Does Insurance for Treatment Crowd Out Prevention? Evidence from Diabetics' Insulin Usage*

Daniel Kaliski†

December 12, 2018

Abstract

I provide new evidence that health insurance can discourage investment in health. I find that, in the United States before 2006, 13-30% of female diabetics who used insulin to manage their condition stopped using insulin once they turned 65 and became eligible for health insurance via Medicare. I reconcile these results with those from other studies by developing a model of the trade-off between prevention and treatment. The model explains the large effect sizes in this paper via two mechanisms. First, individuals substitute prevention efforts away from periods when the price of treatment is low and toward periods when the price of treatment is high. Second, this effect is stronger for preventive measures that have larger effects on health. The model also shows that the long-term crowding out of prevention is at least as large as the shift in the timing of prevention estimated in this paper. The introduction of more generous subsidies for insulin under Medicare Part D in 2006 eliminated this effect, saving up to $487 million per annum in forgone health care costs.

JEL Codes: H51, I12, I13, I18, J14.

Click here for the latest version of this paper.

*I am grateful to my advisors, Michael P. Keane, Ian Crawford, and Olivia S. Mitchell, for their invaluable guidance. I have also benefited greatly from the comments and advice of Jason Abaluck, Johannes Abeler, Abi Adams, Abby Alpert, Juan-Pablo Atal, Maria Balgova, James Best, Steve Bond, Martin Browning, Minsu Chang, Norma B. Coe, Frank DiTraglia, Rob Garlick, Olga Gdula, Ian Jewitt, Hanming Fang, Jesus Fernandez-Villaverde, Isaac (Zac) Gross, Michal Hodor, Rury Holman, Johannes Jaspersen, Michal Kolesar, Ines Lee, Hamish Low, Dan Polsky, Simon Quinn, Victor Rios-Rull, Paul Sangrey, Molly Schnell, Daniela Scur, Andrew Shephard, Petra Todd, Hanna Wang, and Peter Zweifel, as well as seminar participants at the 2017 National Tax Association Meetings (and my discussant there, David Powell), the 2018 European Health Economics Association (EuHEA) meetings, the 2018 European Economics Association meetings, the 2018 European Winter Meetings of the Econometric Society, the Munich Risk and Insurance Centre, Oxford and the University of Pennsylvania. A significant part of the work on this paper was done while visiting the University of Pennsylvania; a special thanks to Jesus Fernandez-Villaverde for inviting me there (twice). I gratefully acknowledge the support of the Rhodes Trust and the Department of Economics at Oxford during the course of my doctorate. Any and all errors are my own.

†Department of Economics and Nuffield College, University of Oxford. Email: daniel.kaliski@nuffield.ox.ac.uk.

1 Introduction

Recent estimates suggest that increasing the use of preventive care in the United States could prevent up to 2 million years of life being lost per annum and save up to 3.7 billion U. S. dollars (Maciosek et al., 2010). Despite this, health insurance in the United States has often paid for treatment but not preventive care. Tenure with employer-based insurers is short, and so the benefit of investing in beneficiaries' health is likely to accrue to another insurer later on (Fang and Gavazza, 2011). Before 2006, the United States federal government program that covers over-65s, Medicare, provided no general prescription drug benefit. These measures seem likely to discourage individuals from purchasing preventive care themselves, since at the margin the return to prevention is lower when the cost of treatment is covered.

This paper challenges the conventional wisdom (Kenkel, 2000, Card, Dobkin and Maestas, 2008, Simon, Soni and Cawley, 2017) by providing new evidence that these negative "crowding out" effects can be large and important. Previous studies have suffered from one of two biases. First, studies that examine unanticipated changes in insurance status - such as the RAND, Oregon or Massachusetts health insurance experiments - underestimate the effect of unexpectedly becoming richer (the income effect) that offsets any negative effect of insurance for treatment on prevention. Second, studies that overcome this limitation by using anticipated changes tend to focus on behaviors that do not have a strong connection to the eventual costs of treatment. The marginal cigarette or glass of wine will be unlikely to significantly change one's costs of care, and so covering those costs is unlikely to produce large behavioral responses.

By contrast, I focus on behaviors aimed at preventing complications of diabetes, which have strong links with individuals' costs of care even at older ages. I develop a novel empirical strategy for bounding the effect of coverage on prevention that exploits two sources of variation in the Health and Retirement Study (HRS) data, combined with a cleaned version of the data known as the RAND HRS (Chien et al., 2013). Since the data were collected in the United States, the respondents to the survey gain access to universal coverage for treatment at age 65 under Medicare - but in most cases do not gain coverage for many kinds of prevention, including prescription medications, before 2006. This allows for a regression-discontinuity design (RDD) comparing individuals either side of the cutoff. Since this change is anticipated well in advance, there is no income effect that counteracts the negative effect of insurance for treatment on prevention. I am also able to use the panel structure of the HRS to examine changes within the same individuals before and after qualifying for Medicare coverage at age 65. Though the evidence for discontinuities in other outcomes at age 65 is weak, the regression-discontinuity estimator may still suffer from biases from comparing non-comparable individuals either side of the age cutoff. I am able to argue in the context of this paper that these biases are likely to be negative, while those present in the first-differences estimator are positive. I therefore use the first-differences estimate as a lower bound for the negative effect of coverage on prevention, and the panel RDD estimate as an upper bound.

My first main finding is that a drug that is used as prevention by the majority of diabetics (Type IIs) - insulin - was used less often when American diabetics qualified for Medicare coverage at age 65 before 2006. Before 2006, diabetic women are between one and a half and twice as likely as diabetic men to have no form of health insurance at ages 60-64. I find a reduction of between 4.4 and 7.9 percentage points in the proportion of diabetic women who report using insulin to manage their diabetes when they

qualify for Medicare in this period, a relative reduction of between 13% and 30%. I find no evidence of offsetting increases in other preventive behaviors, such as diet, exercise, or use of oral medication. After 2006, the Medicare program was expanded to include a prescription drug benefit (Part D) which included private plans with more generous coverage for insulin, which was only covered in special cases - or with prohibitive coinsurance rates of up to 50% - before 2006. The second main finding of this paper is that this expansion of the program cancelled out the *ex ante* moral hazard effect of providing coverage for treatment. I am unable to reject the hypothesis that this offset was twice as large as the original negative effect, so that qualifying for Medicare coverage from 2006 onward had a net *positive* effect on the likelihood of insulin usage.

Expanding Medicare to include prescription drug coverage is likely to have saved up to $487 million per annum in health care costs among female diabetics. This is partly due to a forgone increase of 4.6 percentage points in heart disease in this group, for whom I find the strongest evidence for *ex ante* moral hazard in insulin usage. This is in line with research which shows that heart disease is the most common complication of diabetes, and that diabetic women's risk for cardiovascular complications is much greater relative to non-diabetic women than diabetic men's risk is relative to non-diabetic men (Juutilainen et al., 2004). My calculations indicate that these cost savings are up to 36% as large as those that would result from a similarly effective tobacco control program. Since the latter is widely believed to be the most effective method of improving population health in the developed world, the findings in this paper imply that insulin is in the first rank of effectiveness among public health initiatives. At the same time, the return on investment for insulin subsidies may be significantly lower in the United States than in the past relative to elsewhere in the developed world due to the rapid increase in the price of insulin in that country since 2006.

My results cast doubt on the current consensus that expanded coverage unambiguously improves population health (Sommers, Gawande and Baicker, 2017). While this is likely to be true in the aggregate, jointly providing coverage for treatment and prevention can mask the moral hazard effects of coverage for treatment, which can crowd out investments in health at the margin. It may still be that universal health care regimes are better at incentivising investments in health by being more likely to pay for them, since this is one of the main methods by which they hold down overall costs. The main policy implication of this paper's results is that policymakers have nonetheless underestimated the extent to which these incentives are necessary even where they are already provided.

I focus on diabetics in particular for three main reasons. First, diabetes is one of the fastest-growing noncommunicable diseases in the world. Recent forecasts have estimated that the global diabetic population will have more than doubled between 2000 and 2030, from 171 million to 366 million people, even if obesity rates remain constant (Wild et al., 2004). In the United States, the proportion of the population with diabetes has been estimated at between 12 and 15 percent (Menke et al., 2015).

Second, diabetics typically have high medical expenses that are closely linked to how well they are able to manage their condition. There are individual actions, such as injecting insulin, that are closely tied to their eventual health outcomes, which is not true of most individuals, or even most individuals with chronic conditions. This allows me to circumvent the usual problems encountered by studies of moral

hazard in health behaviors, where any single practice typically has a limited marginal contribution to the costs of care.

This also means that incentives for better or worse self-management among this group matter for the eventual costs of their health care, which typically have a high social cost. In the United States, most of their medical expenses paid for by the Medicare program after age 65. The global burden of diabetes has recently been estimated at $1.31 trillion U.S. dollars, or 1.8% of global gross domestic product (GDP) (Bommer et al., 2017), and 65% of those costs were estimated to result from the direct costs of maintaining the health of diabetics or treating complications due to their condition.

The third reason is that prior to the passage of the Patient Protection and Affordable Care Act (ACA) in 2010, diabetics were routinely ineligible for privately purchased insurance in the United States due to their pre-existing condition. As a result, treatment for medical complications arising from insufficient control of their condition was not just uninsured but *uninsurable* if they didn't have access to insurance via their or their spouse's employer or the low-income health insurance program Medicaid. Hence their risk of incurring large medical expenses at age 65 changed from "background risk" to insurable (and insured) risk. A large literature spanning both macroeconomics and microeconomics focuses on the different implications for behavior of uninsurable and insurable risk (Aiyagari, 1994, Carroll, Dynan and Krane, 2003, Curcuru et al., 2010, Eeckhoudt, Gollier and Schlesinger, 1996, Guerrieri and Lorenzoni, 2017). Differences in diabetics' behavior when faced with background risk and insured risk shed light on the relative importance of the two for choice under uncertainty. They afford us an answer to the question "what would happen if we converted uninsurable background risks to risks against which agents had insurance?".

This paper makes three methodological contributions. First, this is the first paper to use the distinction between the Marshall, Hicks, and Frisch elasticities of a decision with respect to a price change in order to analyze *ex ante* moral hazard in prevention.[1] The model from which I derive these elasticities has three functions. First, it allows me to reconcile my results with those from other studies. Second, it sheds light on the distributional effects of crowding out prevention - those for whom prevention matters most are the same individuals for whom it is crowded out the most. Third, it makes quantitative predictions regarding other responses such as the income elasticity of investments in health.

The Hicks elasticity corresponds to the pure substitution effect due to the change in relative prices resulting from an unexpected change in the price of health care relative to other spending. The Marshallian elasticity corresponds to the pure substitution effect of the Hicks elasticity plus a countervailing income effect - since individuals are richer, they buy more prevention. The Frisch elasticity is the elasticity of intertemporal substitution: it is the effect on differences in the usage of preventive care across periods of differences in the price of treatment across periods, holding the marginal utility of lifetime wealth constant. These distinctions allow me to explain differences in results between the literature on experimental results (which contain income and wealth effects, and hence recover the Marshallian elasticity) and the literature that uses Medicare eligibility as part of a regression-discontinuity design (which recovers Frisch elasticities, since the estimated responses are long-anticipated reactions to eligibility, and hence the result

[1] See Keane (2011) for a review of the literature on estimating these quantities for the response of labor supply to changes in wages and/or taxes.

of intertemporal substitution). I leave the estimation of Hicks elasticities of prevention with respect to the price of treatment to future research.

The second methodological contribution is to obtain bounds on the effect of gaining insurance at age 65 by comparing cross-sectional regression discontinuity results, which are potentially biased downward and away from zero, to first-differences "event-study" results, which are biased upward and towards zero. The first reason RDD results have negative bias is due to differential mortality. Minimizing the mean squared-error (MSE) of the estimator necessitates using wider bandwidths - especially with a discrete running variable, as only a small number of individuals are located at any one age (Calonico, Cattaneo and Titiunik, 2014, Kolesár and Rothe, 2018). This leads to the inclusion of individuals in the sample who will have had to have been healthier in order to survive to older ages. They will therefore be less likely to need insulin to manage their condition, spuriously producing fewer insulin users to the right of the cutoff.

Similarly, due to the accelerating approach of the end of life for older individuals, there will be spuriously weaker trends towards insulin usage among diabetics who are continuously insured by Medicare than those who have just qualified at age 65, biasing a first-differences estimator that compares those who have just aged into the program with their older, continuously insured counterparts towards zero. The bias due to this second mechanism is also negative for the regression-discontinuity estimator: in the regression discontinuity design, the "treatment" group is older than the "control" group (individuals to the right vs. the left of the cutoff age), whereas in the first-differences regressions the "treatment" group are "switchers" who have just qualified for Medicare, and are younger than the "control" group of those who are continuously insured by Medicare. This produces biases in different directions for the two estimators even given the same underlying mechanism.

To my knowledge, this is the first paper to use panel data in conjunction with an RDD to obtain these bounds. Obtaining partial identification via a combination of economic theory and specific institutional features constitutes a third way in between methods that use as few assumptions as possible to derive bounds (Manski, 1990, 1997, Manski and Pepper, 2000) and methods that derive bounds on parameters of interest from structural economic models (see Ho and Rosen (2015) for a broad overview of this literature). The specific methods in this paper can be fruitfully applied to any data set such as the Health and Retirement Study data where there are both repeated observations of the same units over time and exogenous variation based on a variable that changes deterministically with time.

The third methodological contribution is to extend the standard regression-discontinuity design to account for persistence in individuals' decisions. Persistence can arise endogenously through forward-looking investment in human capital, where there is complementarity between present, past, and future investments. A wide range of other economic models also generate persistence, such as those with exogenous habits formed via routine or external advice being positively serially correlated over time.

To my knowledge, this is the first paper that identifies the coefficient on a lagged dependent variable via institutional variation. The main previous paper to use a dynamic RDD is Cellini, Ferreira and Rothstein (2010). In that paper, the *treatments* are correlated over time, so that being exposed to a program in the past decreases the probability of receiving it in the present. By contrast, in my setting the program

eligibility is deterministically linked across time since the eligibility criteria depend on age, while it is individuals' *decisions* that are correlated across time periods. This deterministic link between program eligibility in one period and eligibility in subsequent periods provides variation in the data that allows me to identify the persistence parameter: some individuals above the cutoff will have gained coverage only in the last survey period, while others will have already crossed the threshold that provides them with coverage. With no persistence, there should be no difference in the two groups' differences with the average outcomes of the uninsured. With persistence, older individuals will have their decisions affected by both having crossed the threshold last period and being insured in the present period, while those just above the threshold will only exhibit the effect of being covered in the current period. Common problems that plague the best currently available dynamic panel data methods - which also arise when they are applied to this setting, as I explore in the Appendix - are those of weak instruments or "too many" instruments, for which recent remedies such as those of Roodman (2009b) can be inadequate. Similarly, though I am able to show that the persistence parameter can be identified in principle using a dynamic regression discontinuity design, its finite-sample properties are poor. At least one dynamic specification is able to recover the main results while being unable to estimate the extent of persistence in insulin usage.

The rest of this paper is organised as follows. Section 2 presents the medical and institutional features that define diabetics' incentives and constraints in the U.S. health care system in the period 1998-2010. Section 3 outlines the identification arguments for the standard regression-discontinuity framework, the difference-in-discontinuities approach used to recover the effect of Part D, and the difference-in-differences regressions used to recover the aggregate effects on health and health care costs. Section 4 describes the data. Section 5 presents the empirical results obtained by applying the methods described in Section 3 to the data described in Section 4. Section 6 develops the partial identification strategy that allows me to provide bounds on the effect of interest in case point identification fails. Section 7 develops a life-cycle model of prevention to explain the differences between this study's results and those in the rest of the literature, as well as the results' relationship to longer-term effects on prevention. Section 8 concludes.

2 Medical and Institutional Background

Diabetes is a disorder where the cells of the body do not respond to insulin (insulin resistance). Insulin's function is to regulate blood sugar levels. Since insulin decreases blood sugar levels, inability to absorb insulin results in both higher levels of blood sugar and higher volatility of blood sugar levels, both of which are corrosive to the blood vessels within the human body. As a result, diabetics are more likely to experience both disorders of the major blood vessels ("macrovascular" complications) such as heart attacks and strokes and disorders of the small blood vessels ("microvascular" complications) such as retinopathy (which results in blindness), neuropathy (nerve damage, which can cause ulcers and often necessitates limb amputation), and nephropathy (kidney failure). The latter category of disorders - microvascular complications - is observed at a much higher frequency among diabetics than individuals with

other chronic conditions.

Injecting insulin is a form of preventive care for the vast majority of diabetics. For Type I diabetics, roughly 10% of the total diabetic population, poor control of their blood sugar levels will quickly result in life-threatening complications. For Type II diabetics insulin usage is a forward-looking behaviour where the short-term cost of purchasing insulin and blood glucose strips to monitor blood sugar levels is weighed against the longer-term costs of hospitalisation and medical complications. Type Is are typically diagnosed in childhood, while Type IIs develop the disease from middle age onwards. The rate of progression of the disease is highly individual-specific, and depends in part on adherence to prevention regimens that aim to reduce the level and volatility of blood glucose. Unlike Type Is, Type IIs are most likely to be recommended to use insulin only once their disease has progressed to the point where intermediate methods for controlling blood sugar levels such as dieting or oral medication have become ineffective.[2] This will be significant for the empirical strategy in this paper, since healthier diabetics are both more likely to have avoided needing insulin to manage their condition and more likely to survive to older ages.

Importantly for the results in this paper, there are also notable physiological differences between male and female diabetics (Kautzky-Willer, Harreiter and Pacini, 2016). Most women are less likely to suffer from cardiovascular disease than men. Diabetic women still have lower rates of heart disease than their male diabetic counterparts, but the relative risk of heart disease compared to their non-diabetic counterparts is higher for women than for men (Kautzky-Willer, Harreiter and Pacini, 2016). As a result, we should expect that if diabetic women's preventive behavior changes and diabetic men's does not, the main differences in health outcomes between the genders should be cardiovascular. This is in fact what I find in this paper (Section 5.5).

I first focus on the period 1998-2006 in this paper for three reasons. First, a sizeable number of individuals enrolled in Medicare in this period did not have prescription drug coverage. In 1997, only 44 percent of Medicare beneficiaries had some form of prescription drug coverage (Soumerai and Ross-Degnan, 1999); by 2005, this had fallen to 35 percent (Soumerai et al., 2006). Piette, Heisler and Wagner (2004) found in a 2002 survey of diabetics that 28% reported forgoing essential purchases such as food to pay for their medications, with 19% reporting nonadherence due to the high cost of their medications. Since there has never been a generic drug that can substitute for branded insulin, these figures are likely even higher for insulin than for diabetes medications such as metformin. In 2006, by contrast, prescription drug coverage was made available to Medicare Beneficiaries with the rollout of Medicare Part D, which also provided plans with more generous coinsurance rates for insulin than had previously been available, covering up to 100% of the cost of insulin purchases in some cases.

Without coverage for insulin, purchasing it independently could be prohibitively expensive, in part because there is no generic form of insulin. For example, Eli Lilly's fast-acting insulin, Humalog, cost $34.81 per vial (which would typically contain a month's worth of insulin) in 2001. This would amount to a yearly cost of $416.72. This is a modest estimate since many diabetics will require more than one vial's worth of insulin per month. Diabetics who use a "basal-bolus" regime, so called because it combines

[2] Unfortunately, I cannot distinguish between Type I and Type II diabetics in the data used for this paper.

a baseline daily dose of insulin (the "basal" part) with regular injections before mealtimes (the "bolus" part), will require 9 vials every two months on average. This amounts to a yearly cost of $1879.74 in 1998 dollars. In 1998 200% of the federal poverty line (which would exclude the possibility of qualifying for Medicaid) outside of Alaska and Hawaii for a two-person household was $21 700. Therefore an uninsured married couple with one diabetic member could expect to spend 7% of total household income on insulin alone if they were at 200% of the federal poverty line in 1998. Moreover, even with coverage that included a prescription drug benefit, insulin's dual status as a non-generic drug and an "injectable" often led to small rates of reimbursement for insulin on health insurance plans (discussed in more detail below).

Second, in 1997 the United States passed the Balanced Budget Act (BBA), which contained several adjustments to the treatment of private health insurers that offered Medicare beneficiaries different packages of coverage as an alternative to traditional fee-for-service Medicare. Starting the sample period in 1998 therefore allows for relative stability in the Medicare program over the pre-2006 portion of the data. Third, the Health and Retirement Study (HRS) added several cohorts in 1998, greatly increasing the sample size, which is particularly useful when examining a subset of the full sample.

As a result, the sample period can be divided into two separate health care regimes. The first is the one which prevailed in 1998-2006, after the Balanced Budget Act of 1997 but before the 2006 implementation of the Medicare Modernization Act (MMA). The second is the one that prevailed in 2006-2010, after the Medicare program had been expanded to include prescription drug coverage under Medicare Part D, but before the passage of the Patient Protection and Affordable Care Act (ACA) in 2010. In Section 5.5, the aggregate results include data from the years 2010-2014. In that Section, I discuss why the passage of the ACA does not pose a significant threat to my ability to attribute the observed aggregate changes to the rollout of Part D.

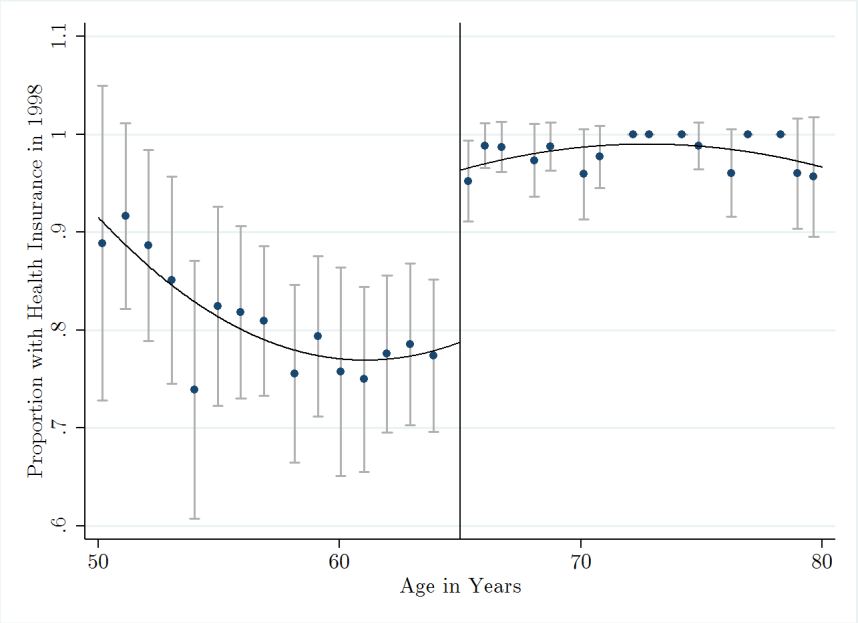
Table 1: Percentage of Diabetics Uninsured Ages 60-64 By Gender (Excl. Medicaid Recipients), 1998-2008

	1998	2000	2002	2004	2006	2008
Men	15.98	14.65	9.70	7.11	7.74	12.50
Women	25.87	26.36	15.32	18.60	16.33	14.29

Notes: Each cell in the top row displays the percentage of diabetic men who are not enrolled in Medicaid, the public health insurance program for low-income low-asset United States citizens, for a given wave of the Health and Retirement Study (HRS). Each cell in the bottom row reports the same percentages for women.

Consider an American diabetic who is younger than 65 in 1998. Her health insurance options will depend upon the severity of her illness. In the worst case scenario, where her disease has already progressed to End-Stage Renal Disease (ESRD), also known as kidney failure, she will qualify for Medicare, which is normally only available to over-65s. If she has made a successful application for disability benefits (SSDI) and been collecting them for two years, she will also qualify for Medicare despite being under 65. If her income and assets are low enough, she can qualify for her state's Medicaid program, which will give her both coverage for hospitalisations and subsidies for insulin, blood glucose strips and other sup-

Figure 1: Regression Discontinuity Plot - Insurance Status of Diabetics Before and After Age 65 in 1998



Notes: The figure represents binned data by age group with a second-order polynomial fit either side of the cutoff. The kernel used is the Uniform kernel. Medicaid recipients are excluded from the calculations. The dependent variable is an indicator variable equal to one if an individual reports having health insurance through any of the following sources: their employer, their spouse's employer, their union, veterans' agencies (Tricare), Medicaid, Medicare or a privately purchased plan.

plies necessary to manage her condition. If she is not eligible for Medicaid, and is well enough to work, she will be reliant on her employer, her spouse's employer, or (in some cases) her trade union to enrol her in a health insurance plan. Private insurance plans outside of employer-based plans will almost certainly deny her coverage on the basis that she has a pre-existing condition (employer-based plans could only do this for a year after an employee is hired; thereafter, they become part of the group-based insurance plan offered by their employer's insurance provider). This shows up in the data as a much larger gain for female diabetics than male diabetics in access to coverage at age 65, which cannot be discerned in the aggregate change displayed in Figure 1, but can be observed in Table 1. Female diabetics who are not enrolled in Medicaid, the publicly provided health insurance program for low-income populations in the United States, are consistently between one and a half and twice as likely to be uninsured before they qualify for Medicare coverage at age 65, with the gap only narrowing significantly in the last survey year before the Affordable Care Act (ACA) is passed in 2010. This is likely due to the lower attachment of these cohorts of women to the labor market before age 65. It is for this reason that the remainder of this paper focuses on changes in female diabetics' health behaviors.

Once she turns 65, she will become eligible for Medicare Parts A and B, which will cover her for treatment and doctor's appointments. (Part A is for "inpatient" services such as hospital stays, whereas

Part B is for "outpatient" services such as doctor's visits, X-rays, outpatient surgeries and laboratory work). Traditional Medicare will not, however, cover her insulin unless she is one of the rare individuals who is recommended by her doctor to use an insulin pump, in which case 80% of the cost of the pump and its insulin will be paid for by Medicare Part B. In the overwhelming majority of cases she will be liable for the costs of preventing a serious medical episode, but not for the costs of treating her once it occurs. Diabetics could opt to receive their Medicare coverage through a privatised plan on the Medicare+Choice programme (renamed Medicare Advantage in 2003), which did typically provide coverage for prescription medications, but access to insulin would still face the following obstacles. First, a growing proportion of these plans in 1998-2006 - 26% in 2002 (Christian-Herman, Emons and George, 2004) - would restrict prescription drug benefits to generics, which excludes insulin since there was and is no generic form of insulin. Second, these plans were mostly available to urban Medicare beneficiaries, since firms offering Medicare Advantage plans have operated in almost no rural counties (McGuire, Newhouse and Sinaiko, 2011), and more rural states such as Alabama tended (and still tend) to have the highest incidences of diabetes. Third, insulin is usually classed differently from most prescription medications as it is an "injectable" and if covered typically involves higher co-payments than other prescription medications (Joyce et al., 2007, Boland, 1998). For example, Boland (1998) gives the example of one of the largest Health Maintenance Organisations (HMO) in New York, Independent Health, increasing its coinsurance rate for injectables to 50% in 1997, before the beginning of the sample period. Moreover, though there was an active effort in the United States Congress' Balanced Budget Act of 1997 to encourage take-up of Medicare+Choice plans, initial enrollment was low and declined as many insurers exited the Medicare+Choice market (McGuire, Newhouse and Sinaiko, 2011). It is for these reasons that it is likely that the "intensive-margin" effects of more generous prescription drug coverage on Medicare are likely to be small for insulin usage over the period studied. Further evidence that another intensive-margin effect - lowering the price of generic medications paid for by a Medicare+Choice plan - did not induce significant substitution towards using generic oral medications to manage diabetes is presented below. There appears to be no corresponding upward spike in self-reported usage of oral diabetic medications at age 65, which one would expect if the subsidies provided by Medicare+Choice plans were an important countervailing factor.

Before 2006, the previously uninsured who gained coverage via the Medicare program could also purchase supplemental insurance only offered to Medicare beneficiaries, commonly known as Medigap (since it fills the "gaps" in Medicare coverage). The number and type of Medigap plans varied (and varies) from state to state, but prior to 2006 only 3 out of 10 Medigap plans included prescription drug coverage, and even those that did were likely subject to similar restrictions on insulin coverage as above. In sum, only low-income, low-asset diabetics who qualified for Medicaid coverage in the period before 2006 were guaranteed full coverage for the costs of using insulin.

In the next section, I describe the strategy I employ to estimate the effect of coverage on insulin usage.

3 Empirical Strategy

The first central goal of this paper is to provide evidence for a substantial negative effect of insurance for treatment on prevention. In this section, I will describe the regression-discontinuity design that I use to identify this effect, the difference-in-discontinuities design that allows this effect to change when the policy regime changes, and the quasi-difference-in-differences method I use for estimating the effects of this policy change on aggregate health outcomes.

3.1 The (Fuzzy) Regression-Discontinuity Estimator

In this subsection, I will outline the usual conditions for consistency of the regression discontinuity (RD) estimator.

Suppose we have a panel of observations with individuals indexed $i \in \{1,...,N\}$ in periods indexed $t \in \{1,...,T\}$ for some outcome Y_{it}, a vector of covariates X_{it}, a running variable R_{it} with some discontinuous change in program assignment at $R_{it} = \bar{R}$, time-invariant unobserved heterogeneity η_i, idiosyncratic unobserved shocks v_{it}, some (usually polynomial) functions $f(.)$ and $g(.)$, which are continuous in R_{it} at $\bar{R}$ and have parameter vectors γ_0 and γ_1 respectively, and a dummy indicator variable for assignment to "treatment" (in this setting, coverage for medical treatment) D_{it}, and denote by h the bandwidth - the absolute distance from the cutoff that determines whether an observation is included in the sample or not, and by $K(.)$ some kernel function, both of which are chosen at the discretion of the econometrician,

$$Y_{it} = \beta_0 + \beta_1 1[R_{it} \geq \bar{R}] + f(R_{it}, \gamma_0) + g(R_{it}, \gamma_1) \times 1[R_{it} \geq \bar{R}] + \delta X_{it} + \zeta t + \eta_i + v_{it} \text{ for } K\left(\left|\frac{R_{it} - \bar{R}}{h}\right| < 1\right); \tag{1}$$

I use local linear regression (so that $f(.)$ and $g(.)$ are linear) and the Uniform kernel throughout, so that $K(.)$ is just an identity function, and so the regressions are restricted to observations for which $\left|\frac{R_{it}-\bar{R}}{h}\right| < 1$ (see the empirical specification, Equation 4 in Section 5). Local linear regression has the advantage of putting the least weight, of all local polynomial regressions, on observations far from the cutoff (Gelman and Imbens, 2018). The Uniform kernel does not have this same advantage - the Edge (Triangular) kernel places more weight than it does on observations near the cutoff, for example - but does have the advantage of transparency, since the weights that are placed on different observations by other kernels are often difficult to interpret, which leads to difficulty interpreting differences across results that use different kernel weighting functions. It is for this reason that Lee and Lemieux (2010) recommend using the Uniform kernel and presenting a variety of results using different bandwidths, h, to make the empirical analysis easier to assess. I present evidence in Section 6 on the effects of varying the bandwidth on the results. I do not, however, use this analysis for bandwidth selection, as this introduces pre-test bias.[3] Instead, I use the mean-squared error (MSE)-optimal bandwidth derived by Calonico, Cattaneo and Titiunik (2014).

The (sharp) regression discontinuity (RD) estimand is a comparison of outcomes just above the cutoff,

[3] Armstrong and Kolesár (2017) derive critical values that are robust to this bias.

for some $R_i^+ > \bar{R}$, and just below the cutoff for some $R_i^- < \bar{R}$ (omitting the covariates for simplicity),

$$\lim_{R_{it}^+ \to R_{it}^-} [E[Y_{it}|R_{it} = R_{it}^+] - E[Y_{it}|R_{it} = R_{it}^-]]$$

$$= \beta_1 + \lim_{R_{it}^+ \to R_{it}^-} [f(R_{it}^+) - f(R_{it}^-) + g(R_{it}^+) - g(R_{it}^-) + [E[\eta_i + v_{it}|R_{it} = R_{it}^+] - E[\eta_i + v_{it}|R_{it} = R_{it}^-]]$$

which recovers β_1, the difference in average outcomes between those who are exposed to the treatment just above the cutoff and those who are just below the cutoff, if and only if (since $f(.)$ and $g(.)$ are

$$\lim_{R_{it}^+ \to R_{it}^-} [E[\eta_i + v_{it}|R_{it} = R_{it}^+] - E[\eta_i + v_{it}|R_i = R_i^-]] = 0$$

- equivalently, $E[\eta_i|R_i]$ is continuous in R_i at the cutoff $\bar{R}$. Intuitively, if no other unobserved characteristics change discontinuously at the cutoff, then the observed change in outcomes can be attributed to the observable change in policy at the cutoff. For example, if some other behavior changes discontinuously at the cutoff, then the observed difference in outcomes could be due to that behavior rather than the observed difference in treatment status. For example, if retired individuals are more likely to use insulin (due to the greater time available to them for the management of their disease), then in a cross section a spurious discontinuous increase in the proportion of those in work at age 65 in a cross-section would exaggerate the effect of gaining insurance.

If some individuals have access to the treatment of interest (gaining coverage, say) without being eligible for the program of interest, then the regression discontinuity design is said to be "fuzzy" instead of "sharp", and since it compares outcomes across individuals it is conventional to scale the difference in mean outcomes between the two groups for the proportion of individuals who change status at the cutoff. In this case the regression discontinuity estimand results from two-stage least squares (2SLS) estimation with D_{it} replacing the assignment indicator $1[R_{it} \geq \bar{R}]$ in Equation 1 and $1[R_{it} \geq \bar{R}]$ used to instrument for D_{it}. Define $Z^+ = \lim_{R_{it} \to \bar{R}} E[Z_{it}|R_{it} \geq \bar{R}]$, and $Z^- = \lim_{R_{it} \to \bar{R}} E[Z_{it}|R_{it} < \bar{R}]$ for any variable Z_{it}. Then the "fuzzy" treatment effect recovered by the 2SLS estimator of β_1, scaled for the difference in the proportion of individuals treated above and below the cutoff, is

$$\frac{Y^+ - Y^-}{D^+ - D^-},$$

if we define p to be the sample fraction of individuals who have access to "treatment" below the cutoff, so that $p = \hat{D}^-$, and all individuals above the cutoff are treated, we obtain that this denominator is $1 - p$, which in this context is the fraction of uninsured individuals before age 65. When comparing the fuzzy regression discontinuity estimates with the first-differences estimates, I will need to rescale them so as to obtain the implicit "treatment-on-the-treated" difference in means, which will be $1 - p$ multiplied by the estimate.

In practice, for a given bandwidth h, the estimators of Y^+ and Y^- for some kernel function $K(.)$ are (Hahn, Todd and Van der Klaauw, 2001)

$$\hat{Y}^{+}(\bar{R}) = \frac{1}{n}\sum_{i=1}^{n} Y_{it}\,1[0 < K\left(\frac{R_{it}-\bar{R}}{h}\right) < 1],\ \hat{Y}^{-}(\bar{R}) = \frac{1}{n}\sum_{i=1}^{n} Y_{it}\,1[0 < K\left(\frac{\bar{R}-R_{it}}{h}\right) < 1];$$

which are consistent estimators of Y^{+} and Y^{-}.

3.2 Difference-in-Discontinuities

I also estimate the effect of Part D, the 2006 expansion of the Medicare program that provided prescription drug coverage at age 65, on insulin usage among diabetic Medicare beneficiaries. More generous coverage options following the implementation of Part D in 2006 are likely to have decreased the price that Medicare-eligible diabetics faced for insulin. This will require a version of Equation 1 augmented to allow for the change at the cutoff to differ by regime period, viz.:

$$\begin{aligned} Y_{it} = \beta_0 + \beta_1 D_{it} + f(R_{it},\gamma_0) + g(R_{it},\gamma_1)\times 1[R_{it}\geq \bar{R}] + \beta_2 1[t\geq 2006] + \beta_3 D_{it}\times 1[t\geq 2006] \\ + \delta X_{it} + \zeta t + \eta_i + v_{it}; \end{aligned} \tag{2}$$

with, as before, the included observations satisfying $K(|\frac{R_{it}-\bar{R}}{h}| < 1)$. Interest will then focus on β_3, the marginal effect of the regime change in 2006, and $\beta_1+\beta_3$, the total effect in 2006. The difference in discontinuities at age 65 between the pre- and post-Part D era will require the following identifying assumption:

$$\begin{aligned} &\lim_{R_{it}^{+}\to R_{it}^{-}} E[\eta_i + v_{it}|R_{it}^{+}, t\geq 2006] - E[\eta_i + v_{it}|R_{it}^{-}, t\geq 2006] \\ &= \lim_{R_{it}^{+}\to R_{it}^{-}} E[\eta_i + v_{it}|R_{it}^{+}, t< 2006] - E[\eta_i + v_{it}|R_{it}^{-}, t< 2006] \end{aligned}$$

This is a weaker identifying condition than the original RDD assumption or traditional difference-in-differences (DID) assumptions. We only require that the discontinuity in the unobservables in the period before the policy change is the same as the discontinuity in the unobservables after the policy change. This allows for discontinuities in the unobservables at $\bar{R}$ in each period (and hence is weaker than the standard RD assumptions) and does not require that the unobservables are conditionally independent of the interaction between the period indicator and the treatment indicator (and hence is weaker than the standard DID assumptions). This identifying assumption would fail if, for example, another policy change coincided with the observable change in discontinuities so that the effect of crossing the threshold differs between the periods for multiple reasons, or the composition of individuals who cross the threshold changed between the two periods (i.e. a cohort effect coincidentally lined up with the implementation of the new policy regime). It is clear that checking for violations of this condition is more intricate precisely because it is only violated in more elaborate scenarios. I include robustness checks for whether the change in behavior can be attributed to one of (i) changes in the location of the retirement spike between the two eras or (ii) changes in patterns of selection into Medicare Advantage plans (which also changed as a result of the Medicare Modernization Act of 2003 that created Part D (McGuire, Newhouse and Sinaiko, 2011)).

Despite these weaker identifying assumptions, the difference-in-discontinuities estimator may suffer from finite-sample problems that are common to the dynamic RDDs in the preceding subsections. These problems are inherent to two-stage-least-squares (2SLS) estimators of heterogeneous effects. In the difference-in-discontinuties design, the effects differ between the pre-2006 and post-2006 periods; in the dynamic RDD (see Appendix), the effects differ according to how long ago individuals became eligible for treatment. Since in the heterogeneous effects case there is more than one endogenous variable, we need more instruments so that there are at least as many instruments in the first stage as endogenous variables in the second stage. Since $1\{R_i \geq \bar{R}]$ will be used to instrument for insurance status pre-2006 (say) and $1[R_i \geq \bar{R}] \times 1[t = 2006]$ to instrument for $D_{it} \times 1[t = 2006]$, this leads to the problem of insufficient *independent* variation across instruments in the first stage, which leads to weak identification (Shea, 1997, Feir, Lemieux and Marmer, 2016). In practice, this turns out to be a more serious problem for the dynamic RDDs than for the regressions investigating the effect of Part D. I address this and other problems of both finite-sample bias and bias in standard errors in the next subsection.

3.3 The Quasi-Difference-in-Differences Design

To examine the effect of Part D on the aggregate health of diabetics, I rely on a quasi-"difference-in-differences" design. This is a modification of the standard difference-in-differences design where all units are exposed to treatment, but the extent of treatment varies exogenously across groups. In this setting, this corresponds to the larger effect of prescription drug coverage post-2006 on diabetic women's insulin usage. This will be seen to result from the larger proportion of diabetic women relative to men pre-65 who are uninsured - likely as a result of their lower participation in the labor market. The estimating equation is

$$Y_{it} = \beta_0 + \beta_1 1[Female = 1]_{it} + \beta_2 1[t \geq 2006]_{it} + \beta_3 1[Female = 1]_{it} \times 1[t \geq 2006]_{it} + \psi_i + \xi_{it}, \qquad (3)$$

with β_3 as the parameter of interest. A key challenge for identification - having established that women respond differently to men to qualifying for Medicare coverage at age 65, and hence to changes in the composition of Medicare benefits - is to explain why differences in levels in say, heart disease between the genders can coexist with parallel trends in that same outcome (Kahn-Lang and Lang, forthcoming). I address these issues in more detail in Section 5.

4 Data

I use data from two principal sources: the Health and Retirement Study, a nationally representative longitudinal survey administered by the Institute for Social Research at the University of Michigan, as well as a cleaned version of a subset of the data called the "RAND HRS" dataset (Chien et al., 2013). Attention is restricted to 3 043 individuals diagnosed with diabetes from the 1998 wave of the HRS. In 1998, the indi-

viduals are drawn from four birth cohorts: the Oldest Old (born pre-1924), the Children of the Depression (born 1924-31), the original cohort from 1992 (born 1931-41) and the War Babies (born 1942-47), plus their co-habitants in the households in which they resided at the time of the survey. The HRS followed up respondents every two years, and so I also have data on surviving individuals from the 1998 wave in 2000, 2002, 2004, and 2006. The demographic characteristics of the sample are summarised in Table 2 (below).

Table 2 shows that the sample is primarily composed of either high school graduates or high school dropouts. The sample has roughly two-thirds as many college graduates as the full HRS sample in 1998 (11.53% among diabetics versus 16.97% for the full sample). Strikingly, Table 3 reports that a majority of diabetics under 65 in the sample are not employed; this is in contrast to the nearly two-thirds of the full sample who report working for pay in 1998. This has adverse consequences for these diabetics' access to health insurance. Less than 1% of the sample report having some form of private insurance that isn't provided by an employer plan, which likely reflects the reluctance of insurers to accept enrollees with pre-existing conditions in the pre-ACA era. Baseline usage of insulin is in line with previous estimates for the total diabetic population in the United States, at 29.13% of under-65s and 25.99% of over-65s, a relative difference of 10.77%. (Compare Saaddine et al. (2002), who find that 30.9% of diabetics in the Third National Health and Nutrition Examination Survey (NHANES III) report using insulin to manage their condition). Those who receive Medicare before age 65 are equally split between individuals who report receiving Social Security Disability Insurance (SSDI) and those who do not. The latter are likely receiving Medicare for the treatment of End-Stage Renal Disease (ESRD), as this is the main alternative for accessing Medicare before age 65 for diabetics, but this is not asked in the HRS survey.

Table 2: Characteristics of the Sample at Baseline - Demographics, Health and Health Behaviors

Characteristic	Age < 65 ($N = 1308$)	Age ≥ 65 ($N = 1735$)
Age - Mean	57.99	74.40
BMI - Mean	30.73	27.86
Male (%)	45.34	45.82
High school graduate (%)	46.71	41.11
College graduate (%)	13.91	9.74
White (%)	65.62	76.89
Black (%)	28.10	18.96
Married (%)	68.46	54.56
Medical History (%)		
Current Smoker	19.88	8.18
Cancer	7.19	14.50
Heart Disease	25.61	39.32
Stroke	8.49	16.26
Using (%)		
Insulin	29.13	25.99
Medication	59.22	62.32
Diet	62.84	59.18
Vigorous Exercise	36.73	27.95

Notes: Drawn from the 3 043 self-reported diabetics in the 1998 Health and Retirement Study. All health conditions except for the "Current Smoker" indicator, which only applies to those who report smoking in 1998, are coded as "1" if a respondent reports that a doctor has ever diagnosed them with that condition, and "0" otherwise. The last four rows correspond to questions only asked of diabetics regarding their methods for managing their diabetes. Respondents who neither report their race as "White" or "Black" in the HRS are coded as "Other", and comprise the balance of the sample.

Table 3: Characteristics of the Sample at Baseline - Insurance and Employment Status

Coverage via:	Age < 65 ($N = 1308$)	Age ≥ 65 ($N = 1735$)
Medicaid (%):	13.0	16.5
Employer (%):	40.3	19.8
Spouse's Employer (%):	19.7	10.2
Union (%):	21.7	9.2
Medicare (SSDI) (%):	9.1	-
Medicare (not on SSDI) (%):	9.1	-
Private Insurance (%):	0.9	-
Not Covered (%):	21.5	2.1
Retiree Benefits? (%):	67.4	-
Working (%):	46.6	12.1
Considers Self Retired (%):	41.2	91.3

Notes: Drawn from the 3 043 self-reported diabetics in the 1998 Health and Retirement Study. I distinguish between individuals who report having access to Medicare and are on Social Security Disability Benefits (SSDI), and those who are not. The latter group almost certainly has access to Medicare via being in End Stage Renal Disease (ESRD) or kidney failure, which is not specifically recorded in the 1998 HRS but is one of the few routes to accessing Medicare before age 65 and a long-term consequence of diabetes.

Table 4: Incidents (%) of First Insulin Use vs. Insulin Cessation, Ages 65-66, 2000-2008

	2000	2002	2004	2006	2008
Ceased	12.50	17.07	11.90	1.89	3.45
Began	4.32	6.25	9.00	13.58	13.64

Notes: Each cell in the top row displays the percentage of respondents who reported using insulin at ages 63-64 two years prior who no longer report using insulin in the survey year in the header at ages 65-66. Each cell in the bottom row displays the percentage of respondents who reported not using insulin at ages 63-64 two years prior who report using insulin in the survey year in the header at ages 65-66. Author's own calculations from the Health and Retirement Study data, waves 4-9.

Table 4 shows that the percentage of diabetics who report not using insulin prior to receiving Medicare who begin reporting insulin usage once they qualify for Medicare coverage increases year-on-year between 2000 and 2006. In 2006, the percentage of newly qualified Medicare beneficiaries who report using insulin at ages 63-64 and no longer report using insulin once they are 65 or older falls precipitiously, from an average of 13.82% across 2000-2004 to 1.89%. This descriptive evidence appears to make a *prima facie* case that Part D's introduction of more generous coverage for insulin significantly offset the crowding out of insulin usage by Medicare's coverage for treatment.

Moreover, when examining the percentages using insulin across the genders, a pattern emerges for

the pooled sample in 1998-2004 that suggests that the crowding out effect of insurance is of first-order importance. Among non-Medicaid-dependent non-smokers aged 60-64 before 2006, female diabetics are between one and a half and twice as likely to be uninsured as male diabetics (see Section 2). At the same time, 25% of female diabetics with the same qualifiers report using insulin before age 65, compared with only 21% of men, a relative difference of nearly 25%. After age 65, these figures are 23.2% for women and 22.6% for men, a relative difference of approximately 2.6%. The relative difference in insulin usage between men and women is therefore smaller by nearly a factor of ten after age 65, when nearly 100% of individuals have access to insurance for treatment, relative to pre-age 65, when there are more uninsured female diabetics than male diabetics.

The following section gives the results of applying the empirical strategy explained in Section 3 to this data.

5 Results

In this section, I first present the results from the panel RDDs on the crowding out of prevention by insurance for treatment. I then extend the analysis to examine the offsetting effect of Medicare Part D, which made prescription drug coverage available at age 65, including more generous coverage for insulin. Lastly, I examine whether Medicare Part D affected aggregate health outcomes. The conclusions in this section are not altered in the dynamic regression-discontinuity designs, for reasons discussed alongside their presentation in the Appendix.

Throughout my analysis I exclude diabetic Medicaid recipients as they are eligible for full subsidies for their insulin by 1997 (with some mild restrictions in some states on the purchase of auxiliary medical equipment such as blood glucose strips). I also exclude smokers, who have a weaker response to treatment with insulin and altered metabolism compared to the majority of diabetics (Eliasson, 2003). The regression-discontinuity and difference-in-discontinuity results are compared with their first-differenced counterparts to obtain lower and upper bounds for the effect of Medicare eligibility on health behaviors.

In the RDD results, I exclude Medicare Advantage and Medigap recipients who are over age 65 to deal with the "multiple treatments" problem (Caetano, Caetano and Escanciano, 2017, Card, Dobkin and Maestas, 2008). Not only do the previously uninsured gain insurance for the first time at age 65, the continuously insured also gain access to more generous coverage than before. The effect at the cutoff will therefore be a combination of these separate effects. Male diabetics do not exhibit the large changes in insulin usage that diabetic women do, and are also significantly less likely to be uninsured prior to age 65 when compared with female diabetics. This provides some evidence that the main results are due to changes at the extensive margin from being uninsured to being insured rather than at the intensive margin from less generous to more generous insurance. I also estimate regressions restricted to individuals who receive health insurance via their own or their spouse's employer, reported in the Appendix - the null results recorded there provide further evidence that changes in the composition and/or generosity of

employer-provided coverage at age 65 is not the main mechanism behind the results. In addition, few of those in the sample who are uninsured before age 65 purchase supplemental insurance after age 65. Of those Medicaid-ineligible diabetics in the period 1998-2004 in the HRS who report buying supplemental insurance in the first two years of their Medicare eligibility, only 10.5% report having no source of health insurance two years prior. (By contrast, the first-differences regressions in Section 6 exploit variation within individuals rather than across individuals, and so can track what happens to individuals before and after qualifying for Medicare, or enrolling in supplemental Medicare insurance (Medigap) or a Medicare Advantage plan. Instead of dropping these observations for the first-differences regressions, I therefore control for the first difference of enrolment in Medigap or Medicare Advantage).

In the Part D results, matters are further complicated by the fact that aftter 2006 Medicare Advantage plans were required by the Medicare Modernization Act (MMA) to offer prescription drug coverage that was at least equivalent to what could be obtained in a private Part D plan (McGuire, Newhouse and Sinaiko, 2011). As a result, diabetics already enrolled in Medicare Advantage plans before age 65 may lead to underestimates of the extent to which the effect of crossing the age 65 threshold changes post-2006. In consequence, I exclude Medicare Advantage enrollees at all ages for the regressions that use the 1998-2008 sample. The resulting loss of observations is compensated by the greater sample size due to the addition of two waves of the HRS data.

To summarize: Medicaid recipients and smokers are present in none of the samples used for estimation. Supplemental insurance and Medicare Advantage enrollees are excluded if over 65 for the regression discontinuity design estimates in 1998-2004, and at all ages for the regression discontinuity design estimates in for 1998-2008 that determine the effect of Medicare Part D. In the first-differences regressions in Section 6, the first difference of indicators for enrolment in Medicare supplemental insurance (Medigap) or Medicare Advantage are included in one of the specifications as part of a set of control variables.

5.1 Crowding Out of Prevention by Insurance for Treatment: 1998-2004

In this subsection I document that the strongest evidence for *ex ante* moral hazard in insulin usage comes from female diabetics. The likely source of this difference is the much larger proportion of diabetic women who report having no source of health insurance relative to men prior to age 65, a difference of ten percentage points.

I now turn to evidence from panel RDDs, pooling together the years 1998-2004 (avoiding the policy regime change of 2006). This involves estimating the empirical counterpart of Equation 1 by two-stage least squares, which, with a Uniform kernel and local linear regression (used throughout this paper), has the second-stage equation

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \gamma_0(R_{it} - \bar{R}) + \gamma_1(R_{it} - \bar{R}) \times 1[R_{it} \geq \bar{R}] + \delta X_{it} + \zeta t + \eta_i + v_{it} \text{ for } \left| \frac{R_{it} - \bar{R}}{h} \right| < 1, \quad (4)$$

where the subscript indicates an observation is for individual i in period t. I distinguish between the time-invariant unobserved "fixed effect" η_i and the time-varying idiosyncratic error v_{it}. X_{it} is a vector of

covariates and h denotes the bandwidth, chosen to minimize the mean-squared-error (MSE) criterion of Calonico, Cattaneo and Titiunik (2014). I cluster standard errors at the individual level to account for the joint presence of persistence in treatment status and the error term (Bertrand, Duflo and Mullainathan, 2004).

Table 5: Unrestricted Panel RDDs, 1998-2004, Diabetic Women: Labor Market Outcomes

(1) Employed	(2) Retired	(3) Partly Retired	(4) Hours	(5) Earnings	(6) Social Security
0.03	-0.03	0.09	-0.31	731.71	-0.00
(0.76)	(-0.72)	(1.01)	(-0.13)	(0.87)	(-0.11)

t statistics in parentheses

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

Notes: Standard errors are clustered at the individual level. All specifications use local linear regression with age in months as the running variable and a Uniform kernel. Bandwidth used is 83.33, selected by the MSE criterion of Calonico, Cattaneo and Titiunik (2014). Individuals enrolled in Medicaid at any age, or enrolled in supplemental insurance (Medigap) or a Medicare HMO (Medicare Advantage) after age 65, are excluded.

Table 6: Panel RDDs, 1998-2004, Diabetic Women: Other Health Behaviors and Outcomes

(1)	(2)	(3)	(4)
Any Hospital Stay	Nights in Hospital	Any Doctor Visit	No. Doctor Visits
-0.04	3.09	0.14*	2.34
(-0.21)	(0.66)	(2.43)	(0.29)
Kidney Problems	Poor Health	Diabetes Diagnosis	BMI
0.25	-0.09	-0.02	3.80
(1.89)	(-0.46)	(-0.37)	(1.67)

t statistics in parentheses

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

Notes: Standard errors are clustered at the individual level. All specifications use local linear regression with age in months as the running variable and a Uniform kernel. Regression discontinuity design is sharp. Bandwidth used is 83.33.

I now turn to testing for other changes at age 65 that might explain any differences in behavior other than Medicare eligibility. One potential threat to internal validity would be the coincidence of retirement at 65 with Medicare eligibility. There is some controversy over whether the spike in retirement status at

65 has disappeared in the United States (Card, Dobkin and Maestas, 2008, Von Wachter, 2002, Johnson, Smith and Haaga, 2013). This is the position of recent papers that use the Medicare eligibility age in a regression-discontinuity design such as Card, Dobkin and Maestas (2008, 2009). Whether or not this prevails for the U.S. population in general, a striking number of the diabetics in the sample under 65 are not employed (see Table 2). This likely contributes to my findings that there is little evidence of a spike in retirement at age 65 for diabetics (see Table 5). Another reason for this absence may be that the employed among this group are likely to retire later than at age 65. This is due to their condition giving them especially strong incentives to retain any benefits their employer-provided coverage may offer that are not provided on traditional Medicare. Several studies have found that retaining health insurance benefits provided by employer-based plans are a significant influence on the timing of retirement in the United States, such as French and Jones (2011), Blau and Gilleskie (2008), Blau and Gilleskie (2006) and Rust and Phelan (1997). In 1997, around 32 percent of private-sector employers offered their employees retiree coverage (Buchmueller, Johnson and Lo Sasso, 2006).[4] In the sample, a higher fraction of employed diabetics have retiree benefits than the general working population, but that nonetheless leaves a significant fraction do not, at around 32.6% (see Table 2). This suggests that the 1998 sample is relatively polarized between those not in work prior to age 65 and those who both work and have retiree coverage. These two forces will lead to a distribution that is polarized between individuals who retire at the Social Security claiming age of 62 on the one hand, and those who delay retirement as much as possible on the other, which will lead to a much less pronounced spike in retirement status at 65.

I also test for discontinuities in eight other behaviors and outcomes at age 65 (Table 6). The only statistically significant change is a discontinuous increase in the probability of reporting having had a doctor's appointment in the past two years, in line with the results found in Card, Dobkin and Maestas (2008) and Dave and Kaestner (2009).[5] Most significantly, there is no discontinuous change in the diagnosis of diabetes at age 65, which would otherwise potentially explain any negative effect as an increase in newly diagnosed diabetics who were not in need of insulin to manage their condition.

The results for the unrestricted panel are summarised in Table 7.[6] The diet and exercise variables are either not available for the entire period 1998-2004 or, in the case of the exercise variable, are changed so as to make comparisons across time difficult. I nonetheless find little evidence of substitution towards these alternative investments in health as a result of crossing the age 65 threshold (see Appendix). The null hypothesis that there is no substitution towards oral medication to manage diabetes at age 65 is not rejected. By contrast, there is consistently strong evidence in favor of a decrease in insulin usage at age 65 across the years 1998-2004. Given the first stage estimate of 0.24 for insurance status, the smallest coefficient of -0.33 implies a $-0.33 \times 0.24 = -7.92$ percentage point decrease in insulin usage among

[4] For a more recent treatment of the effects of retiree coverage on precautionary behaviour before the Affordable Care Act, see Clark and Mitchell (2014).

[5] It is unlikely that these doctors' appointments can explain the reductions in insulin usage in this paper. Once a patient is already using insulin, the "therapy of last resort", it would be contrary to the official guidelines for physicians (Nathan et al., 2009) to recommend that they discontinue using insulin to manage their condition. This would be more consistent with both the argument in Card, Dobkin and Maestas (2008) that a small reduction in smoking at age 65 may be attributed to this greater frequency of doctor's appointments and the medical literature on adherence to insulin, where the goal of most health providers is to encourage adherence to insulin once prescribed (cf. Weinger and Beverly (2010)).

[6] Separate results restricted to the original 1998 cohort can be found in the Appendix.

Table 7: Unrestricted Panel RDDs, 1998-2004, Diabetic Women: Insulin and Oral Medication Usage

	(1)	(2)	(3)	(4)
	Mean 60-64			
Insulin				
	0.26	-0.33*	-0.34*	-0.35*
		(-2.28)	(-2.42)	(-2.46)
Oral Medication				
	0.66	0.15	0.18	0.19
		(0.87)	(1.13)	(1.15)

t statistics in parentheses

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

Notes: Standard errors are clustered at the individual level. Column (2) reports estimated discontinuities from specifications without covariates apart from the age in months and age in months interacted with the treatment indicator; Column (3) reports results including time dummies; Column (4) reports results including both time dummies and health status, marital status, work status and education fixed effects. All specifications use local linear regression with age in months as the running variable and a Uniform kernel. Bandwidth used is 83.33, selected by the MSE criterion of Calonico, Cattaneo and Titiunik (2014). Individuals enrolled in Medicaid at any age, or enrolled in supplemental insurance (Medigap) or a Medicare HMO (Medicare Advantage) after age 65, are excluded.

female diabetics at age 65 from a baseline of 26% at ages 60-64, a relative reduction of 30.5%.

5.2 Mechanisms: Crowding Out via *Ex Ante* Moral Hazard

There are two mechanisms which could produce a discontinuous change in behavior in the month of Medicare eligibility. One of these, which I do not model explicitly in this paper, is a precautionary motive. Hospitalizations do not increase discontinuously, in line with the findings of Card, Dobkin and Maestas (2009) (see Table 6), so the incentive to use insulin is unlikely to come from avoiding emergency treatment. It is more likely that the change in incentives for insulin usage comes from reduced uncertainty regarding the threat of a costly medical incident that, while difficult to defer, need not require immediate medical attention. This is due to the unpredictability of the timing of complications due to poorly managed diabetes. One of the consequences of episodes of abnormally high blood sugar levels (hyperglycemia) is direct damage to the cardiovascular system (Barrett-Connor et al., 2004), which raises the likelihood of an adverse cardiovascular event such as a heart attack. This is supported by the results in Section 5.5, which finds a forgone increase of 4.6 percentage points in the rate of heart disease among diabetic women due to the provision of prescription drug coverage under Medicare Part D in 2006.

Insulin usage may play the same role as precautionary behavior in this context. Studies of precautionary saving find similar results for exogenous variation across individuals in the replacement rate provided by unemployment insurance (Engen and Gruber, 2001) and access to a consumption floor via social insurance (Hubbard, Skinner and Zeldes, 1995). A discontinuous change in the month of Medicare eligibility, similarly, is consistent with an exogenous reduction in uncertainty regarding the risk of being liable for medical expenses for complications of diabetes at that time.

As individuals approach the threshold of Medicare eligibility, the risk of liability for large medical expenses decreases. As a result, if this is the mechanism behind the discontinuous decrease at the threshold, there could in principle also be a continuous downward trend before age 65 (cf. De Preux (2011) on "anticipatory moral hazard"). However, a consistent estimator of this age effect is difficult to obtain in practice; in the regression-discontinuity design, the running variable is not exogenous and we cannot obtain consistent estimators of the age profile of the outcome of interest (otherwise there would be no need for the discontinuity in the first place). In addition, using panel data presents the problem of distinguishing age effects from period-specific and cohort-specific effects. I do not attempt to obtain a consistent estimator of the extent of anticipatory moral hazard in this paper, since the existence of per-period reductions leading up to age 65 is not mutually exclusive with the existence of a discontinuous change in individuals' incentives at age 65.

One test for whether there is a discontinuous change in the risk of medical expenses is to test for discontinuous reductions in the dispersion of medical expenses or the mean medical expenditures at age 65. This would then track the precautionary motive that diabetics have to use insulin, and would shed further light on the mechanism at work. It may be that the sequential reduction of uncertainty month-by-month approaching the date of Medicare eligibility is too small to be distinguished from noise, but that the discontinuous reduction in uncertainty in the month of Medicare eligibility is large enough to produce statistically significant changes in behavior. Barcellos and Jacobson (2015) find both a discontinuous 53 percent reduction in the 95th percentile of medical expenses at age 65 in the Medical Expenditure Panel Survey and similar declines in medical expenditures risk in the HRS, the same data set used in this study. In the Appendix, I examine similar evidence for changes in two measures of financial and medical expenditure risk at age 65, but obtain less conclusive results, likely because of the smaller sample size in this study.

It is also possible to explain the change in behavior as intertemporal substitution in a model without uncertainty. In the model in Section 7, I show that if prevention lowers the marginal utility of medical expenses in the future (i.e. it decreases future demand for healthcare, and is a substitute for treatment *ex post*), then an anticipated reduction in the price of treatment in the next period will lower the optimal amount of prevention in that period. The responses found in this paper can therefore arise in an environment of pure certainty as well (as there is no uncertainty in the aforementioned model). It is sufficient that self-insurance *ex ante* and insurance against financial losses *ex post* are substitutes. This mechanism also has the advantage of allowing me to reconcile the results in this paper with those in other studies. I leave further discussion of the model to Section 7.

5.3 The Effect of Prescription Drug Coverage on Ex Ante Moral Hazard and Aggregate Outcomes

I now turn to analysing the effect of introducing generous subsidies for the purchase of insulin under Medicare Part D in 2006. The key result in this subsection is that Medicare Part D appears to have increased the demand for insulin among diabetic women by enough to more than offset the *ex ante* moral hazard effect of traditional fee-for-service Medicare. These results perform four separate functions in this paper. First, they buttress the initial results on the negative impact of coverage without insulin subsidies on usage by showing that this effect is reversed just when subsidies are introduced. Second, they suggest a method for combating *ex ante* moral hazard - lower the expected price of health-preserving behaviors in tandem with lowering the expected price of health care. Third, strong changes in oral medication usage are not observed with the onset of Part D, likely because of their significantly lower cost and better coverage options before 2006 (see Section 2) compared to those available for insulin. This allows me to attribute the prevention of an increase in heart disease among diabetic women found in Section 5.5 (below) to the insulin subsidies available on Medicare Part D specifically, rather than its broader coverage for other medications.

In the "difference-in-discontinuities" regressions, I estimate the empirical counterpart of Equation 2,

$$\begin{aligned} Y_{it} = \beta_0 + \beta_1 D_{it} + \gamma_0(R_{it} - \bar{R}) + \gamma_1(R_{it} - \bar{R}) \times 1[R_{it} \geq \bar{R}] + \beta_2 1[t \geq 2006] + \beta_3 D_{it} \times 1[t \geq 2006] \\ + \delta X_{it} + \zeta t + \eta_i + \nu_{it} \end{aligned} \tag{5}$$

In Section 3, I discuss the weaker identifying assumptions necessary to identify the effects of interest than in the preceding subsection. Although two-stage least squares regressions that use highly correlated instruments are more susceptible to weak identification (Shea, 1997), in practice I find strong evidence against the null hypothesis that the set of instruments is weak.

There are three additional empirical challenges in this subsection. The first results from the fact that the passage of the Medicare Modernization Act (MMA) was in 2003, so there is at least one survey year (2004) in which individuals' behavior may have already been affected due to their anticipation of the availability of prescription drug coverage two years thereafter (Alpert, 2016). I discuss how much of the results can be accounted for by this mechanism in Section 5.4. Second, the MMA also changed the regulations governing private plans on Medicare Advantage (Part C), and there was a corresponding rapid increase in the take-up of these plans relative to their decline in the period 1997-2003 (McGuire, Newhouse and Sinaiko, 2011). As per the discussion in Section 3 (above), I include robustness checks for changes in enrolment in Medicare HMOs (Medicare Advantage), retirement behavior, and frequency of diagnosis, as well as time dummies to absorb anticipatory behavior by those not yet eligible for Medicare (Table 8). Third, as in the case of the dynamic equations (see Appendix), heterogeneous effects of two-stage least squares require at least as many sources of exogenous variation as endogenous variables. Although the identification conditions are weaker than for two-stage least squares (see above), the finite-sample issues are the same. Since the variation in $1[R_i \geq \bar{R}]$ is similar to that in $1[R_i \geq \bar{R}] \times 1[t = 2006]$,

Shea's R^2 may be low, reflecting little independent variation in the first-stage (Shea, 1997). Since this variation is monotonically increasing in the sample size, the optimal bandwidth for the purposes of maximizing Shea's R^2 is $h = \infty$. It turns out that in this case there is little evidence of weak identification even at the MSE-optimal bandwidth of $h = 66.23$, as the Cragg-Donald statistics used to test for the presence of weak instrument sets all exceed conventional critical thresholds used to reject the null hypothesis that the set of instruments is weak (Stock and Yogo, 2005).[7]

Table 8: Differences in Discontinuities in Medicare Advantage Enrolment and Employment Measures, Pre- and Post-2006

	(1)	(2)	(3)	(4)
	Med. Adv.	Employed	Retired	Partly Ret.
Women				
	0.08	-0.03	-0.00	-0.04
	(1.37)	(-1.27)	(-0.00)	(-1.16)
Men				
	0.10	-0.03	0.06*	-0.06
	(1.56)	(-1.12)	(2.24)	(-1.79)
	Hours	Earnings	Soc. Sec.	Diagnosis
Women				
	-0.11	-574.21	0.05***	0.05**
	(-0.11)	(-0.63)	(4.54)	(2.69)
Men				
	-0.31	-2252.86	0.05***	0.05**
	(-0.27)	(-1.24)	(4.60)	(2.69)

t statistics in parentheses

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

Notes: Estimates are from the interaction term between the indicator for the year 2006 and the treatment indicator for a sharp regression discontinuity design. Standard errors are clustered at the individual level. "Medicare Advantage" results are from regressions where the dependent variable is equal to 1 if survey respondents answer "Yes" to the question "Do you receive your Medicare through an HMO?" and 0 otherwise. Earnings are measured in constant 1998 dollars. All results are from local linear regressions using the Uniform kernel. Bandwidth used is 66.23, selected by the MSE criterion of Calonico, Cattaneo and Titiunik (2014).

It appears that there are mild differences in the effect of reaching age 65 on employment outcomes and diagnosis of diabetes between the period 2006-08 and 1998-2004 (Table 8). The latter is a reasonable response to gaining health insurance at age 65. As pointed out by Kenkel (2000), not all forms of

[7] Further discussion of weak identification in regression-discontinuity designs can be found in Feir, Lemieux and Marmer (2016).

prevention are the same: investments in health, as in the bulk of this paper, are substitutes for having coverage *ex post*. By contrast, screenings are complementary to having coverage, since knowledge of one's condition is more useful if one can pay to treat it once it is discovered. Providing more generous coverage for prescription drugs will therefore provide an extra incentive to screen for conditions such as diabetes, due to individuals' increased ability to pay for the maintenance of one's health on discovering latent diabetes.

Of the changes in retirement and Social Security claiming behavior, none are particular to women and not to men (the increased rate of retirement post-2006 is among men only, and the rise in Social Security claims occurs across both sexes). The change in diabetes diagnosis is close to identical in both sexes. It seems unlikely - with the exception of increased retirement among men post-2006 - that these changes can explain disparities between the genders in their responses to qualifying for Medicare and the change in the Medicare program after 2006. There are two further reasons to suspect that these differences in discontinuities do not explain changes in insulin usage instead of differences in insurance status and the composition of available insurance packages. First, as in Card, Dobkin and Maestas (2008), the change in retirement status amongst men is too small to explain a dramatic change in the behavior of women either relative to men or in absolute terms. Second, to the extent that increased frequency of diagnosis of diabetes introduces bias into the estimator of the coefficient on $\widehat{D}_{it} \times 1[t \geq 2006]$, $\widehat{\beta}_3$, this bias is likely to be towards zero, since newly diagnosed diabetics are not typically prescribed insulin as it is a therapy of last resort (see Section 2). As a result, the effect of Part D will be *under*stated if the difference-in-discontinuities in diagnosis of diabetes plays a large role in producing the results to follow.

Table 9 shows the effect of Part D on the change in the effect of qualifying for Medicare. I use a bandwidth of 66.23, selected by the MSE criterion derived in Calonico, Cattaneo and Titiunik (2014) as before. The effect of qualifying for Medicare in 2006-08 is found to have a significantly more positive net impact on insulin usage than in the pre-2006 part of the sample, as examination of Table 3 in Section 5 would suggest. A Wald test cannot reject either the hypothesis that $\beta_1 + \beta_3 = 0$ or that $\beta_3 = -2\beta_1$. In sum, it appears that Part D increased the demand for insulin to an extent that completely offset the *ex ante* moral hazard effect of coverage for treatment on Medicare Parts A and B. The precise size of this effect is difficult to ascertain, but the results leave room for the possibility that it not only completely offset the negative "crowding out" of insulin usage at age 65 but also led to an equally large increase in uptake at 65 post-2006.

5.4 Mechanisms: Prescription Drug Coverage Under Part D Post-2006

In this subsection I discuss potential alternative mechanisms that can account for the net positive effect of qualifying for Medicare coverage on insulin usage post-2006.

There are two potential challenges to interpreting these results against which I can find no direct evidence in the data. The first is that a new long-acting (requiring only once-daily usage) insulin, insulin detemir (Levemir), was approved by the United States Food and Drug Administration in 2005, the year before Part D was implemented. At least one other long-acting insulin compound, insulin glargine (Lantus), had been available since April 2000. Since these types of insulin were not differentially available

Table 9: Difference-in-Discontinuities, Diabetic Women: Effect of Part D on Insulin and Oral Medication Usage

	(1)	(2)	(3)
Insulin			
$\widehat{\beta}_1$	-0.34*	-0.35**	-0.35**
	(-2.50)	(-2.62)	(-2.62)
$\widehat{\beta}_3$	0.68*	0.68*	0.60*
	(2.41)	(2.42)	(2.28)
Cragg-Donald Stat.	33.70	35.22	40.77
Oral Medication			
$\widehat{\beta}_1$	0.08	0.10	0.11
	(0.48)	(0.65)	(0.72)
$\widehat{\beta}_3$	0.03	0.03	0.05
	(0.11)	(0.12)	(0.17)
Cragg-Donald Stat.	33.66	35.27	40.63

t statistics in parentheses

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

Notes: Standard errors are clustered at the individual level. Column (1) reports estimated discontinuities from specifications without covariates apart from the age in months and age in months interacted with the treatment indicator; Column (2) reports results including time dummies; Column (3) reports results including both time dummies and work status, marital status, education, and health status indicators. All results are from local linear regressions using the Uniform kernel. Bandwidth used is 66.23, selected by the MSE criterion of Calonico, Cattaneo and Titiunik (2014). Individuals enrolled in Medicaid, supplemental insurance (Medigap), or a Medicare HMO (Medicare Advantage) at any age are excluded.

to over-65s, accounted for a small share of the market for insulin over the sample period, and require large implicit non-monetary costs of insulin usage to explain the results here, it seems unlikely that their introduction is responsible for the results in this subsection. Another reason to be sceptical that this can explain a large share of the difference in the treatment effect in 2006 is that the more significant innovation of long-acting insulin had taken place three years before the passage of the Medicare Modernization Act and six years before the rollout of Medicare Part D. The approval of insulin detemir was a signif-

icantly smaller contribution to the therapeutic options available to diabetics relative to the invention of long-acting insulin, which was already available at the time.

The second is that take-up of Part D was generally slow and made difficult by a chaotic sign-up process. There are three potential explanations for the results in this section even with this observation. First, the subset of individuals who did sign up for Part D were precisely those with a high enough demand for coverage for insulin to outweigh the moral hazard effect. Second, there is evidence that the effect of Part D went beyond the direct effect on prices faced by enrollees and had spillover effects that lowered the prices faced by other Medicare beneficiaries (Duggan and Scott Morton, 2010). Third, the 2006 wave of the Health and Retirement Study was collected between March 2006 and February 2007; while the beginning of the survey period had some overlap with the period during which enrolment in Part D had been lower than anticipated, by July 2006 nearly 22.5 million senior citizens had enrolled in Part D (Cubanski and Neuman, 2007). Hence both the 2006 and 2008 waves of the Health and Retirement Study are likely to have been gathered after significant problems with the rollout of Part D had been resolved.

One explanation for the post-2006 results is intertemporal substitution, similar to the explanation in this paper for the large effect sizes pre-2006 (cf. Section 7). Alpert (2016) studies the effect of Part D's prescription drug coverage on the demand for non-essential medications and finds that those close to the Medicare eligibility age and those who qualify for Medicare coverage before 2006 strategically delay purchases of prescription drugs until they become cheaper once subsidies under Part D are rolled out. This accords with statistically significant negative coefficients on the Post-2006 indicator variable in the RDD results (though its counterpart is not significant in the first-differences regressions). In principle, this could mean that there is no net effect adherence to insulin therapy; instead, there could be merely a redistribution of the timing of initiating insulin therapy so that the negative effect found in 1998-2004 is not offset at all by the subsidies for insulin available after 2006 on Medicare Part D. There are three reasons to believe that the lifetime increase in insulin usage exceeds the measured intertemporal substitution effect. First, individuals' lifetime income was increased by the passage of the Medicare Modernization Act which created Part D, which would increase the lifetime demand for insulin even absent any price effects. Not only were there no tax increases implemented to pay for Medicare Part D (that could have in principle neutralized this effect), the Tax Increase Prevention and Reconciliation Act of 2005 extended the horizon to which the tax cuts of 2001 and 2003 applied, effectively until the end of the working lives of those near Medicare eligibility. Second, as mentioned when discussing the problems with the rollout of Part D (above), Part D lowered the prices of covered prescription drugs after its implementation more generally (Duggan and Scott Morton, 2010). Third, the net effect is likely to be understated due to the increased frequency of diagnoses of diabetes upon qualifying for Medicare post-2006 (see discussion in previous subsection).

In sum, it seems extremely unlikely that Part D did not increase lifetime insulin usage among diabetics. In the next section, I examine the implications of Part D encouraging insulin usage for aggregate health outcomes and health care costs.

5.5 Aggregate Effects of Prescription Drug Coverage: 1998-2014

Despite the large changes in behavior documented above that ought to affect blood sugar levels and fluctuations, which are known to damage both the large and small blood vessels, I can only discern evidence in the HRS data for a reduction among diabetic women in the most common complication of diabetes, which is heart disease. In this section I document a relative decrease of 4.6% in the trend in heart disease rates among female diabetics over 65 relative to their male counterparts in the post-Part D era. This also accords with the results of the previous section, where the largest behavioral effects are observed for women. I then calculate a conservative estimate of forgone health costs based on this decrease in heart disease, as well as a larger estimate based on a previous study of cost containment attributable to reductions in blood sugar levels among diabetics.

Accordingly, the main difference that can be attributed to higher take-up of insulin following Part D is an improvement in the rate of heart disease: though diabetic men saw their rate of heart disease rise by 4.6% in 2006-2014, the proportion of diabetic women who had contracted heart conditions remained constant. In this subsection I graph the year-on-year deviations from the 1996 average of heart disease for diabetics over and under 65 separately for each gender, and report a difference-in-difference specification to quantify the extent of the differences after 2006. Throughout this subsection, I exclude Medicaid recipients for the same reasons as in the preceding subsections.[8]

Figure 2 shows the differences in rates of cardiovascular disease between men and women over and under the age of 65 relative to the year 1996. We need evidence for two assumptions to attribute a given change in disease trends to Part D. The first is that there is no corresponding change in trends among under-65s who are not affected by changes to Medicare. The second is that men and women's trends before 2006 moved in parallel, so that their trends would have been parallel in the counterfactual where Medicare Part D was not introduced. The bottom panels of Figure 2 show flat trends for diabetic under-65s, supporting the first assumption. The second assumption is supported by the flat trends in cardiovascular disease for both genders among over-65s prior to 2006, when the rates for men start increasing relative to 1996 at an increasing rate. Table 10 estimates the difference-in-differences between men and women before and after 2006, and finds a statistically significant forgone increase in heart disease of 4.6% among diabetic women. The estimating equation is Equation 3, viz.:

$$Y_{it} = \beta_0 + \beta_1 1[Female = 1]_{it} + \beta_2 1[t \geq 2006]_{it} + \beta_3 1[Female = 1]_{it} \times 1[t \geq 2006]_{it} + \psi_i + \xi_{it}, \qquad (6)$$

The main competing explanation for these changes is the passage of the Affordable Care Act (ACA) in 2010, which mandated changes in the United States' public provision of health insurance over the period 2010-2014. There are three reasons for scepticism that this can explain the patterns observed in Figure 2 and Table 10. The first is that we should observe similar differences among under-65s, to whom the Affordable Care Act - unlike Medicare reform - applied to the same extent. The second is that the

[8] In the Appendix, I present evidence that there are no significant differences in the trends of diagnosis of diabetes or take-up of Medicaid among diabetics that can be attributed to Part D. In the first case, the trend is positive and not significantly different post-2006; in the second case, all of the increase in Medicaid take-up among diabetic men can be attributed to the Affordable Care Act (ACA). I discuss elsewhere in this section why the ACA cannot explain the main difference in trends pre- and post-2006.

main expansions of insurance under the Affordable Care Act were expansions of the Medicaid program, whose recipients are excluded from the analyses in this and the preceding sections. The third is that the timeline of the changes implemented by the Affordable Care Act cannot cannot explain either the modest divergence in trends between men and women in 2006-2010, before the passage of the ACA, or the larger divergence by 2012. In 2012, the United States Supreme Court ruled that the ACA was constitutional in July and President Obama was re-elected that November. Prior to those events, the implementation of the main portion of the ACA - the creation of health insurance exchanges backed by an individual mandate to purchase insurance - was in significant doubt due to the scale of political opposition to the Act. These would go on to be implemented in 2014, by which time the divergence in trends documented in this section had already arisen. The part of the ACA most relevant to diabetics pre-2012 is the provision of pre-existing condition plans (PCIPs), which had low overall enrolment. Frean, Gruber and Sommers (2017) only find modest effects of the ACA on access to health insurance in 2012-3, with larger effects in 2014-5. Even their largest estimate of increase in enrolment in 2014-5 - 10.8 percentage points for single adults - is smaller than the percentage changes at age 65 estimated in this section in the percentage of diabetic women who are insured. These changes would also apply with equal force to under-65s, and so still cannot explain the differential trends between under- and over-65s found in this subsection.

Table 10: Heart Disease Rates Among Diabetics and Non-Diabetics by Gender, Pre- and Post-2006

	Diabetics		Non-Diabetics	
	(1)	(2)	(3)	(4)
	Over 65	Under 65	Over 65	Under 65
Female	-0.0539**	-0.0569**	-0.0909***	-0.0565***
	(-2.89)	(-3.03)	(-11.31)	(-11.25)
Post-2006	0.0474**	-0.0227	0.0479***	-0.00502
	(3.25)	(-1.32)	(6.41)	(-0.84)
Female × Post-2006	-0.0457*	0.00959	-0.0142	0.0229**
	(-2.26)	(0.43)	(-1.50)	(3.18)
Constant	0.416***	0.260***	0.305***	0.127***
	(31.19)	(18.23)	(48.08)	(30.62)

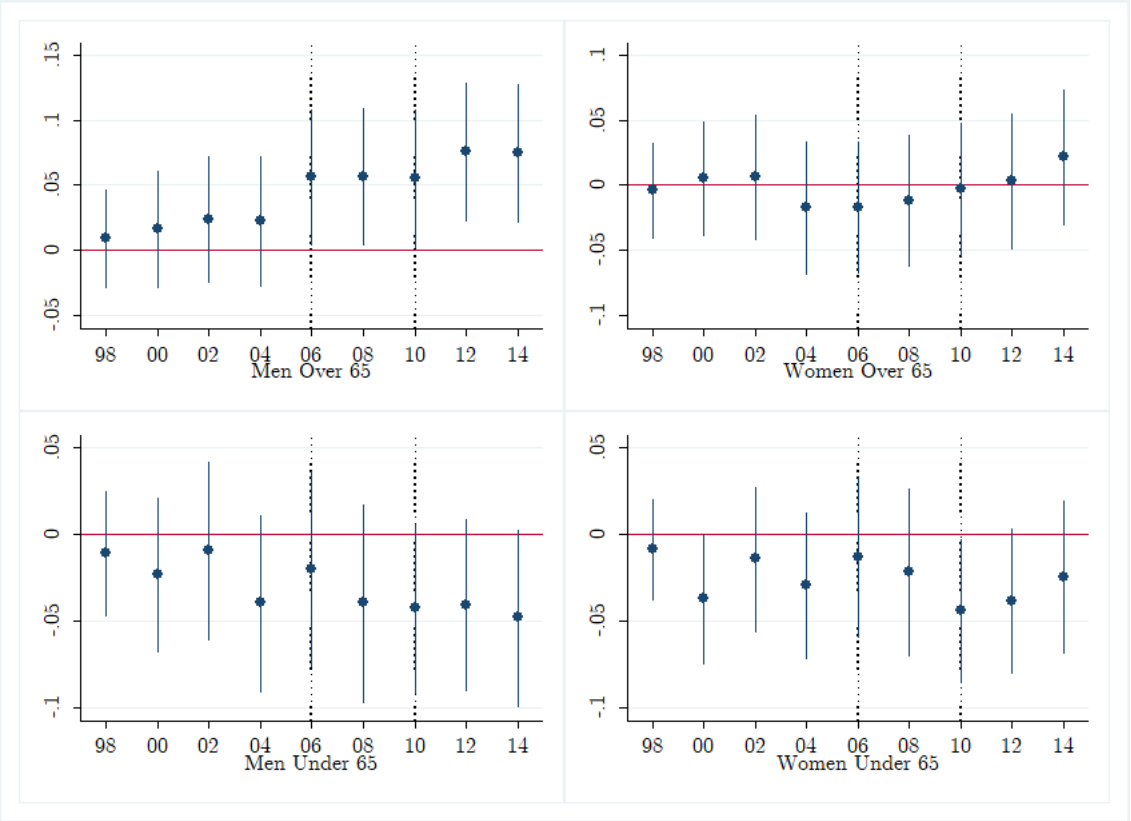
t statistics in parentheses

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

Notes: Standard errors are clustered at the individual level. Dependent variable is an indicator for whether an individual responds "Yes" to the question "Has a doctor ever told you that you had a heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems?". Medicaid recipients are excluded.

The difference in mean rates of heart disease between the genders may still be cause for concern. Kahn-Lang and Lang (forthcoming) argue that this calls counterfactuals in difference-in-differences and

Figure 2: Trends in Heart Disease Relative to 1996 Among Diabetics in the HRS (Excl. Medicaid Recipients), 1998-2014



Notes: Plots are of coefficients from pooled OLS regressions of the outcome (proportion responding "Yes" to the question "Has a doctor ever told you that you had a heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems?") on time dummies for the years 1998-2014, with 1996 as the reference category. Since there is overlap among individuals across the different waves of the survey, standard errors for these regressions are clustered at the individual level. Dotted lines correspond to major health care reforms - Medicare Part D being implemented in 2006, and the Affordable Care Act being passed in 2010. Medicaid recipients are excluded.

shift-share analyses into question, since one has to explain why the mechanism that produces (or allows for) the difference in levels does not also produce a difference in trends prior to the policy change. The counterfactual in this context is a similar upward trend for male and female diabetics' rates of heart disease over time. Some evidence that this would have occurred for diabetic women is given by columns 3 and 4 of Table 10, which provide evidence on trends in heart disease among non-diabetics. Over-65s' rates of heart disease are higher post-2006 for both genders without a differential trend, with a larger difference in levels between the sexes. This latter fact reflects previous findings that diabetes decreases women's advantage in vulnerability to heart disease (Juutilainen et al., 2004). Among non-diabetics under 65 (column 4 of Table 10) we can observe convergence in rates of heart disease post-2006 (which are unlikely to be attributable to Medicare, since they are typically ineligible). These observations together with the flat trends for both genders before 2006 (Figure 2) should strengthen our confidence that diabetic women would have shared in their male counterparts' increased propensity for heart disease after 2006 were it not for Part D.

A reduction in heart disease rates of this size is likely to have large cost savings. Using relatively conservative modelling assumptions, Barton et al. (2011) calculate that a 1% reduction in cardiovascular disease in the United Kingdom would result in cost savings to that country's National Health Service (NHS) of at least $48 million (£30 million at 2011 rates) per year in 2011 dollars. It is necessary to inflate the figures found in that study by a factor of 2 to obtain comparable numbers for the United States, since it has been found that the U. S. health care system pays roughly twice as much on average for comparable procedures to those in the rest of the OECD (Papanicolas, Woskie and Jha, 2018). The population of American diabetics over age 65 is somewhat smaller, at around 20% of 46 million over-65s with diagnosed diabetes Centers for Disease Control and Prevention (2017), while the overall population of over-65s in the United States is similar in size to the overall population of the United Kingdom. Since Medicare tends to reimburse cardiovascular procedures at relatively high rates , preventing an increase in cardiovascular disease among diabetic women of 4.6% over 8 years is likely to have saved roughly $4.6 \times 0.2 \times 2 \times \$48$ million per year $= \$88.32$ million per year in 2011 dollars over the same period.

This may yet underestimate of the net effect of lower blood sugar levels on health care costs. The HRS data may be underpowered to capture significant effects on other health outcomes, which have been found in other settings (i.e. Gilmer et al. (1997)). It is relatively easy to find larger estimates of forgone health care costs for even mild improvements in control of blood sugar levels, a broader criterion than examining changes in a specific adverse outcome such as heart disease. Using the lowest of Gilmer et al. (1997)'s estimates of $670 per person per annum in forgone health care costs of a 1 percentage point decrease in fasting blood sugar levels (at their data's average blood sugar levels) and the numbers of diabetics changing their behavior as a result of Part D, one can obtain forgone health care costs due to better control of blood sugar levels of up to $487 million per annum. Suppose we take the largest estimate of the effect of Part D of a net change of 15.8% (since I cannot reject the null hypothesis that the positive effect of Part D was twice as large as the negative "crowding out" effect). This gives 15.8 percent of the female diabetic population (23 million $\times 0.2 \times 15.8\%$ = approx. 726 800 people) each forgoing $670 per annum in health care costs from better glycaemic control, yielding approximately $487 million per

annum in forgone health care expenditures.

Given the number of people involved, providing better coverage for insulin under Part D appears to reduce health care costs by between one-tenth and one-third as much as discouraging a similar number of people from smoking cigarettes. Choi, Dave and Sabia (2016) calculate that reducing the number of cigarette smokers by 2.5 million translates into a $4.6 billion reduction in health care costs. This is equivalent to a reduction of $1.3 billion in health care costs for a population of 726 800. Since tobacco control is considered one of the most cost-effective methods of improving population health, this suggests that encouraging insulin usage to the extent that prescription drug coverage under Part D did is among the more effective methods of holding down health care costs. A broad estimate of the savings indicates that the forgone costs may be up to 36% as large as smoking cessation in an equivalently large population. Given that smoking cessation is considered among the most cost-saving measures in public health, this places subsidising insulin in the first rank of policies aimed at containing health care costs.

The return on investment in insulin in the United States has likely decreased year-on-year as the price of insulin in that country has increased dramatically between 2006 and the time of writing. The Medicare Modernization Act of 2003 prohibited the U. S. federal government from using its size as a purchaser of pharmaceuticals to bargain the prices of prescription medications downwards, as is done in most countries that provide prescription drug benefits. This component of the legislation may explain why average real annual expenditures per insulin user on insulin nearly tripled over the period 2006-2013, while average annual quantities of insulin demanded only increased by one-seventh (Hua et al., 2016). Given the cost savings calculated above, the returns to subsidizing insulin are likely to be higher in countries such as the Netherlands, where increases in the price of insulin have been less dramatic.

5.6 Summary of Findings

The empirical exercises above have three aims. First, they provide evidence that the negative crowding-out effect of insurance for treatment on prevention is strictly negative. Second, they estimate the extent to which this effect is counteracted if prevention is itself subsidized. Third, they provide evidence on the extent to which counteracting the crowding-out effect matters for health outcomes and spending on health care. The answers provided in this section were, first, that 30.5% of female insulin users, who undergo much larger changes in the proportion of uninsured individuals than males, stop using insulin when they qualify for Traditional Medicare coverage at age 65 in the pre-2006 sample. Second, this effect is either offset exactly or to the extent that there are equally large *positive* responses to qualifying for Medicare coverage post-2006, when that coverage included subsidies for insulin. Third, we can attribute a 4.6 percentage point forgone increase in heart disease among female diabetics and up to $487 million per annum in forgone health care costs to the change in behavior induced by the change in the Medicare program in 2006.

In the next section, I provide a new method for bounding the effect of qualifying for Medicare coverage by exploiting the panel dimension of the data. The bounds turn out to be relatively tight. Even when sacrificing point identification for partial identification, the negative effect of insurance for treatment on prevention remains large in absolute terms.

6 Partial Identification Via Within-Individual Variation

In this section, I will exploit the fact that I have repeated observations on the same individuals in the data to compare the same individuals' insulin usage behavior before and after turning 65. I will use a combination of economic theory and institutional variation to show that a first-differences estimator that makes this within-individual comparison is biased upwards and towards zero by age effects, and that a corollary of this is that the regression-discontinuity estimator, which makes comparisons across individuals, is biased downward and away from zero (if at all).[9]

In this section, I first discuss the first-differences estimator and why it is likely to be biased upward (i.e. towards zero if the true effect is negative). Second, I discuss finite-sample biases in the standard regression-discontinuity design and why the arguments for the first-differences estimator being biased upward imply that the regression-discontinuity estimator is biased downward (i.e. away from zero if the true effect is negative). Third, I present simulations of the main theoretical mechanisms - differential mortality and shorter horizons towards the end of life - that provide further evidence for the arguments in the first two subsections. The fourth subsection presents the first-differences results and uses them to bound the effect of qualifying for Medicare coverage at age 65.

It will be useful to define the hypothetical outcomes Y_{it} for a given individual i in a period t when covered for treatment (denoted $D_{it} = 1$) and not covered (denoted $D_{it} = 0$), as

$$Y_{it} = \beta_0 + \delta X_{it} + \zeta t + \eta_i + v_{it} \text{ if } D_{it} = 0, \tag{7}$$

$$Y_{it} = (\beta_0 + \beta_1) + \delta X_{it} + \zeta t + \eta_i + v_{it} \text{ if } D_{it} = 1, \tag{8}$$

where X_{it} is a vector of covariates including age, R_{it}, t is a linear time trend, η_i is individual-specific time-invariant unobserved heterogeneity, v_{it} is an idiosyncratic shock in period t. From this we can derive the equation in levels

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \delta X_{it} + \zeta t + \eta_i + v_{it}, \tag{9}$$

and assuming individuals stay in the program once eligible, so that we can only have $\triangle D_{it} = 1$ or $\triangle D_{it} = 0$, the equations in first differences

$$\triangle Y_{it} = (\beta_1 + \zeta) + \delta \triangle X_{it} + \triangle v_{it} \text{ if } \triangle D_{it} = 1, \tag{10}$$

$$\triangle Y_{it} = \zeta + \delta \triangle X_{it} + \triangle v_{it} \text{ if } \triangle D_{it} = 0, \tag{11}$$

[9] In a previous version of this paper, I also estimated adjusted confidence intervals for the RDD results according to the method of Kolesár and Rothe (2018). The results (available on request) are robust to this modification for reasonable values of the tuning parameter required by their method. A key advantage of using the current method relative to both this method and "bias-corrected" estimators such as that of Calonico, Cattaneo and Titiunik (2014) is that each of these either explicitly or implicitly makes some assumption regarding the curvature of the "true" conditional expectation function (CEF) (see Armstrong and Kolesár (2018) for a proof). The method presented in this section eschews making any such assumption but forgoes point identification by doing so.

from which we can derive the equation

$$\triangle Y_{it} = \zeta + \beta_1 \triangle D_{it} + \delta \triangle X_{it} + \triangle v_{it}. \tag{12}$$

The assumption that allows both the equation in levels and the equation in differences to recover the same underlying effect is that the true difference in behavior between the covered and uncovered states is a fixed proportion. This assumption comes is supported by the fact that the change in agents' constraints is anticipated in advance, so that there is no shift of the lifetime participation profile either side of the age threshold of interest (cf. Section 7 and/or MaCurdy (1981)), and so the movement along that profile recovered by the first-differences estimator is the same as the movement along the participation profile measured using the cross-sectional differences between individuals either side of the age threshold.

6.1 Positive Bias in the First-Differences Estimator

In this subsection, I will show that the first-differences estimator in my setting is biased upward and towards zero for a negative effect of interest. This will be for two reasons. First, the "treatment" group of those who switch status to being covered by Medicare at age 65 is younger than the "control" group of those who are continuously insured by Medicare in both adjacent periods, who are 67 or older since the Health and Retirement Study surveys are taken every two years. Second, given that the "control" group is older, they are less likely to begin using insulin, and more likely to stop using insulin if they are already using it. This will be because a greater proportion of the older group will be approaching the end of life, and so will have progressively smaller returns to investing in their health as they approach the end of their investment horizon (compare the model in Grossman (1972)). In the next subsection, I will show that both this argument and an argument from differential mortality will imply that the regression-discontinuity estimator is biased in the opposite direction, downward and away from zero.

Since I have repeated observations of the same individuals, I can estimate Equation 12, with an additional sample selection criterion that will buttress the arguments in this section,

$$\triangle Y_{it} = \zeta + \beta_1 \triangle D_{it} + \delta \triangle X_{it} + \triangle v_{it} \text{ if } \triangle D_{it} = 0 \implies R_{it} \geq \bar{R},$$

where $\triangle D_{it}$ is equal to 1 if individual i qualified for coverage under Medicare in period t (and only 0 otherwise, since no one who becomes eligible ages out of the program), $\triangle X_{it}$ is a vector of first-differenced time-varying individual characteristics, $\triangle v_{it}$ is an idiosyncratic error term, and the condition $\triangle D_{it} = 0 \implies R_{it} \geq \bar{R}$ implies that the only individuals with $\triangle D_{it} = 0$ are continuously insured by Medicare at ages 67 or older (where R_{it}, age, exceeds the age of eligibility, $\bar{R}$). The effect of qualifying for coverage on the change in the probability of using insulin between the uncovered and the covered period is β_1, which is identified if and only if

$$E[\triangle v_{it} | \triangle D_{it}] = 0 \text{ (Strict exogeneity)}, \tag{13}$$

We can obtain relatively tight bounds on the effect of qualifying for Medicare by restricting the sample so that violations of strict exogeneity are likely to create bias in the opposite direction of finite-sample

bias for the regression-discontinuity estimator. If we restrict the sample to individuals who have just aged into the program (who have $\triangle D_{it} = 1$) and individuals who are older than them who are continuously enrolled in the program (so that $\triangle D_{it} = 0$), then the "treatment group" is younger than the "control group". On the other hand, regression-discontinuity designs will compare individuals for whom $D_{it} = 1$ with individuals younger than them with $D_{it} = 0$ so that the "treatment group" is older than the "control group". It is therefore likely that if we can establish the sign of the bias for one of these estimators, the bias for the other estimator will be of the opposite sign.[10]

Suppose that the true conditional expectation functions (CEF) for the trends in insulin usage $\triangle Y_{it}$ among those who switch status ($\triangle D_{it} = 1$) and those who do not ($\triangle D_{it} = 0$) is given by (omitting covariates for simplicity):

$$E[\triangle Y_{it} | \triangle D_{it} = 0] = \zeta, \tag{14}$$

$$E[\triangle Y_{it} | \triangle D_{it} = 1] = \zeta + \beta_1, \tag{15}$$

then

$$\beta_1 = E[\triangle Y_{it} | \triangle D_{it} = 1] - E[\triangle Y_{it} | \triangle D_{it} = 0]; \tag{16}$$

Now consider that $\triangle Y_{it} \in \{-1, 0, 1\}$ for all i, t. Let n^T be the number of individuals for which $\triangle D_{it} = 1$, n^C the number of individuals for which $\triangle D_{it} = 0$, n_k^T, n_k^C the number of individuals with $\triangle Y_{it} = k$ for $\triangle D_{it} = 1$ and $\triangle D_{it} = 0$ respectively. Then it is easy to show that

$$\beta_1 = \left(\frac{n_1^T - n_{-1}^T}{n^T} \right) - \left(\frac{n_1^C - n_{-1}^C}{n^C} \right), \tag{17}$$

and the sample counterpart of β_1, $\widehat{\beta}_1$, is obtained by replacing n^T, n^C, n_k^T, n_k^C with their sample counterparts (for each value of k in the latter two cases). Hence any discrepancy between the sample counts of these groups and the counts that would be obtained from an experiment with true random sampling will lead to bias in the first-differences estimator. Intuitively, the first-differences estimator compares the numbers of those who initiate insulin therapy between ages 65-67 after reporting not using it in the last wave with the number of those who cease using insulin at ages 65-67 after having used it in the last wave of the survey. This net initiations minus cessations figure is then compared with the same figure for the "control" group, the rest of the sample. If the control groups figure for initiations - cessations is not a consistent estimator for that of the treatment group's counterfactual, strict exogeneity is violated and the estimator is biased overall.

If the bias is unambiguously signed, this will be because of an unambiguously signed difference in the number of individuals at the margin of switching their insulin usage status for whom $\triangle D_{it} = 1$ and

[10] Note that it is not logically impossible for the biases to be of the same sign. If decisions exhibit negative serial correlation, then higher probabilities of an outcome in a given period also produce higher probabilities of reversal in the subsequent period. In this case, equations in levels and equations in differences may still exhibit opposite biases if the treatment and control groups are the same for both. The evidence from both the medical literature (Brown et al., 1999, Rajagopalan et al., 2003) and the Appendix, where I estimate bounds for the effect of past decisions on present decisions, indicates that decisions are positively serially correlated in this context.

those for whom $\triangle D_{it} = 0$. If those with $\triangle D_{it} = 0$ do not have the same distribution of unobserved heterogeneity as those who have just qualified and have $\triangle D_{it} = 1$, then they are a control group that differs systematically from the treatment group for reasons that are not related to being treated (provided with coverage). Rewriting β_1 as the difference in means between groups, we obtain

$$\beta_1 = \left(\frac{n_1^T}{n^T} - \frac{n_1^C}{n^C}\right) - \left(\frac{n_{-1}^T}{n^T} - \frac{n_{-1}^C}{n^C}\right), \tag{18}$$

so that the bias is unambiguously positive if and only if we underestimate the proportion of individuals who would have begun insulin in the absence of qualifying for Medicare $\left(\frac{n_1^C}{n^C}\right)$ and overestimate the number of individuals who would have stopped insulin therapy in the same counterfactual $\left(\frac{n_{-1}^C}{n^C}\right)$. In this case $\widehat{\beta}_1$ will be an underestimate of β_1 if $\beta_1 < 0$ and an overestimate if $\beta_1 > 0$. This would imply that the control group for whom $\triangle D_{it} = 0$ has a weaker trend towards insulin usage than in the true counterfactual, which therefore leads us to underestimate the strength of the departure from trend upon qualifying for Medicare.

I argue that since those with $\triangle D_{it} = 0$ are older than 67 by construction, these conditions are likely to hold.

Individuals who have already begun insulin therapy will have weaker incentives to continue using insulin over time as they approach the end of life. In the limit, total certainty about not surviving to the next period reduces the return to investing in health to zero (Grossman, 1972).[11] The difference between those with $\triangle D_{it} = 0$ and a randomly assigned control group is that the first-differences estimator's "control" group has a spuriously negative trend (or spuriously weak positive trend) and this will lead to an underestimate of the true counterfactual trend.[12] I will assume that the effects of insulin on lengthening life are second-order (as evidence in the literature seems to suggest (Weinger and Beverly, 2010)). Thus the first-order effect of aging within insulin adherents will be to reduce the return to using insulin. This is because as individuals approach the end of life, the expected return from using insulin in the future decreases (Grossman, 1972, 2000).[13]

It may be argued that insurance for treatment also reduces mortality, and so increases individuals' horizons and hence their propensity to invest in their health. There are two problems with this argument. First, since Medicare is a long-expected change rather than an unexpected change, it will already have been built into individuals' subjective mortality expectations, and hence will not increase life *expectancy*

[11] This effect of expected lifespan on investment in health is termed the "Mickey Mantle effect" in Fang et al. (2007) and Khwaja (2009).

[12] Results from first-differences regressions that include individuals younger than 65 with $\triangle D_{it} = 0$ are available on request from the author. In line with the arguments in this subsection, the results are more negative than in the regressions that restrict the sample to those who change status and older individuals who have $\triangle D_{it} = 1$.

[13] In fact, we require that the net present value of investing in health decreases at an increasing rate towards the end of a finite lifespan. This is true under relatively mild conditions in human capital models (Grossman, 2000), and follows from the fact that investment in a single period raises the stock of human capital in every subsequent period. If there is only one period in which to reap a return on investment in health capital, this is larger than the return of zero from investment in health in the last period of life, and smaller than the return from investment with two periods left, and so forth. A proof of this argument (Proposition 3) can be found at the end of the Appendix.

per se. Second, the effects of insurance on mortality in the literature tend to be small. The largest such effects are found within a small window of Medicare eligbility by Card, Dobkin and Maestas (2009), and that paper finds a reduction in one-week mortality of between 0.8 and 1 percentage points for a subset of 12% of their sample. It seems unlikely that this extends individuals' horizons more generally to the extent that it has a first-order effect on their decisions regarding investment in health.

I formalize these ideas in the following Proposition:

Proposition 1. Suppose that $\left(\frac{n_1^T - n_{-1}^T}{n^T}\right)$ is estimated consistently by its sample counterpart but that

$$\frac{\hat{n}_1^C}{\hat{n}^C} < \frac{n_1^C}{n^C} \text{ and } \frac{\hat{n}_{-1}^C}{\hat{n}^C} > \frac{n_{-1}^C}{n^C},$$

then $\widehat{\beta}_1 > \beta_1$, implying $E[\triangle v_{it} | \triangle D_{it} = 0] < E[\triangle v_{it} | \triangle D_{it} = 1]$, hence that the first-differences estimator $\widehat{\beta}_1$ is biased upward due to a failure of strict exogeneity.

Proof: Write

$$\begin{aligned}
\widehat{\beta}_1 &= \left(\frac{\hat{n}_1^T - \hat{n}_{-1}^T}{\hat{n}^T}\right) - \left(\frac{\hat{n}_1^C - \hat{n}_{-1}^C}{\hat{n}^C}\right) \\
&= \left\{\left(\frac{n_1^T - n_{-1}^T}{n^T}\right) - \left(\frac{n_1^C - n_{-1}^C}{n^C}\right)\right\} + \left\{\left(\frac{n_1^C}{n^C} - \frac{\hat{n}_1^C}{\hat{n}^C}\right) - \left(\frac{n_{-1}^C}{n^C} - \frac{\hat{n}_{-1}^C}{\hat{n}^C}\right)\right\} \\
&> \left(\frac{n_1^T - n_{-1}^T}{n^T}\right) - \left(\frac{n_1^C - n_{-1}^C}{n^C}\right) = \beta_1,
\end{aligned}$$

where the last line follows from the assumptions that $\frac{\hat{n}_1^C}{\hat{n}^C} < \frac{n_1^C}{n^C}$ and $\frac{\hat{n}_{-1}^C}{\hat{n}^C} > \frac{n_{-1}^C}{n^C}$,. Hence $E[\widehat{\beta}_1 - \beta_1] = E[\triangle v_{it} | \triangle D_{it} = 1] - E[\triangle v_{it} | \triangle D_{it} = 0] > 0$, so the bias in the estimators of $\frac{n_1^C}{n^C}$ and $\frac{n_{-1}^C}{n^C}$ can be explained as a violation of strict exogeneity■.

Hence selection on unobservables works to bias the first-differences estimator towards zero. Intuitively, the counterfactual trend towards insulin usage is underestimated for the treatment group that qualifies for coverage at age 65. I exploit this in the rest of the paper in order to bound the effect of qualifying for Medicare coverage between the first-differences results and the cross-sectional RDD results. The reason for the difference in the sign of the bias for the two methods, given the same institutional variation, is that the "treatment" group in the RDD regression is older than the "control" group (right-hand side vs. left-hand side of the cutoff age of 65), whereas the "treatment" group in the first-differences regressions (those who switch status, aged 65-66) is younger than the "control" group (individuals continuously insured by Medicare, 67 or older).

6.2 Negative Finite-Sample Bias in the Regression Discontinuity Estimator

In practice regression-discontinuity designs rely on comparing average outcomes for individuals "close" to the cutoff, within some bandwidth h. A recent literature has focused on how best to choose this bandwidth so as to optimally trade off finite-sample bias and inefficiency in the resulting estimator (Calonico, Cattaneo and Titiunik, 2014, Imbens and Kalyanaraman, 2012, Armstrong and Kolesár, 2017). If one focuses on individuals too close to the cutoff, one risks having too few individuals to have sufficient power to reject the null hypothesis, especially with a discrete running variable which has only so many individuals at each integer value (Kolesár and Rothe, 2018). Examine individuals too far away from the cutoff, and there is a greater risk that non-comparability due to unobserved traits drives differences in average outcomes rather than the discontinuity of interest.

Suppose we have finite-sample bias in the regression discontinuity design. This will result in estimators biased downward and away from zero in this context. Since the running variable is age, individuals far enough to the right of the cutoff in a given cross-section will have had to survive for longer to be observed than those far to the left of the cutoff. It has been known for some time that estimating life-cycle profiles with cross-sectional data will be subject to bias from "differential mortality" - those who survived had characteristics that produce spurious patterns in the data that are likely to differ from the actual life-cycle profiles we could construct if we could observe the same individuals' actions over time (Attanasio and Hoynes, 2000, De Nardi, French and Jones, 2010). In the current context, healthier diabetics are more likely to survive and less likely to use insulin (since it is a preventive measure "of last resort"). If we could see which diabetics ought to be using insulin as a subset of diabetics (which I cannot see in the data), we could distinguish non-users from non-adherents. Since we cannot do this, non-users are more likely to be healthier and to survive for longer than insulin users. (*Among the subset of diabetics who are recommended insulin*, by contrast, those who adhere to insulin will have fewer health complications than non-adherents). Hence as we move to the right of the cutoff age of 65, we encounter individuals who are less and less likely to benefit from using insulin due to the relative mildness of their disease, and so we will spuriously estimate a lower average for insulin usage among this group, biasing the estimator downward and away from zero.

Moreover, the argument from the previous subsection regarding individuals who are closer to the end of life being less likely to start or continue using insulin applies to the proportion using insulin as well as the trend. The latter was used to argue that the bias of the first-differences estimator was upward and toward zero, but here leads to bias in the opposite direction, which reinforces the effect of differential mortality on the negative bias of the estimator. The reason for this is that in the case of the first-differences estimator, the "treatment" group are individuals aged 65-66 who have switched status, and the "control" group comprises older individuals who are continuously insured by Medicare. By contrast, the regression-discontinuity estimator, which examines outcomes in levels, compares a "treatment" group of individuals older than 65 with a "control" group of individuals younger than 65. As a result, the same argument for why the first-differences estimator is biased towards zero is also an argument for why the RDD estimator

is biased away from zero, since the two estimators have the opposite ranking for whether the "treatment" or the "control" group is older.

With a discrete running variable, we are caught between an RDD estimator with large efficiency losses due to the small sample size when restricting attention to observations close to the cutoff, and an RDD estimator biased downward and away from zero when using a wider bandwidth. By contrast, if the estimator based on variation within individuals over time is biased upward and towards zero (as argued above), and still gives statistically significant results, then we can bound the size of the effect of gaining coverage at age 65.

In sum: older individuals need to survive to be observed; this will bias the levels of their insulin usage towards zero, hence the effect of Medicare away from zero, since being healthier they will have lower usage on average. Hence I will assume that the finite-sample bias of the RDD is such that

$$E[\eta_i|R_{it} \geq \bar{R}] < E[\eta_i|R_{it} < \bar{R}], \tag{19}$$

In the following subsection, I separately simulate the effects of the two mechanisms cited in this and the previous subsection (differential mortality and shrinking horizons towards the end of life). It turns out that the biases in the two estimators have the expected signs argued for in this and the previous subsection.

6.3 Monte Carlo Evidence: Biases in First-Differences and Regression Discontinuity Estimators

This subsection simulates the effects of differential mortality and approaching the end of life on the regression discontinuity and first-differences estimators. The effects on the biases of the estimators are as they should be according to the arguments in the preceding two subsections.

I simulate data from 1000 draws for each bandwidth in the set $\{1,2,3,...,10\}$ of a Standard Normal random variable labeled $\varepsilon_i \sim N(0,1)$. This is used to draw the initial distribution of disutilities of usage for a cohort of 50 individuals, who start all at the same age. Higher values of ε_i will correspond to higher disutilities of usage and therefore lower likelihoods of usage. I then generate a panel of 21 time periods for these individuals with ε_i constant in each period, but with $R_{it} = t - 11$ - so that the "age" running variable is normalized to 0;

$$Y_{it} = 1[\varepsilon_i - \frac{R_{it}}{20} \leq 1.645] \text{ if } R_{it} < 0 \text{ (equiv. } t < 10); \tag{20}$$

$$Y_{it} = 1[\varepsilon_i - \frac{R_{it}}{20} \leq 0] \text{ if } R_{it} \geq 0 \text{ (equiv. } t \geq 10); \tag{21}$$

so we have an outcome that shifts in expectation by 0.45 at the threshold value of $R_{it} = 0$ (since 1.645 is the 95^{th} percentile of the standard Normal distribution). The first set of simulations use this set-up to estimate the sharp regression discontinuity design with a local linear regression

$$Y_{it} = \beta_0 + \beta_1 1[R_{it} \geq 0] + \beta_2 R_{it} + \beta_3 R_{it} \times 1[R_{it} \geq 0] + \eta_i + v_{it} \text{ if } |R_{it}| < h \tag{22}$$

where $h \in \{1,2,3,...,10\}$, η_i is time-invariant unobserved individual-level heterogeneity due to unobserved permanent differences in the disutilities of usage, and v_{it} is functional form misspecification error, which results from estimating a linear probability model when the true model is a Probit.

In the second specification, I introduce differential mortality by ε_i, so that individuals with lower disutilities of usage (lower ε_i) are selected out of the sample to the right of the cutoff of $R_{it} = 0$. To ensure that the selection process does not affect the true effect of crossing the threshold of $R_{it} = 0$, I restrict the process to individuals who have $R_{it} \geq 1$. "Deaths" are drawn according to the Accelerated Failure Time model, where T_i is the last period of life for individual i,

$$lnT_i = \theta + \varepsilon_i; \tag{23}$$

I set $\theta = 3.5$. The results are displayed in Figure 3. I do not display confidence intervals since I have enough observations in the simulations for relatively small confidence intervals even at smaller bandwidths, thus abstracting away from the efficiency problems that arise in this paper. The local linear estimates without selection are on average the value of the true change in probabilities at the cutoff. When individuals with lower disutilities of usage get selected out of the sample, the RDD estimator is progressively more negatively biased the wider the bandwidth used. Note that the magnitude of the biases depends not only on the rapidity of differential selection but also the distribution of reservation utilities. A distribution of reservation utilities that is more densely concentrated so that a minority of individuals accounts for the majority of cases where $Y_{it} = 1$ will result in larger biases if that minority attrites from the sample than in the case where reservation utilities are Uniformly distributed. Figure 4 shows the empirical counterpart of Figure 3. As expected, estimates produced with observations from within 24 months of the cutoff are closer to zero than those obtained with larger bandwidths. One difference between the figures is that the estimates from the HRS data are asymptotic to roughly -0.5, which is not predicted by the simpler framework used in the Monte Carlo experiments.

The second mechanism is that while diabetics will be more likely to require insulin over time, as they approach the end of life they will also have shorter horizons and hence lower returns from using insulin. In order to examine the effect of this mechanism on first-differences and regression-discontinuity estimates, I change the latent index model equations to

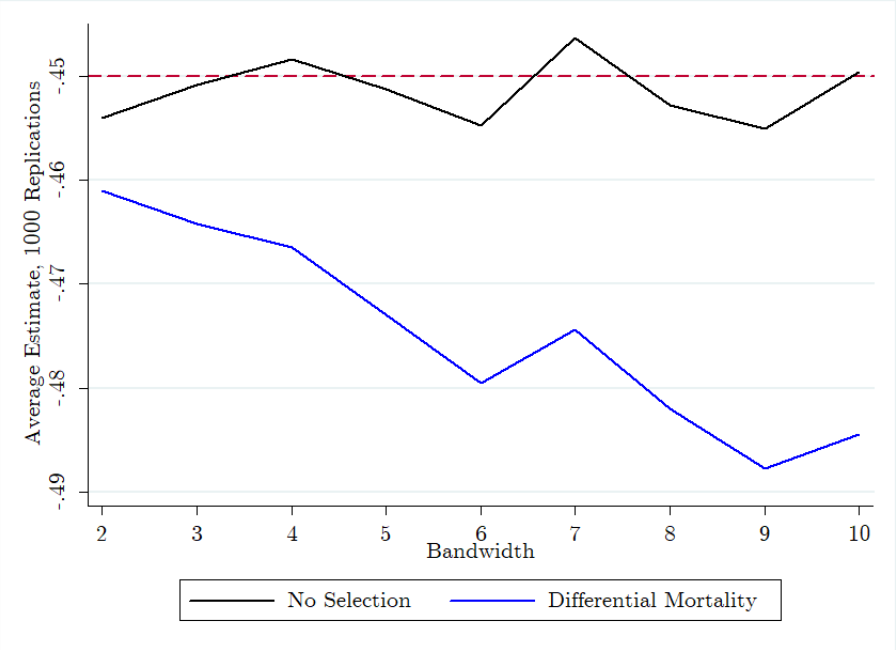
$$Y_{it} = 1[\varepsilon_i - \frac{R_{it}}{20} + \frac{R_{it}^2}{10} \leq 1.645] \text{ if } R_{it} < 0; \tag{24}$$

$$Y_{it} = 1[\varepsilon_i - \frac{R_{it}}{20} + \frac{R_{it}^2}{10} \leq 0] \text{ if } R_{it} \geq 0; \tag{25}$$

In order to examine the effect of this mechanism in isolation, I no longer make use of differential sample selection according to individual-specific mortality. I also introduce results from estimating the first-differences specification

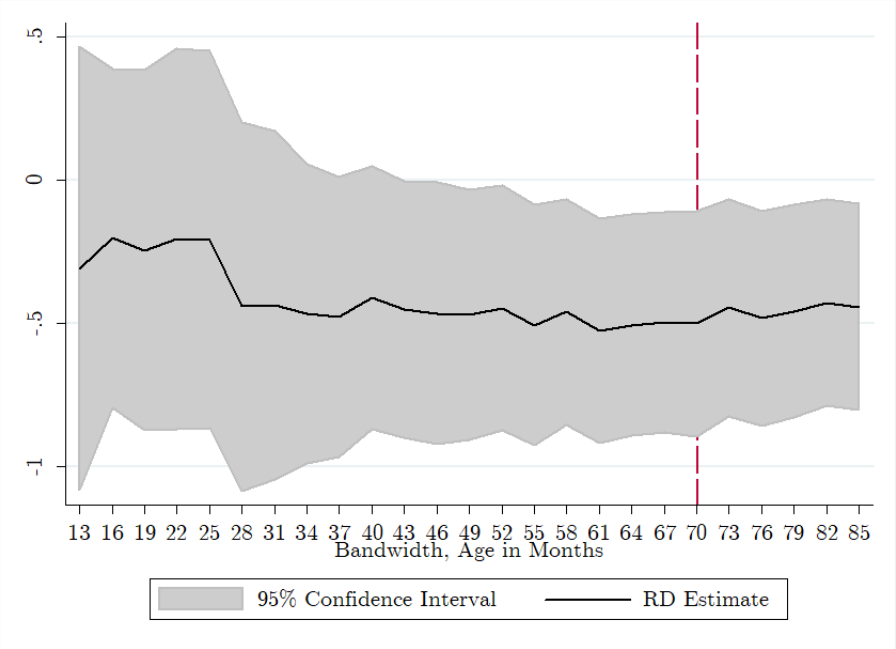
$$\triangle Y_{it} = \beta_1 \triangle 1[R_{it} \geq 0] + \beta_2 \triangle R_{it} + \beta_3 \triangle R_{it} \times 1[R_{it} \geq 0] + \triangle v_{it}, \tag{26}$$

Figure 3: Simulated Effects of Differential Mortality on Bias in RDD Estimates



Notes: Lines connect average estimates of $\widehat{\beta}_1$ for Equation 22 where $\beta_1 = -0.45$ for 1000 simulated processes of aging and extensive-margin usage under differential sample selection. Individuals with higher disutilities of usage are more likely to stay in the sample. Both lines correspond to a local linear regression with a Uniform kernel, using the bandwidth displayed on the x-axis, but the blue line corresponds to simulations where differential sample selection is present and the black line to simulations where it is absent.

Figure 4: RD Estimates for Effect of Medicare Eligibility on Insulin Usage by Bandwidth, 1998-2004 - Original 1998 Cohort

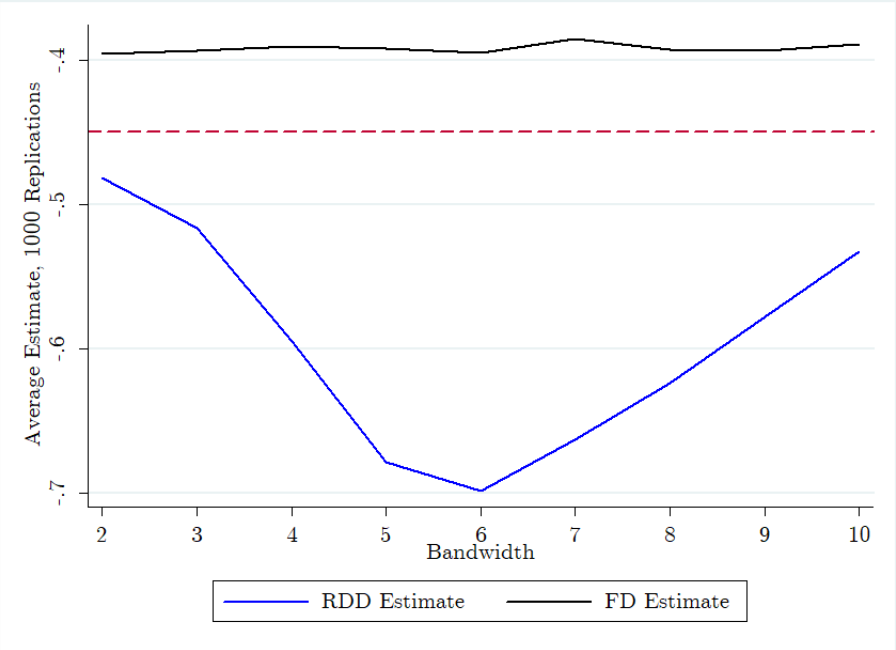


Notes: All fuzzy regression-discontinuity results are obtained excluding those on Medicaid, Medicare Advantage or Medicare supplemental coverage (Medigap), use a Uniform kernel with local linear regression, and include time dummies, health status, and education fixed effects. Standard errors are clustered within individuals. Red dashed line corresponds to the optimal bandwidth of 70.07.

while restricting the sample to $R_{it} \geq -1$. Note that this condition guarantees that the idiosyncratic error $\triangle v_{it}$ is different for individuals who are in the "treatment" group versus the "control" group, since the first difference of some i.i.d. functional form misspecification error will differ for observations around the cutoff of $R_{it} = 0$ and the set of observations for which $0 < R_{it} \leq 10$ (compare Figure 6). The results are displayed in Figure 5. As argued in the preceding subsections, the first-differences estimator is biased upward and towards zero. The RDD estimator is biased downward and away from zero over the range of the running variable considered in the simulations, but the bias appears to decrease for bandwidths wider than $h = 6$. This shows that the relative strength of differential mortality and the within-person horizon effect matter for the extent of the bias in the RDD estimator (and potentially whether it is even biased downward at all). The reason the bias starts decreasing at higher bandwidths when there is no differential mortality is that if the "hump-shaped" life-cycle profile of usage is symmetric about the cutoff, then widening the bandwidth will result in the inclusion of a greater number of zeroes to the left as well as the right of the cutoff (see Figure 6). For the purposes of this paper it seems reasonable to assume that the pattern seen to the left of $h = 6$ in Figure 3 holds over the range of bandwidths considered, since the combined effects of differential mortality and shorter horizons to the right of the cutoff are unlikely to be matched by similarly rapid changes at younger ages.

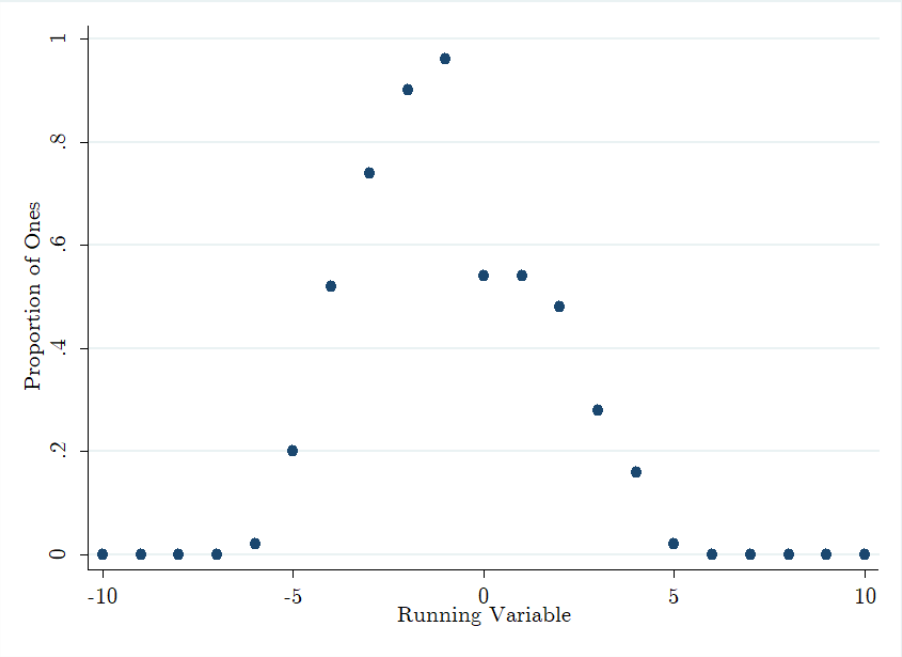
In the next subsection, I estimate first-differences equations to provide a lower bound for the effect of qualifying for Medicare coverage.

Figure 5: Simulated Effects of Shrinking Horizons on Bias in RDD and FD Estimates



Notes: Lines connect average estimates of $\widehat{\beta}_1$ for Equations 22 (blue) and 26 (black) where $\beta_1 = -0.45$ for 1000 simulated processes of aging and extensive-margin usage. Both lines correspond to a local linear regression with a Uniform kernel, using the bandwidth displayed on the x-axis, but the true conditional expectation function (CEF) is quadratic since within a given individual, the disutility of usage first rises, and then falls with "age". The blue line corresponds to simulations of a local linear regression with a Uniform kernel and the bandwidth listed on the x-axis. The black line corresponds to simulations that use the same simulated data to estimate a model in first-differences which only includes data for $R_{it} \geq -1$ that would recover the same underlying population parameter as the regression-discontinuity estimates if both estimators were unbiased.

Figure 6: RDD Graph, One Draw of Simulated Shrinking Horizons, Quadratic Linear Index Model



Notes: The figure represents binned data by simulated "age" from a simulated model where the disutility of "usage" first rises, then falls with "age". The discontinuity at the centered value of the "age" variable is -0.45 by construction. The true conditional expectation function (CEF), as can be seen in the figure, is quadratic.

6.4 First-Differences Results

In this subsection, I present the empirical results from first-differences regressions that allow me to obtain partial identification even if point identification fails. The estimating equation becomes the empirical counterpart of Equation 26, augmented to allow pre- and post-2006 effects to differ,

$$\triangle Y_{it} = \zeta + \beta_1 \triangle 1[R_{it} \geq \bar{R}] + \beta_2 1[t \geq 2006] + \beta_3 \triangle 1[R_{it} \geq \bar{R}] \times 1[t \geq 2006] + \delta \triangle X_{it} + \triangle v_{it}, \qquad (27)$$

The sample is restricted to those 63 and over to ensure that the "control group" of those continuously insured by Medicare is older than the "treatment group" of those who change status at ages 65-66 (since the HRS survey is taken every two years). $\triangle X_{it}$ includes changes in health status, earnings, work status, marital status, and choice of supplemental insurance (Medigap) or choice of Medicare+Choice/Medicare Advantage/Medicare HMO. The results are displayed in Table 11. As expected, the estimated effects are smaller than those estimated using a regression-discontinuity design, but are nonetheless statistically significant. Common to both sets of results is the failure to find significant changes in usage of oral medication. The smallest among these results implies a 4.4 percentage point decline following qualifying for Medicare (note that the first-differences results are on a different scale compared to the fuzzy RD

results, since they make direct comparisons within individuals before and after qualifying for Medicare and so control flexibly for other changes in available insurance at 65 as well). If we are especially conservative and use the average usage of 34 percentage points between ages 60 and 64 from the original 1998 cohort (see Appendix), then this translates into a relative difference of 12.9% between those just under and just over 65 in insulin usage. This produces the lower bound quoted in the Introduction. Reassuringly, the effect of Part D is similar to the one reached from the RDD results, and inspection appears to indicate that the magnitude of the relative offset is also similar. Formally, a Wald test cannot reject the null hypotheses that $\beta_1 = -\beta_3$ or that $\beta_1 = -2\beta_3$ in this context either.

Table 11: First-Differences Results, Diabetic Women: Insulin and Oral Medication Usage

	(1)	(2)	(3)
△Insulin			
$\widehat{\beta}_1$	-0.04*	-0.04*	-0.06*
	(-2.26)	(-2.27)	(-2.23)
$\widehat{\beta}_3$	0.08**	0.08**	0.10**
	(2.99)	(2.92)	(2.93)
△Oral Medication			
$\widehat{\beta}_1$	0.04	0.04	0.05
	(1.44)	(1.39)	(1.34)
$\widehat{\beta}_3$	0.01	0.01	-0.01
	(0.34)	(0.29)	(-0.27)

t statistics in parentheses

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

Notes: Standard errors are clustered at the individual level. Sample is restricted to individuals 63 and older. Column (1) reports estimated pre-post differences from specifications without covariates; Column (2) reports results including time dummies; Column (3) reports results including first-differences in work status, health status (an indicator equal to 1 if an individual reports being in "Fair" or "Poor" health, marital status, and enrolment in either Medicare Advantage/a Medicare HMO or supplementary coverage under Medicare (Medigap).

To summarize: this section has provided a method for obtaining bounds on the effect of qualifying for Medicare coverage on insulin usage. Implementing this method results in a lower bound of a 4.4 percentage point decrease pre-Part D, and an upper bound of a 7.9 percentage point decrease. The relative offset of this negative effect due to Part D subsidies for insulin post-2006 is similar in magnitude in both cases. In the next section, I provide a simple theoretical model that allows me to interpret the results in

this paper and reconcile them with the rest of the literature on *ex ante* moral hazard in health behaviors.

7 Theoretical Framework: A Model of the Intertemporal Allocation of Prevention

In this section, I introduce a model where an agent chooses how to allocate her lifetime expenditures among consumption, prevention *ex ante*, and spending on non-preventive medical services. It turns out that all that is needed for the model to be able to rationalize the large effect sizes in this paper is a sufficiently strong degree of substitutability between prevention and other medical spending. I conclude the section by drawing out the model's implications for which individuals are at the margin in this setting (and hence the distributional impact of the crowding-out effect) as well as its implications for the magnitudes of quantities that are not investigated in this paper, particularly the income elasticity of demand for prevention.

Consider a two-period model where an agent decides between the allocation of her expenditures between consumption C_1 in period 1, medical services M_1 in period 1, and their period 2 counterparts, as well as continuous amounts of prevention ϕ_1, ϕ_2. Her lifetime utility that results from her choices is

$$U(C_1) + V(M_1, \phi_1) + \beta\{U(C_2) + V(M_2, \phi_2)\} \tag{28}$$

where the sub-utility functions satisfy the usual conditions. Denote, by (for example) V_ϕ the derivative of $V(.)$ with respect to ϕ in a given period (so that the sub-utility functions are the same for both periods). I will assume the derivatives of $V(M_t, \phi_t)$ with respect to ϕ_t have the following signs for $t = 1, 2$:

$$V_\phi > 0; \tag{29}$$

$$V_{\phi\phi} < 0; \tag{30}$$

$$V_{\phi M} < 0; \tag{31}$$

ϕ, prevention, has two roles: first, it is intrinsically valuable (i.e. for its role in producing health), and exhibits diminishing marginal utility as do the other arguments of the objective function. Second, the third derivative shows that the marginal utility of medical services is lower when ϕ is higher. This captures the fact that the demand for medical services, and the share of the budget spent on medical services relative to consumption, are lower when the agent is in better health.[14] Setting the interest rate equal to zero for simplicity's sake, the lifetime budget constraint with initial assets A_1 is

$$C_2 + P^M M_2 + P^\phi \phi_2 = A_1 - C_1 - P^M M_1 - P^\phi \phi_1; \tag{32}$$

[14] Note that this is because I use "prevention" interchangeably with "health investments" in this paper. Screenings for conditions can also be referred to as "prevention", but would have the opposite sign for the third derivative in this setting since screenings are complementary to health expenditures: it makes more sense to purchase medical services when one is aware of a condition than unaware, and being better able to purchase treatment makes the return to *information* regarding one's eventual health status higher.

To obtain the elasticity of intertemporal substitution of prevention with respect to the price of treatment P^M, which will give agents' willingness to substitute prevention away from periods when the price of treatment is low and toward periods when the price of treatment is high, we have to hold the marginal utility of wealth (here denoted by μ) constant. This is because agents will move along their lifetime profile of treatment prices which, having been anticipated in advance, involves executing planned changes in prevention efforts conditional on the agent's lifetime resources. This means taking $\frac{\partial \mu}{\partial P^M} = 0$ when implicitly differentiating the first-order condition with respect to ϕ_2. Doing this yields the simple expression for the (Frisch) elasticity of intertemporal substitution

$$\varepsilon^F_{\phi_2,P^M} \equiv \left(\frac{P^M}{\phi_2}\right)\left(\frac{\partial \phi_2}{\partial P^M}\right)\Bigg|_{\frac{\partial \mu}{\partial P^M}=0} = -\left(\frac{P^M}{\phi_2}\right)\left(\frac{V_{\phi M}\left(\frac{\partial M_2}{\partial P^M}\right)}{V_{\phi\phi}}\right), \tag{33}$$

which is positive, since:[15]

$\left(\frac{\partial M_2}{\partial P^M}\right) < 0$ due to the Law of Demand;

$V_{\phi\phi} < 0$ due to diminishing marginal returns to prevention;

$V_{\phi M} < 0$ since prevention *ex ante* and treatment *ex post* are substitutes (more prevention lowers the marginal utility of treatment).

It follows that $\varepsilon^F_{\phi_2,P^M} > 0$: an anticipated fall in the price of medical services unambiguously decreases the incentives to use prevention, and an anticipated rise in the price of medical services unambiguously increases the incentives to use prevention.

Since almost all Americans expect to be entitled to Medicare upon turning 65, the marginal utility of wealth does not change when they qualify for Medicare. Medicare RDD papers therefore estimate this elasticity, which is unambiguously positive (all else equal), unlike the experimental studies which estimate the Marshallian elasticity. Similarly, the main empirical exercises in this paper estimate an extensive-margin elasticity of intertemporal substitution. This illustrates an advantage of using a Medicare RDD rather than an experiment to study this behavior. Note that, as per footnote 15, this change in the timing of preventive efforts is a *lower* bound for the "long-run" or "lifetime" change in the use of prevention, which is measured by the Hicks elasticity.

[15] In contrast with the elasticity of intertemporal substitution of labor supply, which is larger than the Hicks elasticity, in this case the intertemporal substitution response is *smaller* than the Hicks elasticity, which is $\left(\frac{P^M}{\phi_2}\right)\left(\frac{P^\phi U_{CC}\left(\frac{\partial C_2}{\partial P^M}\right) - V_{\phi M}\left(\frac{\partial M_2}{\partial P^M}\right)}{V_{\phi\phi}}\right) > 0$ since $U_{CC}, V_{\phi\phi} < 0$ and $\left(\frac{\partial C_2}{\partial P^M}\right) > 0$ since consumption and medical spending are substitutes. For more details on these elasticities' relative magnitudes in the labor supply context, see Keane (2011). The reason for this difference is that in the labor supply model, "hours worked" are a "bad" rather than a "good", and so the utility function has to be convex with respect to hours of work, whereas here investment in health is a "good", and so $V_{\phi\phi} < 0$. If we flip the sign of $V_{\phi\phi}$ in the elasticities in this paper, we obtain that the elasticity of intertemporal substitution is larger than the Hicks just as in the labor supply case. The intuition is that for labor supply, if lifetime consumption possibilities can be altered by working more hours, diminishing marginal utility of consumption will dampen the response to an increase in the wage since higher consumption is traded for the "bad" of less leisure time. This dampening effect is not present if the marginal utility of wealth is constant. By constrast, in this case, prevention expands lifetime consumption possibilities (by depressing the marginal utility of treatment) and is also valued in itself, and so there is even more reason to use it if the price of treatment rises and the marginal utility of wealth is not held constant.

The usual interpretation of the elasticity of intertemporal substitution provides some intuition for its role in this context. Individuals move along their life-cycle consumption profiles each period according to their planned allocation of their lifetime resources across different periods. In the labor supply context, the elasticity of intertemporal substitution measures the proportional planned increase in hours worked for a proportionally higher wage relative to other periods with proportionally lower wages (since wages vary over the life cycle). In this context, the elasticity measures the response of the planned division between spending on prevention in the current period versus spending on treatment in a subsequent period to anticipated variation in the price of treatment over the life cycle. Over the life cycle, individuals will want to allocate more prevention to periods in which the price of treatment is high and less to those in which the price of treatment is low. Given some estimate of how the price of treatment differs in the Medicare-eligible portion of the life cycle versus the uninsured portion of the life cycle, we can estimate how much variation over the life cycle in prevention spending individuals are willing to undertake to track the price of treatment. Since moving across the Medicare eligibility age is anticipated by almost all United States residents, it is this elasticity for which Medicare RDD results are relevant.[16]

The simple expression for the Frisch elasticity reconciles previous results and commentary on the literature on prevention with the results found in this paper. The size of the effect depends on the relative sizes of the second derivative of utility with respect to ϕ itself - or the rate at which marginal utility from health diminishes - and the strength of the link between health investments and medical expenses, $V_{\phi M}$. The usual explanation for why $\varepsilon^{F}_{\phi_2,P^M}$ may be small is that $V_{\phi\phi}$ is large - individuals are highly risk averse with respect to their health (Cutler and Zeckhauser, 2000, Kenkel, 2000). In this paper, even if this holds, this can be outweighed by a sufficiently tight link between medical expenses and health

[16]One reason this link might fail is if the population in question cannot borrow to smooth expenditures across periods. Then consumption will track changes in the budget constraint each period, instead of changes in the lifetime budget constraint (Deaton, 1992). In that case, Medicare RDD papers would also estimate a Marshallian elasticity, income effects and all, rather than an intertemporal elasticity of substitution. Whether the estimates in this paper underestimate or overestimate the true elasticity of intertemporal substitution depends on two opposing mechanisms. (In a similar vein, Keane and Wolpin (2001) argue that the presence of liquidity constraints has led to underestimates of the elasticity of intertemporal substitution in consumption). First, the income effect increases prevention at the age of Medicare eligibility, offsetting the negative cross-price effect. Second, liquidity constraints strengthen the precautionary motive Deaton (1992), since agents need to have a larger buffer against consumption fluctuations if they cannot borrow in a crisis. Therefore agents who are unable to borrow against future income will have a stronger motivation to use prevention, and exhibit larger decreases in prevention in the face of an exogenous decrease in risk. Hence it is ambiguous whether the true effect of Medicare eligibility on prevention is over- or under-estimated due to the presence of credit constraints. The argument that the responses in this paper are Marshallian elasticities rather than elasticities of intertemporal substitution creates more of a puzzle, since it implies both that the estimators in this paper recover the same elasticity as in the RAND or Oregon health insurance experiments, and yet that the extent of *ex ante* moral hazard in those cases is significantly smaller.

Smoothing consumption across covered and uncovered periods requires access to a store of liquid wealth that can be used to finance more expensive medical care in the uncovered period. To examine the evidence for this, I calculate the amount of liquid wealth held by households in which uninsured diabetic women reside at ages 60-64. The Health and Retirement Study surveys contain a large number of questions regarding pension wealth, assets, and debts. To proxy for a lack of access to liquidity, I calculate the number of uninsured diabetic women aged 60-64 in each wave of the HRS between 1998 and 2006 who live in households with neither housing wealth that could serve as collateral for a loan nor a positive amount of non-housing financial wealth. I use the definition of non-housing financial wealth in the RAND HRS data, which comprises stocks, bonds, checking accounts and certificates of deposit minus debts, and excludes the value of IRAs, Keogh plans, real estate, business wealth, and vehicles (Chien et al., 2013). 18% of this group are liquidity constrained according to this definition. If this forces consumption to track income instead of being smoothed over the age of 65, then we should see evidence of deviations from the permanent income hypothesis in other categories of expenditure as well. In the Appendix, I report regression discontinuity results for the non-medical expenditures on durable goods of households inhabited by female diabetics who are not insured by Medicaid. I am unable to reject the null hypothesis that non-medical durable consumption does not change discontinuously at age 65.

investments, hence greater substitutability between prevention *ex ante* and treatment *ex post* (i.e. $V_{\phi M}$ is large in magnitude), as exists among diabetics.[17]

This model is useful for its ability to shed light on heterogeneous effects of changing the price of medical services in a subsequent period on current-period incentives to use prevention (here, as in the rest of this paper, used as a synonym for investment in health capital *ex ante*). Agents whose intrinsic utility from preventive care is more concave (larger $V_{\phi\phi}$) have weaker preventive care responses to the price of medical services. That is, agents that are more risk-averse with respect to their future health have smaller responses in their preventive behavior to a reduction in the price of medical services. This effect is counteracted by the relative effectiveness of prevention in reducing demand for future medical services in the second period, $V_{\phi M}$. The larger this term is - and so the tighter is the link between prevention and future demand for medical services - the larger is the reduction in prevention for a given fall in the price of future medical services.

This framework allows us to reconcile previous findings of small changes in health investment behavior due to the provision of insurance and the large effects found in this paper. Though smokers and binge drinkers may be less risk-averse with respect to their future health status than others (and so have smaller $V_{\phi\phi}$), the link between their behavior and their future medical expenses at the margin at age 65 is likely to be relatively weak (and so they also have smaller $V_{\phi M}$). The link between investment in health and future medical expenses is much stronger for diabetics than most other subsets of the population, and so given that $V_{\phi M}$ is relatively large for this group, we also see larger responses to a reduction in the price of medical services in preventive behavior in this group.

We can also use this model to address a previous explanation for the weak *ex ante* moral hazard effects found in the literature. Cutler and Zeckhauser (2000) and Kenkel (2000) pointed out that even if agents are insured against the financial losses of illness, they are in general not insured against the expected utility losses of ill health. This is captured in the model by a higher risk aversion over health (larger $V_{\phi\phi}$) leading to a smaller response of prevention to the price of medical services. The previous theoretical explanation in this model corresponds to high risk aversion with respect to health across individuals. This model shows that if the connection between investment in health and eventual medical expenses is particularly strong (high $V_{\phi M}$), as it is for diabetics to a far greater extent than for the general population, this previous explanation can still be valid for the broader near-elderly population without ruling out large responses of the kind found in this paper.

In sum, a simple two-period model that introduces prevention as a choice variable that affects the marginal utility from medical services can reconcile the following observations: (1) Agents receive utility from medical services, and demand them when sick; (2) Even given the value that they place on medical services, if prevention reduces their future demand for medical services, lowering the price of medical

[17] One alternative to the approach in this section to reconciling the twin facts that individuals both seek medical care when ill and choose to increase their probability of falling ill in the future is to use hyperbolic discounting or, in the limit, the model proposed in Banerjee and Mullainathan (2010). In a version of that model derived by the author and considered for this paper, individuals care about their current health and so seek medical care, but not their future health, and so spend their income on consumption rather than prevention. The model presented in its stead has the advantages of (i) more parsimonious assumptions, (ii) a straightforward ability to link the empirical results in this paper with those in the rest of the literature, and (iii) quantitative predictions for behavioral responses that have yet to be studied.

services will reduce the incentive to use prevention; (3) The stronger the link between prevention and future demand for medical services, the stronger is the crowding-out effect.

We can therefore explain the weak effects found in previous studies, as well as why diabetics are a subset of the population among whom we would expect to find strong effects, with a simple two-period model of investment in health.[18]

I close this section with two further remarks that may be of use for future work. First, if $V_{\phi M}$ varies across individuals, this framework shows that the responsiveness of prevention to the price of medical services is strongest among those individuals with the strongest link between prevention and their medical expenses. From a policy perspective, this means that the crowding out of prevention by coverage for treatment is greatest among those individuals whose medical expenses are likeliest to increase to a large extent as a result. The adverse effects of coverage are concentrated among precisely the individuals that a policymaker would least want to discourage from using prevention. The model in this section therefore allows me to make qualitative statements regarding the marginal individuals for whom prevention is crowded out by insurance for treatment.

Second, this model can also show how Marshallian elasticities of prevention with respect to insurance that are exactly zero can coexist with substantial magnitudes for the Hicks and Frisch elasticities. Note that by the Slutsky equation, the Marshallian derivative $\left(\frac{\partial \phi_2}{\partial P^M}\right)^M$ can be written

$$\begin{aligned}
\left(\frac{\partial \phi_2}{\partial P^M}\right)^M &= \left(\frac{\partial \phi_2}{\partial P^M}\right)^H - \frac{\partial \phi_2}{\partial A_1} M_2, \\
\implies \frac{P^M}{\phi_2}\left(\frac{\partial \phi_2}{\partial P^M}\right)^M &= \frac{P^M}{\phi_2}\left(\frac{\partial \phi_2}{\partial P^M}\right)^H - \frac{P^M}{\phi_2}\left(\frac{A_1}{A_1}\right)\frac{\partial \phi_2}{\partial A_1} M_2, \\
\implies \varepsilon^M_{\phi_2, P^M} &= \varepsilon^H_{\phi_2, P^M} - \frac{P^M M_2}{A_1}\varepsilon_{\phi_2, A_1},
\end{aligned}$$

so that the Marshallian elasticity is equal to the Hicks elasticity less the product of the income elasticity of prevention $\varepsilon_{\phi_2, A_1}$ and future medical expenses' share of lifetime wealth $\frac{P^M M_2}{A_1}$. This is analogous to the case of the static labor supply model, except instead of a term depending on the ratio of labor income to lifetime wealth, we have a term depending on the ratio of future medical expenses to lifetime wealth. The main difficulty in recovering the Hicks elasticity from a given Marshall elasticity then comes from estimating $\frac{P^M M_2}{A_1}$ and $\varepsilon_{\phi_2, A_1}$. For the sake of argument, suppose the Marshallian elasticity is 0, so that we can recover $\varepsilon^H_{\phi_2, P^M} = \frac{P^M M_2}{A_1}\varepsilon_{\phi_2, A_1}$. Banks et al. (2016) estimate that by age 70, medical expenses are on average 20% of household spending in the United States, which gives $\frac{P^M M_2}{A_1} = 0.2$.[19] There is as yet no

[18] If we had data that allowed us to calculate the magnitude of the average fall in the price of insulin for the marginal diabetic upon qualifying for Medicare post-2006, we could also use the model to rationalize the relative sizes of the cross-price effect due to the price of treatment falling at age 65 and the own-price effect due to the price of insulin falling at age 65 in the post-Part D era.

[19] Since the 20% budget share of health care pertains to over-70s, one objection to its use is that it is an overestimate of the ratio of medical expenses to *lifetime* income. As a robustness check, I examine the distribution of the ratio of out-of-pocket medical expenses to household income for female diabetics under 65 not enrolled in Medicaid. The mean share is 23%, with an upper quartile of 16%. If I use the smaller figure of 16% for the calculations above, I obtain a Hicks elasticity of 0.512, only slightly above the upper bound for the Frisch elasticity (which is nonetheless consistent with the argument that the larger estimates are from

consensus on the income elasticity of prevention, but if we take the upper end of estimates of the income elasticity for dental care - assuming that most dental care is preventive rather than palliative - from a survey of the literature by Getzen (2000), we obtain $\varepsilon_{\phi_2,A_1} = 3.2$ from Silver (1972). This gives a Hicks elasticity of $\varepsilon^H_{\phi_2,P^M} = 0.2 \times 3.2 = 0.64$. This is a much larger Hicks elasticity of health investments with respect to the price of health care than the majority of studies of labor supply find for the response of hours worked to wages (Keane, 2011). Failure to reject the null hypothesis of no *ex ante* moral hazard effect of health insurance on health investments in an experimental setting is consistent with large *ex ante* moral hazard effects after taking the income effect of providing coverage into account.

To relate the estimates in this paper to the Frisch elasticity of prevention with respect to the price of treatment, first take the estimates which find between a 12.9% and 30.5% relative decrease in the proportion of insulin users.[20] We need to find the expected relative percentage decrease for a 1% difference in the price of health care when 65 relative to pre-65. Medicare Parts A and B typically cover between 60% and 80% of beneficiaries' health care costs. So the results in this paper imply that $\varepsilon^F_{\phi_2,P^M}$ is in the interval $[\frac{0.129}{0.8}, \frac{0.305}{0.6}] = [0.161, 0.508]$. In line with the theoretical model presented in this section, the elasticity of intertemporal substitution is smaller than the Hicks elasticity (see footnote 15, above). This comes with the caveats that the two estimates do not come from the same data, and I do not present a consistent estimator of the income elasticity of insulin in this paper. The lower bound for the Frisch elasticity of 0.161 also allows for the possibility that the Hicks elasticity is smaller than 0.64 while maintaining the result that the Marshallian elasticity is zero. It therefore also tells us that, holding the budget share of future medical expenses constant, the smallest income elasticity of demand for insulin consistent with the results in this paper, assuming a Marshallian elasticity of zero, is $\frac{0.161}{0.8 \times 0.2} = 1.01$. Hence we need insulin to be a luxury good on average to be consistent with the results found in this paper. This is greater than the midpoint of the preferred range for the income elasticity of healthcare between 0 and 1.5 used by Finkelstein et al. (2012). There are nonetheless good reasons to believe this number is larger for the subcategory of preventive medicine than for health care overall; individuals' ability to delay usage is much greater for prevention than for treatment, for example. In addition, if we could observe which diabetics were Type I diabetics - who need insulin to survive - as opposed to Type IIs, the vast majority, we could separately estimate income elasticities of demand for the two subgroups. The income elasticity for Type I diabetics with respect to insulin is likely to be significantly lower than for Type IIs for the reasons cited above.

an estimator biased away from zero). The minimal income elasticity consistent with my empirical results increases to 1.26, which is still within the range of elasticities considered by Finkelstein et al. (2012). This is subject to the caveat that current income may systematically differ from permanent income to different extents over the life cycle (Haider and Solon, 2006).

[20] The 12.9% figure comes from dividing the smaller first-differences result of a 4.4 percentage point decrease by the original cohort 60-64 average insulin usage of 34 percentage points, yielding a conservative estimate of the effect.

8 Conclusion

The effect of insurance on health outcomes and behaviors depends on what is covered and to what extent. This paper has provided evidence that insuring individuals against health risks can worsen those risks to a larger extent than previously thought. Before 2006, Medicare Parts A and B insured 60-80% of previously uninsured female diabetics' health care costs, but did not subsidize insulin. As a result, between 12.9% and 30.5% of the insulin users in this group would forego using insulin in the month of Medicare eligibility, when their risk of incurring large medical expenses would drop discontinuously. This paper also provides an organizing theoretical framework to explain why previous studies have encountered more difficulty with recovering *ex ante* moral hazard in health behaviors. When the connection between a health behavior and health care costs is weaker, the underlying *ex ante* moral hazard effect is smaller. Individuals with the strongest responses will "hide in the herd" in data which contain a large number of individuals with weak responses. In addition, if the insurance is provided via randomized assignment, income effects are likely to be even larger for prevention than for other forms of health care, which further masks the negative effect of insurance for treatment on prevention.

The main policy implication of this paper's findings is that policymakers have underestimated the need for stronger incentives to use preventive care. In a universal health care system, it is nearly impossible - in the absence of invoking strong assumptions regarding counterfactual behaviour - to see the effect of coverage on prevention since there is no control group. In the United States, universal coverage for those above 65 and non-universal coverage for those below 65 provides some evidence on how coverage can crowd out prevention at the margin. Though many universal health care systems spend more on preventive services on average than public health programs in the United States, the evidence provided in this paper suggests that the level of this spending may be an underestimate of what is necessary to minimise overall health care costs. This is due to the unseen crowding out effect which cannot be estimated in a universal health care system since there is no "control" group of uninsured individuals.

The results in this paper suggest at least two avenues of future research. First, other studies that use dynamic models - as in, say, the labor supply literature - can use discontinuities to identify intertemporal spillover effects in behavior. For example, Fu and Gregory (2017) use discontinuities in an institutional rule together with a structural model to identify spatial externalities of rebuilding after Hurricane Katrina. This paper's results allow the identification of similar spillovers - but within individuals, instead of across them - without using a structural model. In the "fuzzy RDD" case, both structural and reduced-form approaches would have to deal with the finite-sample issues raised by having to instrument for closely related changes such as a treatment indicator and its lags. Less difficulty in applying the identification arguments in this paper is likely to be found in sharp regression discontinuity designs, which are not estimated by two-stage least squares and hence are not vulnerable to weak identification. Second, this paper does not estimate the Marshall or Hicks elasticities of prevention with respect to the price of health care. With rich enough data, one could estimate all of the Marshall, Hicks and Frisch elasticities using a single data set. This would be an advance over comparisons of these elasticities across papers since

differences across papers may also be due to unknown differences in data or estimation choices.

References

Aiyagari, S Rao. 1994. "Uninsured Idiosyncratic Risk and Aggregate Saving." *Quarterly Journal of Economics*, 109(3): 659–684.

Alpert, Abby. 2016. "The Anticipatory Effects of Medicare Part D on Drug Utilization." *Journal of Health Economics*, 49: 28–45.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *The Review of Economic Studies*, 58(2): 277–297.

Armstrong, Timothy B, and Michal Kolesár. 2017. "A Simple Adjustment for Bandwidth Snooping." *The Review of Economic Studies*, 85(2): 732–765.

Armstrong, Timothy B, and Michal Kolesár. 2018. "Optimal Inference in a Class of Regression Models." *Econometrica*, 86(2): 655–683.

Attanasio, Orazio P, and Hilary Williamson Hoynes. 2000. "Differential Mortality and Wealth Accumulation." *Journal of Human Resources*, 1–29.

Baltagi, Badi. 2008. *Econometric Analysis of Panel Data.* John Wiley & Sons.

Banerjee, Abhijit, and Sendhil Mullainathan. 2010. "The Shape of Temptation: Implications for the Economic Lives of the Poor." Unpublished manuscript.

Banks, James, Richard Blundell, Peter Levell, and James P Smith. 2016. "Life-Cycle Consumption Patterns at Older Ages in the US and the UK: Can Medical Expenditures Explain the Difference?" National Bureau of Economic Research.

Barcellos, Silvia Helena, and Mireille Jacobson. 2015. "The Effects of Medicare on Medical Expenditure Risk and Financial Strain." *American Economic Journal: Economic Policy*, 7(4): 41–70.

Barrett-Connor, Elizabeth, Elsa-Grace V Giardina, Anselm K Gitt, Uwe Gudat, Helmut O Steinberg, and Diethelm Tschoepe. 2004. "Women and Heart Disease: the Role of Diabetes and Hyperglycemia." *Archives of Internal Medicine*, 164(9): 934–942.

Barton, Pelham, Lazaros Andronis, Andrew Briggs, Klim McPherson, and Simon Capewell. 2011. "Effectiveness and Cost Effectiveness of Cardiovascular Disease Prevention in Whole Populations: Modelling Study." *British Medical Journal*, 343: d4044.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119(1): 249–275.

Blau, David M, and Donna B Gilleskie. 2006. "Health Insurance and Retirement of Married Couples." *Journal of Applied Econometrics*, 21(7): 935–953.

Blau, David M, and Donna B Gilleskie. 2008. "The Role of Retiree Health Insurance in the Employment Behavior of Older Men." *International Economic Review*, 49(2): 475–514.

Boland, Beth. 1998. "The Evolution of Best-in-Class Pharmacy Management Techniques." *Journal of Managed Care Pharmacy*, 4(4): 366–373.

Bommer, Christian, Esther Heesemann, Vera Sagalova, Jennifer Manne-Goehler, Rifat Atun, Till Bärnighausen, and Sebastian Vollmer. 2017. "The Global Economic Burden of Diabetes in Adults Aged 20–79 Years: a Cost-of-Illness Study." *The Lancet Diabetes & Endocrinology*, 5(6): 423–430.

Brown, Jonathan Betz, Gregory A Nichols, Harry S Glauber, and Alan Bakst. 1999. "Ten-Year Follow-Up of Antidiabetic Drug Use, Nonadherence, and Mortality in a Defined Population with Type 2 Diabetes Mellitus." *Clinical Therapeutics*, 21(6): 1045–1057.

Buchmueller, Thomas, Richard W Johnson, and Anthony T Lo Sasso. 2006. "Trends in Retiree Health Insurance, 1997–2003." *Health Affairs*, 25(6): 1507–1516.

Caetano, Carolina, Gregorio Caetano, and Juan Carlos Escanciano. 2017. "Over-Identified Regression Discontinuity Design." Unpublished manuscript.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica*, 82(6): 2295–2326.

Card, David, Carlos Dobkin, and Nicole Maestas. 2008. "The Impact of Nearly Universal Insurance Coverage on Health Care Utilization and Health: Evidence from Medicare." *American Economic Review*, 98(5): 2242–2258.

Card, David, Carlos Dobkin, and Nicole Maestas. 2009. "Does Medicare Save Lives?" *Quarterly Journal of Economics*, 124(2): 597–636.

Carroll, Christopher D, Karen E Dynan, and Spencer D Krane. 2003. "Unemployment Risk and Precautionary Wealth: Evidence from Households' Balance Sheets." *Review of Economics and Statistics*, 85(3): 586–604.

Cellini, Stephanie Riegg, Fernando Ferreira, and Jesse Rothstein. 2010. "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design." *Quarterly Journal of Economics*, 125(1): 215–261.

Centers for Disease Control and Prevention. 2017. "National Diabetes Statistics Report, 2017." *Atlanta, GA: Centers for Disease Control and Prevention, US Dept of Health and Human Services*.

Chetty, Raj. 2008. "Moral Hazard Versus Liquidity and Optimal Unemployment Insurance." *Journal of Political Economy*, 116(2): 173–234.

Chetty, Raj, and Adam Szeidl. 2007. "Consumption Commitments and Risk Preferences." *Quarterly Journal of Economics*, 122(2): 831–877.

Chien, Sandy, Nancy Campbell, Orla Hayden, et al. 2013. *RAND HRS Data Documentation, Version M.* Santa Monica, CA: RAND Center for the Study of Aging.

Choi, Anna, Dhaval Dave, and Joseph J Sabia. 2016. "Smoke Gets in Your Eyes: Medical Marijuana Laws and Tobacco Use." Unpublished manuscript.

Christian-Herman, Jennifer, Matthew Emons, and Dorothy George. 2004. "Effects of Generic-Only Drug Coverage in a Medicare HMO." *Health Affairs*, 23: 455–468.

Clark, Robert L, and Olivia S Mitchell. 2014. "How does Retiree Health Insurance Influence Public Sector Employee Saving?" *Journal of Health Economics*, 38: 109–118.

Cubanski, Juliette, and Patricia Neuman. 2007. "Status Report on Medicare Part D Enrollment in 2006: Analysis of Plan-Specific Market Share and Coverage." *Health Affairs*, 26(1): w1–w12.

Curcuru, Stephanie, John Heaton, Deborah Lucas, and Damien Moore. 2010. "Heterogeneity and Portfolio Choice: Theory and Evidence." *Handbook of Financial Econometrics*, 337–382.

Cutler, David M, and Richard J Zeckhauser. 2000. "The Anatomy of Health Insurance." In *Handbook of Health Economics*. Vol. 1, 563–643. Elsevier.

Dave, Dhaval, and Robert Kaestner. 2009. "Health Insurance and Ex Ante Moral Hazard: Evidence from Medicare." *International Journal of Health Care Finance and Economics*, 9(4): 367.

Deaton, Angus. 1992. *Understanding Consumption.* Oxford University Press.

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

De Preux, Laure. 2011. "Anticipatory Ex Ante Moral Hazard and the Effect of Medicare on Prevention." *Health Economics*, 20(9): 1056–1072.

Dong, Yingying, and Arthur Lewbel. 2015. "Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models." *Review of Economics and Statistics*, 97(5): 1081–1092.

Duggan, Mark, and Fiona Scott Morton. 2010. "The Effect of Medicare Part D on Pharmaceutical Prices and Utilization." *American Economic Review*, 100(1): 590–607.

Eeckhoudt, Louis, Christian Gollier, and Harris Schlesinger. 1996. "Changes in Background Risk and Risk Taking Behavior." *Econometrica*, 683–689.

Eliasson, Björn. 2003. "Cigarette Smoking and Diabetes." *Progress in Cardiovascular Diseases*, 45(5): 405–413.

Engen, Eric M, and Jonathan Gruber. 2001. "Unemployment Insurance and Precautionary Saving." *Journal of Monetary Economics*, 47(3): 545–579.

Fang, Hanming, and Alessandro Gavazza. 2011. "Dynamic Inefficiencies in an Employment-Based Health Insurance System: Theory and Evidence." *American Economic Review*, 101(7): 3047–77.

Fang, Hanming, Michael Keane, Ahmed Khwaja, Martin Salm, and Dan Silverman. 2007. "Testing the Mechanisms of Structural Models: The Case of the Mickey Mantle Effect." *American Economic Review*, 97(2): 53–59.

Feir, Donna, Thomas Lemieux, and Vadim Marmer. 2016. "Weak Identification in Fuzzy Regression Discontinuity Designs." *Journal of Business & Economic Statistics*, 34(2): 185–196.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group. 2012. "The Oregon Health Insurance Experiment: Evidence from the First Year." *Quarterly Journal of Economics*, 127(3): 1057–1106.

Frean, Molly, Jonathan Gruber, and Benjamin D Sommers. 2017. "Premium Subsidies, the Mandate, and Medicaid Expansion: Coverage Effects of the Affordable Care Act." *Journal of Health Economics*, 53: 72–86.

French, Eric, and John Bailey Jones. 2011. "The Effects of Health Insurance and Self-Insurance on Retirement Behavior." *Econometrica*, 79(3): 693–732.

Fu, Chao, and Jesse Gregory. 2017. "Estimation of an Equilibrium Model with Externalities: Post-Disaster Neighborhood Rebuilding." Unpublished manuscript.

Gelman, Andrew, and Guido Imbens. 2018. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics*, 1–10.

Getzen, Thomas E. 2000. "Health Care is an Individual Necessity and a National Luxury: Applying Multilevel Decision Models to the Analysis of Health Care Expenditures." *Journal of Health Economics*, 19(2): 259–270.

Gilmer, Todd P, Patrick J O'connor, Willard G Manning, and William A Rush. 1997. "The Cost to Health Plans of Poor Glycemic Control." *Diabetes Care*, 20(12): 1847–1853.

Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy*, 80(2): 223–255.

Grossman, Michael. 2000. "The Human Capital Model." In *Handbook of Health Economics*. Vol. 1, 347–408. Elsevier.

Guerrieri, Veronica, and Guido Lorenzoni. 2017. "Credit Crises, Precautionary Savings, and the Liquidity Trap." *Quarterly Journal of Economics*, 132(3): 1427–1467.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica*, 69(1): 201–209.

Haider, Steven, and Gary Solon. 2006. "Life-Cycle Variation in the Association Between Current and Lifetime Earnings." *American Economic Review*, 96(4): 1308–1320.

Ho, Kate, and Adam M Rosen. 2015. "Partial Identification in Applied Research: Benefits and Challenges." Unpublished manuscript.

Hua, Xinyang, Natalie Carvalho, Michelle Tew, Elbert S Huang, William H Herman, and Philip Clarke. 2016. "Expenditures and Prices of Antihyperglycemic Medications in the United States: 2002-2013." *Journal of the American Medical Association*, 315(13): 1400–1402.

Hubbard, R Glenn, Jonathan Skinner, and Stephen P Zeldes. 1995. "Precautionary Saving and Social Insurance." *Journal of Political Economy*, 103(2): 360–399.

Hurd, Michael, Susann Rohwedder, Joanna Carroll, Joshua Mallett, and Colleen McCullough. 2015. "RAND CAMS Data Documentation." *Version D, Release*, 2: 11.

Imai, Susumu, and Michael P Keane. 2004. "Intertemporal Labor Supply and Human Capital Accumulation." *International Economic Review*, 45(2): 601–641.

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies*, 79(3): 933–959.

Johnson, Richard W, Karen E Smith, and Owen Haaga. 2013. "How Did the Great Recession Affect Social Security Claiming?" *Program on Retirement Policy Brief*, 37.

Joyce, Geoffrey F, Dana P Goldman, Pinar Karaca-Mandic, and Yuhui Zheng. 2007. "Pharmacy Benefit Caps and the Chronically Ill." *Health Affairs*, 26(5): 1333–1344.

Juutilainen, Auni, Saara Kortelainen, Seppo Lehto, Tapani Rönnemaa, Kalevi Pyörälä, and Markku Laakso. 2004. "Gender Difference in the Impact of Type 2 Diabetes on Coronary Heart Disease Risk." *Diabetes Care*, 27(12): 2898–2904.

Kahn-Lang, Ariella, and Kevin Lang. forthcoming. "The Promise and Pitfalls of Differences-in-Differences: Reflections on "Sixteen and Pregnant" and Other Applications." *Journal of Business & Economic Statistics*.

Kaplan, Greg, and Giovanni L Violante. 2014. "A Model of the Consumption Response to Fiscal Stimulus Payments." *Econometrica*, 82(4): 1199–1239.

Kautzky-Willer, Alexandra, Jürgen Harreiter, and Giovanni Pacini. 2016. "Sex and Gender Differences in Risk, Pathophysiology and Complications of Type 2 Diabetes Mellitus." *Endocrine Reviews*, 37(3): 278–316.

Keane, Michael P. 2011. "Labor Supply and Taxes: A Survey." *Journal of Economic Literature*, 49(4): 961–1075.

Keane, Michael P. 2016. "Life-Cycle Labour Supply with Human Capital: Econometric and Behavioural Implications." *The Economic Journal*, 126(592): 546–577.

Keane, Michael P, and Kenneth I Wolpin. 2001. "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment." *International Economic Review*, 42(4): 1051–1103.

Kenkel, D. 2000. "Prevention." In *Handbook of Health Economics*. Vol. 1, , ed. Anthony J. Culyer and Joseph P. Newhouse, Chapter 31, 1675–1720. Elsevier, North Holland.

Khwaja, Ahmed. 2009. "Investment in Human Capital, Longevity and Moral Hazard in a Stochastic Life-Cycle Model of Demand for Health." Unpublished manuscript.

Kolesár, Michal, and Christoph Rothe. 2018. "Inference in Regression Discontinuity Designs with a Discrete Running Variable." *American Economic Review*, 108(8): 2277–2304.

Lee, David S, and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature*, 48(2): 281–355.

Maciosek, Michael V, Ashley B Coffield, Thomas J Flottemesch, Nichol M Edwards, and Leif I Solberg. 2010. "Greater Use of Preventive Services in US Health Care Could Save Lives at Little or No Cost." *Health Affairs*, 29(9): 1656–1660.

MaCurdy, Thomas E. 1981. "An Empirical Model of Labor Supply in a Life-Cycle Setting." *Journal of Political Economy*, 89(6): 1059–1085.

Manski, Charles F. 1990. "Nonparametric bounds on treatment effects." *The American Economic Review*, 80(2): 319–323.

Manski, Charles F. 1997. "Monotone Treatment Response." *Econometrica*, 1311–1334.

Manski, Charles F, and John V Pepper. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica*, 68(4): 997–1010.

McGuire, Thomas G, Joseph P Newhouse, and Anna D Sinaiko. 2011. "An Economic History of Medicare Part C." *Milbank Quarterly*, 89(2): 289–332.

Menke, Andy, Sarah Casagrande, Linda Geiss, and Catherine C Cowie. 2015. "Prevalence of and Trends in Diabetes among Adults in the United States, 1988-2012." *Journal of the American Medical Association*, 314(10): 1021–1029.

Nathan, David M, John B Buse, Mayer B Davidson, Ele Ferrannini, Rury R Holman, Robert Sherwin, and Bernard Zinman. 2009. "Medical Management of Hyperglycemia in Type 2 Diabetes: A Consensus Algorithm for the Initiation and Adjustment of Therapy: a Consensus Statement of the

American Diabetes Association and the European Association for the Study of Diabetes." *Diabetes Care*, 32(1): 193–203.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica*, 49: 1417–1426.

Papanicolas, Irene, Liana R Woskie, and Ashish K Jha. 2018. "Health Care Spending in the United States and Other High-Income Countries." *Journal of the American Medical Association*, 319(10): 1024–1039.

Piette, John D, Michele Heisler, and Todd H Wagner. 2004. "Problems Paying Out-of-Pocket Medication Costs Among Older Adults with Diabetes." *Diabetes Care*, 27(2): 384–391.

Rajagopalan, R, A Joyce, D Ollendorf, and FT Murray. 2003. "PDB5: Medication Compliance in Type 2 Diabetes Subjects: Retrospective Data Analysis." *Value in Health*, 6(3): 328.

Roodman, David. 2009b. "A Note on the Theme of Too Many Instruments." *Oxford Bulletin of Economics and Statistics*, 71(1): 135–158.

Rust, John, and Christopher Phelan. 1997. "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets." *Econometrica: Journal of the Econometric Society*, 781–831.

Saaddine, Jinan B, Michael M Engelgau, Gloria L Beckles, Edward W Gregg, Theodore J Thompson, and KM Venkat Narayan. 2002. "A Diabetes Report Card for the United States: Quality of Care in the 1990s." *Annals of Internal Medicine*, 136(8): 565–574.

Shea, John. 1997. "Instrument Relevance in Multivariate Linear Models: A Simple Measure." *Review of Economics and Statistics*, 79(2): 348–352.

Silver, Morris. 1972. "An Economic Analysis of Variations in Medical Expenses and Work-Loss Rates." In *Essays in the Economics of Health and Medical Care.* , ed. H.E. Klarman, 97–118. NBER.

Simon, Kosali, Aparna Soni, and John Cawley. 2017. "The Impact of Health Insurance on Preventive Care and Health Behaviors: Evidence from the First Two Years of the ACA Medicaid Expansions." *Journal of Policy Analysis and Management*, 36(2): 390–417.

Sommers, Benjamin D., Atul A. Gawande, and Katherine Baicker. 2017. "Health Insurance Coverage and Health - What the Recent Evidence Tells Us." *The New England Journal of Medicine*, 377(6): 586–593.

Soumerai, SB, and D Ross-Degnan. 1999. "Inadequate Prescription-Drug Coverage for Medicare Enrollees–a Call to Action." *The New England Journal of Medicine*, 340(9): 722–728.

Soumerai, Stephen B, Marsha Pierre-Jacques, Fang Zhang, Dennis Ross-Degnan, Alyce S Adams, Jerry Gurwitz, Gerald Adler, and Dana Gelb Safran. 2006. "Cost-Related Medication Nonadherence Among Elderly and Disabled Medicare Beneficiaries: a National Survey 1 Year Before the Medicare Drug Benefit." *Archives of Internal Medicine*, 1829–1835.

Stock, James, and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. , ed. James Andrews, Donald & Stock. Cambridge: Cambridge University Press.

Von Wachter, Till. 2002. *The End of Mandatory Retirement in the US: Effects on Retirement and Implicit Contracts.* Center for Labor Economics, University of California, Berkeley.

Weinger, Katie, and Elizabeth A Beverly. 2010. "Barriers to Achieving Glycemic Targets: Who Omits Insulin and Why?" *Diabetes Care*, 33: 450–452.

Wild, Sarah, Gojka Roglic, Anders Green, Richard Sicree, and Hilary King. 2004. "Global Prevalence of Diabetes: Estimates for the Year 2000 and Projections for 2030." *Diabetes Care*, 27(5): 1047–1053.

Wooldridge, Jeffrey M. 2005. "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity." *Journal of Applied Econometrics*, 20(1): 39–54.

Appendix

Robustness Checks for Section 5

Table A.1: Original Cohort RDDs, 1998-2004, Diabetic Women: Insulin and Oral Medication Usage

	(1)	(2)	(3)	(4)
	Mean 60-64			
Insulin				
	0.34	-0.51*	-0.44*	-0.47*
		(-2.55)	(-2.33)	(-2.49)
Oral Medication				
	0.64	0.13	0.15	0.16
		(0.62)	(0.78)	(0.82)

t statistics in parentheses

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

Notes: Standard errors are clustered at the individual level. Column (2) reports estimated discontinuities from specifications without covariates apart from the age in months and age in months interacted with the treatment indicator; Column (3) reports results including time dummies; Column (4) reports results including both time dummies and health status, marital status, work status and education fixed effects. All specifications use local linear regression with age in months as the running variable and a Uniform kernel. Bandwidth used is 70.07, selected by the MSE criterion of Calonico, Cattaneo and Titiunik (2014). Individuals enrolled in Medicaid at any age, or enrolled in supplemental insurance (Medigap) or a Medicare HMO (Medicare Advantage) after age 65, are excluded.

Table A.2: RDD Results with Age 62 as the Cutoff, 1998-2004, Diabetic Women: Insulin and Oral Medication Usage

	(1)	(2)	(3)	(4)
	Mean 60-64			
Insulin				
	0.23	-0.53	-0.52	-0.38
		(-0.40)	(-0.39)	(-0.33)
Oral Medication				
	0.69	-0.44	-0.53	-0.56
		(-0.30)	(-0.36)	(-0.43)

t statistics in parentheses

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

Notes: Results treat 62 as the cutoff for the regression discontinuity design instead of 65 as in the rest of this paper. Standard errors are clustered at the individual level. Column (2) reports estimated discontinuities from specifications without covariates apart from the age in months and age in months interacted with the treatment indicator; Column (3) reports results including time dummies; Column (4) reports results including both time dummies and health status, education fixed effects, and indicators for whether individuals are enrolled in a Medicare HMO (Medicare+Choice/Medicare Advantage, depending on whether pre- or post-2003) and whether they have purchased supplemental insurance (Medigap). All specifications use local linear regression with age in months as the running variable and a Uniform kernel.

Table A.3: RDD Results, 1998-2004, Continuously Insured Diabetic Women: Insulin and Oral Medication Usage

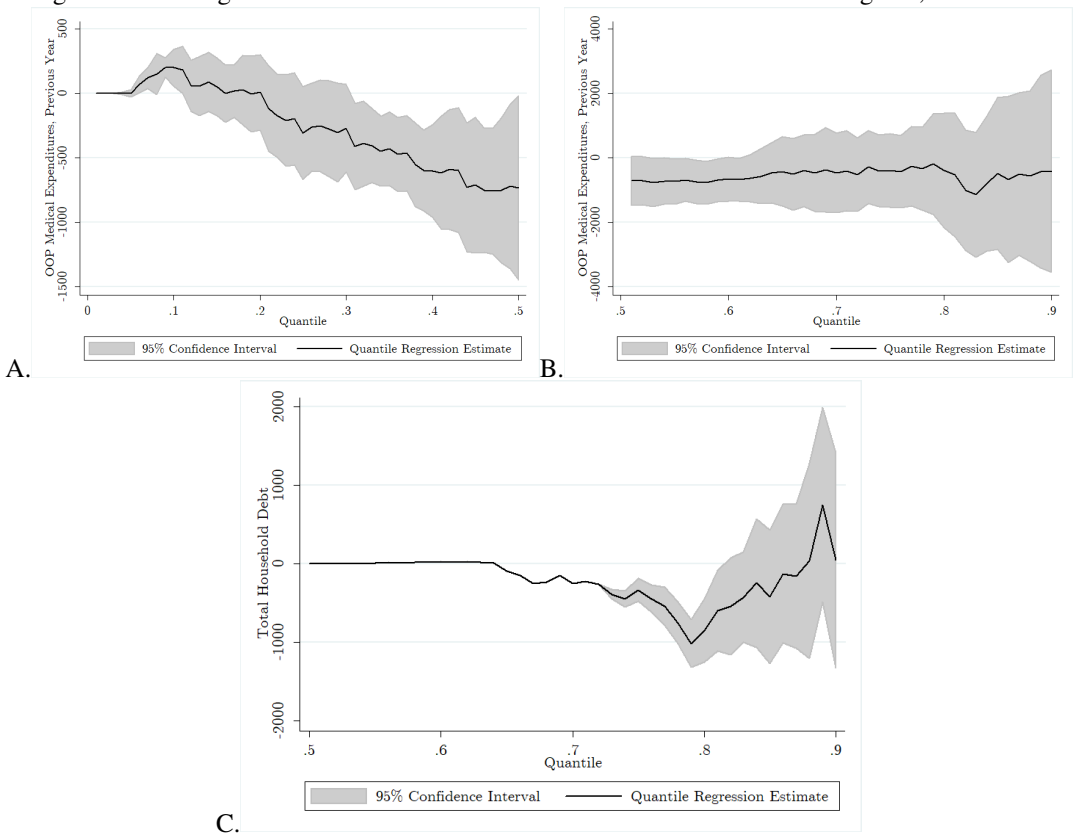
	(1)	(2)	(3)	(4)
	Mean 60-64			
Insulin				
	0.24	-0.02	-0.02	-0.02
		(-0.41)	(-0.44)	(-0.40)
Oral Medication				
	0.67	0.05	0.05	0.05
		(1.07)	(1.01)	(1.07)

t statistics in parentheses

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

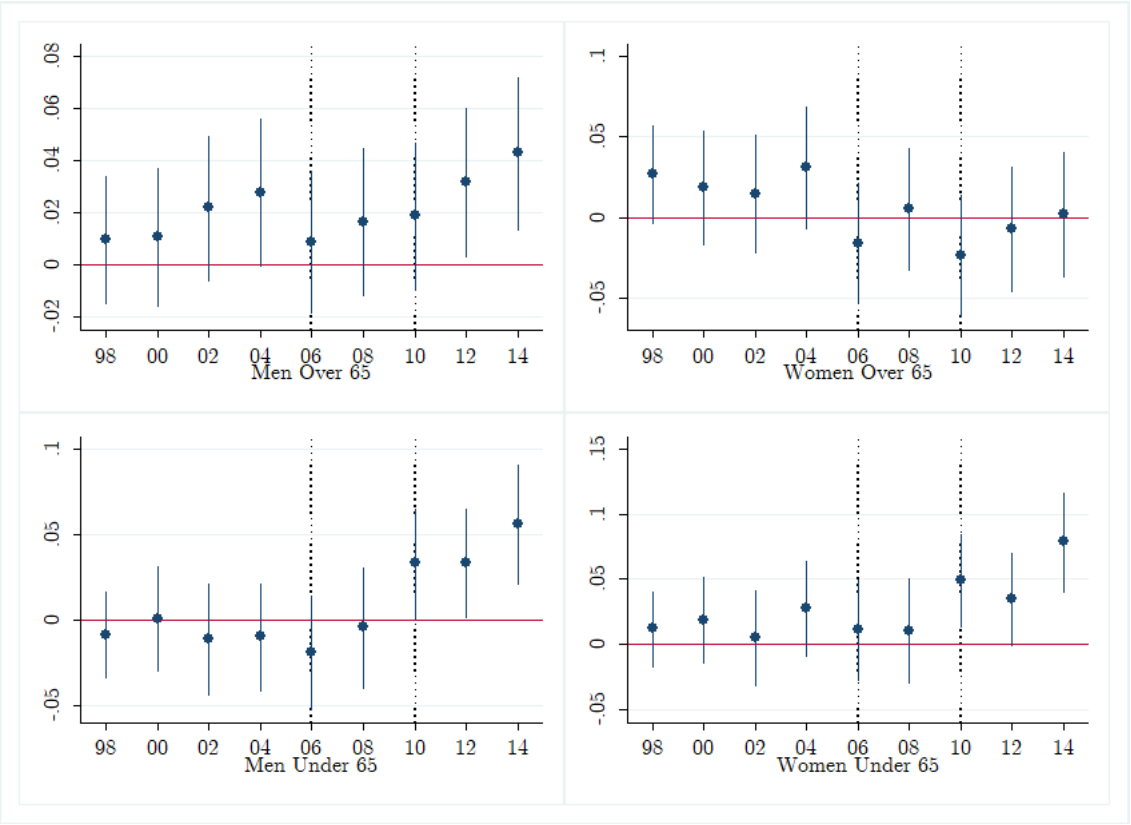
Notes: Sample is restricted to individuals who are insured via either their or their spouse's employer. Standard errors are clustered at the individual level. Column (2) reports estimated discontinuities from specifications without covariates apart from the age in months and age in months interacted with the treatment indicator; Column (3) reports results including time dummies; Column (4) reports results including both time dummies and health status, education fixed effects, and indicators for whether individuals are enrolled in a Medicare HMO (Medicare+Choice/Medicare Advantage, depending on whether pre- or post-2003) and whether they have purchased supplemental insurance (Medigap). All specifications use local linear regression with age in months as the running variable and a Uniform kernel.

Figure A.1: Changes in Measures of Financial Risk for Female Diabetics at Age 65, 1998-2004



Notes: Figure A.1. displays point estimates and confidence intervals for quantile regression versions of the panel data RDD equation with the quantile τ varying over the set $\{0.01, 0.02, ..., 0.89, 0.90\}$. The dependent variable is out-of-pocket medical expenses from the preceding twelve months for Panels (A) and (B) and total household debt for Panel C.
All sharp regression-discontinuity results are obtained excluding those on Medicaid, Medicare Advantage or Medicare supplemental coverage (Medigap), use a Uniform kernel with local linear regression, and include time dummies, health status, and education fixed effects. The magnitude of the effects is generally smaller than those found in Barcellos and Jacobson (2015), likely because of the smaller sample size.

Figure A.2: Trends in Medicaid Enrolment Relative to 1996 Among Diabetics in the HRS, 1998-2014



Notes: Plots are of coefficients from pooled OLS regressions of the outcome (reporting receipt of Medicaid) on time dummies for the years 1998-2014, with 1996 as the reference category. Since there is overlap among individuals across the different waves of the survey, standard errors for these regressions are clustered at the individual level. Dotted lines correspond to major health care reforms - Medicare Part D being implemented in 2006, and the Affordable Care Act being passed in 2010. Medicaid recipients are excluded.

Table A.4: First-Differences Results, 1998-2004, Diabetic Women: Exercise and Diet

	(1)	(2)	(3)
△Exercise			
	0.06	0.06	0.07
	(1.17)	(1.18)	(0.89)
△Diet			
	-0.03	-0.03	-0.01
	(-0.61)	(-0.57)	(-0.08)

t statistics in parentheses

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

Notes: Standard errors are clustered at the individual level. Column (1) reports estimated pre-post differences from specifications without covariates; Column (2) reports results including time dummies; Column (3) reports results including first-differences in earnings, work status, health status (an indicator equal to 1 if an individual reports being in "Fair" or "Poor" health, marital status, and enrolment in either Medicare Advantage/a Medicare HMO or supplementary coverage under Medicare (Medigap).

Dynamic Panel Regression Discontinuity Design: Theory and Evidence

There are at least four reasons to believe that past decisions affect present decisions in health capital investment decisions. First, advice from medical professionals on whether to use a specific type of preventive medicine is likely to be correlated over time, so that individuals who follow a past recommendation are more likely to be recommended to continue using their regular method of prevention, absent some exogenous shock, and vice versa. Second, the difficulty of making adjustments to an individual's routine may generate habits in their demand for prevention, as in standard models of habits in consumption. Third, if individuals exhibit anticipation effects (as in Alpert (2016) or De Preux (2011)), then $Pr(Y_{it}|D_{it}, Y_{it-1}) \neq Pr(Y_{it}|D_{it})$ since the contemporaneous probability $Pr(Y_{it}|D_{it})$ will ignore the impact of past anticipatory actions (for example, delaying investment until a period when it becomes less expensive). Fourth, in general, simple models of investment in human capital over the life cycle, investments in one period are complementary to investments in all other periods, and so the past history of decisions will affect each period's optimal choice for a given individual (as in Imai and Keane (2004) or Keane (2016)).

If we augment the standard panel regression-discontinuity design with a lagged dependent variable to account for the interdependence of decisions made at different times, the estimating equation becomes

$$Y_{it} = \rho Y_{it-1} + \beta_0 + \beta_1 D_{it} + f(R_{it}, \gamma_0) + g(R_{it}, \gamma_1) \times 1[R_{it} \geq \bar{R}] + \delta X_{it} + \zeta t + \eta_i + v_{it}. \tag{34}$$

Table A.5: Unrestricted Panel RDD Results, 1998-2004 1998-2008, Diabetic Men, Insulin and Oral Medication Usage

	(1)	(2)	(3)	(4)	(5)	(6)
Insulin						
$\widehat{\beta}_1$	0.20	0.19	0.14	0.06	0.08	0.05
	(0.58)	(0.56)	(0.47)	(0.21)	(0.25)	(0.16)
$\widehat{\beta}_3$				0.35	0.38	0.21
				(0.74)	(0.79)	(0.48)
Cragg-Donald Stat.				10.98	10.79	13.07
Oral Medication						
$\widehat{\beta}_1$	-0.75	-0.73	-0.70	-0.55	-0.54	-0.56
	(-1.77)	(-1.75)	(-1.84)	(-1.49)	(-1.48)	(-1.62)
$\widehat{\beta}_3$				0.10	0.16	0.13
				(0.21)	(0.33)	(0.27)
Cragg-Donald Stat.				10.97	10.79	13.05

t statistics in parentheses

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

Notes: Standard errors are clustered at the individual level. All specifications use local linear regression with age in months as the running variable and a Uniform kernel and exclude Medicaid recipients and smokers. Columns (1)-(3) are for the period 1998-2004; Columns (4)-(6) are for the period 1998-2008. Columns (1) and (4) report estimated pre-post differences from specifications without covariates; Columns (2) and (5) report results including time dummies; Columns (3)and (6) report results including first-differences in earnings, work status, health status (an indicator equal to 1 if an individual reports being in "Fair" or "Poor" health, marital status, and enrolment in either Medicare Advantage/a Medicare HMO or supplementary coverage under Medicare (Medigap).

Table A.6: First Differences Results, 1998-2004 1998-2008, Diabetic Men, Insulin and Oral Medication Usage

	(1)	(2)	(3)	(4)	(5)	(6)
△Insulin						
$\widehat{\beta}_1$	0.02	0.02	0.03	0.02	0.02	0.04
	(1.08)	(1.12)	(1.36)	(1.08)	(1.12)	(1.57)
$\widehat{\beta}_3$				-0.01	-0.01	-0.03
				(-0.25)	(-0.28)	(-0.92)
△Oral Medication						
$\widehat{\beta}_1$	0.00	0.00	0.01	0.00	0.00	0.01
	(0.15)	(0.16)	(0.18)	(0.15)	(0.16)	(0.48)
$\widehat{\beta}_3$				0.00	0.00	0.01
				(0.10)	(0.08)	(0.29)

t statistics in parentheses

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

Notes: Standard errors are clustered at the individual level. All specifications use local linear regression with age in months as the running variable and a Uniform kernel and exclude Medicaid recipients and smokers. Columns (1)-(3) are for the period 1998-2004; Columns (4)-(6) are for the period 1998-2008. Columns (1)and (4) report estimated pre-post differences from specifications without covariates; Columns (2) and (5) report results including time dummies; Columns (3) and (6) report results including first-differences in earnings, work status, health status (an indicator equal to 1 if an individual reports being in "Fair" or "Poor" health, marital status, and enrolment in either Medicare Advantage/a Medicare HMO or supplementary coverage under Medicare (Medigap).

Table A.7: Medicaid Take-up Among Diabetics by Gender, Pre- and Post-2006

	Including 2010-14		Excluding 2010-14	
	(1)	(2)	(3)	(4)
	Over 65	Under 65	Over 65	Under 65
Female	0.133***	0.0901***	0.133***	0.0901***
	(11.38)	(7.34)	(11.38)	(7.34)
Post-2006	0.00765	0.0320**	-0.00349	-0.00606
	(1.04)	(3.20)	(-0.47)	(-0.52)
Female × Post-2006	-0.0358**	-0.00398	-0.0218	0.00373
	(-2.90)	(-0.26)	(-1.75)	(0.21)
Constant	0.0927***	0.0827***	0.0927***	0.0827***
	(13.87)	(11.47)	(13.87)	(11.47)

t statistics in parentheses

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

Notes: Standard errors are clustered at the individual level. Dependent variable is an indicator for whether an individual reports receipt of Medicaid. First two columns include the years 2010-2014, which follow the passage of the Patient Protection and Affordable Care Act (ACA), while columns (3) and (4) restrict the Post-2006 observations to the 2006 and 2008 waves of the Health and Retirement Study.

Table A.8: Diagnosis of Diabetes by Gender, Pre- and Post-2006

	(1) Over 65	(2) Under 65	(3) Over 65	(4) Under 65
Female	-0.0316***	-0.0197***	-0.0461***	-0.0279***
	(-5.05)	(-3.44)	(-7.30)	(-4.95)
Post-2006	0.0821***	0.0596***	0.0804***	0.0563***
	(14.81)	(9.59)	(14.16)	(9.06)
Female × Post-2006	-0.00800	0.000928	-0.00803	-0.000878
	(-1.12)	(0.12)	(-1.10)	(-0.11)
Constant	0.195***	0.148***	0.191***	0.142***
	(40.14)	(33.32)	(38.46)	(32.02)

t statistics in parentheses

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

Notes: Standard errors are clustered at the individual level. Dependent variable is an indicator for whether an individual reports diagnosis of diabetes. First two columns pertain to the full sample; third and fourth columns restrict the analysis to the subsample that does not report being enrolled in Medicaid.

Let, for any random variable W, $lim_{R\to\bar{R}}E[W|R=R+\varepsilon]=W^+$ and $lim_{R\to\bar{R}}E[W|R=R-\varepsilon]=W^-$. We obtain

$$\begin{aligned}
&(Y_t^+ - Y_t^-)\\
&= \beta_1(D_t^+ - D_t^-) + \rho(Y_{t-1}^+ - Y_{t-1}^-)\\
&= \beta_1(D_t^+ - D_t^-) + \rho\beta_1(D_{t-1}^+ - D_{t-1}^-) + \rho^2(Y_{t-2}^+ - Y_{t-2}^-)\\
&= ...\\
&= \sum_{s=0}^{t} \rho^s \beta_1(D_{t-s}^+ - D_{t-s}^-) + \rho^t(Y_0^+ - Y_0^-).
\end{aligned}$$

From the above, it appears at first glance that we may require the stringent condition

$$(Y_0^+ - Y_0^-) = 0$$

but here the fact that the running variable is age, which advances deterministically with time, allows for weaker identifying assumptions. Note that the fact that the running variable, though continuous, differs between any two periods by an integer increment, implies that for any $s \in \mathbb{Z}^+$, there is some $m \in \mathbb{Z}^+$ such that

$$R_{it-s} = R_{it} - sm, \tag{35}$$

If age is measured in years, and the data is collected taken annually, we can take $m=1$, which just says that age in $t-s$ is, mechanically, age in period t minus s. For age measured in months, and data collected every two years, we have $m=24$. Note that this doesn't require that the running variable R_{it} is itself an integer, but that it can be divided into a set of ordered intervals according to the integers. Equation (3) can be modified to accommodate any combination of periodicity of the data and increment of the running variable between waves of the data. Now in practice, regression-discontinuity designs rely on restricting attention to observations within some bandwidth h of the cutoff (cf. Hahn, Todd and Van der Klaauw (2001)). This restricts attention to all data points such that $|\frac{R_{it}-\bar{R}}{h}|<1$. This, together with equation (6), implies that, for a given bandwidth h, there exists some k such that

$$k = inf_{k\in\mathbb{R}^+}\{R_{it-k} : R_{it-k} < \bar{R} | 0 < \tfrac{R_{it}-\bar{R}}{h} < 1\},$$

which is formalised in Lemma 1 below. For both Lemma 1 and Proposition 2, I assume without loss of generality that $m=1$ in Equation (3) above.

Lemma 1. Let $R_{it} \in \mathbb{R}$, indexed by $i \in \{1,...,N\}$ and $t \in \{1,...,T\}$. Let h be the bandwidth such that for a given period t, $|R_{it}-\bar{R}|<h$. If $R_{it-s}=R_{it}-s$ for all t and $s \in \mathbb{Z}^+$, (i) there is a set $\Omega := \{q : 0 \leq \frac{R_{it}-\bar{R}}{h} < 1 \implies R_{it-q}-\bar{R}<0\}$, and (ii) $\exists k := inf\Omega \in \mathbb{Z}^+$.

Proof: Since by definition $|R_{it} - \bar{R}| < h \forall R_{it}$, $\exists R_{it}^{max}$ s.t. $0 \leq R_{it}^{max} - \bar{R} < h$ but $R_{it}^{max} + 1 - \bar{R} \geq h$. Similarly, $\exists R_{it}^{min}$ s.t. $R_{it}^{min} - \bar{R} > 0$ but $R_{it}^{min} - 1 - \bar{R} < 0$. Take $q \in \mathbb{Z}^+$ s.t. $q \geq 1 + (R_{it}^{max} - \bar{R})$. Then

$R_{it-q}^{max} = R_{it}^{max} - q = \bar{R} - 1 < \bar{R}$

$R_{it-q}^{max} - 1 = R_{it}^{max} - q - 1 = \bar{R} - 2 < \bar{R}$

...

$R_{it-q}^{min} = R_{it}^{min} - q < R_{it}^{min} - 1 < \bar{R}$

So $\Omega = \{q : q \geq 1 + (R_{it}^{max} - \bar{R})\}$ and $k = inf\Omega = 1 + (R_{it}^{max} - \bar{R}) \in \mathbb{Z}^+ \blacksquare$

Note that the running variable R_{it} can be continuous or discrete, but changes deterministically with the integer index t so that the difference between values of R_{it} for the same individual in between two periods is an integer. The above Lemma

Proposition 2. Let k be defined as in Lemma 1. Then if (ignoring covariates and common time trends for simplicity):

(i) $Y_{it} = \rho Y_{it-1} + \beta_1 D_{it} + f(R_{it}, \gamma_0) + g(R_{it}, \gamma_1) \times 1[R_{it} \geq \bar{R}] + \eta_i + v_{it}$,

(ii)

$$(Y_{t-k}^+ - Y_{t-k}^-) = 0,$$

with v_{it} i.i.d. and

(iii) $E[\eta_i + v_{it} | R_{is}]$ and $f(R_{it}, \gamma_0), g(R_{it}, \gamma_1)$ are continuous in R_{is}, R_{it} for all s, t,

then the numerator for the fuzzy RD estimand becomes

$$Y_t^+ - Y_t^-$$

$$= \beta_1 (D_t^+ - D_t^-) + \rho \beta_1 (D_{t-1}^+ - D_{t-1}^-) + ... + \rho^{k-1} \beta_1 (D_{t-k+1}^+ - D_{t-k+1}^-)$$

$$= \beta_1 \sum_{s=0}^{k-1} \rho^s (D_{t-s}^+ - D_{t-s}^-).$$

Proof: Note that

$$Y_t^+ - Y_t^-$$

$$= \beta_1 (D_t^+ - D_t^-) + \rho (Y_{t-1}^+ - Y_{t-1}^-)$$

$$= \beta_1 (D_t^+ - D_t^-) + \rho \beta_1 (D_{t-1}^+ - D_{t-1}^-) + \rho^2 \beta_1 (D_{t-2}^+ - D_{t-2}^-)$$

$$+ ... + \rho^{k-1} \beta_1 (D_{t-k+1}^+ - D_{t-k+1}^-) + \rho^k (Y_{t-k}^+ - Y_{t-k}^-)$$

$$= \beta_1 \sum_{s=0}^{k-1} \rho^s (D_{t-s}^+ - D_{t-s}^-) + \rho^k (Y_{t-k}^+ - Y_{t-k}^-)$$

$$= \beta_1 \sum_{s=0}^{k-1} \rho^s (D_{t-s}^+ - D_{t-s}^-),$$

where line two follows from a combination of (i) and (iii), the lines preceding the last line follow from repeated application of (i) and (iii), and the last line follows from (ii)■.

That is, given a set of individuals in period t whose ages are above the threshold $\bar{R}$, but lie within the interval $[\bar{R}, \bar{R}+h]$ defined by the chosen bandwidth h, there is some greatest lower bound k for the set of positive integers that, if subtracted from all the ages between $\bar{R}$ and $\bar{R}+h$, would produce a set of ages that are less than $\bar{R}$. Equivalently, k picks out the period $t-k$ in which all the individuals above the cutoff - but within the chosen distance from the cutoff h - were below the cutoff. The reason this is useful is that in any period $t-q$ such that $q \geq k$, $D_{it-q} = 0$ for all the individuals within h of the cutoff by definition. That is, all of the individuals in the treatment group were in the control group for periods $t-k, t-k-1, ..., 0$. Therefore if we are only interested in the effect of D_{it}, β_1, we only need the extra restriction that

$$(Y^+_{t-k} - Y^-_{t-k}) = 0,$$

with k defined as above. This is a regression-discontinuity analog of the usual restrictions on initial conditions necessary in dynamic panel data models (for which see Baltagi (2008). It requires that there are no other discontinuties that would result from an age-based cutoff at an age k survey periods in the past. Otherwise, we will confound the persistent effect of this past discontinuity with the other effects on the behavior of the treatment group. For example, in my context, individuals who are 66, with a survey taken every other year, with $k = 2$, must not have undergone a discontinuous jump in behavior at age 62 - which may well happen with U.S. data since this is the first age at which Americans can claim Social Security benefits. Fortunately, this can be tested as per the usual checks that "placebo" cutoffs do not have the same effect as the cutoff of interest. Identification of the persistence parameter ρ then comes from comparing individuals who switched status last period, so that $R_{it} \geq \bar{R}$ but $R_{it-1} < \bar{R}$, who exhibit no persistent effect of treatment since they were not in the treatment group in the last period, with those further away from the cutoff. To illustrate this, let $\triangle D_{it} = 1[R_{it} \geq \bar{R}] - 1[R_{it-1} \geq \bar{R}]$ pick out this group of individuals (since there are no cases where $R_{it-1} \geq \bar{R}$ and $R_{it} < \bar{R}$ by design), $D_{it-1} = 1[R_{it-1} \geq \bar{R}]$, and suppose that $k = 2$. We can rewrite the regression-discontinuity estimand as

$$\begin{aligned} &\beta_1(D^+_t - D^-_t) + \rho\beta_1(D^+_{t-1} - D^-_{t-1}) + \rho^2(Y^+_{t-2} - Y^-_{t-2}) \\ &= \beta_1 + \rho\beta_1(D^+_{t-1} - D^-_{t-1}) \end{aligned}$$

by the definitions of k and h.

Now if we rewrite the estimating equation to include D_{it-1}, so that we can partial out $\beta\rho$ from β, we obtain

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \beta_1\rho D_{it-1} + f(R_{it}, \gamma_0) + g(R_{it}, \gamma_1) + \delta X_{it} + \zeta t + \eta_i + v_{it} \tag{36}$$

$$= \beta_0 + \beta_1 \triangle D_{it} + \beta_1(1+\rho) D_{it-1} + f(R_{it}, \gamma_0) + g(R_{it}, \gamma_1) + \delta X_{it} + \zeta t + \eta_i + v_{it}, \tag{37}$$

The advantage of rewriting equation (7) as equation (8) is that it highlights the role of intra-treatment group comparisons in identifying ρ. β is identified from the difference between the group that switched

status this period and those below the cutoff in the same period, and is the contemporaneous impact on individuals' choices above the cutoff. Individuals who had already switched by the last period undergo a contemporaneous effect β_1 and an effect carried over from when they switched status $\rho\beta_1$, which together is $\beta_1(1+\rho)$.

The ability to obtain a consistent estimate of ρ comes from not having to rely on more traditional exogeneity assumptions on account of the regression discontinuity design. The usual assumption that $E[\eta_i + v_{it}|y_{it-1}] = 0$ will fail since $E[y_{it-1}]$ contains η_i by construction, and first-differencing does not do better since $E[\triangle v_{it}|\triangle y_{it-1}] = 0$ will be violated for the same reason. Identification in the regression discontinuity design only depends on the continuity of these conditional expectation functions at the cutoff, with the only additional restriction for this case being no discontinuity in the period $t-k$ in which all individuals in the treatment group are in the control group. This is less demanding than the traditional exogeneity assumptions necessary for standard dynamic panel data estimators.

Here, instead of a restriction on the distribution of the initial conditions (as in Wooldridge (2005)) or the even stronger requirement that they are exogenous, we require that there is no discontinuity in the period $t-k$ in which all the individuals treated in period t are below the cutoff. Fortunately, in my scenario this involves a test for a discontinuity at or near the Social Security claiming age of 62, which would be a necessary robustness check even if I were not aiming to identify ρ. In practice identifying ρ in the panel RDD simply requires more robustness checks for placebo cutoffs at earlier ages of the kind already usually reported in standard RDDs.

There is a simple intuition for why we can identify the persistence parameter ρ using an RDD with a bandwidth that includes more years than elapse between survey years. Some individuals who are included to the right of the cutoff will already have switched status in the last survey year. As a result, the net effect on their behavior is a combination of the effect from the last survey year persisting and the effect in the current survey year. If we can pin down the effect of having changed status between the last survey and the current year by focusing on the subset of individuals who changed, we can back out the effect in the current year and hence the persistent component of previous treatments by subtraction.

In this scenario, estimating ρ becomes a special case of instrumental variables estimation with heterogeneity in the second-stage coefficients. The first effect is the contemporaneous effect of Medicare coverage, β_1. Individuals sufficiently far from the cutoff will also exhibit a spillover effect from having switched status in the previous period. Comparing individuals who only switched this period and so only exhibit the contemporaneous effect with their peers who exhibit both the contemporaneous and persistent treatment effects allows us to recover the extent of persistence ρ[21]. This is analogous to the scenario in the previous subsection where we can pick out individuals who were previously uninsured and enrol in traditional Medicare and compare them with individuals who were covered prior to age 65 and purchase supplemental insurance. Only one additional identifying assumption - the restriction on the initial conditions - is necessary to identify ρ.

Note that the only part of the above identification arguments that requires panel data is information on

[21] A structural model is used to identify spillover effects due to spatial externalities in an RDD in Fu and Gregory (2017). The analysis in this subsection could in principle be applied to similar cases where an individual's actions are linked to those of other individuals due to social networks or geographical location, rather than to their own past actions.

outcomes in previous periods. In principle, this information could be recorded in a single cross section - at no point in this subsection have the identification arguments had to rely on estimating equations in first differences or fixed-effects estimation, which are only possible with repeated observations of the same individuals. Two drawbacks of cross-sectional data will prevent identification according to the above arguments from succeeding in practice. First, few cross-sectional data sets contain detailed histories of past outcomes that are required for the approach in this subsection. Second, differential mortality in a cross-section will select for individuals with longer *ex ante* expected lifespans and hence higher returns from using prevention. This will lead to spurious upward drift in the probability of using prevention with age due to those with weaker incentives for usage being selected out of the sample. In a balanced panel, this concern no longer holds since we use the realised life-cycle profiles of usage to estimate the model. In a cross-section differential mortality will bias the estimator of ρ downward if β_1 is negative since the effect will only be observed to persist for a short period of time before exhibiting mean reversion. This in turn will lead to an underestimate of the longer-term effect at the cutoff for life-cycle decision-making. This problem can be expected to be especially acute when using the Uniform kernel, since it weights observations within the chosen bandwidth equally (though the order of the polynomial used for the local polynomial regressions either side of the cutoff will weight observations differently (Gelman and Imbens, 2018). By contrast, the Edge (Triangular) kernel gives more weight to observations closer to the cutoff, and so will suffer from this problem to a lesser extent. In a panel the Triangular kernel may tilt the scales too aggressively in favor of larger values of ρ, since it will give less weight to observations further from the cutoff. For this reason I exclusively use the Uniform kernel for the panel RDDs in this paper.

Static and Dynamic Treatment Effects

In a fuzzy cross-sectional RDD, the effect of interest is often defined as the Local Average Treatment Effect given by

$$\frac{Y^+ - Y^-}{D^+ - D^-},$$

(cf. Hahn, Todd and Van der Klaauw (2001)). Conveniently, if we take these limits for equation (1), we obtain

$Y^+ - Y^- = \beta_1(D^+ - D^-) + E[f(R^+,\gamma) - f(R^-,\gamma)] + E[\eta_i|R = R^+] - E[\eta_i|R = R^-] = \beta_1(D^+ - D^-)$ by continuity of $f(.)$ and η in R at $\bar{R}$, and so $\beta_1 = \frac{Y^+ - Y^-}{D^+ - D^-}$, the treatment effect of interest. In the sharp design, $D^+ = 1$ and $D^- = 0$ so this simplifies to $\beta_1 = Y^+ - Y^-$. With dynamic decision-making and the assumption from the last section that decisions from before period $t - k + 1$ do not matter for the RD estimand, we have

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \rho\beta_1 D_{it-1} + ... + \rho^{k-1}\beta_1 D_{it-k+1} + \sum_{j=0}^{k-1} \rho^j \eta_i + \sum_{j=0}^{k-1} \rho^j v_{it-j}, \tag{38}$$

so that while the static effect is given by β, which would be the only effect of treatment if $\rho = 0$, the cumulative effect from the previous $k-1$ periods is given by $\beta(\frac{1-\rho^k}{1-\rho})$. The individual coefficients are

interpretable as net effects of treatment given that there was a discontinuity in treatment status in the previous period. Taking the limits as in the static case above for $k = 2$, we obtain

$$Y_t^+ - Y_t^- = \beta_1(D_t^+ - D_t^-) + \rho\beta_1(D_{t-1}^+ - D_{t-1}^-) \implies \beta_1 = \frac{Y_t^+ - Y_t^-}{(D_t^+ - D_t^- + \rho(D_{t-1}^+ - D_{t-1}^-))},$$

which shows that β_1 is a downward-biased estimator of the average effect for all individuals "near" the cutoff, if individuals "near" the cutoff have heterogeneous effects of being treated due to spillovers from previous periods. Suppose that the one-period spillover is the only additional treatment effect. Then the cumulative effect is

$$\frac{Y_t^+ - Y_t^-}{D_t^+ - D_t^-} + \frac{Y_t^+ - Y_t^-}{Y_{t-1}^+ - Y_{t-1}^-}\frac{Y_{t-1}^+ - Y_{t-1}^-}{D_{t-1}^+ - D_{t-1}^-} = \frac{Y_t^+ - Y_t^-}{D_t^+ - D_t^-} + \frac{Y_t^+ - Y_t^-}{D_{t-1}^+ - D_{t-1}^-},$$

now notice, moreover, that if we had just written the AR(1) process for Y_{it}, ignoring any effects of treatment status, and taken the limits as above, we would obtain

$$Y_t^+ - Y_t^- = \rho(Y_{t-1}^+ - Y_{t-1}^-) \implies \rho = \frac{Y_t^+ - Y_t^-}{Y_{t-1}^+ - Y_{t-1}^-},$$

so that the cumulative effect of interest is given by

$$\frac{Y_t^+ - Y_t^-}{D_t^+ - D_t^-} + \frac{Y_t^+ - Y_t^-}{D_{t-1}^+ - D_{t-1}^-} = \frac{Y_t^+ - Y_t^-}{D_t^+ - D_t^-} + \frac{Y_t^+ - Y_t^-}{Y_{t-1}^+ - Y_{t-1}^-}\frac{Y_{t-1}^+ - Y_{t-1}^-}{D_{t-1}^+ - D_{t-1}^-} = \beta_1 + \rho\beta_1, \tag{39}$$

given that β_1 captures the contemporaneous, static effect of treatment, ignoring spillovers from previous periods. It is trivial to extend this to the case where $k > 2$, where the sum of the coefficients on the treatment indicators $D_{it}, ..., D_{it-k+1}$ captures the dynamic effect of treatment for a group of individuals with discontinuities in treatment status across periods $t, ..., t-k$.

Estimation of a Dynamic RDD

In this subsection I examine the identification of dynamic treatment effects in an RDD. The exposition in this section closely follows that of Dong and Lewbel (2015). Given a bandwidth h in which the observations lie, local linear regression with a Uniform kernel is equivalent to dropping observations for which the running variable lies outside of the interval $[\bar{R} - h, \bar{R} + h]$ and estimating the equation (ignoring covariates and common time trends for simplicity):

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 D_{it-1} + \gamma_0(R_{it} - \bar{R}) + \gamma_1(R_{it} - \bar{R}) \times 1[R_{it} \geq \bar{R}] + \gamma_2(R_{it} - \bar{R}) \times 1[R_{it-1} \geq \bar{R}] + \eta_i + v_{it}, \tag{40}$$

Note that a linear approximation to the function $E[Y|\bar{R} \leq R < \bar{R} + \varepsilon, R_{it-1} < \bar{R}]$ is given by

$$\beta_0 + \beta_1 + (\gamma_0 + \gamma_1)(R_{it} - \bar{R}), \tag{41}$$

while a linear approximation to $E[Y|\bar{R} - \varepsilon \leq R < \bar{R}, R_{it-1} < \bar{R}] = E[Y|\bar{R} - \varepsilon \leq R < \bar{R}, R_{it-1} \geq \bar{R}]$ is given by

$$\beta_0 + \gamma_0(R_{it} - \bar{R}), \tag{42}$$

so that in a hypothetical sharp RDD, the treatment effect for individuals with $R_{it-1} < \bar{R}$ (those who switched between the current survey period and the last survey period) is given by $plim(\widehat{\beta}_1)$. Moreover, we have that a linear approximation to $E[Y|\bar{R} \leq R < \bar{R} + \varepsilon, R_{it-1} \geq \bar{R}]$ is given by

$$\beta_0 + \beta_1 + \beta_2 + (\gamma_0 + \gamma_1 + \gamma_2)(R_{it} - \bar{R}), \tag{43}$$

so that the treatment effect for individuals with $R_{it-1} \geq \bar{R}$ is given by $plim(\widehat{\beta}_1 + \widehat{\beta}_2)$. If there is no spillover effect from having already crossed the threshold in the previous period, then these treatment effects are the same, and so $plim(\widehat{\beta}_2) = \beta_2 = 0$. $\widehat{\beta}_1$ corresponds to the static treatment effect and $\widehat{\beta}_1 + \widehat{\beta}_2$ the dynamic treatment effect. In a fuzzy design, we have separate regressions for the treatment indicators D_{it}, D_{it-1} on indicators of assignment to treatment status $1[R_{it} \geq \bar{R}], 1[R_{it-1} \geq \bar{R}]$ (in this paper's scenario, whether an individual is older than 65 in the subscripted period):

$$\begin{aligned} D_{it} = \pi_{0t} + \pi_{1t} 1[R_{it} \geq \bar{R}] + \pi_{2t} 1[R_{it-1} \geq \bar{R}] + \pi_{3t}(R_{it} - \bar{R}) + \pi_{4t} 1[R_{it} \geq \bar{R}] \times (R_{it} - \bar{R}) \\ + \pi_{5t} 1[R_{it-1} \geq \bar{R}] \times (R_{it} - \bar{R}) + w_{1it} \end{aligned}, \tag{44}$$

$$\begin{aligned} D_{it-1} = \pi_{0t-1} + \pi_{1t-1} 1[R_{it} \geq \bar{R}] + \pi_{2t-1} 1[R_{it-1} \geq \bar{R}] + \pi_{3t-1}(R_{it} - \bar{R}) + \pi_{4t-1} 1[R_{it} \geq \bar{R}] \times (R_{it} - \bar{R}) \\ + \pi_{5t-1} 1[R_{it-1} \geq \bar{R}] \times (R_{it} - \bar{R}) + w_{2it} \end{aligned}, \tag{45}$$

The estimated coefficients on treatment status $\widehat{\pi}_{1t}, \widehat{\pi}_{2t-1}$ are then used as divisors for $\widehat{\beta}_1$ and $\widehat{\beta}_2$, so that the static effect $\beta = \frac{Y_t^+ - Y_t^-}{D_t^+ - D_t^-}$ is given by $\frac{\widehat{\beta}_1}{\widehat{\pi}_{1t}}$ and the dynamic effect $\beta + \rho\beta = \frac{Y_t^+ - Y_t^-}{D_t^+ - D_t^-} + \frac{Y_t^+ - Y_t^-}{Y_{t-1}^+ - Y_{t-1}^-} \frac{Y_{t-1}^+ - Y_{t-1}^-}{D_{t-1}^+ - D_{t-1}^-}$ by $\frac{\widehat{\beta}_1}{\widehat{\pi}_{1t}} + \frac{\widehat{\beta}_2}{\widehat{\pi}_{2t-1}}$. Note that R_{it} is used in the regression for D_{it-1} and not R_{it-1} since R_{it} is the running variable in the main estimating equation and not R_{it-1}.

Dynamic RDD Results

I first presage the Dynamic RDD results with results from more conventional methods, contrasting the OLS and Fixed-Effects estimators of the coefficient on the lagged dependent variable[22]. I then present dynamic models for two separate specifications (Table A.10., below). In practice, a fuzzy regression-discontinuity design appears to be suboptimal for the identification of the coefficient on the lagged dependent variable. This is because a growing number of endogenous variables and instruments that are highly correlated with one another leads to weak identification, worsening the finite-sample properties of the estimators. With larger bandwidths, we might mitigate this problem but run into the separate problem that such bandwidths may violate the assumptions under which the coefficient on the lagged dependent variable is identified under Proposition 2.

The OLS and Fixed-Effects estimators are useful for bounding the value of the persistence parameter ρ since the OLS estimator $\widehat{\rho}_{OLS}$ is biased upward and the FE estimator $\widehat{\rho}_{FE}$ is biased downward (Nickell,

[22] I also estimated Arellano-Bond or "Difference GMM" regressions (Arellano and Bond, 1991), which are more commonly used to identify the coefficient on a lagged dependent variable. In all of the specifications the set of instruments appeared to be invalid by virtue of producing estimates higher than the OLS estimator (which is known to be biased upward) or by virtue of the Sargan test finding strong evidence against the null hypothesis that every instrument is valid.

1981). They appear to produce somewhat different conclusions regarding the extent of persistence in the data depending on whether the 1998-2004 or 1998-2008 sample is used (Table A.9., below). In both cases, the OLS estimator finds that the persistence parameter is less than one, and is around 0.8 (roughly at the upper end of the findings of prospective cohort studies of persistence in insulin usage (Brown et al., 1999)). For 1998-2004, the case where $\rho = 0$ cannot be ruled out by the FE estimator, whereas the estimate from 1998-2008 suggests that $\rho > 0$. The suggested interval for the values that ρ might take on is considerably wider than that suggested by studies in the medical literature of adherence to insulin therapy (cf. Brown et al. (1999), Rajagopalan et al. (2003)), which is an interval of $[0.64, 0.8]$.

Table A.9: OLS and FE Estimates of AR(1) Persistence Parameter, Insulin Usage

	1998-2004		1998-2008	
	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)
Y_{it-1}	0.86***	-0.05	0.87***	0.23***
	(66.85)	(-1.62)	(97.56)	(12.50)

t statistics in parentheses

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

Notes: Standard errors are clustered at the individual level. All specifications use local linear regression with age in months as the running variable and a Uniform kernel, exclude Medicaid recipients and smokers, include no controls other than age and age interacted with eligibility, and otherwise follow the sample selection rules described in the main text.

In conclusion, it seems that even though the persistence parameter ρ can be identified in principle, in practice it is difficult to do so with a fuzzy regression-discontinuity design. More success is likely in settings where either the sample size is significantly larger or the design is sharp, so that the problem of weak identification does not arise (Feir, Lemieux and Marmer, 2016). The only remaining obstacle in that case would be the potential lack of agreement between the bandwidth necessary to satisfy the identification arguments and the optimal bandwidth selected by procedures such as those found in Imbens and Kalyanaraman (2012) or Calonico, Cattaneo and Titiunik (2014).

Evidence on (the Lack of) Credit Constraints

If individuals find it difficult to bring future income forward, they may struggle to exchange more prevention this year for less prevention next year. This would cast doubt on the power that intertemporal substitution has to explain the difference between this study and studies that do not use Medicare eligibility in an RDD. In those circumstances, the same agents would also find it difficult to bring income forward to purchase durable goods that can be financed on credit, and so there would be a spike in durable goods purchases as agents became eligible for Medicare.

I use the RAND CAMS dataset in this subsection. This is a cleaned version of the mail survey data

Table A.10: Dynamic Panel RDDs, Diabetic Women: Insulin Usage

	(1)	(2)	(3)	(4)	(5)	(6)
$\widehat{\beta_1}$	-0.34*	-0.09	-0.85*	-0.35**	-0.13	-0.84*
	(-2.42)	(-0.38)	(-2.46)	(-2.63)	(-0.65)	(-2.40)
$\widehat{\beta_1\rho}$		-0.30	0.23		-0.17	0.09
		(-0.76)	(0.55)		(-0.69)	(0.32)
$\widehat{\beta_1\rho^2}$			0.10			0.23
			(0.27)			(0.85)
$\widehat{\beta_3}$				0.69*	0.45	1.40***
				(2.44)	(1.05)	(3.45)
Bandwidth	83.33	48	72	66.23	48	72
Cragg-Donald Stat.	78.46	4.05	1.10	35.12	4.53	1.35

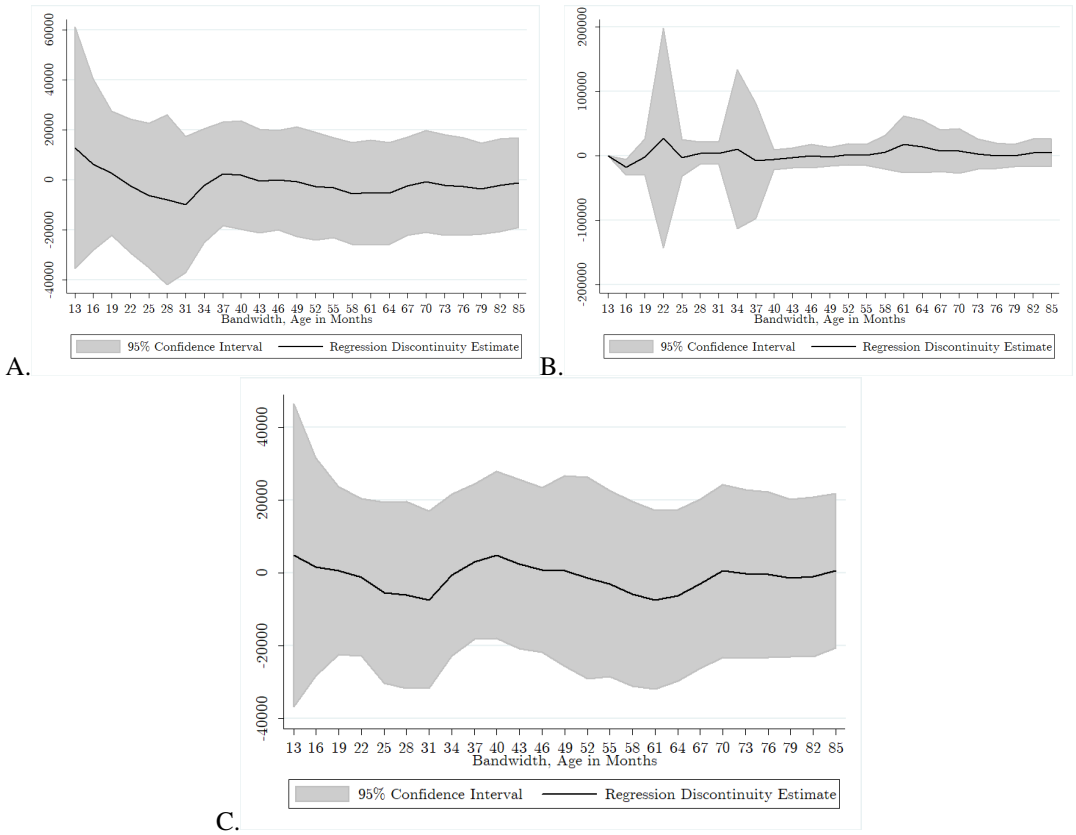
t statistics in parentheses

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

Notes: Standard errors are clustered at the individual level. All specifications use local linear regression with age in months as the running variable and a Uniform kernel, exclude Medicaid recipients and smokers, use only time dummies as controls, and otherwise follow the sample selection rules described in the main text.

run by the Health and Retirement Study (HRS) known as the Consumption and Activities Mail Survey (CAMS) (Hurd et al., 2015). I aggregate together total spending on durable goods, which comprises spending on refrigerators, washing machines, dishwashers, televisions, and computers, and add total spending on vehicles. I then replace insulin with total durable spending in a regression-discontinuity design using the years 1998-2004. I examine first the subsample of non-Medicaid-eligible diabetics who respond to the CAMS, then the subsample of that subsample that have nonpositive financial wealth and no housing wealth (hence no home equity with which to secure a loan), then the subsample that has positive housing wealth and so may exhibit "wealthy hand-to-mouth" behavior (as in Chetty (2008). There appears to be little evidence from these regressions that credit constraints are binding in the HRS data (see Figure A.3., below). This conclusion comes with the caveat that the overlap between the subset of HRS respondents who filled out the CAMS surveys and the subset who report being diabetic overlap very slightly, and so direct evidence on credit constraints among diabetics is difficult to obtain using these data.

Figure A.3: RD Estimates for Effect of Medicare Eligibility on Durable Goods Expenditures by Bandwidth, 1998-2004



Notes: All fuzzy regression-discontinuity results are obtained excluding those on Medicaid and use a uniform kernel with local linear regression. Standard errors are clustered within individuals. Sample consists of both male and female diabetics. Panel (A) corresponds to the broader subsample of diabetics who respond to the CAMS survey. Panel (B) corresponds to the subsample further restricted to individuals with nonpositive financial wealth and no housing wealth. Panel (C) follows Chetty (2008) and restricts the sample to individuals with positive housing wealth in case diabetics exhibit "wealthy hand-to-mouth" (Kaplan and Violante, 2014, Chetty and Szeidl, 2007) behavior. None of the specifications find significant increases in durable goods purchases at the age of Medicare eligibility.

Discrete Investments in Health and the First-Differences Estimator

In this section, I will show that the probability of making a discrete investment in human capital falls at an increasing rate towards the end of a finite lifetime, and that this implies that a first-differences estimator that compares an older "control" group with a younger "treatment" group can be expected to have weaker positive trends (or stronger negative trends) in health investment in the older control group due to unobservable differences in the net returns to investment in health. Throughout this subsection I shall assume there are no differences in long-run endowments of health or lifespan (which are differenced out in a first-differences specification), no heterogeneity in the net return to investment in health apart from that induced by distance from the end of life, and nondecreasing costs to forgoing investment in health with age.

Proposition 3. Suppose we have, as in Grossman (2000), a law of motion for health capital, with H_t the stock of health capital in period t, $\delta_t \in (0,1)$ the rate of depreciation between t and $t+1$, and I_t gross investment in health capital in period t,

$H_{t+1} - H_t = I_t - \delta_t H_t$,

with investment occuring over a finite lifespan so that $t = 1, ..., T$. Then the net present value of the return on gross investment in health capital in period t, defined

$NPV_t = \sum_{s=t}^{T-1} \frac{\partial H_{s+1}}{\partial I_t}$,

falls at an increasing rate as $t \to T$.

Proof. It is easy to show that $NPV_t = 1 + \sum_{r=t}^{T} \prod_{s=t}^{r} (1-\delta_s)$. Working back from period T, we can find that $NPV_{T-1} = 1$, and $NPV_{T-2} = 1 + (1-\delta_{T-2})$, so that $NPV_{T-1} - NPV_{T-2} = -(1-\delta_{T-2}) < 0$, and $NPV_{T-2} - NPV_{T-3} = -(1-\delta_{T-2})(1-\delta_{T-3}) < 0$, which is smaller in absolute value than the decrease in net present value from $T-2$ to $T-1$ since $\delta_t \in (0,1)$ for all t. The rest follows by induction.■

In consequence, we know that if we denote a general function $g(t)$ to mean the net present discounted value of investing in health at age t, assuming that the costs of investment in health are nondecreasing in t, then $g'(t) < 0$ and $g''(t) < 0$. Now suppose we have some latent reservation utility for individual i denoted $\varepsilon_i \sim N(0,1)$ for simplicity and the decision rule for whether we observe an individual making a particular investment in health at time t is $Y_{it} = 1[\varepsilon_i \leq -a_0 - a_1 D_{it} - g_i(t)]$, so that, denoting the cumulative distribution function of ε by F_ε,

$Pr(Y_{it} = 1) = F_\varepsilon(-a_0 - a_1 - g_i(t))$ if $D_{it} = 1$ and $F_\varepsilon(-a_0 - g_i(t))$ otherwise. Clearly $Pr(Y_{it} = 1)$ declines in t at an increasing rate regardless of the value of D_{it}. It follows that if, as in the set-up in the text, we have that $t_0 < t < t_k$ for those who have $D_{it} = 1$ and $D_{it-1} = 0$ and $t_k < t_0' < t < t_k'$ for all those with $D_{it} = D_{it-1} = 1$ (the only case by construction of the sample where $\triangle D_{it} = 0$), then the unobserved net present discounted value of investing in health will lead to a smaller value of $E[\triangle Y_{it} | \triangle D_{it} = 1] - E[\triangle Y_{it} | \triangle D_{it} = 0]$ if this difference is negative.