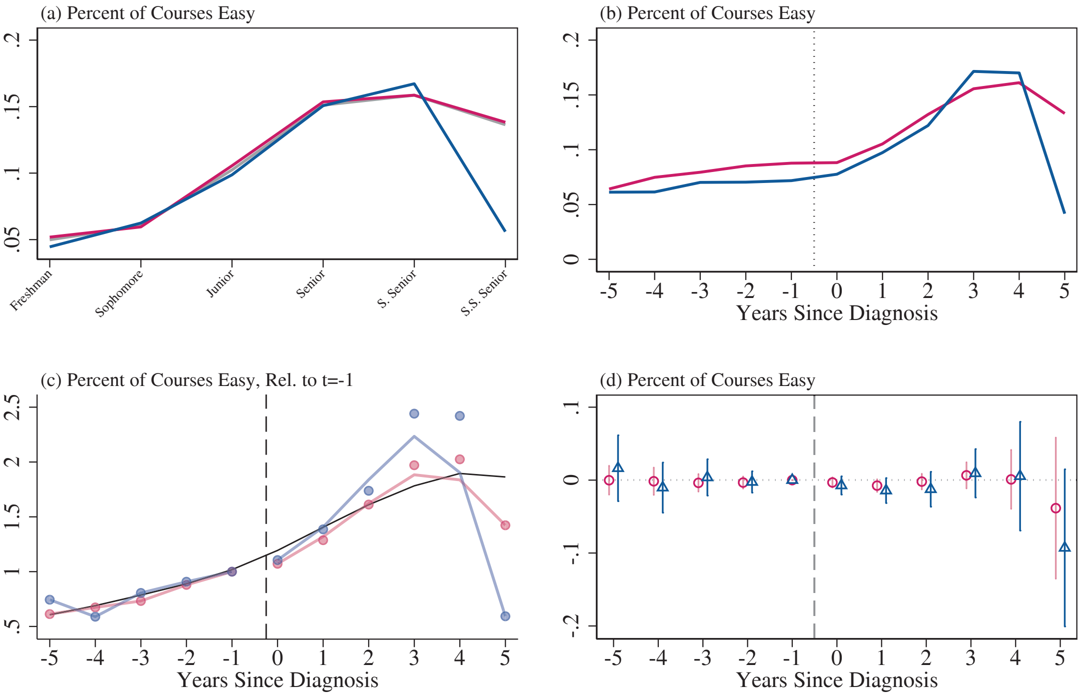

Figure 3.11: Difference in differences results - percent of easy courses enrolled

Note: This figure presents four insights into understanding the relationship of cancer diagnosis and easy course enrollment for the college student. The measure is defined as the count of courses in which the course is categorized as easy. I define an easy course as one where 75% of students enrolled in the course receive a GPA above the mean GPA of the college campus. The measure allows for students to be enrolled at another Ohio public college. In panel (a), the percent for three groups of students are plotted over college tenure, agnostic to cancer timing. The grey line depicts the group of students who will never have a household member that is diagnosed with cancer. The cyclamen and blue lines represent the groups of students who will experience non-fatal and fatal household cancer diagnoses, respectively. In panel (b), trends are depicted in event time, and are limited to the group of students who experience non-fatal household cancer diagnosis (cyclamen), or fatal household cancer diagnosis (blue). Panel (c) plots the ratio of the measure at each period in event time relative to t-1. The black line is the local regression for the non-treated counterfactual. The cyclamen line in panel (c) represents all cancer diagnoses, and the blue line represents only fatal cancer diagnoses. The cyclamen and blue lines are the local linear regression for the respectively-colored points. Panel (d) is the event study difference in differences model, as outlined in the Empirical Specification section. The results come from a stacked event study difference in differences, where stacks are created for each cohort of student, 2015-2020, at each public college in Ohio.

3.7 Heterogeneity

In this section, I consider four types of heterogeneity in students that seem like reasonable sources where we could see differences in education outcomes. Because the samples get thin in the tails, as evidenced by several of the confidence intervals getting large, I focus the analysis on Years Since Diagnosis 0-3.

3.7.1 Household income

In particular, I consider differences in outcomes of students who come from households where the maximum wage earner in the household earns annual wages in the bottom 25% of the sample, compared to the top 75%. If there is an effect driven by an inability for households to contribute to paying for college, it seems likely it would be concentrated in this group. Moreover, because income for the household is relatively lower, the ability to pay for care is likely reduced, and so the emotional and time burdens for students of these households may be greatest too.

In enrollment by household income, shown in Figure 3.12, there is little effect of household diagnosis in either the advantaged (relatively wealthy, i.e. above the 25 percentile) or disadvantaged (relatively poor, i.e. below the 25 percentile) students. We do see that the decline in enrollment in the first year of diagnosis for severe cancers that we observed in aggregate above is driven by the students from relatively advantaged households, and not by the students from relatively disadvantaged households. In the fourth year of cancer diagnosis, there is a slight increase in enrollment for the advantaged group relative to the comparison group (never treated), though this difference is not statistically different from the disadvantaged group. There is also a statistically significant increase in enrollment relative to pre-treatment periods

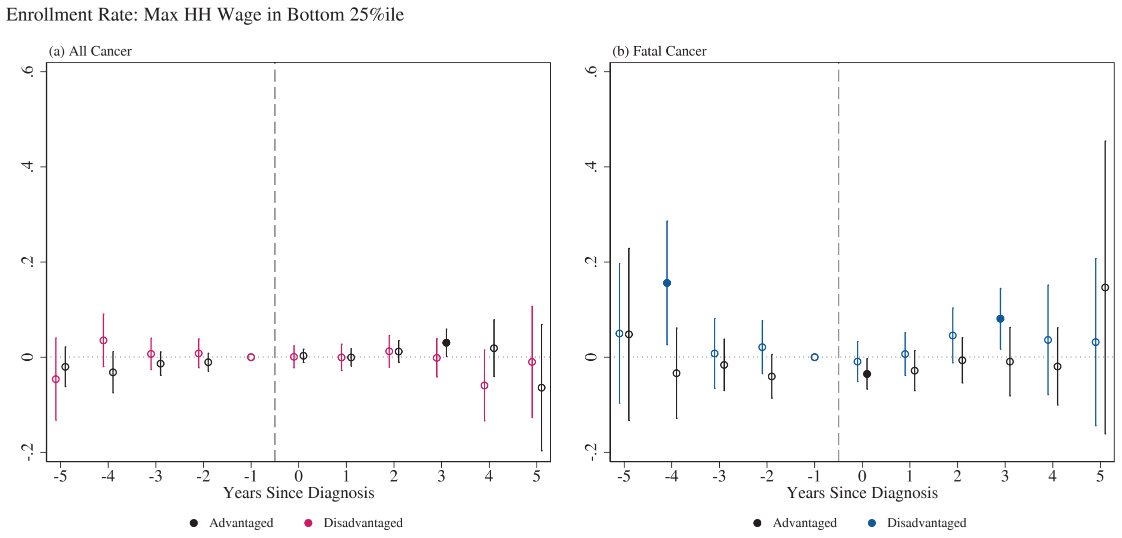

Figure 3.12: Enrollment by household income

Note: This figure presents heterogeneity in the main results by household income. A student from a disadvantaged household in this figure is defined by one who comes from a household where the maximum household wage earner is in the bottom 25% of wages across the full sample of household members of the analytic sample. The results come from the stacked event study difference in differences model described in the empirical methods section, and includes an interaction for the group. Results may be interpreted in relation to the other group.

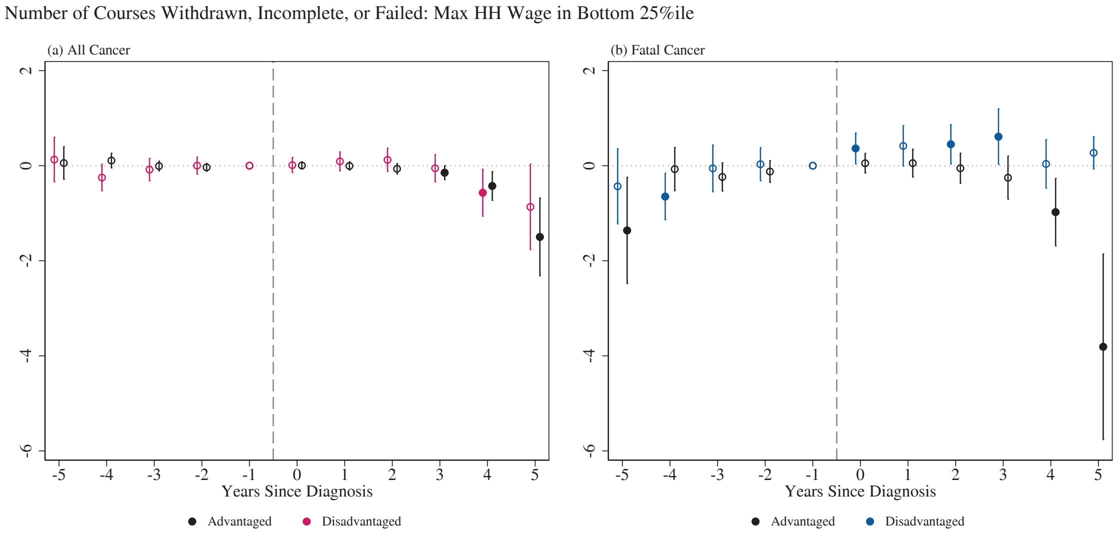

Figure 3.13: Number of courses withdrawn, incomplete, or failed by household income

Note: This figure presents heterogeneity in the main results by household income. A student from a disadvantaged household in this figure is defined by one who comes from a household where the maximum household wage earner is in the bottom 25% of wages across the full sample of household members of the analytic sample. The results come from the stacked event study difference in differences model described in the empirical methods section, and includes an interaction for the group. Results may be interpreted in relation to the other group.

for the disadvantaged group. Moreover, the direction of the effect in the groups is counter-intuitive because it suggests an increase in enrollment after cancer diagnosis. As stated in the aggregate results, one point of caution in this result is that those in the fourth year after diagnosis would be limited to seniors affected in freshman year, super seniors affected in sophomore year, or sixth year seniors affected in their junior year.

Figure 3.13 tells a slightly different story for the group of students from households of severe cancer. In particular, it shows an immediate and sustained increase in the number of courses failed, left incomplete, or withdrawn for students from households below the 25 percentile of income. There is no comparable change in the wealthy households (that is, students from households above the 25 percentile of income). These results suggest that household income is not a moderating factor in *enrollment*, but provides suggestive evidence that in economically disadvantaged households where cancer diagnosis is severe, that students have adverse education *outcomes*. When I consider these results for GPA in the academic year, I find no difference in outcomes, and so this suggests that students are likely leaving courses incomplete or withdrawing.

3.7.2 Distance from home

Second, I consider differences in outcomes for students who live in the 75 percentile or greater from their college (i.e. 100 miles or 160 km). Because these are all college students who attend in-state colleges, non-enrollment may be less likely than for students who attend college out of state. Most students in this sample reside within a few hour's drive of college, and so may still find it easy enough to regularly visit their loved one. For students attending college far from their family, perhaps this is an added emotional toll.

Figures 3.14 and 3.15 plot heterogeneity in the enrollment and the number of failed, incomplete, or withdrawn courses by distance from university. While the results of the former suggest an increase in the enrollment relative to baseline for the disadvantaged group, the results are linear over event time, and suggest that the comparison group in this model is a poor counterfactual for the treated group. A similar result is shown for the group of students with severe cancer diagnoses. Anyway, the results are not statistically different from the advantaged group, which suggests that there is no heterogeneity in the main result by distance from home. In Figure 3.15, we do

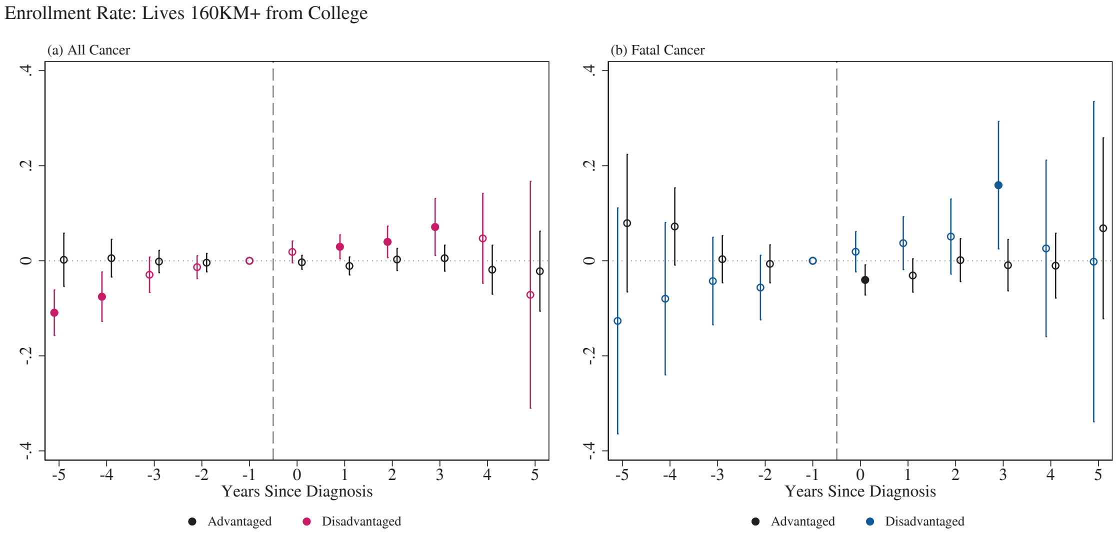

Figure 3.14: Enrollment by distance from home

Note: This figure presents heterogeneity in the main results by distance from home. A student from a disadvantaged household in this figure is defined by one who comes from a household where the zip code is the 75% of distance (160km) from the campus. Distance is calculated using the Haversine formula. The results come from the stacked event study difference in differences model described in the empirical methods section, and includes an interaction for the group. Results may be interpreted in relation to the other group.

not observe any differences in the number of courses that are failed, incomplete, or withdrawn in either subgroup relative to each other or to the counterfactual group.

3.7.3 Student loan balance

Third, I consider differences in students who have high levels of student loans. The intuition of this distinction is similar to the first, and aims to isolate differences in students who are financial disadvantaged. The difference here is that it groups students who have no student loans because they come from wealthy households with students who have no student loans because they come from impoverished households (and hence receive financial assistance). The source of heterogeneity isolates the middle group of students who in some ways have the fewest resources.

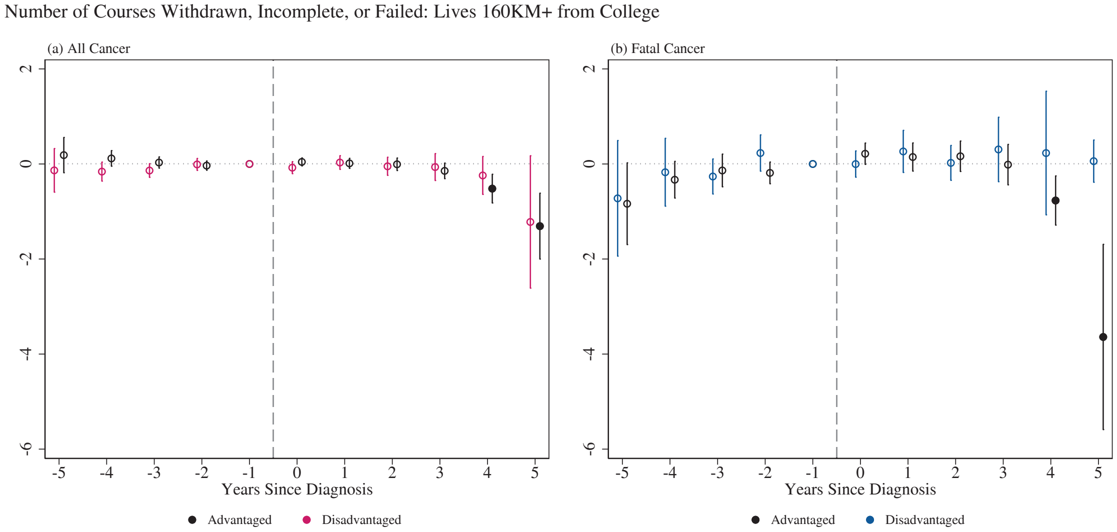

Figure 3.15: Number of courses withdrawn, incomplete, or failed by distance from home

Note: This figure presents heterogeneity in the main results by distance from home. A student from a disadvantaged household in this figure is defined by one who comes from a household where the zip code is the 75% of distance (160km) from the campus. Distance is calculated using the Haversine formula. The results come from the stacked event study difference in differences model described in the empirical methods section, and includes an interaction for the group. Results may be interpreted in relation to the other group.

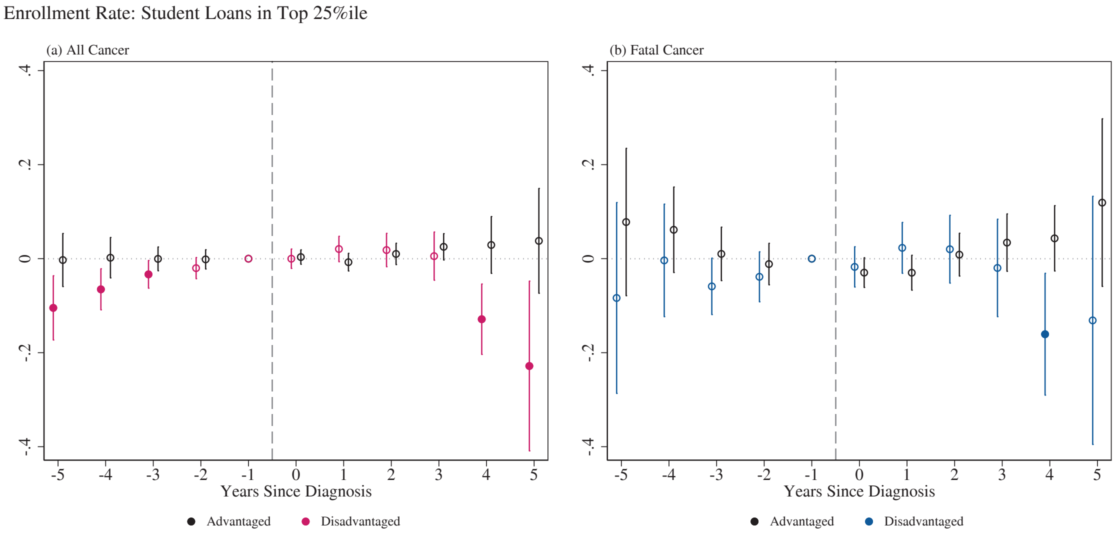

Figure 3.16: Enrollment by baseline student loan balance

Note: This figure presents heterogeneity in the main results by student loan balance. A student from a disadvantaged household in this figure is defined by one who has student loan debt in the top 25% of students. The results come from the stacked event study difference in differences model described in the empirical methods section, and includes an interaction for the group. Results may be interpreted in relation to the other group.

Similar to the results for students living close or far from home, I find no evidence that student loan balance moderates an effect in enrollment in general, or for severe cancer diagnosis. Figure 3.16 shows an increasing pre-diagnosis trend that is interrupted in the fourth year after diagnosis, and then declines sharply. As noted, caution is needed when understanding the sample that this result in the 5th year of cancer diagnosis speaks to. A similar effect is seen in the right panel of Figure 3.16, though not much meaningful inference can be drawn from either. Congruently, I also find no difference in the number of courses withdrawn, left incomplete, or failed for either group of students relative to each other or to the counterfactual group, as shown in Figure 3.17.

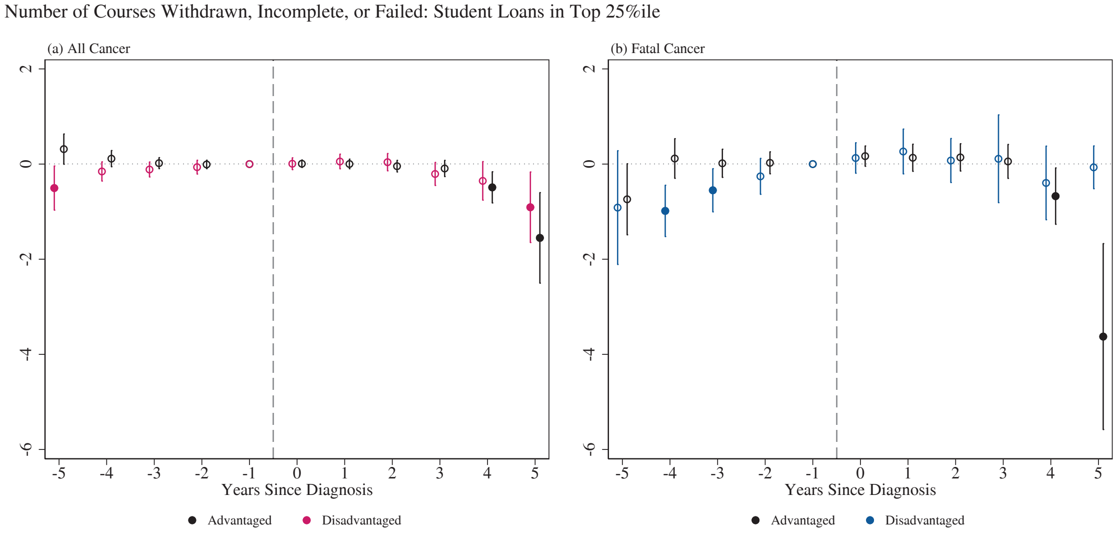

Figure 3.17: Number of courses withdrawn, incomplete, or failed by baseline student loan balance

Note: This figure presents heterogeneity in the main results by student loan balance. A student from a disadvantaged household in this figure is defined by one who has student loan debt in the top 25% of students. The results come from the stacked event study difference in differences model described in the empirical methods section, and includes an interaction for the group. Results may be interpreted in relation to the other group.

In sum, it appears that student loan debt does not moderate the effect of household cancer diagnosis. This is also true for the other outcomes that I analyzed.

3.7.4 Household labor supply

Finally, the fourth source of heterogeneity that I employ is in the household type of the student. In particular, I group students who come from households where more than one household member works and students where there is not more than one working household member. While this cut is perhaps most applicable to Chapter 4, its congruence to the Added Worker Effect could have implications to education outcomes, too.

I find no difference in enrollment when analyzing heterogeneity in the result by household labor supply, shown in Figure 3.18. However, similar to the earlier result in household income, when we compare the outcomes for students who are affected by severe illness and come from a household where fewer than two people are in the labor force, we see a prolonged increase in the number of courses failed, left incomplete, or withdrawn, relative to the counterfactual, but not statistically distinguishable from the [relatively] advantaged group, shown in Figure 3.19.

Like the earlier result, this speaks to the idea that household income may not have much impact on enrollment, but that there is some reason to believe that in cases of severe cancers, it can have an impact on academic performance. To echo the comment above, the lack of difference in GPA for the academic year suggests that students are either increasing the number of courses left incomplete or withdrawing from more courses than their peers.

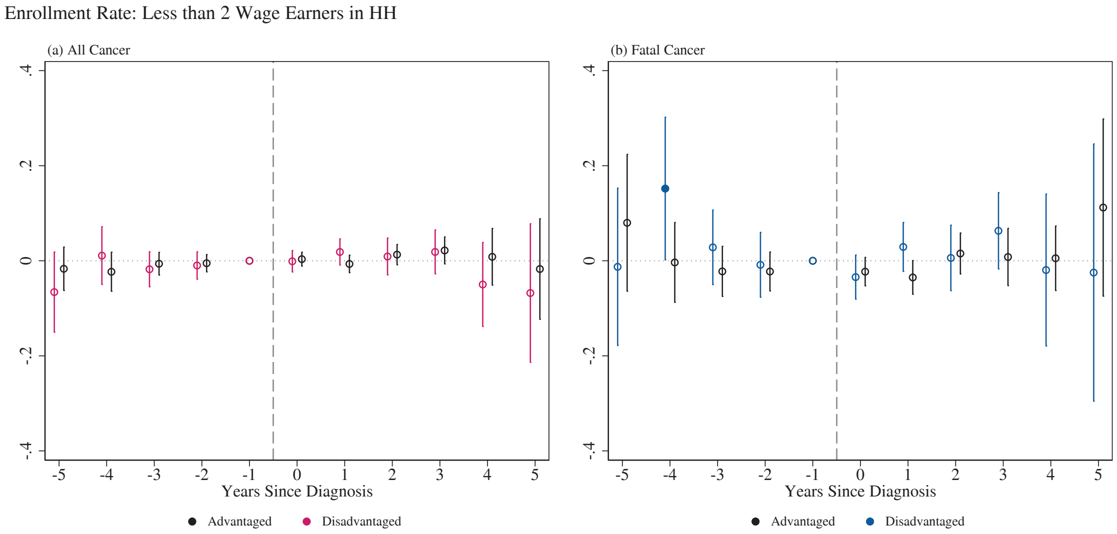

Figure 3.18: Enrollment by household labor supply

Note: This figure presents heterogeneity in the main results by household labor supply. A student from a disadvantaged household in this figure is defined by one who fewer than two workers in his baseline household. This may include himself as the only worker. The results come from the stacked event study difference in differences model described in the empirical methods section, and includes an interaction for the group. Results may be interpreted in relation to the other group.

3.8 Conclusion

In the classical theory of the Human Capital Model, an individual compares the stream of expected costs and expected benefits to evaluate the decision to invest in additional years of education. When the future stream of benefits to higher education are reduced, say by means of a recession or technological revolution, students may reduce their likelihood of investing in higher education. In this chapter, I investigate the extent to which an increase in the cost of education through chronic household illness, in this case, cancer affects investment in higher education. I first motivated this situation by exploring the extent to which cancer besets the household's finances. The literature on financial toxicity has shown that cancer increases adverse financial

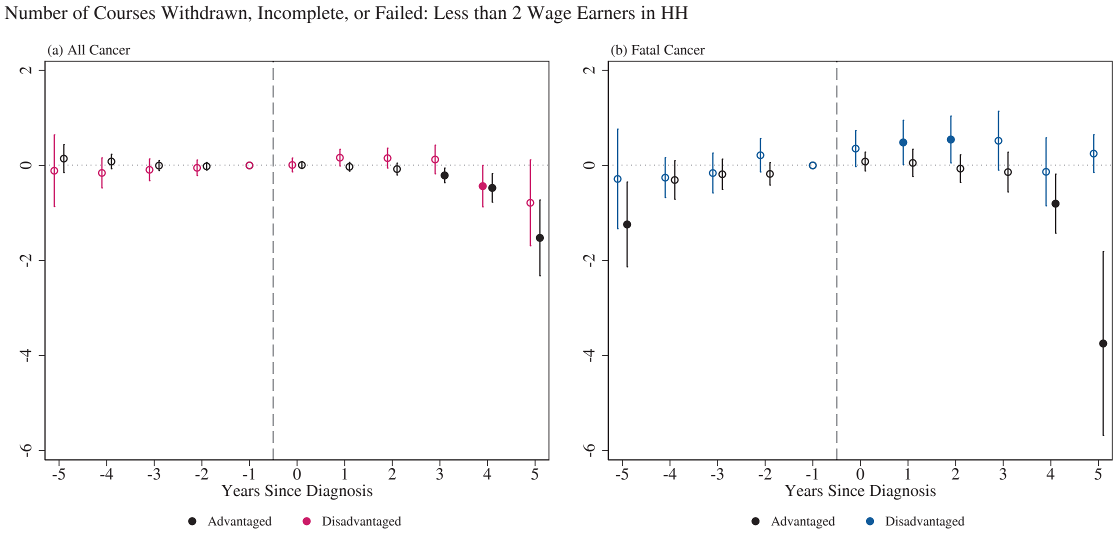

Figure 3.19: Number of courses withdrawn, incomplete, or failed by household labor supply

Note: This figure presents heterogeneity in the main results by household labor supply. A student from a disadvantaged household in this figure is defined by one who fewer than two workers in his baseline household. This may include himself as the only worker. The results come from the stacked event study difference in differences model described in the empirical methods section, and includes an interaction for the group. Results may be interpreted in relation to the other group.

events, increases rates of bankruptcy, and quickly depletes the savings of cancer patients (Shankaran et al., 2022; Gupta et al., 2018; Ramsey et al., 2013; Gilligan et al., 2018). When I match the credit panel data used in this analysis with the Ohio cancer registry, an intriguing picture of the household's finances arises, suggesting a negative effect of cancer on a household's ability to manage its finances. The results of these visualizations suggest that households are immediately beset by increased collections debt and rates of delinquency and charge offs, and by a slowing of growth in credit score. For households where cancer is most severe, the effects are event more pronounced. In addition to the features that households with cancer exhibit on average, households with the most severe cases of cancer show a notable increase in household debt (primarily from housing debt, perhaps explained by Gupta et al. (2018)) and reductions in credit scores. Empirically, it is quite convincing that households are experiencing financial distress.

The secondary question might then be raised – to what extent do parent's finances affect a student's ability to persist through higher education? Statistics from Sallie Mae, which is one of the nation's leading student loan providers, suggests that in the academic year 2022, over 55% of the cost of attending public four-year college was financed by parents and relatives. Additionally, the majority of this assistance was paid for through parental savings, rather than through financing. Only a minority share of the costs of attending public college is actually financed by the student. This suggests that enrollment, which is contingent on an ability to pay tuition, is integrally connected to a household's finances.

To assess this relationship directly, I compile a unique dataset that is comprised of the cancer registry for the state of Ohio, matched to student academic records from the Ohio Department of Higher Education and to credit records from Experian Credit Bureau. I explore the hypothesis that investment in education will be reduced by household cancer, both through decreased ability for the household to offer financial support and through increased non-monetary costs for the student (e.g. emotional costs).

I use a stacked event study difference in differences model that compares students affected by household cancer to students never affected, holding fixed their academic cohort and college campus. While I do find some small exceptions to the general finding, in particular slight evidence of increased rates of summer enrollment in the first year of diagnosis and increased rates of enrollment in pass/fail courses, which may signal a strategy to reduce non-monetary burden by spreading out course difficulty over a longer time horizon, the results generally suggest that students affected by household cancer invest in higher education at consistent rates with students who are not affected, and that academic performance between the two groups is equivalent. I find no difference in GPA across the academic year, no evidence that students are failing, withdrawing, or leaving courses incomplete in aggregate. The exception to this finding is perhaps for students from households experiencing the most severe forms of cancer. For these students, I find a statistically significant 3% decrease in enrollment in the first year of diagnosis, and a slightly smaller, non-statistically significant, reduction in the second year. When I explore heterogeneity in the result by groups that are particularly disadvantaged, I find that students from severely-affected households that are also financial disadvantaged do have an increase in the number of courses that are failed, left incomplete, or withdrawn. This suggests that students in the most severe situations may be in the most need of academic assistance.

Additionally, I run a cross-sectional logistic regression and find that students who are from households of severe cancer are about 80% as likely to graduate from college after four years as students who are not affected. This suggests that perhaps the cumulative effect of household cancer diagnosis is important to understand for doctors, college leaders, guidance counselors, and policymakers.

These findings are perhaps surprising. If the cost of college is largely supported by the household, and the household is severely beset by financial distress, are there explanations that may rationalize decisions of students to remain enrolled in similar rates to their peers? A few caveats may be important.

First, because the data in this essay are limited to students who attend university in Ohio and are in-state students by definition, perhaps the emotional toll is less severe than if we were to observe students enrolled at other colleges across the country. While the situation of cancer diagnosis is always tough, it is reasonable to think that the effect is perhaps mitigated in some sense by living within relatively close proximity to home. While the study period is pre-covid, this time period already saw high rates of technology incorporated into the classroom, and so the time costs of being enrolled may have also been reduced relative to prior years.

Second, this analysis only considers students enrolled in four year colleges. The advantage of this is because there is often a great deal of heterogeneity in the motives of students enrolled in two year colleges or community colleges. Some students are enrolled in two year colleges as a way to build experience and confidence at a post-secondary institution before enrolling in a four year college. Other students are enrolled in two year colleges for technical education to increase human capital for trade-related jobs. While the outcomes of these students are perhaps equally interesting, their assessment deserves an analysis of its own.

Third, students in this analysis have already made a decision to enroll in college. While the Human Capital Model would suggest that students continuously update their decisions, perhaps students act irrationally in this regard. An analysis of students whose household members are affected by cancer diagnosis in high school – that is before making the decision to enroll in college – might sort out the behavioral component of this question. Financing decisions and arrangements have likely already been made, and so perhaps the financial strain on students in terms of tuition assistance is not as great as expected. One potential piece of evidence for this was shown in Figure 3.2, where we saw an increase in housing debt post-diagnosis. As noted, Gupta et al. (2018) shows that individuals who have the ability to finance the treatment of their disease through home equity have superior treatment outcomes. Hence, it is possible that households are not necessarily re-directing savings from tuition assistance to finance medical treatment, but are instead expanding their use of credit through alternative means. And moreover, the question about whether tuition burden is a major contributor to academic dropout is actually not as clear as the HCM suggests it might be. Indeed, as I noted, while Stinebrickner and Stinebrickner (2008) find that credit constraints do increase dropout, the authors note that credit constraints were not the reason for dropout for 70% of students who dropped out of Berea College, where tuition is largely free. The authors note that differences in empirical approaches to the relationship have left unclear and sometimes conflicting answer to the question of how credit constraints impact higher education attainment.

Fourth, the vast majority of cancer diagnoses in this analysis have high five year survival rates, which perhaps indicate that the financial and emotional burdens may be integrated into a student's life without noticeable effect on his academic investment and performance. Where an effect is observed in this analysis, it is unfortunately in the most severe situations, lending some credence to this supposition.

Finally, it is well-documented that students enrolled in four year universities tend to over-estimate their future earnings (Smith and Powell, 1990; Jerrim, 2015; Rouse, 2004; Betts, 1996), perhaps rationalizing the continued enrollment even with added tuition expenses. If, then, (behaviorally) an enrollment decision is seen, more or less, as sticky, then readjustment of other time allocations may need to be made. In Chapter 4, I examine this possibility by exploring labor market decisions.

Because an assessment of the relationship between household cancer and college enrollment may be important both for policy-making and academic accommodations, surprising results such as these may require additional robustness to confirm a null effect. Future work on this should incorporate an alternative estimation strategy, and consider a broader assessment of changes in the time allocation of students.

Appendix E: Explanation of the modeling strategy, Chapter 3

As noted, the "stacked diff in diff" style model has been used previously in the literature. Popularly, Cengiz et al. (2019) used this to study the effect of changes to minimum wage requirements on employment by comparing employment changes in minimum wage within small wage bins. Here, I do something similar, by comparing treated students to students from their same college cohort at the same college. Each stack includes students who began college in the same year at the same college, and either includes individuals who will never be treated or students who were treated n years after beginning college.

Figure E.1 visualizes the empirical strategy for student who began college in academic year 2016. Each panel shows the trends for treated and comparison groups across college tenure, but below each x-axis is also listed the event time within each stack. In panel (a), the comparison is made between students who were treated in year 1 and students who were never treated. Though we do not observe pre-treatment for this cohort, we can see visually how the trends compare for students who began in the same year and were never treated, versus students treated in freshman year. In panel (b), we have a new comparison. This time, I compare students who were treated in year 2 with the same group of non-treated students shown in panel (a). In panel (b), we can see one year of pre-diagnosis data (event time *t-1*), and we can see how the trend compares before and after the event. In both panels (a) and (b), we can use this information to create within-stack event study visualizations for a treated and comparison groups by defining the event at period 1 in panel (a) and period 2 in panel (b). While panel (a) and (b) are done in aggregate, agnostic of the college, in this analysis, I add an additional dimension by considering also the college where student began. In panels (c) and (d), I have selected two different large colleges: Ohio State University Main Campus, and University of Akron Main Campus. In both panels (b) and (c), I have selected students who are treated in year 2 and began in 2016. However, we can see that enrollment behavior, even for the untreated students, is dramatically different between Freshman and Senior years. As a result, the added dimension of considering the campus where the student began college allows us to compare individuals in a more meaningful way. In ways, this strategy has aspects of a coarse and exact matching strategy, which future work may implement for result robustness. Because each and every cohort has relative event time within its stack,

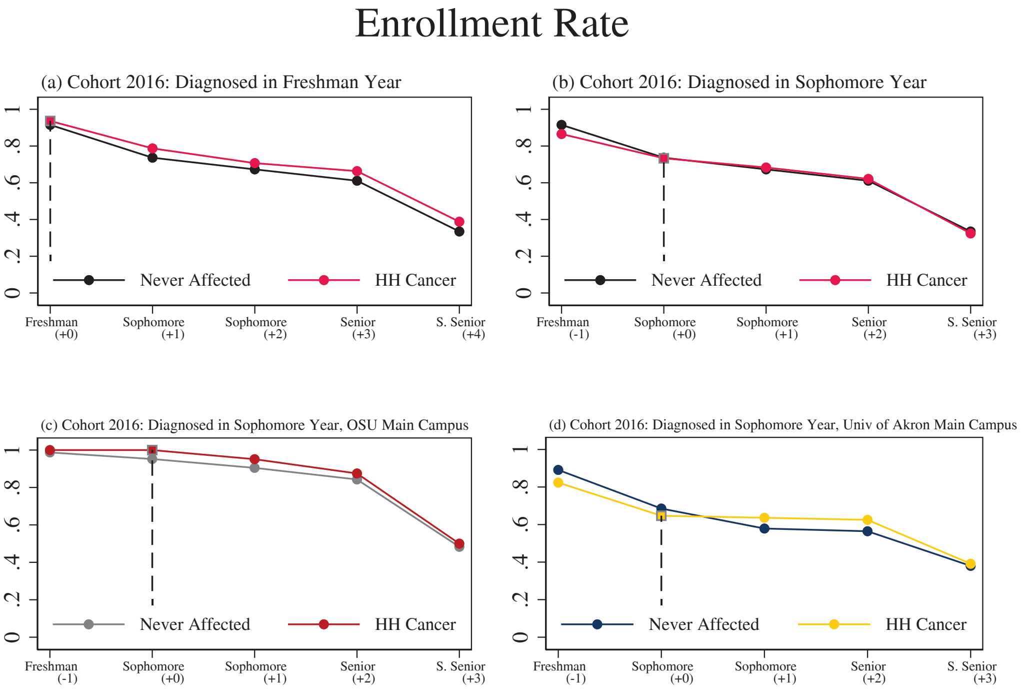

Figure E.1: Illustrative example of stacked approach

Note: Panels (a) and (b) are agnostic to the college campus that the student attended, and depicts only how two separate cohorts are created for students who began college in AY2016. Panels (c) and (d) add an additional dimension to the analysis, by further limiting the cohorts to students who began college at a particular college. In this case, we can see that the enrollment rates are quite different between Ohio State University, Main Campus and the University of Akron, Main Campus. Below the x-axis are the relative event times within each stack.

the datasets can be appended together into one large dataset for the analysis, which allows us, amongst other things, to assess the extent to which parallel trends may or may not exist.

Appendix F: Corresponding tables to HEI event study models, main results, Chapter 3

	Enrolled	Year GPA	Num. Courses	Num. Pass/Fail	Num. Fail, Incomplete, Withdraw	Per Easy Courses	Enrolled in Summer Term
-5	-0.0285	-0.0400	-0.207	-0.0360	0.0826	-0.000115	-0.00569
	(-1.20)	(-1.19)	(-1.01)	(-0.48)	(0.51)	(-0.01)	(-0.21)
-4	-0.0156	-0.0324	-0.225	-0.0445	0.0283	-0.00155	-0.0104
	(-0.80)	(-1.25)	(-1.67)	(-1.02)	(0.44)	(-0.16)	(-0.63)
-3	-0.00876	-0.00652	-0.0525	-0.0382	-0.0228	-0.00371	-0.00302
	(-0.77)	(-0.35)	(-0.49)	(-1.19)	(-0.48)	(-0.60)	(-0.23)
-2	-0.00618	-0.00875	0.00472	0.0115	-0.0259	-0.00335	0.0105
	(-0.70)	(-0.64)	(0.06)	(0.42)	(-0.66)	(-0.79)	(0.95)
0	0.00231	-0.00308	0.0338	-0.000473	0.00646	-0.00317	0.00427
	(0.33)	(-0.27)	(0.50)	(-0.02)	(0.18)	(-0.89)	(0.49)
1	-0.000669	-0.0144	0.0302	0.0167	0.0172	-0.00752	0.0221^{*}
	(-0.08)	(-1.04)	(0.33)	(0.52)	(0.39)	(-1.70)	(2.10)
2	0.0119	0.00817	0.257^{*}	0.0614	-0.0202	-0.00202	0.0125
	(1.15)	(0.43)	(2.52)	(1.84)	(-0.36)	(-0.37)	(0.88)
3	0.0208	0.0615^{*}	-0.0939	0.124^{**}	-0.124	0.00656	0.0385
	(1.64)	(2.23)	(-0.59)	(2.61)	(-1.76)	(0.72)	(1.90)
4	-0.00441	0.146^{**}	-1.046^{*}	-0.00664	-0.457^{***}	0.00102	-0.0974^{**}
	(-0.17)	(3.07)	(-2.57)	(-0.10)	(-3.36)	(0.05)	(-2.63)
5	-0.0349	0.0897	-2.448	-0.162	-1.290^{***}	-0.0384	0.113
	(-0.69)	(0.35)	(-1.94)	(-0.93)	(-4.14)	(-0.78)	(0.88)
Obs	4114873	3221043	3256908	3256908	3256908	3256908	3256908
Students	182926	179530	181621	181621	181621	181621	181621

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table F.1: Tabular results of aggregate HEI models

	Enrolled	Year GPA	Num. Courses	Num. Pass/Fail	Num. Fail, Incomplete, Withdraw	Per Easy Courses	Enrolled in Summer Term
-5	0.0325	0.138	-0.277	0.00574	-0.803^{*}	0.0163	-0.00604
	(0.49)	(1.23)	(-0.58)	(0.03)	(-2.43)	(0.71)	(-0.08)
-4	0.0421	0.0472	0.127	0.106	-0.291	-0.0103	0.0137
	(1.04)	(0.73)	(0.44)	(1.14)	(-1.83)	(-0.59)	(0.49)
-3	-0.00855	0.0426	-0.0380	0.0465	-0.176	0.00369	-0.0139
	(-0.37)	(0.88)	(-0.15)	(0.89)	(-1.35)	(0.29)	(-0.51)
-2	-0.0188	-0.0129	-0.138	0.0783	-0.0703	-0.00255	0.0306
	(-1.01)	(-0.37)	(-0.87)	(1.58)	(-0.72)	(-0.34)	(1.36)
0	-0.0264^{*}	-0.00879	-0.154	0.00148	0.153	-0.00734	-0.0100
	(-2.01)	(-0.31)	(-1.08)	(0.03)	(1.69)	(-1.15)	(-0.51)
1	-0.0161	-0.00461	-0.0463	-0.0738	0.175	-0.0143	-0.0332
	(-1.02)	(-0.14)	(-0.28)	(-1.49)	(1.41)	(-1.64)	(-1.43)
2	0.0119	0.0134	0.176	-0.0551	0.120	-0.0125	-0.0269
	(0.60)	(0.32)	(0.92)	(-0.95)	(0.97)	(-1.03)	(-1.04)
3	0.0246	0.0242	-0.0556	0.0961	0.0676	0.00937	0.000291
	(0.91)	(0.44)	(-0.17)	(1.02)	(0.38)	(0.55)	(0.01)
4	0.000394	0.198^{*}	-0.896	0.0553	-0.609^{*}	0.00540	-0.0643
	(0.01)	(2.02)	(-1.51)	(0.34)	(-2.44)	(0.14)	(-0.99)
5	0.0486	0.723^{*}	0.110	-0.401^{***}	-2.403^{*}	-0.0929	0.0720
	(0.55)	(2.34)	(0.03)	(-4.63)	(-2.07)	(-1.69)	(0.41)
Obs	4091089	3203068	3238749	3238749	3238749	3238749	3238749
Students	177161	174627	176669	176669	176669	176669	176669

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table F.2: Tabular results of severe cancer HEI models

Appendix G: Corresponding tables to labor supply event study models, main results, Chapter 4