Cancer Diagnoses and Household Debt Overhang*

Arpit Gupta
Columbia Business School

Edward R. Morrison
Columbia Law School

Catherine R. Fedorenko and Scott Ramsey
Fred Hutchinson Cancer Research Center

March 23, 2015

Preliminary. Do not cite.

Abstract

This paper explores the role of capital structure in determining how households respond to unanticipated shocks. We draw on a novel dataset linking individual cancer records to high-quality administrative data on personal mortgages, bankruptcies, foreclosures, and credit reports. We find that cancer diagnoses induce a substantial increase in various measures of financial stress regardless of whether the patient carries health insurance. Foreclosure rates, for example, increase 65 percent during the five years following diagnosis. The effect, however, is concentrated among patients with low levels of housing equity. Highly leveraged households default, undergo foreclosure, or file for bankruptcy; less-levered households cope with health shocks by drawing on home equity and other sources of liquidity. These results point to the critical role of capital structure in determining how households respond to severe shocks. They also suggest that household leverage may merit as much policy attention as health insurance.

*We are grateful to workshop participants at the University of Amsterdam, the University of Chicago, Columbia University, and the Fred Hutchinson Cancer Research Center for helpful comments. We also thank Equifax, BlackBox Logic, Zillow, the SNR Denton Fund, and the Fred Hutchinson Cancer Research Center for research support. Albert Chang provided excellent research assistance.

1 Introduction

This paper explores the role of capital structure in determining how households respond to unanticipated shocks. The relationship between shocks, capital structure, and economic outcomes has received sustained examination in the corporate finance literature (see Rajan & Zingales (1995), Fazzari, Hubbard, & Petersen (1988), and Ivashina & Scharfstein (2009)). The relationship is not as well understood in the context of households. To be sure, household financial fragility is a commonly studied phenomenon. Long-term income shocks can lead to reductions in consumption (Cochrane 1991). Short-term expense shocks can induce households to draw on savings or sell assets, work longer hours, turn to friends or family for assistance, or access credit markets (Lusardi, Schneider & Tufano 2011; Shankaran, Jolly, Blough & Ramsey 2012). But the literature has not explored the role of leverage in determining household financial fragility. Like corporations, a household's response to shocks almost certainly depends on its capital structure, including its ratio of debt to assets (leverage) and the maturity structure of its debt contracts (short versus long term).

It seems clear that capital structure *must* must matter to households with incomplete insurance against shocks. For them, capital markets provide an important source of liquidity and consumption-smoothing. As obvious as that proposition seems, we know relatively little about the extent to which access to capital markets matters for the typical household, or how household capital structure affects their responses to shocks. For example, does access to capital markets matter less for households that carry private or public insurance against commonly occurring shocks, such as health problems and auto accidents?

This question is highly relevant to public policy. A large and growing literature on "household financial fragility" has prompted a number of policy proposals, including subsidies to the formation of emergency savings accounts (Lusardi, Schneider, & Tufano 2011), improving household financial literacy (Lusardi & Mitchell 2013), and expanding insurance coverage for important sources of shocks, such as medical care (Mazumder & Miller 2014). All of these proposals assume (implicitly) that households do not "leverage up" in response to the reforms. Emergency savings accounts, for example, are useful buffers against shocks only if households have not incurred substantial debt. With high leverage, the household may have effectively (or explicitly) pledged the accounts to creditors.[1] Alternatively, if households already carry debts at high interest rates (for instance on credit cards), a forced savings plan earning a lower interest rate may be welfare-reducing. The implication is that understanding how households manage the credit instruments available to them is essential to understanding how households respond to shocks and to evaluating public policy programs aimed at relieving financial distress.

We study the implications of leverage for household fragility using data on cancer diagnoses. Health problems are thought to be an important category of financial shocks because they affect a large fraction of the population, are often unexpected, and can generate substantial, uninsured out-of-pocket costs. Many bankruptcies, for example, are thought to be caused by medical expenditures (the phenomenon is sometimes called "medical bankruptcy"). Cancer diagnoses are a particularly important health problem because the resulting uninsured, out-of-pocket costs can be significant even for individuals with public or private insurance. Among Medicare beneficiaries, for example, these costs average \$4,727 annually (Davidoff et al. 2012). Among non-elderly cancer patients, Didem et al. (2011) find that 13% of individuals incurred out of pocket costs exceeding 20% of annual income.

We are not the first to study the effects of health shocks on financial outcomes. Our paper departs from prior work by using high-quality, individual-level data that allows us to trace out the effects of severe shocks on a wide range of financial measures. We begin with a comprehensive database of all cancer diagnoses in western Washington State Using and then link link this individual-level database to property transaction records (from DataQuick), mortgage payment histories (from BlackBox), and consumer credit reports (from Equifax). These linkages allow us to study how credit decisions and defaults evolve before and after a cancer diagnosis. Our observation window includes the five years before and after diagnosis. Importantly, the cancer data include information about patients' insurance status and the property transaction records allow us to calculate household leverage. We can therefore test whether a household's response to health shocks varies with its capital structure.

In the context of a standard event-study framework, which censors patient data upon death, we find that cancer diagnoses generate a long-term increase in foreclosure probabilities.

[1] Indirect evidence of a "leveraging up" phenomenon has been observed in related contexts. Matsa et al. (2012) find that credit costs decline as unemployment insurance benefits increase.

Across all cancers, the three-year (cumulative) probability of foreclosure increases from 0.69% to about 0.93% (a 35% increase). The five-year probability of foreclosure increases from 1% to 1.65% (a 65% increase). The foreclosure probabilities are highest for the most advanced cancers ("distant" and "unstaged"), which increase by .0.89% in magnitude, an 89% increase relative to the foreclosure rate during the five years prior to diagnosis.

These effects are driven almost entirely by households with high pre-diagnosis leverage, as measured by home mortgage loan-to-value ratios. Patients with relatively low loan-to-value ratios (measured at origination) do not experience an increase in foreclosure rates (indeed, in some specifications, the rate declines post-diagnosis). The opposite is true for those with loan-to-value ratios exceeding 100%. This suggests that, although cancer is a large financial shock, a household's ability to cope with the resulting financial stress (as measured by foreclosure) depends on its access to home equity. Importantly, these results persist when we focus on patients we believe to be well-insured (through either private or public insurance programs), and they continue for at least the five years subsequent to diagnosis.

Consistent with this interpretation, we find that low-leverage patients substantially increase their leverage following a medical shock (by taking on a second mortgage or refinancing an existing one). High-leverage patients are substantially less likely to take on new credit following a shock.

Our empirical analysis contributes to the literature on household financial fragility by highlighting the importance of personal leverage as an important driver of household default decisions. Several related papers examine the financial impact of idiosyncratic health shocks. Hubbard et al. (1995) was an early attempt to understand the effect of health shocks on financial outcomes, particularly among the elderly. French and Jones (2004) estimate that 0.1% of households experience a health shock that costs over $125,000 in present value. Our results also echo findings in the household finance literature. We find that a combination of negative shocks and high leverage best explain default patterns, similar to the "double-trigger" theory of mortgage default (see Bhutta, Dokko, & Shan (2010)). We also highlight the trade-off between risk management and financing current investments in durable goods, such as housing and autos, as analyzed by Rampini & Vishwanathan (2014). That trade-off persists even when households carry health insurance.

This paper also builds on our prior work. Ramsey et al. (2013) find that cancer patients are at higher risk of bankruptcy than those without a cancer diagnosis. Morrison et al. (2013) investigated the causal relationship of car accidents on bankruptcy filings. The latter paper found little evidence that car accidents elevate bankruptcy filings, perhaps because car accidents typically represent smaller shocks than the cancer diagnoses investigated in this paper.

2 Background and Data

2.1 Background

Cancer represents one of the most common and costly health shocks. Roughly 40% of Americans can expect to face a cancer diagnosis over their lifetimes, and 20% of Americans will die due to cancer-related complications (American Cancer Society 2013). Cancer diagnosis rates are projected to increase both internationally and domestically over time due to medical progress in other fields, leaving individuals more susceptible to cancer risk. The cost of treating cancer has also been rising over time even faster then overall healthcare inflation, which in turn has been growing faster than economy-wide prices (Mariotto et al. 2011; Trogdon et al. 2012).

Cancer severity is often measured using "stages." A cancer is "localized" if malignant cells are limited to the organ of origin (e.g., liver). "Regional" and "distant" cancers describe tumors that have extended beyond the organ of origin. A cancer is regional if the primary tumor has grown into other organs of the body; it is distant if the primary tumor has produced new tumors that have begun to grow at new locations in the body. Because of this subtlety, it is well known that the coding of these diagnoses is inconsistent (SEER Training Module 2014); the two categories may describe comparably severe cancers. "Unstaged" cancers are those that were not given a formal staging by the investigating physicians. This often occurs when the cancer has spread so extensively through the patient's body that formal staging is not an informative exercise.

Cancer diagnoses generate direct and indirect costs. Direct cancer costs relate to the cost of treatment and typically represent substantial expenses relative to household income. Cancer treatments typically involve some combination of drugs, surgery, radiation, and hormonal therapy. Formal health insurance should cover many of these treatments, but individuals are also exposed to out-of-pocket costs such as co-pays and deductibles. Prior to 2006, for example, older patients (over 65) often had limited insurance coverage of cancer drugs unless they purchased supplemental Medicare plans (in 2006, this situation changed with the enactment of Medicare Part D). Indirect costs include the time required to undergo screening and therapy, transportation to hospitals and clinics, and child or nursing care. Evidence suggests that 6.5% of cancer expenses among the non-elderly ($1.3 billion) are paid out-of-pocket (Howard, Molinari, & Thorpe 2004). Over 40% of cancer patients stop working after initial treatment (De Boer et al. 2009).

Costs are substantial even among individuals with public or private insurance. Among Medicare beneficiaries, for example, out-of-potcket costs average $4,727 annually (Davidoff et al. 2012). Among non-elderly cancer patients, Didem et al. (2011) found that 13% of individuals incurred out-of-pocket costs exceeding 20% of annual income. The percentage is much higher among individuals with public insurance (24% of income) and those with health insurance not provided by their employer (43%).[2]

2.2 Data Construction

We link cancer diagnosis data to bankruptcy filings, property records, mortgage payment data, and credit reports. Our cancer data are provided by the Cancer Surveillance System of Western Washington, which collects information about all cancer diagnoses in 11 counties in western Washington state. These data are a subset of the National Cancer Institute's Surveillance Epidemiology and End Results (SEER) program. Our data include about 270,000 diagnoses occurring during calendar years 1996 through 2009. About 110,000 of these diagnoses involved patients between ages 24 and 64.

[2]The substantial nature of indirect costs with respect to cancer also suggests that our work may have some applicability to countries with more universal health coverage, to the extent that formal insurance mechanisms are insufficient to fully prevent financial distress resulting from cancer diagnosis.

The cancer data were linked to a dataset on federal bankruptcy records by the Fred Hutchinson Cancer Research Center via a probabilistic algorithm based on the patient's name, sex, address, and last four Social Security Number digits (see Ramsey et al. 2013). The bankruptcy records include any individual bankruptcy filing under chapters 7, 11, or 13 of the Bankruptcy Code.

We further link the cancer data to property records maintained by DataQuick to create a "Property Database." The DataQuick records are transaction-based and provide information about every sale, mortgage, foreclosure, or other transaction affecting a property address during calendar years 2000 through 2011. We link these property records to the cancer data based on the patient's property address. This Property Database can be used to study the relationship between cancer diagnoses and foreclosure starts.

We link the Property Database to mortgage payment data and credit reports for patients with privately securitized mortgages. BlackBox LLC provided the mortgage payment data, which includes information about the balance, LTV, borrower FICO, and other characteristics of the mortgage at origination as well as the borrower's post-origination payment history. These data cover the period January 2000 through July 2014, and are restricted to the universe of private-label securitized loans. Equifax provided credit reports, which include monthly information about the borrower's credit score, utilization of revolving lines of credit (mainly credit cards), total debt burden, and other characteristics. These data cover the period from June 2005 through July 2014.[3] We linked the Property Database to the BlackBox and Equinox records using mortgage origination date, origination balance, zip code fields, and other mortgage fields (mortgage type and purpose) that are common to all datasets.

After linking these databases (SEER cancer registry, bankruptcy filings, DataQuick property records, and the BlackBox and Equinox databases), we subset on individuals between ages 21 and 80 at the time of diagnosis. Younger patients are unlikely to file for bankruptcy; older patients have extremely high mortality rates subsequent to diagnosis. Additionally,

[3]Equifax performed the linkage between its records and the BlackBox data. Because this linkage was imperfect, we retained a linkage only if Equifax reported a "high merge confidence" (based on a proprietary algorithm) or if the BlackBox and Equifax records listed the same property zip code (suggesting a common residence between the subject of the credit report and the holder of the mortgage. Additional information about the BlackBox and Equinox databases, and the merge algorithm, can be found in Mayer et al. (2014) and Piskorski, Seru, & Witkin (2014).

we exclude cancer diagnoses that involving benign and in situ stage cancer diagnoses (early stage cancers that have not spread to surrounding tissue) as well as diagnoses discovered only upon death or autopsy. The former cancers represent trivial health shocks; the latter confound death and diagnosis, making it impossible to infer the impact of diagnosis on financial stability. Finally, a number of patients have multiple cancer diagnoses. If the diagnoses were "synchronous"—occurring within a three month period—we treat them as a single event and assign a diagnosis date equal to the first-diagnosed cancer. Synchronous cancers are frequently manifestations of one underlying cancer.[4] If a patient suffered multiple, non-synchronous cancers (diagnoses occurring over a period longer than three months), we included in our analysis any cancer diagnosis that was not followed by another diagnosis during the subsequent three years.

Appendix A provides a more complete description of the data and information about the merge algorithms.

2.3 Summary Statistics

Table 1 presents summary statistics for the cancer patients in our study. The mean age is 58, with a wide standard deviation: ages 32 through 80 are within two standard deviations of the mean. About two thirds of the patients are married, roughly half are male, and over a third had health insurance through Medicare or Medicaid. Although Table 1 indicates that only 14.7 percent of individuals carried private insurance, health insurance information is missing for nearly half of the sample. Most of the individuals with missing information likely had some form of health insurance. Those age 65 and older are covered by Medicare. Among those aged 18 to 64, prior studies indicate that between 8 and 14 percent had no health insurance coverage (Ferguson & Gardner 2008).

Table I also presents information about the "occupation" of individuals in our sample. This information is included in the SEER database and derived from a hospital intake form that asks patients to describe their occupation, not whether they are currently employed in

[4] We assign these cancers the highest stage among the multiple stages present (localized, regional, or distant). We also assign the site of the cancer to the "Other" category if the sites of the synchronous cancers differ.

that occupation. We interpret this information as a proxy for the patient's human capital investment. Using an algorithm supplied by Washington State, we categorized patient responses into broad categories: Professional, Clerical, Laborer, Other, Unemployed, and Missing. The Unemployed category includes individuals who indicated that they were unemployed at the time they completed the intake form.[5],

Table II shows the annual number of cancer diagnoses by stage at diagnosis. As described above, cancer diagnoses can be staged, from least to most severe, as localized, regional, and distant. We include unstaged cancers in the "distant" category because these cancers tend to have a very high mortality rate. As Table 1 shows, nearly half of diagnoses are localized; regional and distant cancers account for most of the remaining diagnoses.

3 Empirical Strategy

We estimate a standard event-study difference-in-difference (DD) regression:

$$O_{it} = \alpha + \sum_{k=-s}^{s-1} \mu_k \cdot 1[(t - T_i) = k] + X_{it} + \theta_t + \varepsilon_{it} \tag{1}$$

Here, O_{it} is an outcome measure. In most specifications it will be a binary equal to one if patient i exhibits a measure of distress (e.g., foreclosure) during calendar year t. θ_t is a matrix of calendar year fixed effects.[6] The matrix X_{it} includes a variety of controls, which vary with the database used for the analysis. In all regressions, we include patient age, marital status, gender, race, occupation, health insurance status, indicators for whether the patient suffered synchronous cancers or had a previous cancer diagnosis, and county fixed effects. In analysis using the BlackBox or Equifax data, the controls include time from origination, static information taken at time of origination (balance, CLTV, details about the purpose and type of mortgage), and dynamic information updated monthly (such as credit score, estimated income, and interest rate).

[5] We classify individuals as "unemployed" if they fail to indicate an occupation, but do indicate marital status. We assume that, if an individual fails to answer both the occupation and marital status questions, he or she is refusing to complete the form. If the individual indicates marital status, but leaves occupation blank, we think it reasonable to assume that the individual is leaving it blank because he or she is unemployed.

[6] We do not include individual fixed effects because our dependent variable is binary and we are typically studying the first occurrence of an event (such as foreclosure or bankruptcy). In this setting, with non-repeating events, fixed effect analysis is not feasible (Andreβ et al. 2013).

The coefficients of interest are μ_k, which measure the change in the outcome variable during the s calendar years prior to and following the diagnosis in year T_i, where s is typically 5. Years $[-s, -1]$ reflect the s pre-treatment years, while the interval $[0, s-1]$ is the post-treatment window. These coefficients are measured relative to the (omitted) year prior to the diagnosis. Standard errors are clustered by patient.

If outcome O_{it} occurs during year t, data for that patient is censored in all subsequent years. This censoring renders our framework similar to a discrete time hazard model. Additionally, if patient i dies during year t, data for that patient is also censored in all subsequent years. Finally, the model is only estimated during years for which we are confident that the patient lived in the property in question as determined by sale transactions data.

4 Results

We begin by documenting the average effect of cancer diagnoses on household financial outcomes, including foreclosure, bankruptcy, and missed payments. The overall patterns, however, conceal important heterogeneity with respect to household leverage. Households that have untapped liquidity through home equity or credit cards are better able to withstand cancer diagnoses.

4.1 Average Effects

Table III estimates our event-study model using foreclosure as the outcome measure. The model is estimated using the Property Database. The coefficients of interest, μ_k, are reported for the five years before and after diagnosis, with the year before as the excluded category. Coefficients for the controls are suppressed to simplify presentation. At the bottom of the table, we report the cumulative estimated effect for the three years after diagnosis ("Treatment 3 Years") and the five years after diagnosis ("Treatment 5 Years"). Again, these estimates are measured relative to the year prior to diagnosis. Additionally, the bottom of the table reports the average foreclosure rate during the year prior to diagnosis ("Ref. Foreclosure Prob. 1 Year") and the cumulative probability during the five years prior to diagnosis ("Ref. Foreclosure Prob. 5 Years").

In the full Property Database sample, including all cancers, Columns (1) and (2) show a substantial, sustained increase in the probability of foreclosure during the five years following diagnosis. The effect is larger after including controls such as age, gender, and marital status. During the three years post-diagnosis, the filing rate increases .0024 percentage points, a 35 percent increase relative to the foreclosure rate during the three years prior to diagnosis. Widening our window to look at the five years post diagnosis, we find that the foreclosure rate increases .0065 points, a 65 percent increase relative to the rate during the five years preceding diagnosis (.01).

We observe large effects across all cancer stages, though timing varies, as Columns (3) through (5) show. Among distant and unstaged cancers, we observe an increase in foreclosure rates beginning in the second post-diagnosis year. Among less severe cancers (localized and regional), significant effects appear in the third year following diagnosis. Unsurprisingly, the cumulative, five-year effect is largest among the most severe cancers: Foreclosure rates increase 0.0089 percentage points, a nearly 90 percent increase relative to the filing rate during the five years prior to diagnosis. Localized and regional cancers exhibit large but smaller effects (.0062 and .0048, representing 65 and 40 percent increases).

These estimates are based on the full sample, regardless of insurance status. Table IV subsets on patients with verified insurance coverage, public or private. In Column 2, we observe a 51 percent increase in foreclosures during the five years following diagnosis (.0051 increase on a base of .01). Estimated effects for localized cancers are comparable to those for distant/unstaged cancers. Insurance status may attenuate the impact of a cancer diagnosis (a 51 percent increase among insured patients versus a 65 percent increase overall), but the diagnosis remains destabilizing for a substantial fraction of households.

Table V examines the impact of cancer on bankruptcy filings, another extreme financial outcome. Here we look at the full sample, regardless of insurance coverage, but results are comparable when we subset on patients with verified insurance coverage. Estimated effects of the diagnosis are substantially lower for bankruptcy than for foreclosure. In Column 2, for example, we observe a 20 percent increase in the bankruptcy filing rate during the five years post-diagnosis (.004 increase on a base of .02). When we split the sample by cancer staging, estimates are significant and substantial only for localized cancers.

We look at less extreme financial outcomes in Table VI, which estimates the fraction of first mortgage payments made during the calendar years before and after diagnosis. The fraction ranges from 0/12 to 12/12. We subset on the sample that links to the BlackBox sample, which is much smaller than the sample used in the preceding tables. As a result, we expect estimates that are much less precise. Table VI shows an immediate decline in payments. During the year prior to a cancer diagnosis, patients made roughly 11 of the 12 scheduled payments. During the year following the diagnosis, the fraction falls by .026. During the five years following diagnosis, the cumulative effect is a .14 decline, which is roughly equivalent to two missed payments.

We view the foregoing estimates as evidence that cancer diagnoses induce a substantial rise in extreme financial outcomes, especially foreclosure, but less extreme responses are difficult to measure using our data.

4.2 Financial Fragility and Household Leverage

The analysis thus far conceals important heterogeneity across patients. Cancer diagnoses are destabilizing primarily for households with high levels of pre-diagnosis leverage.

Table VII reexamines the effect of cancer diagnoses on foreclosure, but subsets on patients for whom we can verify the origination date and balance of a mortgage in the Property database.[7] Although the sample here is smaller than in Table III, the estimated effects are comparable. In Column 1, for example, the foreclosure rate increases 0.0063 percentage points during the five years following diagnosis, a 42 percent increase. Column 2 subsets further on patients for whom the Property database allows us to compute the patient's cumulative loan to value ratio (CLTV) at mortgage origination. CLTV is equal to total mortgage debt, including both first and second mortgages, divided by the purchase price of the home. We can verify CLTV for a relatively small number of patients (about 10,000) and, in this subsample, we actually observe a decline in foreclosures following diagnosis, as Column 2 shows. This anomaly is explained in Columns 3 and 4, which split the sample by CLTV. We use CLTV=100 as a natural break-point because it is a proxy for the existence of

[7] We cannot observe the origination date and balance of a mortgage originated prior to around 2000. Our data track transactions after that date.

home equity that can be used to fund consumption. Cancer is destabilizing only for patients who have no home equity (CLTV $\geq$ 100) at mortgage origination. Among these patients we observe a very large increase—.05 percentage points—in the foreclosure probability during the five years following diagnosis, a 90 percent increase relative to the baseline (.055). The foreclosure rate declines among patients with home equity at origination (CLTV$<$100).

Table VIII repeats the exercise, but subsets on patients with verified health insurance. The effects here are comparable. Columns 3 and 4, for example, show the same contrast between patients with high and low CLTV: Highly leveraged patients are less able to cope with health shocks, even though they carry health insurance. In unreported regressions, predicting the probability of bankruptcy, we find similar results: Cancer diagnoses cause substantial and significant increases in bankruptcy filing rates only among patients with CLTV above 100.

Together, these results highlight the importance of home equity as a source of insurance. We observe this pattern even if we we look at outcomes less severe than foreclosure or bankruptcy. Recall that, in Table VI, we found relatively small effects of cancer on mortgage payments. When we split the sample by current CLTV ("CCLTV") instead of CLTV at origination, we find substantial, significant effects among patients with high CCLTV (we use 80 as the break-point here due to the limited sample size and because it is commonly thought that homeowners have limited ability to exploit home equity if their CCLTV exceeds 80). We find that patients with a CCLTV above 80 miss about six payments during the five years after diagnosis.

4.3 Access to New Credit

These estimates suggest that home equity—and access to liquidity generally—is an important channel through which patients cope with the financial stress of health shocks. This is true regardless of whether the patient carries health insurance. We can study this channel more directly by looking at patients' use of credit following cancer diagnosis. Table IX predicts the annual probability that a patient refinances a first mortgage or takes on a second mortgage. Although we see a decline in credit use by the average patient during the years following a diagnosis (Column 2), the decline is driven entirely by patients with high levels of leverage (Column 4). We observe comparable patterns when we subset on patients with health insurance, as Table X shows. Among patients with relatively low CLTV, we see a rise in credit usage following diagnosis. The effect among these patients may not be attributable to the diagnosis, however, given pre-existing trend upwards in credit usage.

5 Conclusion

Our results point to the central importance of credit markets as a buffer against health events and other adverse financial shocks. Even households with health insurance face sizeable out-of-pocket costs after a cancer diagnosis. These costs are destabilizing when a household has taken on high pre-diagnosis leverage. The household is effectively priced out of the credit market.

Our research is, however, subject to several caveats. First, we document the patterns of financial distress surrounding severe medical events, but do not make claims about the strategic nature of those defaults. Nor do we make any normative claims about the desirability of foreclosure among affected households. Bankruptcy, default, and foreclosure are commonly viewed as manifestations of severe financial distress, with adverse consequences for debtors and creditors alike. An alternative view might see these outcomes as manifestations of strategical calculations by households. Because a cancer diagnosis reduces a patient's life expectancy, for example, a rational household might strategically default on long-term debts such as mortgages. Under this interpretation, our results on leverage form an analogue to the "double trigger" theory of household default: Default may be the result of both (i) an adverse shock (cancer diagnosis) that reduces ability to pay and (ii) an adverse financial position (negative equity) that limits the household's desire to repay.

Our analysis also leaves unexplored the question of how and why households acquire the capital structures they have. To draw the parallel with corporate finance: We know much more about the overall determinants of corporate leverage decisions than household leverage.

Finally, we look exclusively at the effects of cancer diagnoses on financial management (defaults, foreclosures, bankruptcies). We are unable to test whether cancer diagnoses affect consumption choices.

Our results present a sharp contrast with much of the prevailing literature on household financial fragility and health insurance because we find a limited role of formal insurance in preventing financial default. Highly levered individuals face a higher probability of financial default even in the presence of medical insurance. While medical insurance is clearly an important buffer for households facing severe medical shocks, our results show that household financial fragility depends on much more than the existence of such insurance. Many individuals with insurance file for bankruptcy or experience foreclosure (particularly if they are heavily levered); many individuals without insurance never file for bankruptcy or foreclosure (particularly if they have equity). Household capital structure is, at the very least, an additional, important, and underemphasized driver of default decisions among medically distressed households.

A potential implication of our work is that public policy should focus on household asset-building, both by limiting leverage or by raising savings. Unlike efforts to increase medical coverage, efforts to build household assets have the advantage that accumulated savings may be used to deal with any sort of shock, not just medical ones. Laws that limit household leverage, such as restrictions on recourse mortgages, may also help households preserve assets that can fund out-of-pocket costs.

References

American Cancer Society. 2013. "Lifetime Risk of Developing or Dying From Cancer." available at www.cancer.org.

Andreβ Hans-Jürgen, Katrin Golsch, and Alexander W. Schmidt. 2013. *Applied Panel Data Analysis for Economic and Social Surveys*, Berlin: Springer.

Bhutta, Neil, Jane Dokko, and Hui Shan. 2010. "The depth of negative equity and mortgage default decisions." Finance and Economics Discussion Series 2010-35. Washington, DC: Board of Governors of the Federal Reserve System.

Cochrane, John H. 1995. "Time-Consistent Health Insurance." *Journal of Political Economy* 103(3): 445–73.

Cochrane, John H. 1991. "A Simple Test of Consumption Insurance." *Journal of Political Economy* 99(5): 957–76.

Davidoff, A. J., M. Erten, T. Shaffer, J.S. Shoemaker, I.H. Zuckerman, N. Pandya, M.-H. Tai, X. Ke, and B. Stuart. 2013. "Out-of-pocket health care expenditure burden for Medicare beneficiaries with cancer." *Cancer* 119: 1257–65.

De Boer A.G., T. Taskila, A. Ojajarvi, F.J. van Dijk, and J.H. Verbeek. 2009. "Cancer survivors and unemployment: a meta-analysis and meta-regression." *JAMA* 301(7):753–62.

De Nardi, Mariacristina, Eric French, and John Bailey Jones. 2010. "Why Do the Elderly Save? The Role of Medical Expenses," *Journal of Political Economy* 118(1): 39–75.

Fazzari, Steven M., R. Glenn Hubbard, and Bruce C. Petersen. 1988. "Financing Constraints and Corporate Investment." *Brookings Papers on Economic Activity, Economic Studies Program* 19(1): 141–206.

Ferguson, Deron and Erica Gardner. 2008. "Estimating Health Insurance Coverage Using Hospital Discharge Data and Other Sources." Health Care Research Brief No. 48. Olympia, WA: Washington State Office of Financial Management.

French, Eric and John Bailey Jones. 2004. "On the Distribution and Dynamics of Health Care Costs," *Journal of Applied Econometrics* 19(6): 705–21.

Gross, Tal and Matthew J. Notowidigdo. 2009. "Health Insurance and the Consumer Bankruptcy Decision: Evidence from Expansions of Medicaid." *Journal of Public Economics* 95: 767-778.

Himmelstein, David U., Deborah Thorne, Elizabeth Warren, and Steffie Woolhandler. 2010. "Medical Bankruptcy in the United States, 2007: Results of a National Study." *American Journal of Medicine* 122: 741.

Himmelstein, David U., Elizabeth Warren, Deborah Thorne and Steffie Woolhandler. 2005. "Market Watch: Illness and Injury as Contributors to Bankruptcy,.' *Health Affairs* 63 (February).

Hollingworth, William, Annemarie Relyea-Chow, Bryan A. Comstock, Karen A. Overstreet and Jeffrey G. Jarvik. 2007. "The Risk of Bankruptcy Before and After Brain or Spinal Cord Injury." *Medical Care* 45: 702-11.

Howard, D.H., N.A. Molinari, K.E. Thorpe. 2004. "National estimates of medical costs incurred by nonelderly cancer patients." *Cancer* 100(5): 883–91.

Hubbard, R. Glenn, Jonathan Skinner, and Stephen P. Zeldes. 1995. "Precautionary Saving and Social Insurance." *The Journal of Political Economy* 103(2): 360–99.

Ivashina, Victoria and David Scharfstein. 2010. "Bank Lending During the Financial Crisis of 2008." *Journal of Financial Economics* 97(3): 319–38.

Lusardi, Annamaria and Olivia S. Mitchell. 2013. "The Economic Importance of Financial Literacy: Theory and Evidence," NBER Working Paper No. 18952.

Lusardi, Annamaria, Daniel Schneider, and Peter Tufano. 2011. "Financially Fragile Households: Evidence and Implications." *Brookings Papers on Economic Activity* Spring 2011: 83-134.

Mariotto, Angela B., K. Robin Yabro, Yongwu Shao, Eric J. Feuer, and Martin L. Brown. 2011. "Projections of the Cost of Cancer Care in the United States: 2010-2020." *Journal of the National Cancer Institute* 103(2): 1-12.

Matsa, David A, Brian Melzer, and Joanne W. Hsu. 2012. "Unemployment Insurance and Consumer Credit," working paper.

Mayer, Christopher, Edward Morrison, Tomasz Piskorski, and Arpit Gupta. 2014. "Mortgage modification and strategic behavior: Evidence from a legal settlement with Countrywide." *American Economic Review* 104(9): 2830–57.

Mazumder, Bhash and Sarah Miller. 2014. "The Effects of the Massachusetts Health Care Reform on Financial Distress," Federal Reserve Bank of Chicago Working Paper No. 2014-01.

Morrison, Edward R., Arpit Gupta, Lenora Olsen, Larry Cook, and Heather Keenan. 2013. "Health and Financial Fragility: Evidence From Car Crashes and Consumer Bankruptcy." University of Chicago Coase-Sandor Institute for Law & Economics Research Paper NO. 6565.

Ozkan, Serdar. 2010. "Income Differences and Health Care Expenditures over the Life Cycle." University of Pennsylvania Technical Report.

Piskorski, Tomasz, Amit Seru, and James Witkin. forthcoming. "Asset quality misrepresentation by financial intermediaries: Evidence from RMBS market." *Journal of Finance.*

Rampini, Adriano and S. Viswanathan. 2014. "Household Risk Management." Working Paper (Duke University).

Ramsey, Scott, David Blough, Anne Kirchho, Karma Kreizenbeck, Catherine Fedorenko, Kyle Snell, Polly Newcomb, William Hollingworth and Karen Overstreet. 2013. "Washington State Cancer Patients Found To Be At Greater Risk For Bankruptcy Than People Without A Cancer Diagnosis." *Health Affairs* 31(6).

Rajan, Raghuram G. and Luigi Zingales.1995. "What Do We Know about Capital Structure? Some Evidence from International Data." *Journal of Finance* 50(5):1421–60.

Shankaran, Veena, Sanjay Jolly, David Blough, and Scott D. Ramsey. 2012. "Risk Factors for Financial Hardship in Patients Receiving Adjuvant Chemotherapy for Colon Cancer: A Population-Based Exploratory Analysis." *Journal of Clinical Oncology*, 30(14): 1608-1614.

Trogdon, J.G., F.K. Tangka, D.U. Ekwueme, G.P. Guy, Jr., I. Nwaise, and D. Orenstein. 2012. "State-Level Projections of Cancer-Related Medical Care Costs: 2010 to 2020." *American Journal of Managed Care* 18(9): 525-32.

TABLE I Summary Statistics

	Mean	SD	N
Age	58	12.8	63,893
Married	0.650	0.48	63,893
Marriage Missing	0.096	0.29	63,893
Male	0.497	0.50	63,893
Non-White	0.142	0.35	63,893
Synchronous Cancer	0.018	0.13	63,893
Professional	0.209	0.41	63,893
Clerical	0.184	0.39	63,893
Laborer	0.234	0.42	63,893
Other	0.056	0.23	63,893
Unemployed	0.065	0.25	63,893
Missing	0.252	0.43	63,893
Self-Pay	0.003	0.051	63,893
Private Insured	0.147	0.35	63,893
Medicare	0.340	0.47	63,893
Medicaid	0.011	0.10	63,893
Other	0.008	0.089	63,893
Missing	0.492	0.50	63,893
Previous Cancer	0.035	0.18	63,893

TABLE II Staging Frequency by Year

	Localized	Regional	Distant	Unstaged	Total
1996	1460	600	634	208	2902
1997	1644	660	702	222	3228
1998	1719	666	743	213	3341
1999	1870	757	791	197	3615
2000	2013	832	793	151	3789
2001	2171	991	953	123	4238
2002	2348	1098	1055	87	4588
2003	2464	1137	1086	112	4799
2004	2599	1208	1100	87	4994
2005	2640	1169	1222	113	5144
2006	2784	1135	1209	126	5254
2007	2989	1355	1299	138	5781
2008	3116	1386	1270	92	5864
2009	3269	1394	1336	264	6263
Total	33086	14388	14193	2133	63800
Observations	63800				

t statistics in parentheses

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

TABLE III Panel Regression, OLS—Foreclosure, Censor: 5 Years

	All - No Controls	Controls	Localized	Regional	Distant/Unstaged
Year 5 Before Diagnosis	-0.00033	-0.00040	-0.00021	-0.00094	-0.00029
	(-1.15)	(-1.24)	(-0.48)	(-1.32)	(-0.43)
Year 4 Before Diagnosis	0.000054	0.000034	0.00023	-0.00013	-0.00018
	(0.18)	(0.11)	(0.53)	(-0.18)	(-0.30)
Year 3 Before Diagnosis	-0.0000097	0.000048	0.00017	-0.00030	0.00012
	(-0.03)	(0.16)	(0.44)	(-0.47)	(0.19)
Year 2 Before Diagnosis	-0.00013	0.000016	0.00013	0.00034	-0.00050
	(-0.46)	(0.06)	(0.35)	(0.52)	(-0.93)
Year 1 After Diagnosis	-0.00035	-0.00018	-0.00030	0.000071	-0.00020
	(-1.32)	(-0.69)	(-0.87)	(0.12)	(-0.35)
Year 2 After Diagnosis	0.00040	0.00082**	0.00042	0.00023	0.0021**
	(1.34)	(2.68)	(1.11)	(0.37)	(2.95)
Year 3 After Diagnosis	0.0012**	0.0018**	0.0017**	0.0019*	0.0021**
	(3.35)	(5.14)	(3.59)	(2.44)	(2.79)
Year 4 After Diagnosis	0.00086*	0.0017**	0.0018**	0.00029	0.0029**
	(2.37)	(4.61)	(3.60)	(0.45)	(3.30)
Year 5 After Diagnosis	0.0015**	0.0024**	0.0026**	0.0023*	0.0020*
	(3.49)	(5.55)	(4.52)	(2.51)	(2.24)
Controls	No	Yes	Yes	Yes	Yes
Treatment 3 Years	0.0012	0.0024	0.0018	0.0022	0.0040
S.E.	0.00071	0.00072	0.00093	0.0016	0.0016
Treatment 5 Years	0.0035	0.0065	0.0062	0.0048	0.0089
S.E.	0.0012	0.0012	0.0016	0.0025	0.0026
Ref. Foreclosure Prob. 1 Year	0.0023	0.0023	0.0020	0.0027	0.0026
Ref. Foreclosure Prob. 5 Years	0.010	0.010	0.0095	0.012	0.010
N	466441	466441	242862	105040	117906

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

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

TABLE IV Panel Regression, OLS—Foreclosure, Censor: 5 Years, Insured

	All - No Controls	Controls	Localized	Regional	Distant/Unstaged
Year 5 Before Diagnosis	0.00016	0.000081	0.00078	-0.0014	0.000044
	(0.39)	(0.17)	(1.20)	(-1.37)	(0.05)
Year 4 Before Diagnosis	0.00027	0.00023	0.00031	0.00044	-0.0000087
	(0.68)	(0.54)	(0.56)	(0.38)	(-0.01)
Year 3 Before Diagnosis	0.00036	0.00045	0.00061	-0.00017	0.00066
	(0.91)	(1.10)	(1.14)	(-0.18)	(0.86)
Year 2 Before Diagnosis	-0.00019	0.000045	0.00032	0.00014	-0.00051
	(-0.52)	(0.12)	(0.65)	(0.14)	(-0.79)
Year 1 After Diagnosis	-0.00016	0.00011	0.000060	0.00015	0.000062
	(-0.44)	(0.29)	(0.13)	(0.15)	(0.09)
Year 2 After Diagnosis	0.00064	0.0012**	0.00095	-0.00021	0.0024*
	(1.43)	(2.61)	(1.66)	(-0.23)	(2.52)
Year 3 After Diagnosis	0.00019	0.0013**	0.0018**	0.0018	0.00030
	(0.39)	(2.70)	(2.60)	(1.39)	(0.37)
Year 4 After Diagnosis	-0.00027	0.00089	0.00092	-0.000064	0.0015
	(-0.57)	(1.81)	(1.49)	(-0.06)	(1.42)
Year 5 After Diagnosis	0.00044	0.0016**	0.0019*	0.0012	0.0015
	(0.77)	(2.69)	(2.30)	(0.89)	(1.28)
Controls	No	Yes	Yes	Yes	Yes
Treatment 3 Years	0.00066	0.0026	0.0028	0.0017	0.0028
S.E.	0.00098	0.00100	0.0013	0.0024	0.0019
Treatment 5 Years	0.00084	0.0051	0.0056	0.0029	0.0057
S.E.	0.0016	0.0016	0.0021	0.0038	0.0032
Ref. Foreclosure Prob. 1 Year	0.0021	0.0021	0.0016	0.0030	0.0022
Ref. Foreclosure Prob. 5 Years	0.010	0.010	0.0093	0.013	0.0096
N	215779	215779	104802	46262	64283

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

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

TABLE V Panel Regression, OLS—Bankruptcy, Censor: 5 Years

	All - No Controls	Controls	Localized	Regional	Distant/Unstaged
Year 5 Before Diagnosis	0.00072	-0.00029	-0.00046	-0.00043	0.00025
	(1.61)	(-0.59)	(-0.70)	(-0.39)	(0.26)
Year 4 Before Diagnosis	0.000097	-0.00099*	-0.00087	-0.0019*	-0.00033
	(0.23)	(-2.20)	(-1.44)	(-1.98)	(-0.35)
Year 3 Before Diagnosis	0.00017	-0.00042	-0.0012*	0.00038	0.00058
	(0.41)	(-1.00)	(-2.24)	(0.39)	(0.67)
Year 2 Before Diagnosis	-0.00020	-0.00042	-0.00020	-0.00047	-0.00069
	(-0.51)	(-1.05)	(-0.39)	(-0.53)	(-0.88)
Year 1 After Diagnosis	0.00066	0.00072	0.00091	0.0012	0.000049
	(1.63)	(1.75)	(1.65)	(1.27)	(0.06)
Year 2 After Diagnosis	0.0010*	0.0012**	0.0012*	0.00094	0.0015
	(2.29)	(2.63)	(2.04)	(0.93)	(1.48)
Year 3 After Diagnosis	0.00054	0.00099*	0.0017**	0.00086	-0.0014
	(1.14)	(1.96)	(2.59)	(0.77)	(-1.31)
Year 4 After Diagnosis	-0.000090	0.00062	0.00096	-0.000098	0.00082
	(-0.18)	(1.16)	(1.44)	(-0.09)	(0.56)
Year 5 After Diagnosis	-0.00046	0.00046	0.00026	0.0013	0.00078
	(-0.88)	(0.82)	(0.38)	(1.01)	(0.46)
Controls	No	Yes	Yes	Yes	Yes
Treatment 3 Years	0.0022	0.0029	0.0038	0.0030	0.00018
S.E.	0.0010	0.0011	0.0014	0.0024	0.0022
Treatment 5 Years	0.0017	0.0040	0.0051	0.0042	0.0018
S.E.	0.0016	0.0018	0.0023	0.0040	0.0039
Ref. Bankruptcy Prob. 1 Year	0.0047	0.0047	0.0042	0.0054	0.0049
Ref. Bankruptcy Prob. 5 Years	0.021	0.021	0.019	0.025	0.023
N	443989	443989	241805	100666	100898

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

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

TABLE VI Panel Regression, OLS—Fraction of Payments, by Loan to Value

	All	Below 80 CCLTV	Above 80 CCLTV
Year 5 Before Diagnosis	-0.030	-0.016	-0.37**
	(-1.35)	(-0.78)	(-4.08)
Year 4 Before Diagnosis	-0.0039	0.0042	-0.062
	(-0.25)	(0.28)	(-0.74)
Year 3 Before Diagnosis	-0.00072	-0.0019	-0.0089
	(-0.07)	(-0.17)	(-0.23)
Year 2 Before Diagnosis	0.011	0.012	-0.015
	(1.75)	(1.93)	(-0.78)
Year 1 After Diagnosis	-0.026**	-0.019*	-0.038*
	(-3.66)	(-2.44)	(-1.98)
Year 2 After Diagnosis	-0.033*	-0.020	-0.095*
	(-2.52)	(-1.61)	(-2.47)
Year 3 After Diagnosis	-0.034	-0.019	-0.16**
	(-1.69)	(-1.02)	(-2.79)
Year 4 After Diagnosis	-0.024	-0.0097	-0.18*
	(-0.86)	(-0.36)	(-2.17)
Year 5 After Diagnosis	-0.019	-0.026	-0.11
	(-0.49)	(-0.68)	(-0.97)
Controls	Yes	Yes	Yes
Linear Combination Treatment	-0.14	-0.094	-0.58
S.E.	0.096	0.093	0.27
Ref. Var Prob. 1 Year	0.94	0.94	0.93
Ref. Var Prob. 5 Years	4.78	4.79	4.71
N	6979	5371	1608

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

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

TABLE VII Panel Regression, OLS—Foreclosure, by Mortgage Statistics: 5 Years

	Has Mortgage	Has CLTV	CLTV < 100	CLTV >= 100
Year 5 Before Diagnosis	-0.0011*	0.0022	-0.00071	0.0057
	(-2.08)	(0.85)	(-0.38)	(1.52)
Year 4 Before Diagnosis	-0.00054	0.0028	-0.0011	0.0068*
	(-1.06)	(1.27)	(-0.69)	(2.10)
Year 3 Before Diagnosis	-0.00055	0.0015	0.00045	0.0015
	(-1.17)	(0.82)	(0.28)	(0.60)
Year 2 Before Diagnosis	-0.000022	0.00088	0.000035	0.0015
	(-0.05)	(0.57)	(0.02)	(0.74)
Year 1 After Diagnosis	-0.00052	-0.0031*	-0.0038**	0.00056
	(-1.19)	(-2.34)	(-2.85)	(0.29)
Year 2 After Diagnosis	0.00069	-0.0032*	-0.0052**	0.0064*
	(1.30)	(-2.22)	(-3.80)	(2.43)
Year 3 After Diagnosis	0.0028**	-0.0015	-0.0046**	0.013**
	(4.04)	(-0.87)	(-2.94)	(3.57)
Year 4 After Diagnosis	0.0016*	-0.0015	-0.0028	0.0089*
	(2.17)	(-0.76)	(-1.44)	(2.13)
Year 5 After Diagnosis	0.0018*	0.00096	-0.0050*	0.022**
	(2.12)	(0.36)	(-2.45)	(2.95)
Controls	Yes	Yes	Yes	Yes
Treatment 3 Years	0.0030	-0.0079	-0.013	0.020
S.E.	0.0013	0.0038	0.0036	0.0059
Treatment 5 Years	0.0063	-0.0085	-0.021	0.050
S.E.	0.0022	0.0067	0.0063	0.012
Ref. Foreclosure Prob. 1 Year	0.0032	0.010	0.0090	0.0068
Ref. Foreclosure Prob. 5 Years	0.015	0.061	0.039	0.055
N	233729	56828	53708	24224

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

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

TABLE VIII Panel Regression, OLS—Foreclosure, by Mortgage Statistics: 5 Years, Insured

	Has Mortgage	Has CLTV	CLTV < 100	CLTV >= 100
Year 5 Before Diagnosis	-0.00023	0.0033	-0.00041	0.013*
	(-0.31)	(0.98)	(-0.16)	(2.47)
Year 4 Before Diagnosis	-0.00011	0.0020	0.00027	0.0062
	(-0.16)	(0.75)	(0.12)	(1.62)
Year 3 Before Diagnosis	0.00047	0.0040	0.0035	0.0046
	(0.74)	(1.77)	(1.49)	(1.53)
Year 2 Before Diagnosis	0.00010	0.00065	0.0016	-0.00032
	(0.18)	(0.36)	(0.76)	(-0.14)
Year 1 After Diagnosis	-0.00012	-0.0025	-0.0027	0.00013
	(-0.21)	(-1.49)	(-1.42)	(0.06)
Year 2 After Diagnosis	0.0015	-0.0012	-0.0044*	0.0076
	(1.82)	(-0.54)	(-2.13)	(1.95)
Year 3 After Diagnosis	0.0025*	0.00074	-0.0054**	0.016*
	(2.17)	(0.22)	(-2.62)	(2.01)
Year 4 After Diagnosis	0.0011	-0.00074	-0.0045	0.0063
	(0.99)	(-0.20)	(-1.70)	(0.85)
Year 5 After Diagnosis	0.0019	0.0036	-0.0053	0.030
	(1.22)	(0.65)	(-1.84)	(1.79)
Controls	Yes	Yes	Yes	Yes
Treatment 3 Years	0.0038	-0.0030	-0.013	0.023
S.E.	0.0018	0.0052	0.0050	0.0093
Treatment 5 Years	0.0068	-0.000051	-0.022	0.060
S.E.	0.0030	0.0098	0.0085	0.021
Ref. Foreclosure Prob. 1 Year	0.0027	0.0084	0.0084	0.0052
Ref. Foreclosure Prob. 5 Years	0.014	0.052	0.040	0.048
N	116066	30327	24418	13513

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

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

TABLE IX Panel Regression, OLS—New Credit, by Mortgage Statistics: 5 Years

	Has Mortgage	Has CLTV	CLTV < 100	CLTV >= 100
Year 5 Before Diagnosis	-0.093**	-0.0045	-0.074**	-0.043**
	(-22.74)	(-0.42)	(-8.19)	(-2.84)
Year 4 Before Diagnosis	-0.074**	0.0021	-0.069**	-0.026*
	(-18.97)	(0.23)	(-8.36)	(-1.96)
Year 3 Before Diagnosis	-0.060**	-0.0063	-0.063**	-0.027*
	(-15.97)	(-0.83)	(-8.25)	(-2.28)
Year 2 Before Diagnosis	-0.030**	0.00073	-0.034**	-0.0096
	(-8.06)	(0.11)	(-4.75)	(-0.84)
Year 1 After Diagnosis	-0.021**	-0.0098	0.0020	-0.037**
	(-5.61)	(-1.55)	(0.30)	(-3.37)
Year 2 After Diagnosis	-0.039**	-0.018**	0.023**	-0.10**
	(-10.27)	(-2.61)	(3.07)	(-8.88)
Year 3 After Diagnosis	-0.026**	-0.012	0.034**	-0.082**
	(-5.59)	(-1.42)	(3.79)	(-5.41)
Year 4 After Diagnosis	-0.017**	-0.021*	0.055**	-0.11**
	(-3.31)	(-2.02)	(5.17)	(-5.96)
Year 5 After Diagnosis	0.00044	-0.0095	0.074**	-0.10**
	(0.07)	(-0.75)	(5.62)	(-4.39)
Controls	Yes	Yes	Yes	Yes
Treatment 3 Years	-0.085	-0.040	0.059	-0.22
S.E.	0.0095	0.017	0.018	0.029
Treatment 5 Years	-0.10	-0.071	0.19	-0.44
S.E.	0.017	0.032	0.033	0.054
Ref. Default Prob. 1 Year	0.32	0.29	0.28	0.40
Ref. Default Prob. 5 Years	1.43	1.66	1.27	2.09
N	229801	56870	52668	24504

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

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

TABLE X Panel Regression, OLS—New Credit, by Mortgage Statistics: 5 Years, Insured

	Has Mortgage	Has CLTV	CLTV < 100	CLTV >= 100
Year 5 Before Diagnosis	-0.062**	0.0073	-0.052**	-0.038
	(-10.61)	(0.52)	(-4.01)	(-1.90)
Year 4 Before Diagnosis	-0.042**	0.021	-0.046**	-0.0012
	(-7.61)	(1.72)	(-3.97)	(-0.07)
Year 3 Before Diagnosis	-0.036**	0.00047	-0.042**	-0.024
	(-7.07)	(0.05)	(-4.07)	(-1.55)
Year 2 Before Diagnosis	-0.017**	0.0081	-0.029**	0.0091
	(-3.45)	(0.94)	(-3.09)	(0.61)
Year 1 After Diagnosis	-0.0091	0.010	0.021*	-0.0045
	(-1.92)	(1.24)	(2.31)	(-0.31)
Year 2 After Diagnosis	-0.027**	-0.012	0.017	-0.067**
	(-5.11)	(-1.23)	(1.60)	(-4.07)
Year 3 After Diagnosis	-0.011	0.016	0.052**	-0.052
	(-1.42)	(0.95)	(3.04)	(-1.66)
Year 4 After Diagnosis	-0.025**	-0.025	0.026	-0.092*
	(-2.64)	(-1.26)	(1.27)	(-2.42)
Year 5 After Diagnosis	0.0056	-0.011	0.054*	-0.083
	(0.46)	(-0.41)	(2.08)	(-1.66)
Controls	Yes	Yes	Yes	Yes
Treatment 3 Years	-0.048	0.014	0.090	-0.12
S.E.	0.013	0.026	0.027	0.045
Treatment 5 Years	-0.067	-0.021	0.17	-0.30
S.E.	0.025	0.049	0.051	0.090
Ref. Default Prob. 1 Year	0.25	0.24	0.21	0.34
Ref. Default Prob. 5 Years	1.40	1.53	1.16	1.99
N	114417	30590	23962	13785

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

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

Appendix A: Data Construction

Data Sources

SEER Data Our data are a subset of the National Cancer Institute's Surveillance Epidemiology and End Results (SEER) program, and comprise the Cancer Surveillance System of Western Washington. The data are intended to be a comprehensive catalog of cancer diagnoses occurring between 1996–2009, totaling over 270,000 cases overall. A unique patient id links records together: patients re-enter the dataset for each separate diagnosis.

The data include a rich set of fields detailing the demographic characteristics of the patient (such as race, age, listed occupation, marital status), the nature of the cancer (its type and staging), as well as select treatment decisions taken by the patient.

Bankruptcy Data Our bankruptcy data comprise all federal bankruptcy records from Western Washington state including chapters 7, 11, and 13. These data are readily accessible through PACER and have been frequently used in prior academic scholarship on bankruptcy.

Deeds Data Our Deeds dataset is provided by DataQuick, a vendor which collects public-use transactions information. The data are organized at a property level and are comprehensive of all mortgage transactions which take place from 2000–2011 (foreclosure transactions typically go back further in time). The data list each mortgage transaction—including sales, transfers, new mortgages (first and second liens), and refinancing—which occur on a given property. We use the timing of the sales information to infer when cancer patients were resident in the property, and follow foreclosures for the duration of the time individuals were resident. We additionally use mortgage information dating to the time of the patient's residence to calculate our key leverage statistics.

BlackBox Data BlackBox LLC is a private vendor which has collected the individual mortgage records related to private label securitized bonds (ie, those not securitized by a government-sponsored entity like Fannie Mac or Freddie Mae). Though private label securitization made up only a fraction of total mortgage origination even at its peak before the crisis; our data contain more than 20 million mortgages in total; which is typically either subprime, Alt-A, or jumbo-prime in credit risk.

The BlackBox data contain static information taken at the time of origination, such as origination balance, credit score (FICO score), interest rate, and contract terms. The data are also updated monthly with dynamic information on fields like interest rates, mortgage payments, and mortgage balances. The mortgage payment field is most critical for our analysis, as it allows us to calculate the precise number of payments the household has made, not just whether or not the household has entered foreclosure.

Equifax Data Equifax is a major credit bureau which maintains detailed dynamic monthly credit information on households concerning their balances on mortgage and other debt, as well as credit scores (Vantage score).

Data Merges

A key innovation our of analysis is the use multiple sources of data on individual behavior to track financial outcomes around cancer diagnosis. This requires us to implement complex merges between many datasets which were not originally intended to be linked. Due to privacy restrictions, we are unable to make these data publicly available. However, the code used for all analysis is available upon request and below we document the document the merge process and linking variables which enable us to construct our dataset.

SEER-Bankruptcy The linkage between the SEER and Bankruptcy datasets was performed by the Fred Hutchinson Cancer Research Center via a probabilistic algorithm based on the patients name, sex, address, and last four Social Security Number digits (Ramsey et al. 2013).

SEER-Deeds Data Three match criteria were used to link SEER and Deeds data based on common text address fields:

- A *tight* match was based on full address, street directional (ie, NW), zip or city, and census tract.
- An *intermediate* match was based on house number, the first three letters of the street name, street end (ie, lane or drive), end number (any number in the last position of the address, such as an apartment number), street directional, zip or city, and census tract.
- A *loose* match was based on house number, the first three letters of the street name, street end, end number, zip or city, and census tract. These are all of the match criteria used in the intermediate match, with the exception of street directional.

The match was conducted by first prioritizing tight matches. Intermediate matches not found using the tight match were added next, and finally any loose matches not found using either of the two other methods were added. The vast majority of matches were achieved using the tight match (63,661 records were matched using the tight match; 7,970 using the intermediate match; and 2,065 using the loose match for a total of 73,696 SEER records which matched into a record in the Deeds data.

Deeds Data-BlackBox Though Deeds and BlackBox data were not designed to be linked, they are both administrative datasets containing reliable information on a variety of mortgage fields. We developed a novel a match method to link the two datasets using a training dataset (for which we knew matches exactly) to develop the algorithm. The merge relies on the following common fields:

1. Exact date matches between origination dates of the mortgage are reported in the two datasets (not used if the origination date was likely imputed; ie the date reflected in BlackBox was the first or end of the month.
2. Zip code matches between the two datasets.

3. Matches based on mortgage purpose (ie, refinancing or purchase).

4. Matches based on mortgage type (ie, adjustable-rate or fixed-rate).

5. Matches based on mortgage origination amount (rounded down to the hundred)

We used a *backward* window of 31 days, in which the mortgage origination date reflected in BlackBox was at most 31 days after the date of the mortgage reflected in Deeds; and a *forward* window of 20 days.

The match algorithm worked by first focusing on 1) zip matches and 2) origination amount matches within the backward window (or the forward window if no matches existed in the backward window).

If only one match was found using those criteria, it was kept. If there were multiple matches, we restricted further by iteratively applying the following the following criteria. We first employed a "tight" match which required that the loan match uniquely on day, or (if there were multiple day matches) uniquely on mortgage purpose or type among those that matched on day.

If this did not uniquely identify a match, we next restricted to "looser" matches where there was 1) only one match uniquely on mortgage type and purpose. If no mortgage matched, we moved on to cases where there was 2) one unique match of either mortgage type or purpose with the other field missing; 3) one unique match on mortgage type, and 4) one unique match on mortgage purpose. The merge algorithm proceeded among all matching cases in the order specified above—if a high quality match was found, the mortgage was kept and the procedure only moved on to the other match cases in the order specified if no match was found.

BlackBox-Equifax BlackBox, a mortgage-level dataset, was linked by Equifax to borrower-level information on a variety of debts, including mortgages. The merge algorithm relied on a proprietary code which we cannot access. The vast majority of accounts in BlackBox were linked to a credit account.

To verify the accuracy of the merge, we imposed a restriction samples which make use of Equifax variables. Specifically, we require that the two entries match either on 1) zip code of the borrower (at least once over the life of the loan); or 2) have a match confidence of at least .85. The zip code restriction compares the zip code of the property as listed in BlackBox matches with the address of the borrower as listed in Equifax. A mismatched zip code is not necessarily indicative of a mismatch in loans—it could also suggest the presence of an investor who does not live in the property in question.

In addition to the zip code measure, Equifax provided a measure of match confidence ranging from 0–0.9. Loans at the top end of the confidence score reflect extremely well matched loans, and we allow for a mismatch in zip code so long as it is accompanied by a match confidence score of at least 0.85. Robustness checking based on other common attributes between the two datasets (such as common measures of default) suggest that the two measures of match accuracy we employ are effective in correctly identifying well-matched loans. For further details of the BlackBox-Equifax merge algorithm; see Piskorski, Seru, & Witkin (2014).

Variable Definitions

Occupation The SEER data provide a numerical occupation coding. Using the occupation coding derived from Washington State government at `https://fortress.wa.gov/doh/occmort/docs/OccupationList.pdf`; we classified the following occupation fields: Professional, Clerical, Laborer, Other Occupation, and Occupation Missing.

We impute "Unemployed" individuals as those who: 1) Are listed as "Occupation Missing," and 2) have a marital status at diagnosis which is not missing or listed as "Unknown." We assume that the occupation non-response of such individuals, since it is paired with a response on the martial status form, is indicative of a genuine non-response for occupation (which would have been recorded by the reporting hospital as an occupation had the individual reported an occupation) and is assumed to come from an unemployed individual.

Mortgage Equity For the Property Database, we measure housing equity by estimating the total mortgage amount (of both first and second liens) at origination and comparing with an estimate of house price.

To estimate the house price, we begin with the purchase price if given. Unfortunately, sometimes we lack information on sale prices (but do have data on mortgages if the mortgage was refinanced). In that case, we impute the house price based on other sales on the same property at a different time (including by other owners), and infer the original house price using a zip-level house price index from Zillow.

For the Credit Report Dataset, we use the exact mortgage balances. We combine data on both first liens (data from which is derived from BlackBox) and second liens (from Equifax). We use an estimate of origination house value derived from the reported origination loan-to-value; and adjust the house price at the time of diagnosis using the Zillow index to compute a current loan to value ratio.

Data Cleaning

From the base SEER data, the following cuts were made:

- Benign cancers were dropped.
- Among cancers reported multiple times within the same day, only one cancer entry was kept.
- Synchronous cancers were identified in which multiple cancers presented within a three month interval. Only the first instance of the synchronous cancer was kept; if the stages of the two cancers differed, the maximum stage was taken. If the sites of the two cancers differed, the cancer was classified as "Other."
- In the case of multiple, non-synchronous cancers; the cancer was included if there was at least three years subsequent to diagnosis in which there were no intervening cancer diagnoses. If there was an intervening cancer; the second cancer would be included (provided that there were no subsequent diagnosis in the three years subsequent to that diagnosis), with a dummy variable indicating the presence of a prior cancer.

- We keep patients aged 21–80 at the time of diagnosis.

To connect the SEER data with the DataQuick Deeds records, the DataQuick data were separated on the basis of sale records. If a cancer diagnosis was associated with a record prior to any recorded sale; it is assumed that a real estate transaction took place prior to when the DataQuick records begin (the year 2000) resulting in the move-in of a resident who was subsequently diagnosed with cancer prior to any other sale.

The data were organized in a panel structure based on diagnosis-calendar year. It is possible for the same patient to have multiple cancers and so be repeated in the data for the years surrounding each diagnosis (again, provided a three year window). The panel includes the five calendar years subsequent to diagnosis (counting the year of diagnosis); and five calendar years prior to diagnosis.

Three forms of censoring were applied to the panel data:

- Censoring based on property information. Calendar years prior to the individual moving into the property as reflected in a sale record were excluded, as are calendar years after the person moving out (again as reflecting in a sale record).
- Censoring based on mortality. Our data record the death date of individuals. We censor all calendar year subsequent to death.
- Censoring due to previous episode of financial distress. Given the property-centric nature of out dataset, we can only follow one foreclosure per patient, and so censor all future observations in the calendar year subsequent to financial distress (it is possible for individuals to file for multiple bankruptcies; but such events are more rare due to the statue of limitations imposed after typical bankruptcy filings. We adopt an identical censoring strategy with respect to bankruptcies.

In addition to the other cuts, the Credit Panel Data made the following additional restrictions:

1. We require that the diagnosis take place subsequent to origination.
2. We require sufficient data from our datasets in order to estimate effects. If observations are missing for the entire year of observation, the year is dropped.
3. If more than two BlackBox entries matched a given borrower in the Property Dataset, we dropped the entries. Two were permitted as these frequently coincided with a refinanced mortgage (in which both original and refinanced mortgage were present in the dataset), or a first and second lien.
4. Among entries with two BlackBox entries, entries were dropped if:
   (a) The two BlackBox entries did not share a common id as reported in Equifax. These entries may reflect mismatched loans, rather than different borrowing by the same consumer.
   (b) If the two BlackBox entries were non-overlapping in date (ie, as frequently happens in the case of refinancing), they were kept. If they were overlapping, the entry with the smaller mortgage amount was dropped (frequently, this was a second lien).