Rational Self-Medication*

Michael E. Darden
Johns Hopkins University and NBER

Nicholas W. Papageorge
Johns Hopkins University, IZA and NBER

January, 2019

Abstract

We develop a theory of rational self-medication. The idea is that forward-looking individuals, lacking access to better treatment options, attempt to manage the symptoms of mental and physical pain outside of formal medical care. They use substances that relieve symptoms in the short run but that may be harmful in the long run. For example, heavy drinking could alleviate current symptoms of depression but could also exacerbate future depression or lead to alcoholism. Rational self-medication suggests that, when presented with a safer, more effective treatment, individuals will substitute towards it. To investigate, we use forty years of longitudinal data from the Framingham Heart Study and leverage the exogenous introduction of selective serotonin reuptake inhibitors (SSRIs). We demonstrate an economically meaningful reduction in heavy alcohol consumption for men when SSRIs became available. Additionally, we show that addiction to alcohol inhibits substitution. Our results suggest a role for rational self-medication in understanding the origin of substance abuse. Furthermore, our work suggests that punitive policies targeting substance abuse may backfire, leading to substitution towards even more harmful substances to self-medicate. In contrast, policies promoting medical innovation that provide safer treatment options could obviate the need to self-medicate with dangerous or addictive substances. More broadly, our findings illustrate how the effects of medical innovation operate in part through behavior changes that are not measured in clinical trials.

JEL Classification: I10; I12;

Keywords: Self-Medication; Depression; Mental Health; Dynamic System of Equations

*We thank Robert Kaestner, John Cawley, John Mullahy, Jon Skinner, Tom Mroz, Matthew Harris, Melinda Pitts, Peter Savelyev, and seminar participants at the Atlanta Federal Reserve Bank, University of Tennessee, 2018 Workshop on the Economics of Risky Behavior, the 2018 American Society of Health Economists Conference, the 2018 European Health Economics Association Conference, IUPUI, the Southeastern Health Economics Working Group, and UNC-Chapel Hill. The Framingham Offspring Study (FOS) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the FOS Investigators. This manuscript was prepared using a limited access dataset obtained from the NHLBI and does not necessarily reflect the opinions or views of the FOS or the NHLBI.

1 Introduction

Beginning with Grossman (1972), economists have envisioned health as a form of human capital that increases survival rates, raises productivity, and improves the quality of life. Accordingly, behaviors that can improve health, such as exercise, healthy eating, abstaining from risky behavior, or medication usage, can be viewed as costly investments in human capital. Rational individuals invest in their health until the long-term benefits of doing so cease to outweigh the upfront costs. This basic model has been expanded upon to incorporate the realities of many health-related decisions. Examples include uncertainty and learning about how well a drug will work (Crawford & Shum, 2005), treatment decisions when faced with an acute illness (Gilleskie, 1998), and addiction that encourages use of harmful substances (Darden, 2017), among others.

An overlooked idea is that many individuals, lacking access to *good* medication, may rationally choose to take matters into their own hands, turning to substances that are potentially harmful in the long-run (e.g., alcohol or opioids) in an effort to manage short-run symptoms of illnesses, such as chronic pain or depression. Consistent with the basic Grossman framework, individuals who use harmful or addictive substances can be seen as rationally choosing to *self-medicate*; that is, they optimally make use of available technologies, some of which have drawbacks, in order to alleviate symptoms, albeit at the risk of future poor health, addiction, and other negative consequences.[1] Understanding how, and under what circumstances, people self-medicate is important because self-medication is socially costly, especially if it leads to addiction. However, treating use of dangerous substances as an error in judgment or an act of desperation can lead to the wrong policy conclusions. For example, viewing problem drinking as purely irrational behavior suggests policies to curb drinking. Viewing it as rational self-medication raises the possibility that such policies could backfire if, for example, people substitute to substances that are even more harmful. A better policy response would be to promote treatment innovations that obviate the need to self-medicate and thus induce rational actors to substitute towards less harmful substances.

In this paper we test the rational self-medication hypothesis. In particular, we ask whether the emergence of effective medication obviates the need to self-medicate with riskier substances. In the case we study, we leverage a technological advancement — the 1988 Food and Drug Administration (FDA) approval of Selective Serotonin Reuptake Inhibitors (SSRIs) — as an exogenous change in the choice set for the management of depression. Rational self-medication predicts that following

[1] It is important to note that there are two definitions of self-medication. One encompasses any self-administered medication to alleviate illness and includes, for example, taking aspirin for headaches. A second definition is limited to the use of potentially dangerous substances in order to alleviate symptoms. While the first is often used in medical literature, the second is more aligned to a layperson's notion of self-medication and is also discussed at length in medical and psychological literature. For example, Khantzian (1985) introduces the concept of self-medication, in which an individual manages her ailment outside of formal prescription medicine or therapy.

the introduction of new medications, the use of riskier treatment alternatives should decline. If heavy drinking is in part a form of self-medication, we predict that heavy alcohol consumption should fall following the introduction of SSRIs.[2] If we are unable to detect such substitution patterns as better medications emerge, heavy drinking is less likely to be a form of self-medication. Broadly, this analysis illustrates a central contribution of health economics, which is to move beyond quantifying the direct impacts of new medicines (e.g., treatment effects on health or the harms of risky substances) by incorporating additional factors, such as uptake and compliance decisions along with substitution patterns in other potentially relevant health behaviors. In the context we examine, if alcohol is used to self medicate, a potentially overlooked social benefit of SSRIs is a reduction in heavy drinking.

Depression is an ideal context to study self-medication through alcohol for several reasons. One, it is prevalent. In the United States, Major Depressive Disorder (which we simply refer to as depression unless the meaning is unclear) affects 8.1% of individuals over the age of 18. Two, while alcohol is not recommended for the treatment of depression, it is well-understood to be a highly effective way to treat immediate symptoms of depression, which makes it an attractive option for people who lack alternatives (Khantzian, 1990). Three, depression affects many facets of life, including human capital accumulation, productivity, family structure, risky behaviors, and employment, along with other physical health outcomes, such as cancer, cardiovascular disease, and diabetes. Therefore, it is little surprise that individuals would engage in potentially costly attempts to alleviate their immediate symptoms. Four, there is massive stigma surrounding mental health treatment, which might make self-medication via a socially-acceptable behavior, such as drinking, an attractive option.[3] Finally, and key to our empirical work, there are large changes in treatment options over time, in particular the emergence of SSRIs, which replaced earlier drugs that, while effective, had massively adverse side effects that precluded widespread use.

To begin our analysis of self-medication we formalize the concept with a simple two-period model in which an agent makes health investment decisions, jointly choosing alcohol and antide-

[2] In Section 2, we document a strong correlation between depression and alcohol consumption using NHANES data, and we review the significant literature on alcohol self-medication. For example, Bacolod et al. (2017) study minimum drinking age laws and show that the largest increase in drinking at age 21 (for those in the military) comes from the most depressed.

[3] Another issue, which we do not explore given a relatively homogeneous sample, is that prevalence of depression is heterogeneous across socio-economic groups. Depression is about four times more likely for poor versus non-poor individuals. For those below 100% of the Federal Poverty Line (FPL), the rate was 15.8% between 2013 and 2016, while the rate was only 3.5% for those at or above 400% of the FPL (Brody et al. , 2018). This is especially concerning in the context of self-medication if low-income individuals have less access to medical care, safer medications, or treatment options, such as therapy. Moreover, low-income individuals may face other challenges that encourage use of addictive substances, compounding the risks of self-medication.

pressant medications to maximize utility.[4] Poor mental health generates symptoms which reduce utility. Health investments have contemporaneous affects on symptoms along with inter-temporal effects on the stock of mental health. In our case, alcohol relieves current-period symptoms, but may also exacerbate future mental health problems, which produce symptoms. The model also permits the possibility that substances, such as alcohol, are enjoyable in their own right. The key factor underlying self-medication is a complementarity: the current-period marginal benefit of substances rises with symptoms of illness. One way to achieve this is that alcohol is more effective at reducing more severe symptoms. Alternatively, if the utility cost of symptoms is larger as symptoms increase, the same reduction via alcohol has a larger utility benefit. In either case, the complementarity between alcohol and immediate symptoms of depression generates the following prediction: SSRIs lower symptoms and thus the marginal benefits of improving symptoms, which dis-incentivizes alcohol usage.[5]

To investigate self-medication empirically, we use data from the Framingham Heart Study Offspring Cohort. The data set includes longitudinal information on alcohol, tobacco, and antidepressant consumption, as well as depression measures for roughly 5,000 individuals over a forty-year period. Exploiting the arrival of SSRIs, we estimate a series of differences-in-differences models to provide strong *prima facie* evidence of substitution away from alcohol and towards antidepressants once they come available. Estimates suggest that taking an antidepressant is associated with a statistically significant 3.9 percentage point (12.5%) increase in abstinence from alcohol. Effects are stronger for men and potentially concentrated among individuals with moderate depression. The latter finding underscores the self-medication hypothesis since it suggests that, until better options emerge, alcohol is an effective way to combat depression.[6] Simple regressions ignore potentially important dynamics, including the stock of addiction, which could affect how costly it is to switch from alcohol to SSRIs. They also ignore initial conditions, such as a long history of heavy

[4] Our model formalizes the argument that the type of substance being used depends on the type and severity of mental health ailment (Khantzian, 1985).

[5] The notion that harmful substance use is explained through a complementarity is similar to Becker & Murphy (1988), who model dynamic complementarities in the marginal utility of consumption as a necessary condition for addiction. We discuss further links to this paper below.

[6] In interpreting empirical results, we note that substitution away from alcohol resulting from the emergence of new medications may not reflect use of alcohol as self-medication *per se*, but could instead reflect doctors' recommendations that the two not be used together. If so, substitution away from alcohol amounts to giving up an enjoyable good in order to take medication that relieves symptoms of physical or mental health problems. In the case of depression, however, this alternative interpretation is less clearly distinct from self-medication. Symptoms of depression include sadness and so engaging in behaviors that one enjoys, such as alcohol, is a (risky) way to alleviate sadness. Using alcohol to feel better is therefore not inconsistent with our notion of self-medication. Again, the key idea is that the marginal benefit of using alcohol is higher for those who are depressed. This is another reason why depression is a good context to study self-medication. It is also worth noting that use of both SSRIs and alcohol is widespread, and, for depressed individuals with a strong preference for alcohol, SSRIs may have *increased* alcohol consumption as their interaction is significantly less risky than with previous generation antidepressants.

alcohol use that could lead to depression. This could generate a correlation between alcohol use and depression that is not explained by rational medication. To address these issues, along with others, such as selective attrition and mortality, we augment our analysis to estimate a system of dynamic equations which approximates a more general structural model. Specifically, we estimate dynamic equations for alcohol, tobacco, and antidepressants jointly, along with depression, attrition, and mortality equations.[7]

Estimates from the model incorporating dynamics and unobserved heterogeneity are generally consistent with findings from basic regressions.[8] We use the estimated model to perform two sets of counterfactual policy simulations. First, we impose antidepressants on the entire sample relative to our baseline simulation. Heavy drinking declines by 3.4 percentage points, primarily driven by men. Moreover, while we show that the reduction in heavy drinking is largest in those simulated to be moderately depressed, we find no change in heavy alcohol consumption, in any period, for those simulated to be in the highest tercile of depression. Persistence of alcohol use among the most depressed could reflect additcion. To investigate the role of addiction, our second simulation sets lagged alcohol consumption to zero in the contemporaneous alcohol demand equation, regardless of simulated behavior in the previous period. Overall, regardless of gender or mental health, heavy alcohol consumption drops enormously. Antidepressant usage (which is chosen endogenously in this simulation) increases by 5.5 percentage points by the final exam of FHS, and the magnitude of this substitution is increasing in depression severity. We interpret these results to suggest that alcohol addiction may significantly hinder substitution away from alcohol. Finally, we demonstrate that the simulated reduction in heavy drinking is equivalent to a roughly 10% increase in alcohol prices. Together, our results exploiting a large medical innovation provide compelling evidence of self-medication. When introduced to a new and better medical technology, individuals who self-medicate substitute towards it.

Our work relates to a large medical literature on self-medication, which has generally reported cross-sectional correlations. For example, Bolton et al. (2009) recognizes that the direction of causality between alcohol abuse and mental health problems is unclear. We also contribute to a literature in health economics that moves beyond assessing the direct effects of medical innovation (e.g., lower mortality and better health) to incorporate a more complete set of indirect effects. This type of work is a crucial complement to findings from clinical trials, which measure treat-

[7] We allow for correlation in the permanent component of the error structure across equations to capture unobserved heterogeneity in the joint determination of these behaviors and outcomes (Heckman & Singer, 1984; Mroz, 1999). The empirical framework is similar to the dynamic seemingly unrelated regression (SUR) model in Darden et al. (2018), who use FHS data to study the effect of cigarette smoking on expected longevity.

[8] It is worth noting that there are some important differences in estimated relationships between depression, medication usage and alcohol that affect policy conclusions. These differences are discussed when we present empirical results.

ment effects under controlled conditions, but are ill-suited to analyze additional relevant factors, such as changes in other health behaviors and impacts on longer-run lifecycle outcomes, (e.g. employment), all of which contribute to the full social impact of medical innovation. For example, Papageorge (2016) shows that an important benefit of new HIV treatments emerging in the mid-1990s was to raise productivity and increase labor supply. Conversely, Kaestner et al. (2014) shows evidence of technological substitution away from diet and exercise with the introduction Statin pharmaceuticals introduced to combat cholesterol. Failing to account for this would lead to over-estimation of their net social value. In our case, to the extent that alcohol consumption is a form of self-medication that harms health, the net benefit of SSRIs on long-term mental health has likely been understated because randomized trials do not account for long-term shifts in alcohol consumption.

This paper also contributes to our understanding of addiction. In the seminal paper on rational addiction, Becker & Murphy (1988) posit that under addiction, a person has a low level of utility while addicted but a high marginal utility of usage of addictive substances, which incentivizes continued use. While the model explains why forward-looking and addicted individuals continue to use an addictive substance, it is silent on why they would ever become addicted in the first place. Our paper suggests one possible reason. Initial usage of an addictive substance need not be an error in judgment or due to lack of perfect foresight or a large exogenous shock. An individual in pain may assess the probability of future addiction and rationally medicate her pain with available technology, fully aware that doing so can lead to a Becker-style addictive spiral with some probability. Moreover, providing evidence of rational self-medication has implications for understanding the dramatic increase in mortality rates of white non-Hispanic men since 1998, the so-called "Deaths of Despair" documented in Case & Deaton (2015). However, whereas "despair" technically suggests a lack of hope, self-medication suggests the opposite: heavy alcohol use or addiction may reflect an earlier, rational and hopeful attempt to medicate away pain.[9] If so, the appropriate policy response is to stop punishing people who use risky substances to self-medicate and instead work to develop treatments that are less addictive so that people can rationally substitute away from harmful self-medicating behavior.

This paper proceeds as follows. In Section 2, we provide background on depression and depression treatment, as well as the literature on self-medication. In Section 3, we discuss a simple, two-period theoretical model of rational self-medication. In Section 4, we present our main data, the Framingham Heart Study, and we document empirical evidence of a plausibly causal relation-

[9]According to the online etymology dictionary, "despair" comes from the French-Anglo *despeir*, originally the French *despoir*, referring to "hopelessness" or a "total loss of hope." See https://www.etymonline.com/word/despair.

ship between antidepressants and alcohol consumption. Section 5 presents our dynamic model, as well as parameter estimates, model fit, and simulation results. Section 6 discusses our results and Section 7 concludes.

2 Background

Depression is a chronic mental health condition that, while highly treatable, is the leading cause of disability globally[10]. Depression produces symptoms that include feelings of sadness, pessimism, guilt, and anxiety, while also causing decreased energy, loss of interest in daily activities, and indecisiveness. Clinical diagnosis of Major Depressive Disorder (MDD) includes a set of daily symptoms plus some functional impairment with respect to family and peer relationships, school/work performance, and stress and anxiety level (EA et al. , 2009).[11] In the United States, in any given two-week period between 2013 and 2016, 8.1% of Americans suffered from depression, ranging from 5.5% for men to 10.4% for women, and there exists a strong gradient between depression and income: 19.8% of women earning less that 100% of the Federal poverty line (FPL) exhibit depressive symptoms compared to only 4.8% of women at or above 400% of the FPL (Brody et al. , 2018).

Unsurprisingly, depression is associated with a wide variety of mental and physical ailments, including sleep problems, irritability, persistent physical pain, and risk of suicide (U.S. HHS, 2015). Beck et al. (2011) show that depression is associated with significantly lower fundamental economic building blocks such as workforce productivity, which they measure with the Work Productivity and Activity Impairment Questionnaire, and Berndt et al. (1998) demonstrate that depressed workers have lower levels of perceived at-work productivity and performance. Furthermore, Kessler (2012) shows that depression is associated with low educational attainment, teen pregnancy, martial disruption, unemployment, functional status, early mortality, and suicide. Unsurprisingly, there is a strong correlation between depression and alcoholism.[12] Indeed, Figures 1a. and 1b. present National Health and Nutrition Examination Survey (NHANES) data on heavy alcohol consumption for men and women by the tertile of the Patient Health Questionnaire (PHQ-9) depression score between 2007 and 2013. For both men and women, more severely depressed individuals are persistently and significantly more likely to engage in heavy alcohol consumption.

[10] http://www.who.int/en/newsroom/fact-sheets/detail/depression

[11] In the middle 20th century, anxiety was the leading mental illness in the United States. Horwitz (2010) describes how, through a series of reclassifications, as well as the introduction of SSRIs, anxiety has given way to a focus and prevalence of depression.

[12] For example, see Bolton et al. (2009), who use nationally representative survey data from the National Epidemiologic Survey on Alcohol and Related Conditions to document cross-sectional correlations between alcohol and drug use and a variety of mental health conditions.

Depressed individuals have a clear incentive to manage, maintain, and improve mental health. Antidepressant pharmaceuticals have existed since the initial Monoamine Oxidase Inhibitors (MAOI) developed in the 1950s. Most antidepressants function by preventing or slowing the re-uptake of neurotransmitters (such as Serotonin) in the brain, without which depression is more likely. MAOI antidepressants were effective at relieving symptoms of depression, but these, along with Tricyclic antidepressents (TCA) developed in the 1960s, prevent reuptake of many types of neurotransmitters, not only those that regulate mood, and the associated side effects of MAOIs and TCAs include risk of stroke, cardiovascular ailments,and sexual dysfunction, among others. Reflecting these side effects, which prevented certain groups from using antidepressants (e.g., the elderly), as well as public stigma associated with antidepressants, only 2-3% of Americans used an antidepressant through the middle 1980s.[13]

A major advancement in the management of depressive symptoms came with the 1988 FDA approval of Selective Serotonin Reuptake Inhibitors (SSRIs), which, as the name suggests, effectively inhibit the re-uptake of serotonin selectively, making more serotonin available in the brain without affecting the levels of other neurotransmitters. SSRIs significantly altered the perception of antidepressants, reducing stigma, and expanding the set of individuals for whom an antidepressant is considered safe.[14] As a result, rates of antidepressants have increased dramatically since 1988 — up to 12.7% of Americans were prescribed an antidepressant between 2011 and 2014, and of those taking an antidepressant, 25.3% have been taking an antidepressant for more than 10 years (Brody et al. , 2018). Researchers now use SSRI prescriptions to *gauge* the rates of depression, mental health, and happiness. For example, Blanchflower & Oswald (2016) study the well-known u-shaped well-being curve with respect to age and show a similar pattern between antidepressants and age.

To summarize, major depressive disorder is the most common mood disorder in the United States, affecting over 16.2 million adults in 2016. SSRIs significantly expanded the choice set with respect to the management of depression, which is frequently medicated outside of the medical system with potentially harmful and addictive substances. These endogenous investments into the mental health production function may have important implications for a variety of outcomes, including labor market productivity and long-term health. Finally, while earlier literature has documented that self-medication likely occurs, studies are cross-sectional and generally do not empirically address causality or the dynamic implications of endogenous investments via self-

[13]See Hillhouse & Porter (2015) for an excellent overview of the history on antidepressants.

[14]Despite a significant literature that relates SSRIs to teen suicide, Ludwig et al. (2009) shows that SSRIs actually reduces suicides across 25 countries after controlling for the intuitive selection of depressed individuals into antidepressant use.

medication, such as addiction. These issues are the topic of our study.

3 Theory

A significant body of work in psychology, medicine, and public health studies the management of depressive symptoms outside of formal prescription drugs, a hypothesis known as self-medication. Khantzian (1985) introduced the idea that the kind of substance used to self-medicate is not random but depends on the type of illness, and that those in states of pain will experiment with different types of substances, some of which may lead to addiction. While the application of Khantzian (1985) was on self-medication with hard drugs (heroin and cocaine), Khantzian (1990) extended the notion of self-medication to the consumption of alcohol, which he described as "a means to achieve and maintain self-regulation." This presentation of self-medication connects "intense affects, such as rage, shame, loneliness, and depression" with the "use of drugs and alcohol to cope with these emotions."

In the absence of safe medication for depression (historically), significant cross-sectional survey evidence suggests that depressed individuals consume alcohol to cope with the symptoms of depression. For example, Crum et al. (2013) show that mental illness is a significant rationale for alcohol consumption and that coping with depressive symptoms with alcohol is associated with the development of alcohol use disorders. Indeed, the consumption of alcohol induces short-term anxiolysis, which produces feelings of relaxation. Deykin et al. (1987) were the first to demonstrate that major depressive disorder typically predates alcohol use disorders in adolescents, providing some evidence on the direction of causality for the robust and pervasive correlation between heavy alcohol consumption and depression.

Before proceeding to our empirical analysis, we formalize the Khantzian (1985) notion of self-medication in a two-period model of behavior.[15] Importantly, our model differentiates between the short-term symptoms of depression, for which alcohol may be therapeutic, and the long-run mental health consequences of alcohol consumption. Specifically, in our model, depression causes symptoms which draw from utility in a consumption sense, and antidepressants and alcohol potentially alleviate contemporaneous symptoms.[16] While our model is similar to Kaestner et al. (2014), who study disease-specific (cholesterol drugs) and non-disease-specific (diet and exercise) behaviors after the introduction of Statin pharmaceuticals, we focus on the discrete choice to take antidepressants and the intensive margin of alcohol consumption. Whereas antidepressants do not have any inter-temporal effects in our model, contemporaneous alcohol consumption may worsen

[15] Our model is similar to Becker (2007), who distills the Grossman (1972) model into a two-period framework.

[16] We abstract from any investment rationale for the management of mental health with respect to outside productivity.

future depression. Because SSRIs represent an improvement in the side-effects of antidepressants, rather than an improvement in effectiveness with respect to the symptoms of depression, their introduction encourages use by lowering the marginal cost of antidepressants. Our model demonstrates that a convex cost of symptoms in the utility function is sufficient for SSRIs to induce a reduction in alcohol consumption.

Agents solve a two-period problem, where periods are denoted t and $t+1$. Where possible, we drop time subscripts and denote $t+1$ variables with a "prime". An agent enters period t with state variable M_t, which is the stock of mental health and where lower values of M_t imply worse mental health. Agents choose whether or not to take an antidepressant, denoted $D_t \in \{0,1\}$ and how much alcohol to drink $A_t \in \mathbf{R}^+$. For ease of exposition, we assume that the agent chooses non-zero alcohol consumption.

Agents have preferences over alcohol consumption A and antidepressant consumption D, where the latter includes the price of antidepressants along with side effects, stigma and other non-pecuniary costs of SSRI use. They do not have preferences over mental health *per se*, but instead over symptoms of mental health S. Agents choose A and D to solve:

$$\max_{A_t, D_t} \Big(u(S_t, A_t, D_t) + \beta v(S') \Big) \tag{1}$$

where we assume that S and D enter negatively and A enters positively into both u and v. Period $t+1$ is effectively a "terminal" period in which no decisions are made and $v(S')$ is thus a continuation payoff affected by period-t choices which thus provides dynamic incentives to improve mental health.

Mental health evolves according to the following production function

$$M_{t+1} = f_m(M_t, A_t, D_t) \tag{2}$$

where the argument M_t captures persistence in mental health stock, A_t captures how alcohol usage can have negative impacts on future mental health, perhaps through increases in history of alcohol terms, and D_t captures how antidepressants can improve long-run mental health. Period-t symptoms are a function of the same arguments so that:

$$S_t = f_s(M_t, A_t, D_t) \tag{3}$$

where symptoms are more likely to occur when M_t is lower. Alcohol can improve symptoms, which is the "self-medication" effect, and antidepressants can also improve symptoms.

To characterize self-medicating behavior, we use the model to make the following three points.

First, we show conditions under which $D^\star = 1$. Second, we characterize optimal alcohol usage. Finally, we discuss conditions under which lowering the costs associated with antidepressant usage — through the approval of SSRIs — would lead to decreases in alcohol usage. The third point is consistent with a reduction in self-medication through alcohol when medication becomes a more attractive option.

To show optimal antidepressant usage, denote optimal alcohol consumption A^* and A^{**}, when using antidepressants and not using antidepressants, respectively. Agents use antidepressants when the benefits of doing so exceed the costs:

$$\begin{aligned} u(S(D=1), A^*, D=1) \quad &+ \quad \beta v(S'(M'(D=1))) \geq \\ u(S(D=0), A^{**}, D=0) \quad &+ \quad \beta v(S'(M'(D=0))) \end{aligned} \tag{4}$$

To fix ideas, suppose we make the simplifying assumption on period-t utility that the costs of medication usage are additively separable from other utility components, e.g., $u(S_t, A_t, D_t) = \tilde{u}(S_t, A_t) - \phi(D_t)$ where $\phi(D_t = 1) = \phi$ and $\phi(D_t = 0) = 0$.[17] The agent uses antidepressants if and only if

$$\begin{aligned} \tilde{u}(S(D=1), A^*) + \phi \quad &+ \quad \beta v(S'(M'(D=1))) \geq \\ \tilde{u}(S(D=0), A^{**}) \quad &+ \quad \beta v(S'(M'(D=0))) \iff \\ \tilde{u}(S(D=1), A^*) - \tilde{u}(S(D=0), A^{**}) \quad &+ \quad \beta[v(S'(M'(D=1))) - v(S'(M'(D=0)))] \geq \phi \end{aligned} \tag{5}$$

The last line implies that the benefits must outweigh the costs in order for antidepressant usage to occur, where the benefits include current period utility of fewer symptoms along with discounted $t+1$ reductions in symptoms due to increased mental health stock. For a given level of antidepressant effectiveness, antidepressant usage increases if the flow utility costs decline, e.g., through side effects, stigma or price reductions. Moreover, as long as $\phi > 0$, antidepressant usage only occurs if there are benefits in the form of improved symptoms, either currently or in the future.

Next, we characterize optimal alcohol consumption, in which the relevant first order condition is:

$$\frac{\delta u}{\delta S}\frac{\delta S}{\delta A} + \frac{\delta u}{\delta A} + \frac{\delta v}{\delta S'}\frac{\delta S'}{\delta M'}\frac{\delta M'}{\delta A} = 0 \tag{6}$$

or

$$\frac{\delta u}{\delta A} + \frac{\delta u}{\delta S}\frac{\delta S}{\delta A} = -\beta\frac{\delta v}{\delta S'}\frac{\delta S'}{\delta M'}\frac{\delta M'}{\delta A} \tag{7}$$

[17] Additive separability implies that the marginal utility of alcohol is unaffected by SSRI usage. While this assumption is unrealistic, it simplifies the exposition for optimal SSRI usage, and it does not affect our comparative dynamics analysis presented below.

The left hand side captures the marginal benefits of alcohol use, including both the enjoyment of alcohol along with reduction in symptoms from self-medicating. The right hand side captures marginal costs: higher A reduces M' and lower M' reduces continuation payoffs captured by v. Optimal alcohol usage occurs when the marginal benefit of an additional unit of A is equal to the marginal cost.

Finally, we use our simple model to derive conditions under which antidepressant usage should lead to decreases in alcohol usage. It is convenient to define a function for the marginal utility of side effects for both periods as follows:

$$\frac{\delta v}{\delta S} = \frac{\delta u}{\delta S} \equiv \alpha(S) \tag{8}$$

For example, if $\alpha(S) = \alpha S$ and $\alpha > 0$, then utility is a concave function with increasingly negative marginal utility of S. Having done this, the first-order condition above can be rewritten as:

$$\frac{\delta u}{\delta A} = -\alpha(S)\left[\frac{\delta S}{\delta A} + \beta\frac{\delta S'}{\delta M'}\frac{\delta M'}{\delta A}\right] \tag{9}$$

If alcohol usage decreases with SSRIs, it must be the case that SSRIs lead to a decline in the left-hand-side of the last equation or an increase in the right-hand-side. We do not allow the enjoyment of alcohol to be a function of symptoms, so the left hand side does not change. Thus, for SSRIs to lower alcohol usage, it must be the case that the right hand side rises or that -1 times the right hand side falls. Thus, to understand reduced self-medication in the form of drinking, we examine why the following expression should decline when symptoms decline:

$$\alpha(S)\left[\frac{\delta S}{\delta A} + \beta\frac{\delta S'}{\delta M'}\frac{\delta M'}{\delta A}\right] \tag{10}$$

There are four possibilities:

1. $\alpha(S)$ is lower when $D = 1$. Given that utility is a declining function of S, this suggests that costs of S rise with S. The implication is that medication leads to a decline in symptoms. This reduces the marginal cost of symptoms, which means that the marginal benefit of technology that reduces symptoms is lower.

2. A second possibility is that $\frac{\delta S}{\delta A}$ is lower when $D = 1$. This could occur if alcohol is less productive at reducing symptoms at lower symptom levels.

3. The third possibility is that $\frac{\delta S'}{\delta M'}$ is smaller when $D = 1$. This means that improvements to mental health reduce symptoms more so when mental health is better.

4. Finally $\frac{\delta M'}{\delta A}$ is lower when $D = 1$ which suggests that alcohol reduces future mental health more so if mental health is better.

Which of these is true is difficult to pinpoint using the data we have. This means that there are several possible dynamics that could underlie self-medication. Still, some of our empirical work provides some guidance on which mechanism is more likely to help explain self-medication. We return to this point when discussing our estimates. We now turn to our empirical investigation of self-medication.

4 The Framingham Heart Study

To study self-medication empirically, we turn to the Offspring Cohort of the Framingham Heart Study (FHS). The Offspring Cohort data are ideal for our purposes as they include longitudinal information on alcohol, tobacco, antidepressant medication, and mental health over nine detailed health exams over 40 years. Begun in 1971, the Offspring Cohort includes roughly 5,000 offspring of the FHS Original Cohort, which began in 1948 in Framingham Massachusetts, and their spouses. Both cohorts of individuals have received detailed health examinations at 2-4 year intervals into the 21st century, and both cohorts have made significant contributions to the understanding of cardiovascular disease.[18]

Participants range from 13 to 62 years of age at the first exam, which reflects the wide age variation in the Original Cohort. The Original Cohort restricted its sampling to white residents of Framingham Massachusetts, and, while no restriction was placed on the ethnicity or residency the spouses of the offspring, data are not available on these characteristics. As the FHS was not meant to be representative of any larger population, we restrict our final estimation sample to 2,497 individuals for whom we have consistent exam participation and information.[19] To enter our sample, an individual must have completed exams one through three and must not have skipped exams in the subsequent periods. Following the third exam, individuals may leave the sample through either death or attrition. Because of an eight year gap between exams one and two, and because of data limitations discussed below, we restrict our analysis to exams two through nine. All FHS Offspring participants completed exam two between 1979 and 1983.

Table 1 presents summary statistics of the Offspring Cohort at our initial exam (exam two) by gender and by whether an individual is ever, over the subsequent seven exams, observed to be on any type of antidepressant. Of the 1,241 men in our sample, 12.17% are observed at some point to

[18]See Mahmood et al. (2014) for a detailed history of the Study. See Darden et al. (2018) and Darden (2017) for economic studies of the Original and Offspring Cohorts, respectively.

[19]Kaestner et al. (2014) and Darden (2017) construct very similar samples from FHS Offspring Data.

be taking antidepressants; for women, the percentage ever taking antidepressants is significantly higher at 24.52%. The FHS asks respondents the number of 12oz beers, 5oz glasses of wine, and 1.5oz liquor drinks they typically consume per week. We aggregate these to a drinks per week measure, and we follow the National Institute on Alcohol Abuse and Alcoholism guidelines for light and heavy alcohol consumption based on gender: light drinking is defined as up to seven drinks per week for women and 14 drinks per week for men; heavy drinking is any number above the gender-specific thresholds.[20] At the second exam, men drink more heavily than women (despite the higher threshold for heavy drinking), and rates of heavy drinking are higher for those ever-observed to take an antidepressant (although these differences are not statistically significant). Generally, there are not statistical differences between ever and never antidepressant users, although a notable exception is cancer and mortality incidence for women, which are both statistically higher among the never users, despite the fact that women taking antidepressants are more likely to smoke.

At exam three, Offspring Cohort participants took the Center for Epidemiological Services - Depression (CES-D) test for depression, which aggregates 20 clinically verified depression questions (each on 0 to 3 Likert Scale) into a depression summary score (Radoff, 1977).[21] We break the continuous depression score at exam three into tertiles, and we present the fraction of individuals in each exam three tertile by gender and whether they are ever observed to take an antidepressant in the last three rows of Table 1. Not surprisingly, the fraction of both men and women in higher CES-D tertiles are higher for those who go on to take an antidepressant, but we emphasize the sizable fraction of those in the lowest tertile of depression in exam three who eventually use antidepressants as foreshadowing of the heterogeneity results presented below.[22] Importantly, antidepressants are prescribed for a wide variety of conditions other than depression, including bipolar disorder, bulimia, fibromyalgia, insomnia, PTSD, and social anxiety disorder (CMS, 2013).

Table 2 shows means and proportions of key variables over the eight exams. Each FHS exam was administered within a three to four year window, and, while we do not have information on the date that an individual took an exam, Table 2 displays the year ranges in which all participants completed each exam. Unfortunately, we do not observe antidepressant medication usage at exam two, however, the absence of this information likely stems from the observed trends in their use: at exam three, only 1.0% of men and 2.1% of women used antidepressants. Importantly, exam three was completed prior to 1988, when the FDA approved SSRIs, after which antidepressant medication usage grows considerably within our sample over time for both men and women. Light

[20]NIAAA. Accessed on November 7th, 2018.

[21]The clinically verified threshold for depression is any score at or above 16.

[22]Wulsin et al. (2005) use FHS Offspring Cohort data to relate the exam three CES-D score to future health outcomes. They find that, relative the lowest tertile, CES-D score is statistically related to all-cause mortality but not coronary heart disease.

and heavy alcohol use decline over our sample period and cigarette smoking plummets. Between exams two and nine, we lose roughly 48% and 38% of men and women, respectively, to sample attrition or death.

Figure 2 demonstrates trends in alcohol and smoking behavior over time by whether an individual is ever observed to take an antidepressant. Prior to 1988, antidepressants were quite rare in our data, but, as shown above, following the approval of SSRIs, antidepressant use grew rapidly. Relative to those never taking an antidepressant, Figure 2 demonstrates relatively parallel trends in both alcohol and tobacco consumption prior to 1988 and potentially important deviations from trend after 1988 for alcohol abstinence and light drinking.[23]

To test the rational self-medication hypothesis that consumption of risky goods should decline following an improvement in the choice set of treatment options, we begin by regressing binary indicators for never, light, and heavy drinking on a binary variable for antidepressant usage at a given exam. Equation 11 presents our baseline empirical specification,

$$y_{it} = \mu_i + x_{it}^{'}\beta + \delta d_{it} + \theta_t + \epsilon_{it}, \tag{11}$$

where y_{it} is risky behavior y for person i in year t, μ_i represents an individual specific effect, x_{it} are time-varying individual characteristics characteristics, θ_t are exam binary variables, and ϵ_{it} is an i.i.d. error component. Our variable of interest is d_{it}, which equals one if person i in exam t is taking an antidepressant. Table 3 presents results from Equation 11, in which we estimate separate linear probability models for never, light, and heavy drinking, as well as whether an individual smokes cigarettes.[24] Because antidepressants are unobserved in exam 2, we estimate Equation 11 on data from exams three through nine. For each alcohol measure, Table 3 presents both an estimate of δ, as well as interaction terms between d_{it} and gender and exam three CES-D tertile.[25] Results presented in Table 3 are conditional on age, education, and other health metrics, including blood pressure, obesity, cardiovascular disease, cancer, and exam fixed effects.[26] Standard errors

[23]Trends in behaviors in Table 2 and Figure 2 reflect both changing behavior and the changing composition of the sample, which we emphasize below in our dynamic system of equations model.

[24]Multinomial and ordered logit estimators yield similar results as those from estimation of Equation 11 without the individual fixed effect, μ_i. Because of the incidental parameters problem associated with logit estimators with fixed effects, we focus on linear probability models in this section.

[25]Our focus on the exam 3 CES-D score is for two reasons. First, exam three took place between 1983 and 1987, just before the introduction of SSRIs. Thus, we consider the exam three score to be a baseline metric of depression, prior to the improved technology. Second, unfortunately, FHS only conducted the CES-D test in exams three, six, seven, and nine. Estimates of our dynamic system of equations model proved to be erratic when we we attempted to model the time-varying metric of depression while integrating over missing years. The medium category is associated with a CES-D score between 5 and 10; high category is associated with a score between 11 and 51.

[26]Throughout our paper, our results are not sensitive to the inclusion of these endogenous health outcomes.

are clustered at the individual level.

Table 3 demonstrates preliminary evidence of substitutability between antidepressants and alcohol. The columns labeled with a 1 in Panel 1 of Table 3 present baseline estimates of Equation 11. When we focus our attention on within-individual variation, there exists evidence of substitution: participants are more likely to report no alcohol consumption (3.9 percentage points or 12.4%) when using an antidepressant, but the estimates are not precise enough to be able to distinguish reductions in light versus heavy alcohol consumption. Overall, we find no effect of antidepressants on smoking behavior.

To investigate heterogeneity in our results, column 2 for each respective behavior presents estimates of δ as well as interactions between antidepressants and binary variables for female and medium and high tertiles of the exam three CES-D score. Focusing on panel 1, column 2 results suggest that the overall 3.9 percentage point increase in alcohol abstinence is driven largely by a reduction in heavy drinking among men (8.6 percentage point decline) and a reduction in light drinking among those in the highest tertile of the exam three depression score. Panel 1 of Table 3, shows no significant effect of antidepressants on smoking behavior overall but provides suggestive evidence that antidepressants may prevent smoking cessation in men.[27]

Our fixed effects results take a causal interpretation if there is no time-varying unobserved heterogeneity that affects both the decision to take antidepressants and behavior. While we cannot test this assumption, we interact our time fixed effects with a binary variable for ever being observed to take an antidepressant. Conditional on contemporaneous antidepressant usage, time-varying individual unobserved heterogeneity would likely generate different trends in behavior. Furthermore, in the presence of time-varying unobserved heterogeneity, controlling for differential trends would likely significantly change the estimates on antidepressants. Panel 2 of Table 3 presents estimates of δ and the associated interaction coefficients while controlling for medication specific trends. At the bottom of panel 2, we present the p-values of the F-tests that the interacted trend variables are all zero. Consistent with Figure 2, none of the alcohol or smoking p-values suggest statistically significantly different trends, and the point estimates, while slightly attenuated, are not significantly different from those in panel 1. Panel 3 of Table 3 repeats this exercise but replaces the medication trend with separate sets of interactions between moderate exam three depression ($CES-D \in [5, 10]$) and high exam ($CES-D \in [11, 51]$) three depression and exam

Especially in the dynamic model presented below, we include these outcomes as controls as there is evidence that own health shocks may alter health behaviors. (Arcidiacono et al. , 2007)

[27]We estimate a smoking equation because we model smoking in our dynamic model that jointly estimates a mortality equation below. Using Behavioral Risk Factor Surveillance Survey data, Plurphanswat et al. (2011) document that a.) individuals with mental health problems smoke cigarettes at higher rates and b.) smoking likely causes a deterioration in mental health.

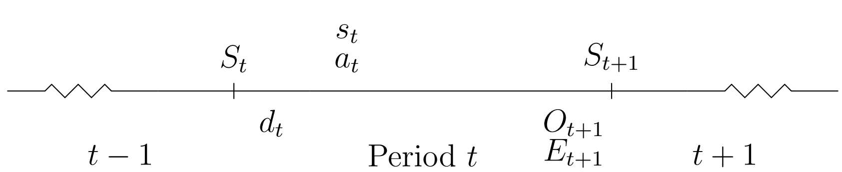
dummies with similar results. Exploiting purely within-individual variation, results from Table 3 demonstrate some evidence of self-medication — during a period in which medication for depression became much better and more common, antidepressants were associated with declines drinking, specifically for heavy drinking by men. [28]

5 Dynamic Empirical Model

Despite providing suggestive evidence of self-medication, results from Table 3 are problematic in several important ways. First, a large and growing empirical literature recognizes the inherent dynamics in addictive goods (Arcidiacono et al. , 2007; Darden, 2017), and Equation 11 is static in the sense that contemporaneous behavior is not allowed to depend on past behavior. The marginal utility of alcohol likely depends on past consumption, and the failure to model the dynamics of these behaviors will likely lead to an overestimate on the effect of antidepressants on behavior. Furthermore, while alcohol consumption may improve contemporaneous mental health, a large literature suggests that heavy alcohol consumption may harm future mental health. Second, the composition of our sample is changing over time through mortality and attrition. Especially because (i) the behaviors being modeled may cause mortality or attrition; and (ii) significant antidepressant medication usage is not observed until the end of our sample period, selective exits may significantly bias our results. Finally, estimation of each equation separately does not allow for correlation in unobserved heterogeneity across equations.

In the spirit of our two-period model presented above, and to address the limitations of our static empirical model, we estimate a dynamic system of equations for antidepressants, alcohol and tobacco consumption, sample attrition, and mortality. The empirical model is an approximation of a more general structural model of behavior and outcomes in which an individual optimally selects a bundle of investments in health, and health, both mental health and mortality, is a function of behavior. In what follows, we briefly outline the timing of our dynamic system of equations.

The following time line presents a representative exam period t of an individual's problem in which we suppress the individual subscript i for ease of notation:



[28]We have also estimated a specification consistent with the rational addiction literature (Becker et al. , 1994) in which we include both a lead and lag of the health behavior being modeled. Results are similar to those in Table 3.

Here, S_t captures the period t state vector, which sufficiently summarizes measures of past behavior. Given her state S_t, an individual begins period t by choosing whether or not to take an antidepressant, d_t. Conditional on d_t, an individual chooses whether to smoke s_t and the intensity of alcohol consumption $a_t \in \{None, Light, Heavy\}$. Alcohol and cigarette decisions follow the antidepressant decision to allow the marginal utility of alcohol and cigarettes to depend on antidepressant consumption.[29] Following these decisions, at the end of period t, a person may attrit from the sample, E_{t+1} or die, O_{t+1}, but conditional on remaining in the sample, the state variable S updates.

While solution of such a model is beyond the scope of this paper, such a solution would generate demand equations for antidepressants, alcohol, and cigarettes, as well as outcome equations for attrition and mortality. Specifically, solution would theoretically yield the following probabilities for each behavior:

$$p(d_t = d) = d(S_t, X_t, c_3, \mu^d, \epsilon_t^d) \tag{12}$$

$$p(a_t = a) = a(S_t, d_t, X_t, P_t, c_3, \mu^a, \epsilon_t^a) \tag{13}$$

$$p(s_t = s) = s(S_t, d_t, X_t, P_t, c_3, \mu^s, \epsilon_t^s) \tag{14}$$

The demand for antidepressants is a function of past behavior (alcohol, cigarettes, and antidepressants), as well as exogenous characteristics X_t. The final two terms, μ^d and ϵ_t^d, represent a permanent, individual specific component and an i.i.d. error component, respectively. The demand for alcohol and cigarettes are chosen simultaneously as a function of the same arguments, including a price vector P_t, lagged behavior, exogenous characteristics, and antidepressants, which again captures the potential for negative interaction effects between these behaviors and antidepressants. Similar to the antidepressant equation, the final two terms, $\mu^\cdot$ and $\epsilon_t^\cdot$, represent permanent, individual specific components and i.i.d. error components, respectively.

The structural framework above suggests that an outcome equation for mental health should be a function of the state vector S_t, which includes lagged mental health, and period t behavior. Unfortunately, we do not consistently observe the CES-D score in the Framingham data.[30] Our solution is to estimate a time invariant measure of depression based on the exam three CES-D

[29] We model alcohol sequentially with antidepressants because clinical guidelines suggest that patients should not combine alcohol and any type of antidepressants due to the potential for negative interaction effects. Because an antidepressant requires a prescription and is therefore a less flexible input, we allow the marginal utility of alcohol (and tobacco) consumption to depend on contemporaneous antidepressant consumption.

[30] We observe the CES-D measure in exams 3, 6, 7, and 9. While estimation of a dynamic production function for the CES-D score is technically possible, the parameter estimates were highly unstable when estimating jointly with other behavioral/outcome equations.

tertiles presented above. Specifically, we estimate:

$$p(c_3 = c) = c(a_2, s_2, X_3, \mu^c, \epsilon_3^c) \tag{15}$$

where $c \in \{Low, Medium, Heavy\}$. Importantly, the exam three CES-D is measured prior to the introduction of SSRIs in 1988; thus, we interpret c_3 as a baseline measure of depression which is predictive of future mental health. Because our baseline measure of mental health may itself be a function of past alcohol and tobacco consumption, we allow the probability of each depression state to be a function of lagged alcohol and tobacco consumption, a_2 and s_2, respectively. Furthermore, as discussed in more detail below, estimating Equation 15 jointly with the demand/outcome system allows us to jointly estimate the distribution of permanent unobserved heterogeneity, μ.

In addition to Equation 15, we estimate equations for sample attrition and mortality, respectively:

$$p(E_{t+1} = e) = e(S_t, a_t, s_t, d_t, c_3, X_t, \mu^e, \epsilon_t^e) \tag{16}$$

$$p(O_{t+1} = o) = o(S_t, a_t, s_t, d_t, c_3, X_t, \mu^o, \epsilon_t^o). \tag{17}$$

Finally, because we observe individuals between the ages of 17 and 72 at exam two, we observe very different initial histories of alcohol and cigarette consumption. Thus, we estimate initial conditions equations for alcohol consumption and cigarette smoking at exam two:

$$p(a' = a) = a(X_2, \mu^{a\prime}, \epsilon^{a\prime}) \tag{18}$$

$$p(s' = s) = s(X_2, \mu^{s\prime}, \epsilon^{s\prime}). \tag{19}$$

Included in X_2 is a coarse cohort control for initially entering our sample over the age of 50. Under the assumption that each ϵ term takes an extreme value type 1 distribution, equations 12 through 19 become a system of dynamic logit equations.[31]

The μ terms represent equation specific permanent unobserved heterogeneity, and we allow the μ terms to be correlated across equations, yielding the familiar seemingly unrelated regression framework. Conditional on the distributional assumption that each ϵ term takes an i.i.d. extreme value distribution, we treat the joint distribution of $(\mu^{a'}, \mu^{s'}, \mu^c, \mu^a, \mu^s, \mu^e, \mu^O)$ non-parametrically. Following Heckman & Singer (1984) and Mroz (1999), we estimate a step-function for an assumed number of points of support for each term. Subject to the normalization that the first point of support is zero in all equations, we jointly estimate each point of support and the probability of each type. While μ takes the form of a random effect (i.e., we are estimating the distribution of the

[31]Equations for alcohol and the exam three CES-D tertile are multinomial logit equations.

permanent component of the error structure), μ is not independent of the *endogenous* right-hand side variables because the latent factor μ helps to determine past realizations of the endogenous behaviors and outcomes.

To estimate the system, we maximize the log-likelihood function with respect to the parameters that dictate initial conditions, exam three depression, behavior, and outcomes. The latent factor approach allows individual characteristics that are unobserved by the researcher to impact all jointly estimated equations (in a non-linear way) and integrates over their distributions when constructing the likelihood function. The weighted-sum of likelihood contributions for each individual i at time t is:

$$\begin{aligned} L_i(\Theta, \mu, \rho) \;=\; & \sum_{k=1}^{K} \rho_k \Bigg\{ \prod_{s=0}^{1} p(s' = s|\mu_k^{s'})^{1\{s'=s\}} \prod_{a=0}^{2} p(a' = a|\mu_k^{a'})^{1\{a'=a\}} \prod_{j=0}^{2} p(c = j|\mu_k^{c})^{1\{c=j\}} \times \\ & \times \prod_{t=3}^{9} \Bigg[\prod_{d=0}^{1} p(d_{it} = d|\mu_k^{d})^{1\{d_{it}=d\}} \prod_{a=0}^{2} p(a_{it} = a|\mu_k^{a})^{1\{a_{it}=a\}} \prod_{s=0}^{1} p(s_{it} = s|\mu_k^{s})^{1\{s_{it}=s\}} \times \\ & \times \prod_{e=0}^{1} p(E_{it+1} = e|\mu_k^{e})^{1\{E_{it+1}=e\}} \prod_{o=0}^{1} p(O_{it+1} = o|\mu_k^{o})^{1\{O_{it+1}=o\}} \Bigg] \Bigg\} \end{aligned} \tag{20}$$

where Θ defines the vector of parameters of the model. Here, ρ_k denotes the probability of the k^{th} mass-point, which is estimated jointly with the k^{th} permanent mass point $\mu_k^{\cdot}$ in each equation. After taking the log of each individual's unconditional likelihood contribution, we add the contributions to form the sample log-likelihood function and we maximize with respect to Θ.

Estimation of our dynamic system of equations uses both within-individual and across-individual variation in behavior and outcomes (as opposed to results in Table 3, which focus only on within-individual variation), which we argue provides a richer test of the rational self-medication hypothesis, in addition to addressing the limitations of the static reduced-form approach listed above. Because of the presence of both sources of variation, identification of the system comes from three sources. First, as a natural set of exclusion restrictions, prices of cigarettes and alcohol appear only in the demand equations for cigarettes and alcohol.[32] The assumption is that any effect of prices

[32] While we do not observe an individual's location, most of our sample remain in Massachusetts, so the only variation in average prices is temporal. We interact all prices with age to generate cross-sectional variation and to allow the price elasticity and cross-price elasticities of demand for alcohol and cigarettes to vary with age. We use the alcohol specific Consumer Price Index for urban consumers from the Federal Reserve Bank of St. Louis's Federal Reserve Economic Data, which is seasonally adjusted and relative to 1982-1984. https://fred.stlouisfed.org. Accessed on April 2nd, 2018. Price data for cigarettes represent the mean cigarette price in Massachusetts in a given year over all cigarette brands. We merge these data to the median year in which an individual may have

on antidepressant behavior and our endogenous outcomes works through alcohol and cigarette behavior, and in what follows, we show that prices significantly shift behavior. Furthermore, following the logic in Arellano & Bond (1991), time-varying exogenous variables such as prices serve as implicit instruments for behavior.

Second, as discussed above, the FDA's 1988 approval of SSRIs dramatically lessened the side-effects of taking an antidepressant and opened antidepressants to new demographic markets (e.g., the elderly). We argue that the full price of antidepressants shifted exogenously between exams three (taken between 1983 and 1987) and four (taken between 1987 and 1991) as a result of this innovation. While we do not observe antidepressant use in exam two (or exam one), the absences of these questions in FHS surveys is likely due to the national prevalence of antidepressants. Indeed, in exam three, only 1.6% of our sample was taking an antidepressant. We model antidepressant usage as a function of past depression, among other things, to capture observable types of individuals most likely to select into subsequent antidepressant usage, and we argue that the diffusion of antidepressants documented in Table 2 was due to SSRIs.

Finally, by estimating Equations 12- 19 jointly, we allow permanent unobserved heterogeneity to influence both initial and per-period behavior and outcomes. We argue that modeling the distribution of unobserved heterogeneity is important because permanent unobserved characteristics such as genetic endowments may affect both behaviors and outcomes and because measurement error, which is always a problem with measures of mental health, may be lessened. For example, if some individuals are less likely to consume alcohol, and thus are observed at exam two not drinking alcohol, but more likely to take an antidepressant later in the sample for permanent unobserved reasons, we allow for this correlation.

Table 4 provides selected estimates from the multinomial logit equation for per-period alcohol consumption relative to the omitted category of not drinking.[33] For example, for light drinking, Table 4 presents the estimated coefficients on selected right-hand-side variables and the associated standard errors for both a model without unobserved heterogeneity (i.e., where we set k=1) and for a model in which we assume four points of support for the joint distribution of μ (subject to the normalization that the first point of support is zero in all equations). While the coefficients are difficult to interpret, the Table demonstrates a negative relationship between antidepressants and both light and heavy drinking. Table 4 also demonstrates the importance of allowing for unobserved heterogeneity. The coefficient on heavy depression at Exam three (i.e. $CES-D \in [11, 51]$) flips from negative to positive and statistically significant with respect to heavy drinking, which suggests

taken each exam. See Darden et al. (2018) for further information. We thank Koleman Strumpf for sharing these data.

[33]Tables 7-10 present the entire set of parameter estimates and standard errors for all estimated equations.

that the marginal utility of heavy alcohol consumption is higher for those with depression.

Table 5 presents the estimated points of support for the joint distribution of μ and the associated probabilities of each "type."[34] Our preferred specification includes four points of support for the distribution of μ, and we normalize the first point in each equation to zero. For example, type four individuals are significantly more likely to be highly depressed at exam three, they are significantly more likely to take antidepressants and smoke, but they are significantly less likely to drink, both lightly and heavily. Because parameters in both Tables 4 and 5 are difficult to interpret on their own, we now turn to simulation exercises to investigate rational self-medication.

Simulation

To evaluate our model, we simulate both the extent to which our model can recover the time path of each behavior/outcome and the extent to which it can capture transitions between behaviors. To proceed, we replicate the baseline sample, complete with their baseline characteristics, 50 times. For men, this implies a simulated sample of 50*2,497=124,850 simulated observations. Using the estimated distribution of μ, we endow each simulated individual with a complete set of draws of the error structure (including all ϵ terms). We begin by using the estimated initial conditions and exam three CES-D equations to simulate starting points for our simulation.[35] Conditional on these and the assigned draws of the error structure, we simulate behavior and outcomes forward from exam two, taking care to update the state vector with endogenous variables and associated interaction terms. For example, when an individual is simulated to drink lightly, his or her next period lagged light drinking variable is updated accordingly, regardless of if the person actually drank lightly.

Figure 3 presents the empirical time path of each behavior/outcome, as well as our simulated time path.[36] In all cases, our model produces the observed patterns quite well. To further demonstrate that our model does a good job in capturing the data, Table 6 presents simulated transitions for each behavior along with the analogous transition proportion in the data for both men and women. For example, conditional on drinking heavily in period $t-1$, 61.7% of individuals are simulated to be drinking heavily in period t. In the data, that percentage is 58.7%. Capturing transitions is more difficult than capturing averages, yet our model does a good job of recovering the transitions in the data. Finally, Figure 4 demonstrates the importance of modeling the unob-

[34]Not reported are the estimated points of support in the initial conditions alcohol and cigarette equations. These are presented in Table 10

[35]Simulating the initial conditions equations prevents us from breaking the link between the initial conditions, the unobserved heterogeneity, and the per-period equations.

[36]In simulation, we assign the median year in the range of years in which each exam could have occurred.

served heterogeneity distribution. For example, Figure 4a shows a significantly higher fraction of Type 1 individuals using antidepressants while these same individuals are much less likely to be drinking heavily. Importantly, despite the fact that each μ term shifts the respective logit equation intercept, the time paths by type are not perfectly parallel. This highlights selection out of the sample by type.

To test our theory of rational self-medication, we simulate our estimated dynamic model under two counterfactual scenarios. As a natural first step, we evaluate a counterfactual in which all sample participants take an antidepressant as soon as SSRIs become available and there onward (i.e., exam 4 through 9). Figure 5a presents results for the entire sample. Heavy drinking declines by approximately five percentage points by the end of the ninth exam. Figures 5b and 5c break the results from Figure 5a by gender, which demonstrates that men are primarily driving our heavy drinking result. Figures 5d, 5e, and 5f break the results from Figure 5a by simulated exam three CES-D tertile. Surprisingly, the reduction in heavy drinking associated with antidepressants is driven by those in the middle tertile, with no reduction in heavy drinking for those simulated to be highly depressed.

One potential explanation for the lack of substitution away from heavy drinking for those simulated to be highly depressed is that with depression comes addiction. If highly depressed individuals face significant reinforcement, tolerance, and withdrawal mechanisms, then alcohol consumption may not change despite improvements in mental health. To investigate, Figure 6 presents results in which we simulate our model assuming that the parameters on all terms reflecting past alcohol consumption are set to zero. Not surprisingly, relative to the baseline simulation heavy alcohol consumption plummets while light drinking remains unchanged — a roughly equal fraction of light drinkers quit as compared to the fraction of heavy drinkers who move to light drinking. Figure 7 presents the simulated time paths of endogenously chosen antidepressant medication under this counterfactual relative to the baseline simulation. Overall, Figure 7a demonstrates a 5.5 percentage point increase in antidepressant usage by the end of the sample. Figures 7b-7f demonstrate that substitution towards antidepressants greater (in percentage terms) for men and is increasing in simulated depression. Figure 7 provides clear evidence that addiction inhibits substitution. These results are consistent with rational self-medication.

Finally, Figure 8b contrasts our main finding in Figure 5a of a roughly five percentage point decline in heavy drinking when antidepressants are imposed on the entire sample with a similar simulation in which we both impose antidepressants and decrease alcohol prices by 10%. Figure 8b shows that the 10% completely nullifies the antidepressant effect by exam nine. The simulation also demonstrates that prices, which serve an important role with respect to identification of our dynamic system, significantly affect even heavy alcohol consumption.

6 Additional Evidence: Smoking

Until now, we have provided evidence of substitution, which provides empirical evidence in support of the rational self-medication hypothesis. Individuals substitute a risky substance, alcohol, for SSRIs. We now turn to smoking.[37] There are three reasons. One, smoking is associated with depression and could therefore be used as a treatment, which would suggest an additional test of the self-medication hypothesis. Two, the dynamics of smoking, in particular, the impact of smoking on future depression, are different from alcohol. As we show, using a framework similar to the theory above, this allows us to draw sharper conclusions about which mechanisms drive self-medication. Three, smoking is an important behavior in a long panel as it can increase mortality rates.

Figure 9 presents simulated time paths of smoking behavior for our first counterfactual in which we impose antidepressants on the entire sample at exam four onward. Consistent with our regression estimates in Table 3, Figure 9 shows a positive relationship between taking an antidepressant and smoking. This effect is driven by men and by those in the lowest simulated tertile of depression, and it entirely reflects a failure to quit smoking, as opposed to smoking initiation. The tobacco results are puzzling, but they may help us to say more about the rational self-medication hypothesis, both with respect to tobacco and alcohol.

To investigate, we return to our theoretical model in Section 3. Rather than develop complementarities between alcohol and tobacco with respect to both preferences and the production of mental health, we simplify the problem by replacing alcohol A with tobacco T.[38] Focusing on tobacco, the model yields a similar equilibrium expression to Equation 7 for optimal consumption, where the marginal cost of tobacco is given as:

$$\alpha(S)\left[\frac{\delta S}{\delta T}+\beta\frac{\delta S'}{\delta M'}\frac{\delta M'}{\delta T}\right] \tag{21}$$

To simplify the problem further, suppose that smoking has no long-run negative effect on mental

[37] Hughes et al. (2014), in a comprehensive review of the epidemiological literature, find no evidence that SSRIs *aid* in smoking cessation. Fluharty et al. (2017) review the longitudinal literature on cigarettes and depression and show that smoking and depression are strongly correlated, but the direction of causation is unclear.

[38] The model developed in Section 3 focuses on the intensive margin of alcohol consumption. For simplicity, we continue on that margin with respect to tobacco, recognizing that our empirical finding is that of a failure to quit smoking.

health, which implies that the marginal cost becomes:

$$\alpha(S)\frac{\delta S}{\delta T} \tag{22}$$

Given our empirical findings that alcohol consumption decreases while tobacco cessation rates drop, it must be the case that

1. $\alpha(S)\left[\frac{\delta S}{\delta A} + \beta\frac{\delta S'}{\delta M'}\frac{\delta M'}{\delta A}\right]$ falls and

2. $\alpha(S)\frac{\delta S}{\delta T}$ does not change or rises.

To explain why alcohol consumption falls when $D = 1$, we argued above that $\alpha(S)$ is smaller for lower S. If $D = 1$ causes $\alpha(S)$ to fall, then $\frac{\delta S}{\delta T}$ must increase with reduced symptoms if the expression $\alpha(S)\frac{\delta S}{\delta T}$ is to remain constant or increase. Formally, this means that:

$$\alpha(S, D = 1)\frac{\delta S}{\delta T}\bigg|_{D=1} > \alpha(S, D = 0)\frac{\delta S}{\delta T}\bigg|_{D=0} \tag{23}$$

If $\alpha(S, D = 1) < \alpha(S, D = 0)$, then there exists some value δ such that $\alpha(S, D = 1)+\delta = \alpha(S, D = 0)$, and the above expression can be rewritten as:

$$\alpha(S, D = 1)\frac{\delta S}{\delta T}\bigg|_{D=1} > (\alpha(S, D = 1) + \delta)\frac{\delta S}{\delta T}\bigg|_{D=0} \tag{24}$$

Which implies that:

$$\frac{\frac{\delta S}{\delta T}\big|_{D=1}}{\frac{\delta S}{\delta T}\big|_{D=0}} > \frac{(\alpha(S, D = 1) + \delta)}{\alpha(S, D = 1)} \tag{25}$$

Thus, the decline in $\alpha(S)$ means that the increase in $\frac{\delta S}{\delta T}$ has a lower bound and must be "large enough". In other words, our results can be rationalized within our model if smoking is more effective for reducing symptoms when symptoms are mild and the change in symptoms is large. Importantly, the overall change in smoking rates in Figure 9a is entirely explained by the relative failure to quit smoking of those simulated to be in the lowest tertile of the CES-Depression metric (Figure 9d).

To summarize, under the assumption that the marginal cost of symptoms rises with symptoms, alcohol should decrease with improved symptoms. If smoking is more useful with respect to symptoms when symptoms are mild, then an improvement in symptoms may have the perverse effect of inhibiting smoking cessation. We argue that our empirical results with respect to both alcohol and tobacco are consistent with the self-medication hypothesis. Moreover, specific empirical patterns are consistent with three underlying mechanisms:

1. Individuals are more prone to self-medicate with more severe symptoms due to the marginal utility cost of symptoms. This is an assumption rather than a direct implication of the model.

2. Individuals use less alcohol when their symptoms are reduced because alcohol is more effective when symptoms are more severe.

3. Smokers fail to quit smoking when their symptoms are reduced since smoking is more effective when symptoms are less severe.

7 Conclusion

We develop a theory of rational self-medication, which suggests a relationship between optimal investments in health and the degree of negative symptoms generated by a stock of health. We test our hypothesis by studying alcohol and tobacco consumption when the choice set for the management of depression expands due to technological advancement (i.e., SSRIs). Using a dynamic system of equations estimator, we show that heavy alcohol consumption decreases for men and for those with moderate depression following the introduction of SSRIs. The dynamic model allows us to simulate antidepressant behavior under the counterfactual that alcohol is less addictive, which shows that antidepressant consumption increases by five to six percentage points, and this increase is increasing in the severity of depression. Finally, we show that SSRIs prevented smoking cessation in those with mild depression, which suggests that different forms of self-medication (e.g., alcohol versus tobacco) have different effects depending on the severity of mental health problems.

To the extent that rational self-medication accurately characterizes behavior, our theory has important implications for addiction and health policy. Given the growing literature on the significant effects of technological innovation on health behaviors, policy should promote treatment innovations that obviate the need to self-medicate and thus induce rational actors to substitute towards less harmful substances.

We acknowledge three main limitations of our work. First, even with forty years of longitudinal data on alcohol, tobacco, and antidepressant consumption, FHS lacks a consistently measured metric of mental health. Ideally, a representative period of our dynamic empirical model would include a time-varying mental health production function which is a function of period t health investments. Second, while our theory has important implications for current policy, FHS is not representative of a larger population, and thus our results may not extend to at-risk populations in other areas of the United States or for underrepresented groups. Finally, our dynamic system of equations abstracts from an explicit forward-looking decision-making process. In a fully structural model, an individual's decision to consume alcohol or tobacco would depend on the present discounted value of being in different possible future states, about which an individual would form expectations conditional on contemporaneous behavior. For example, fully-rational self-medicating agents should consider the possibility of future addiction when considering current management of pain, and we are unable to address these expectations with our current estimator. We leave specification and estimation of such a model for future work.

References

Anton, Raymond, O'Malley, Stephanie, Ciraulo, Domenic, & et al. 2006. Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence. The COMBINE Study: A Randomized Controlled Trial. JAMA, **295**(17), 2003–2017.

Arcidiacono, Peter, Sieg, Holger, & Sloan, Frank A. 2007. Living Rationally Under the Volcano? An Empirical Analysis of Heavy Drinking and Smoking. International Economic Review, **48**(1), 37–65.

Arellano, Manuel, & Bond, Stephen. 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.

Bacolod, Marigee, Cunha, Jesse, & Shen, Yu-Chu. 2017. The Impact of Alcohol on Mental Health, Physical Fitness, and Job Performance.

Beck, Arne, Crain, A. Lauren, Solberg, Leif, & et al. 2011. Severity of Depression and Magnitude of Productivity Loss. Annals of Family Medicine, **9**(4), 305–312.

Becker, Gary. 2007. Health as Human Capital: Synthesis and Extensions. Oxford Economic Papers, **59**(3), 379–410.

Becker, Gary, & Murphy, Kevin. 1988. A Theory of Rational Addiction. Journal of Political Economy, **96**(4), 675–700.

Becker, Gary S., Grossman, Michael, & Murphy, Kevin. 1994. An Empirical Analysis of Cigarette Addiction. American Economic Review, **84**(3), 396–418.

Berndt, Ernst R., Finkelstein, Stan N., Greenberg, Paul E., & et al. 1998. Workplace Performance Effects from Chronic Depression and its Treatment. Journal of Health Economics, **17**, 511–535.

Blanchflower, David G., & Oswald, Andrew J. 2016. Antidepressants and Age: A New Form of Evidence for U-shaped Well-being Through Life. Journal of Economic Behavior and Organization, **127**, 46–58.

Bolton, James M., Robinson, Jennifer, & Sareen, Jitender. 2009. Self-medication of Mood Disorders with Alcohol and Drugs in the National Epidemiologic Survey on Alcohol and Related Conditions. Journal of Affective Disorders, **115**(3), 367 – 375.

Brody, Debra J., Pratt, Laura, A., & Hughes, Jeffery P. 2018. Prevalence of Depression Among Adults Aged 20 and Over: United States, 2013–2016. Data Brief 303. National Center for Health Statistics.

Case, Anne, & Deaton, Angus. 2015. Rising Morbidity and Mortality in Midlife Among White Non-Hispanic Americans in the 21st Century. Proceedings of the National Academy of Sciences, **112**(49), 15078—-15083.

Case, Anne, & Deaton, Angus. 2017. Mortality and Morbidity in the 21st Century. Brookings Papers on Economic Activity, **Spring**, 397–476.

CMS. 2013. Antidepressant Medications: Use in Adults. Report.

Crawford, Gregory S, & Shum, Matthew. 2005. Uncertainty and Learning in Pharmaceutical Demand. Econometrica, **73**(4), 1137–1173.

Crost, Benjamin. 2012. The Effect of Alcohol Availability on Marijuana Use: Evidence from the Minimum Legal Drinking Age. Journal of Health Economics, **31**(1), 112–121.

Crum, Rose M., Mojtabai, Ramin, Lazareck, Lamuel, & et al. 2013. A Prospective Assessment of Reports of Drinking to Self-medicate Mood Symptoms With the Incidence and Persistence of Alcohol Dependence. JAMA Psychiatry, **70**(7), 718–726.

Darden, Michael. 2017. A Dynamic Stochastic Model of Lifetime Smoking Behavior. Journal of Political Economy, **125**(4), 1465–1522.

Darden, Michael, Gilleskie, Donna, & Strumpf, Koleman. 2018. Smoking and Mortality: New Evidence from a Long Panel. International Economic Review, **59**(3), 1571–1619.

Deykin, Eva, Levy, Janice, & Wells, Victoria. 1987. Adolescent Depression, Alcohol and Drug Abuse. American Journal of Public Health, **77**(2), 178–182.

Dinardo, John. 2001. Alcohol, Marijuana, and American Youth: The Unintended Consequences of Government Regulation. Journal of Health Economics, **20**(6), 991–1010.

EA, O'Connor, EP, Whitlock, B, Gaynes, & et al. 2009. Screening for Depression in Adults and Older Adults in Primary Care: An Updated Systematic Review. **115**(3), 367 – 375.

Fluharty, Meg, Taylor, Amy E., Grabski, Meryem, & Munafo, Marcus. 2017. The Association of Cigarette Smoking With Depression and Anxiety: A Systematic Review. Nicotine and Tobacco Research, **19**(1).

Gilleskie, Donna B. 1998. A Dynamic Stochastic Model of Medical Care Use and Work Absence. Econometrica, **66**(1), 1–45.

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

Heckman, James J., & Singer, Burton. 1984. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. Econometrica, **52**(2), 271–320.

Hillhouse, Todd, & Porter, Joseph. 2015. A Brief History of the Development of Antidepressant Drugs: From Monoamines to Glutamate. Experimental Clinical Psychopharmacology, **23**(1), 1–21.

Horwitz, Allan. 2010. How an Age of Anxiety Became an Age of Depression. The Millbank Quarterly, **88**(1), 112–138.

Hughes, JR, Stead, LF, Harmann-Boyce, J, Cahill, K, & Lancaster, T. 2014. Antidepressants for Smoking Cessation (Review). Cochrane Database of Systematic Reviews.

Kaestner, Robert, Darden, Michael, & Lakdawalla, Darius. 2014. Are Investments in Disease Prevention Complements? The Case of Statins and Health Behaviors. Journal of Health Economics, **36**, 151–163.

Kessler, Ronald C. 2012. The Costs of Depression. Psychiatric Clinics of North America, **35**(1), 1–14.

Khantzian, Edward. 1985. The Self-Medication Hypothesis of Addictive Disorders: Focus on Heroin and Cocaine Dependence. The American Journal of Psychiatry, **142**(11), 1259–1264.

Khantzian, Edward. 1990. Self-regulation and self-medication factors in alcoholism and the addictions. Similarities and differences. Recent Developments in Alcoholism, **8**, 225–271.

Ludwig, Jens, Marcotte, Dave E., & Norberg, Karen. 2009. Anti-depressants and Suicide. Journal of Health Economics, **28**, 659–676.

Mahmood, SS., Levy, D., Vasan, RS., & TJ., Wang. 2014. The Framingham Heart Study and the Epidemiology of Cardiovascular Disease: a Historical Perspective. Lancet, **383**(9921), 999–1008.

Mroz, Thomas. 1999. Discrete Factor Approximations in Simultaneous Equation Models: Estimating the Impact of a Dummy Endogenous Variable on a Continuous Outcome. Journal of Econometrics, **92**(2), 233–274.

Papageorge, Nicholas W. 2016. Why Medical Innovation is Valuable: Health, Human Capital, and the Labor Market. Quantitative Economics, **7**(3), 671–725.

Plurphanswat, Nantaporn, Kaestner, Robert, & Rodu, Brad. 2011. The Effect of Smoking on Mental Health. American Journal of Health Behavior, **41**(4), 471–483.

Powell, David, Pacula, Rosalie Liccardo, & Jacobson, Mireille. 2018. Do Medical Marijuana Laws Reduce Addictions and Deaths Related to Pain Killers? Journal of Health Economics, **58**, 29–42.

Radoff, Lenore Sawyer. 1977. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Applied Psychological Measurement, **1**(3), 385–401.

Ruhm, Christopher J. 2018. Deaths of Despair or Drug Problems? NBER Working Paper.

Wulsin, Lawson R., Evans, Jane C., & et al. 2005. Depressive Symptoms, Coronary Heart Disease, and Overall Mortality in the Framingham Heart Study. Psychosomatic Medicine, **67**(1), 697–702.

A Main Tables

Table 1: Baseline Characteristics by Gender and Ever Antidepressant Usage

	Men = 1,241			Women = 1,256		
	Never (87.83%)	Ever (12.17%)	p-value	Never (75.48%)	Ever (24.52%)	p-value
Alcohol Consumption						
Never	0.177	0.205	0.398	0.284	0.286	0.947
Light	0.573	0.556	0.691	0.506	0.529	0.485
Heavy	0.250	0.238	0.767	0.210	0.185	0.347
Smokes	0.417	0.430	0.745	0.296	0.370	0.015
Ever Has Cancer	0.414	0.411	0.941	0.343	0.276	0.030
Ever Has CVD	0.372	0.397	0.540	0.203	0.234	0.243
Dies Before Exam 9	0.336	0.285	0.212	0.214	0.091	0.000
Age	45.025	44.093	0.291	44.872	41.292	0.000
Education						
Less than HS	0.017	0.026	0.440	0.006	0.003	0.528
HS Grad.	0.304	0.272	0.419	0.379	0.390	0.732
Some College	0.423	0.404	0.659	0.461	0.481	0.551
College or More	0.185	0.252	0.053	0.098	0.078	0.290
BMI	26.799	27.170	0.230	24.391	24.509	0.702
Obese	0.162	0.199	0.263	0.114	0.120	0.767
Exam 3 CES-Depression Tertile [Range]						
Low [0,4]	0.397	0.291	0.012	0.350	0.250	0.001
Medium [5, 10]	0.350	0.351	0.990	0.349	0.302	0.128
High [11, 51]	0.252	0.358	0.006	0.301	0.448	0.000

Notes: $n = 2,497$. With the exception of the CES-D score, statistics are calculated from exam 2, which took place between 1979 and 1983. The sample is constructed such that an individual must be present for exams 2 and 3, after which an individual may leave the sample through death or attrition. Rows for never and ever antidepressant usage reflect whether the person was ever observed to take an antidepressant. Depression is measured by the CES-D scale, which is broken into tertiles. Light drinking is defined as seven or fewer drinks per week for women and 14 or fewer drinks per week for men. Heavy drinking is defined as more than seven drinks per week for women and more than 14 drinks per week for men.

Table 2: Sample Behaviors over Time by Gender.

Men, $n = 8,345$								
Exam	Count	Year Range	Age	Antidepressant	Never	Light	Heavy	Smoke
2	1241	1979-1983	44.911	.	0.180	0.571	0.248	0.418
3	1241	1983-1987	49.267	0.010	0.212	0.555	0.233	0.269
4	1198	1987-1991	52.422	0.013	0.264	0.539	0.197	0.234
5	1122	1991-1995	55.603	0.020	0.266	0.546	0.188	0.178
6	1043	1995-1998	59.301	0.036	0.291	0.548	0.161	0.129
7	1005	1998-2001	61.867	0.056	0.276	0.554	0.170	0.116
8	845	2005-2008	67.424	0.088	0.249	0.591	0.161	0.090
9	650	2011-2014	71.462	0.105	0.269	0.554	0.177	0.055
Women, $n = 8,913$								
Exam	Count	Year	Age	Antidepressant	Never	Light	Heavy	Smoke
2	1256	1979-1983	43.994	.	0.284	0.512	0.204	0.314
3	1256	1983-1987	48.362	0.021	0.350	0.473	0.177	0.278
4	1225	1987-1991	51.740	0.036	0.343	0.507	0.150	0.219
5	1183	1991-1995	55.173	0.049	0.332	0.525	0.143	0.174
6	1131	1995-1998	59.034	0.084	0.450	0.417	0.133	0.141
7	1107	1998-2001	61.822	0.112	0.388	0.451	0.162	0.114
8	972	2005-2008	67.418	0.186	0.321	0.515	0.164	0.099
9	783	2011-2014	71.775	0.217	0.354	0.469	0.178	0.056

Notes: $n = 17,258$. Statistics are calculated from eight exams, which took place between 1979 and 2011. The sample is constructed such that an individual must be present for exams 2 and 3, after which some individuals are lost to death or attrition. Light drinking is defined as seven or fewer drinks per week for women and 14 or fewer drinks per week for men. Heavy drinking is defined as more than seven drinks per week for women and more than 14 drinks per week for men.

Table 3: Reduced-Form Estimates of Antidepressants on Behavior

Panel 1: Estimates from Equation 11

	Alcohol Consumption							
	Never		Light		Heavy		Smoking	
	1	2	1	2	1	2	1	2
Antidepressant	0.039**	0.052	-0.026	0.034	-0.013	-0.086***	-0.005	0.050*
	(0.017)	(0.042)	(0.019)	(0.045)	(0.014)	(0.033)	(0.012)	(0.027)
* Female		-0.048		-0.036		0.084***		-0.030
		(0.038)		(0.040)		(0.031)		(0.025)
* CES-D $\in [5,10]$		0.003		0.006		-0.008		-0.049
		(0.047)		(0.051)		(0.034)		(0.030)
* CES-D $\in [11,51]$		0.044		-0.088**		0.043		-0.046
		(0.040)		(0.044)		(0.031)		(0.031)

Panel 2: Estimates from Equation 11 with Separate Trends by Medication

	Alcohol Consumption							
	Never		Light		Heavy		Smoking	
	1	2	1	2	1	2	1	2
Antidepressant	0.030	0.045	-0.015	0.042	-0.015	-0.087***	0.000	0.056**
	(0.019)	(0.043)	(0.021)	(0.046)	(0.013)	(0.033)	(0.013)	(0.029)
* Female		-0.048		-0.036		0.083***		-0.030
		(0.037)		(0.040)		(0.030)		(0.025)
* CES-D $\in [5,10]$		0.001		0.008		-0.008		-0.049
		(0.047)		(0.051)		(0.034)		(0.030)
* CES-D $\in [11,51]$		0.041		-0.082*		0.041		-0.046
		(0.040)		(0.044)		(0.031)		(0.032)
p-value	0.310	0.323	0.123	0.153	0.836	0.881	0.469	0.474
Mean		0.313		0.516		0.171		0.164

Panel 3: Estimates from Equation 11 with Separate Trends by Baseline Depression

	Alcohol Consumption							
	Never		Light		Heavy		Smoking	
	1	2	1	2	1	2	1	2
Antidepressant	0.036**	0.053	-0.024	0.038	-0.012	-0.090***	-0.005	0.047*
	(0.017)	(0.042)	(0.019)	(0.045)	(0.014)	(0.033)	(0.012)	(0.027)
* Female		-0.047		-0.036		0.083***		-0.031
		(0.038)		(0.040)		(0.031)		(0.025)
* CES-D $\in [5,10]$		0.008		-0.002		-0.006		-0.042
		(0.047)		(0.052)		(0.035)		(0.031)
* CES-D $\in [11,51]$		0.032		-0.089*		0.057*		-0.043
		(0.041)		(0.046)		(0.031)		(0.032)
p-value	0.563	0.645	0.983	0.992	0.941	0.878	0.173	0.149
Mean		0.313		0.516		0.171		0.164

Notes: n =14,687 person/year observations in all regressions. All regressions are estimated on data from exams 3 through 9 and include controls for age, education, cardiovascular disease, cancer, body mass index, and exam binary variables. All results are from linear probability models. Interaction effects are the second and third tertile of exam three CES-D score, relative to the lowest tertile. The p-value is with respect to an F-test with null hypothesis that interactions between ever taking a medication (panel 2) or medium and high baseline depression (panel 3) and each exam binary variable are jointly zero. * $p < 0.1$, ** $p < 0.05$ *** $p < 0.01$.

Table 4: Selected Parameter Estimates

	Light Drinking				Heavy Drinking			
	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.
Antidepressant	-0.321	0.219	-0.314	0.303	-0.871	0.336	-1.178	0.444
Antidepressant*								
CES-D$\in [5, 10]$	-0.180	0.237	-0.192	0.319	-0.367	0.375	-0.529	0.500
CES-D$\in [11, 51]$	-0.232	0.222	-0.240	0.303	0.029	0.344	0.331	0.465
Female	0.174	0.190	0.286	0.257	0.765	0.304	1.047	0.400
CES-D$\in [5, 10]$	0.053	0.056	0.183	0.117	0.088	0.078	0.248	0.157
CES-D$\in [11, 51]$	-0.060	0.059	0.482	0.119	-0.147	0.085	1.281	0.237
Female	-0.250	0.049	-0.676	0.085	-0.227	0.070	-0.836	0.116
L. Heavy Drinking	2.476	0.047	1.217	0.070	3.789	0.159	2.567	0.179
L. Light Drinking	2.887	0.092	1.624	0.137	6.795	0.174	4.216	0.198
L. Smoking	-0.122	0.079	0.019	0.111	0.027	0.109	0.083	0.150
Years Smoking	0.001	0.002	-0.006	0.003	0.013	0.003	-0.002	0.004
Years Smoking Cessation	0.008	0.002	0.008	0.003	0.013	0.003	0.021	0.004
Age								
(35, 40]	-0.017	0.178	0.013	0.213	0.401	0.253	0.373	0.293
(40, 45]	-0.135	0.168	-0.151	0.202	0.331	0.238	0.355	0.273
(45, 50]	-0.012	0.176	-0.017	0.210	0.537	0.247	0.695	0.281
(50, 55]	-0.123	0.186	-0.122	0.220	0.572	0.261	0.809	0.296
(55, 60]	-0.060	0.203	-0.111	0.237	0.589	0.287	0.769	0.322
(60, 65]	-0.050	0.228	-0.124	0.264	0.717	0.322	0.877	0.359
(65, 70]	-0.038	0.258	-0.174	0.298	0.751	0.370	0.879	0.412
(70, 75]	-0.194	0.294	-0.461	0.339	0.578	0.423	0.455	0.474
>75	-0.264	0.355	-0.773	0.409	0.377	0.513	-0.083	0.572
Education								
High School	0.185	0.099	0.341	0.171	0.139	0.146	0.246	0.272
Some College	0.410	0.099	0.783	0.170	0.396	0.145	0.699	0.266
College or More	0.530	0.113	1.013	0.192	0.577	0.162	0.905	0.293
(Alcohol CPI * Age)/100	-0.008	0.005	-0.006	0.006	-0.029	0.008	-0.033	0.009
(Cents/cig. Pack * Age)/100	-0.001	0.002	0.001	0.002	0.007	0.002	0.015	0.003
Constant	-1.862	0.228	1.076	0.375	-5.405	0.359	-2.110	0.537
μ_1			0.000	.			0.000	.
μ_2			-1.500	0.271			0.553	0.328
μ_3			-2.116	0.207			-4.841	0.318
μ_4			-4.712	0.247			-4.136	0.346

Notes: $n = 17,258$. Selected parameter estimates are from models estimated on data in exams 2-9.

Table 5: Unobserved Heterogeneity Distribution

	Probability	Medium Dep.	High Dep.	Anti-depressants	Light Drinking	Heavy Drinking	Smoking	Attrition	Death
μ_1	0.241	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
μ_2	0.185	0.070	-1.062***	0.437*	-1.500***	0.553*	0.546***	0.167	0.318
μ_3	0.375	0.081	0.811***	0.269	-2.116***	-4.841***	0.241	-0.314	0.208
μ_4	0.198	0.160	1.084***	0.595***	-4.712***	-4.136***	0.356*	-0.368	0.067

Notes: $n = 17,258$. Selected parameter estimates are from models estimated on data in exams 2-9, with the exception of the multinomial logit for exam 3 depression. Also estimated jointly, but not listed here, are initial conditions equations for drinking and smoking in exam 2. * $p < 0.1$, ** $p < 0.05$ *** $p < 0.01$.

Table 6: Model Fit: Transitions.

	Period t Behavior									
	No Drinking		Light Drinking		Heavy Drinking		Antidepressants		Smoking	
Lagged Behavior $t-1$	Data	Sim.	Data	Sim.	Data	Sim.	Data	Sim.	Data	Sim.
No Drinking	0.671	0.730	0.148	0.156	0.044	0.050	0.372	0.421	0.288	0.304
Light Drinking	0.218	0.259	0.690	0.738	0.295	0.333	0.366	0.440	0.414	0.475
Heavy Drinking	0.008	0.011	0.094	0.106	0.587	0.617	0.118	0.138	0.204	0.220
Antidepressants	0.092	0.117	0.060	0.073	0.065	0.069	0.623	0.748	0.078	0.086
Smoking	0.124	0.139	0.114	0.123	0.195	0.190	0.144	0.152	0.694	0.682

Notes: $n = 17,258$. Results are from models estimated on data in exams 2-9.

Table 7: Antidepressant Parameter Estimates

	Antidepressant Logit Estimates			
	Beta	S.E.	Beta	S.E.
L. Antidepressant	3.814	0.178	3.798	0.188
L. Antidepressant*				
Female	-0.120	0.214	-0.113	0.226
L. Light Drinking * CES-D$\in [5, 10]$	0.453	0.234	0.458	0.238
L. Light Drinking * CES-D$\in [11, 51]$	0.555	0.221	0.573	0.226
L. Heavy Drinking * CES-D$\in [5, 10]$	0.070	0.312	0.091	0.315
L. Heavy Drinking * CES-D$\in [11, 51]$	0.364	0.296	0.496	0.305
L. Light Drinking	-0.680	0.171	-0.445	0.193
L. Heavy Drinking	-0.469	0.222	-0.400	0.285
CES-D$\in [5, 10]$	0.026	0.171	0.011	0.176
CES-D$\in [11, 51]$	0.322	0.159	0.272	0.164
Female	0.657	0.096	0.683	0.102
L. Smoking	0.037	0.145	0.002	0.146
Years Smoking	0.012	0.003	0.012	0.003
Years Smoking Cessation	0.003	0.003	0.003	0.003
Age				
(35, 40]	1.170	1.071	1.156	0.666
(40, 45]	1.800	1.025	1.776	0.586
(45, 50]	2.149	1.018	2.127	0.572
(50, 55]	2.015	1.017	1.988	0.569
(55, 60]	1.747	1.018	1.725	0.568
(60, 65]	1.504	1.020	1.474	0.570
(65, 70]	1.268	1.024	1.238	0.575
(70, 75]	1.460	1.027	1.433	0.578
>75	1.112	1.030	1.095	0.581
Education				
High School	-0.074	0.178	-0.093	0.187
Some College	-0.052	0.178	-0.098	0.188
College or More	0.082	0.204	0.023	0.211
CVD Last Period	0.404	0.213	0.404	0.216
Any History of CVD	-0.037	0.154	-0.029	0.157
Cancer Last Period	0.452	0.191	0.453	0.194
Any History of Cancer	-0.043	0.148	-0.034	0.151
Obese	0.033	0.093	0.033	0.095
Currently Working	-0.261	0.104	-0.264	0.107
Work Missing	0.025	0.109	0.028	0.114
Married	0.371	0.119	0.376	0.124
Married Missing	-0.118	0.170	-0.124	0.182
Exam Trend	0.358	0.030	0.360	0.032
Constant	-7.631	1.044	-8.036	0.646
μ_1			0.000	.
μ_2			0.437	0.244
μ_3			0.269	0.170
μ_4			0.595	0.210

Notes: $n = 17,258$. Selected parameter estimates are from models estimated on data in exams 2-9.

Table 8: Behavior Parameter Estimates

	Light Drinking				Heavy Drinking				Smoking			
	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.
Antidepressant	-0.321	0.219	-0.314	0.303	-0.871	0.336	-1.178	0.444	1.473	0.411	1.452	0.475
Antidepressant*												
CES-D$\in [5, 10]$	-0.180	0.237	-0.192	0.319	-0.367	0.375	-0.529	0.500	-0.904	0.451	-0.944	0.521
CES-D$\in [11, 51]$	-0.232	0.222	-0.240	0.303	0.029	0.344	0.331	0.465	-1.125	0.407	-1.125	0.466
Female	0.174	0.190	0.286	0.257	0.765	0.304	1.047	0.400	-0.719	0.345	-0.707	0.351
CES-D$\in [5, 10]$	0.053	0.056	0.183	0.117	0.088	0.078	0.248	0.157	0.196	0.095	0.206	0.100
CES-D$\in [11, 51]$	-0.060	0.059	0.482	0.119	-0.147	0.085	1.281	0.237	0.386	0.098	0.430	0.105
Female	-0.250	0.049	-0.676	0.085	-0.227	0.070	-0.836	0.116	0.159	0.081	0.162	0.086
L. Light Drinking	2.476	0.047	1.217	0.070	3.789	0.159	2.567	0.179	-0.157	0.092	-0.051	0.119
L. Heavy Drinking	2.887	0.092	1.624	0.137	6.795	0.174	4.216	0.198	0.090	0.110	0.002	0.157
L. Smoking	-0.122	0.079	0.019	0.111	0.027	0.109	0.083	0.150	3.438	0.115	3.430	0.128
Years Smoking	0.001	0.002	-0.006	0.003	0.013	0.003	-0.002	0.004	0.087	0.004	0.087	0.005
Years Smoking Cessation	0.008	0.002	0.008	0.003	0.013	0.003	0.021	0.004	-0.028	0.011	-0.028	0.011
Age												
(35, 40]	-0.017	0.178	0.013	0.213	0.401	0.253	0.373	0.293	-0.074	0.219	-0.090	0.221
(40, 45]	-0.135	0.168	-0.151	0.202	0.331	0.238	0.355	0.273	-0.290	0.215	-0.301	0.215
(45, 50]	-0.012	0.176	-0.017	0.210	0.537	0.247	0.695	0.281	-0.397	0.236	-0.392	0.234
(50, 55]	-0.123	0.186	-0.122	0.220	0.572	0.261	0.809	0.296	-0.685	0.265	-0.669	0.261
(55, 60]	-0.060	0.203	-0.111	0.237	0.589	0.287	0.769	0.322	-0.945	0.305	-0.931	0.300
(60, 65]	-0.050	0.228	-0.124	0.264	0.717	0.322	0.877	0.359	-1.292	0.359	-1.275	0.353
(65, 70]	-0.038	0.258	-0.174	0.298	0.751	0.370	0.879	0.412	-1.466	0.432	-1.446	0.423
(70, 75]	-0.194	0.294	-0.461	0.339	0.578	0.423	0.455	0.474	-1.661	0.508	-1.639	0.501
>75	-0.264	0.355	-0.773	0.409	0.377	0.513	-0.083	0.572	-2.003	0.633	-1.962	0.630
Education												
High School	0.185	0.099	0.341	0.171	0.139	0.146	0.246	0.272	0.021	0.153	0.025	0.162
Some College	0.410	0.099	0.783	0.170	0.396	0.145	0.699	0.266	-0.072	0.154	-0.090	0.164
College or More	0.530	0.113	1.013	0.192	0.577	0.162	0.905	0.293	-0.266	0.185	-0.314	0.196
CVD Last Period	-0.264	0.130	-0.293	0.155	-0.443	0.198	-0.466	0.241	-0.448	0.211	-0.453	0.212
Any History of CVD	-0.132	0.088	-0.262	0.125	-0.123	0.135	-0.361	0.189	0.024	0.149	0.033	0.150
Cancer Last Period	-0.192	0.128	-0.145	0.152	-0.281	0.187	-0.138	0.222	-0.200	0.272	-0.178	0.279
Any History of Cancer	0.185	0.095	0.063	0.126	0.126	0.138	-0.188	0.180	-0.338	0.209	-0.353	0.214
Obese	-0.180	0.052	-0.135	0.076	-0.247	0.077	-0.113	0.112	-0.450	0.088	-0.452	0.094
Currently Working	0.117	0.062	0.142	0.078	-0.032	0.090	-0.032	0.112	-0.057	0.108	-0.058	0.111
Work Missing	0.096	0.064	0.098	0.087	0.134	0.095	0.159	0.125	-0.196	0.111	-0.193	0.114
Married	0.387	0.076	0.513	0.095	0.453	0.113	0.695	0.142	0.387	0.147	0.399	0.154
Married Missing	0.287	0.086	0.347	0.101	-0.010	0.124	0.029	0.149	0.369	0.147	0.374	0.154
Exam Trend	0.247	0.052	0.169	0.061	0.373	0.076	0.204	0.091	0.508	0.101	0.512	0.103
(Alcohol CPI * Age)/100	-0.008	0.005	-0.006	0.006	-0.029	0.008	-0.033	0.009	-0.028	0.009	-0.028	0.010
(Cents/cig. Pack * Age)/100	-0.001	0.002	0.001	0.002	0.007	0.002	0.015	0.003	-0.005	0.004	-0.005	0.004
Constant	-1.862	0.228	1.076	0.375	-5.405	0.359	-2.110	0.537	-4.459	0.372	-4.754	0.415
μ_1			0.000	.			0.000	.			0.000	.
μ_2			-1.500	0.271			0.553	0.328			0.546	0.194
μ_3			-2.116	0.207			-4.841	0.318			0.241	0.164
μ_4			-4.712	0.247			-4.136	0.346			0.356	0.198

Notes: $n = 17,258$. Selected parameter estimates are from models estimated on data in exams 2-9.

Table 9: Outcome Parameter Estimates

	Sample Attrition				Mortality			
	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.
Antidepressant	-0.133	0.435	-0.131	0.598	0.661	0.358	0.647	0.378
Antidepressant*								
CES-D$\in [5, 10]$	0.204	0.459	0.176	0.629	-0.174	0.453	-0.163	0.479
CES-D$\in [11, 51]$	0.097	0.432	0.114	0.586	-0.317	0.421	-0.312	0.450
Female	0.214	0.369	0.206	0.389	-0.403	0.344	-0.392	0.352
CES-D$\in [5, 10]$	0.309	0.135	0.328	0.143	0.146	0.113	0.152	0.115
CES-D$\in [11, 51]$	0.422	0.140	0.509	0.149	0.085	0.123	0.114	0.126
Female	0.107	0.114	0.090	0.119	-0.430	0.104	-0.446	0.109
Light Drinking	-0.024	0.116	-0.197	0.162	-0.390	0.102	-0.390	0.143
Heavy Drinking	-0.061	0.162	-0.466	0.246	-0.260	0.134	-0.362	0.224
Smoking	0.470	0.177	0.477	0.182	0.347	0.144	0.330	0.145
Years Smoking	0.006	0.004	0.004	0.004	0.012	0.003	0.011	0.003
Years Smoking Cessation	-0.003	0.004	-0.002	0.004	-0.007	0.004	-0.006	0.004
Age								
(40, 45]	-0.654	0.348	-0.644	0.386	1.176	0.771	1.170	0.476
(45, 50]	-1.068	0.343	-1.050	0.384	1.770	0.736	1.769	0.410
(50, 55]	-1.371	0.333	-1.350	0.378	1.965	0.727	1.970	0.390
(55, 60]	-1.168	0.314	-1.148	0.363	2.304	0.721	2.300	0.375
(60, 65]	-0.854	0.312	-0.834	0.365	2.534	0.720	2.533	0.371
(65, 70]	-0.949	0.333	-0.932	0.387	2.724	0.725	2.719	0.375
(70, 75]	-0.921	0.350	-0.900	0.407	3.060	0.729	3.057	0.381
>75	0.241	0.341	0.254	0.399	3.706	0.732	3.700	0.387
Education								
High School	-0.052	0.204	-0.028	0.216	-0.177	0.149	-0.177	0.155
Some College	-0.158	0.205	-0.124	0.218	-0.344	0.152	-0.349	0.159
College or More	-0.344	0.244	-0.311	0.258	-0.564	0.198	-0.575	0.205
CVD this period	-0.062	0.172	-0.063	0.178	1.838	0.105	1.843	0.108
Any History of CVD	0.115	0.141	0.112	0.144	0.484	0.105	0.485	0.108
Cancer this period	-0.024	0.166	-0.028	0.167	1.632	0.105	1.638	0.107
Any History of Cancer	-0.340	0.143	-0.359	0.145	0.888	0.111	0.885	0.112
Obese	0.143	0.119	0.152	0.120	-0.081	0.109	-0.085	0.110
Currently Working	0.041	0.140	0.047	0.147	-0.381	0.135	-0.381	0.140
Work Missing	0.150	0.178	0.151	0.193	-0.150	0.133	-0.147	0.135
Married	0.674	0.186	0.691	0.196	-0.140	0.145	-0.132	0.149
Married Missing	0.536	0.248	0.547	0.288	0.115	0.159	0.116	0.163
Exam Trend	0.703	0.057	0.703	0.061	-0.076	0.036	-0.076	0.039
Constant	-7.769	0.455	-7.520	0.522	-5.336	0.764	-5.467	0.488
μ_1			0.000	.			0.000	.
μ_2			0.167	0.231			0.318	0.241
μ_3			-0.314	0.216			0.208	0.209
μ_4			-0.368	0.272			0.067	0.263

Notes: $n = 17,258$. Selected parameter estimates are from models estimated on data in exams 2-9.

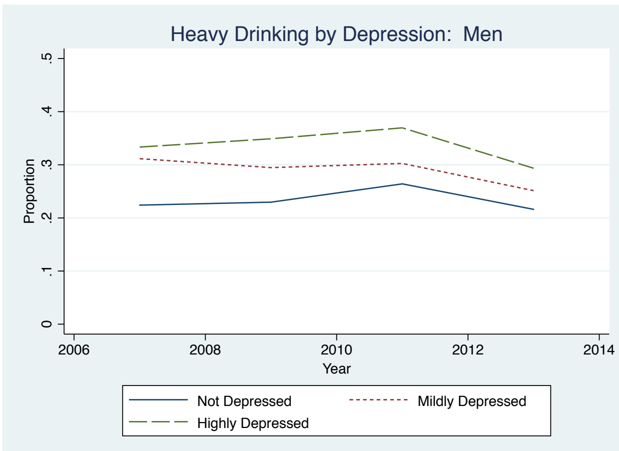
Table 10: Initial Conditions Parameter Estimates

	Light Drinking		Heavy Drinking		Medium Depression		Heavy Depression		Smoking	
	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.
Age	-0.038	0.011	-0.025	0.014	-0.005	0.009	-0.028	0.010	-0.016	0.008
Female	-0.739	0.126	-0.869	0.163	0.150	0.101	0.477	0.108	-0.524	0.087
Education										
High School	0.303	0.242	-0.025	0.334	-0.181	0.218	-0.321	0.222	-0.481	0.172
Some College	0.612	0.242	0.060	0.328	-0.265	0.217	-0.679	0.223	-0.870	0.172
College or More	1.146	0.289	0.219	0.380	-0.253	0.242	-0.827	0.259	-1.283	0.201
Age > 50	0.397	0.215	0.542	0.268	-0.059	0.180	0.116	0.191	-0.067	0.156
					-0.005	0.154	0.059	0.155		
					0.027	0.216	0.936	0.249		
					-0.251	0.146	-0.043	0.153		
					0.015	0.005	0.022	0.005		
					-0.005	0.007	-0.011	0.007		
Constant	3.952	0.583	3.025	0.726	0.227	0.514	0.501	0.544	1.107	0.364
μ_1	0.000	.	0.000	.	0.000	.	0.000	.	0.000	.
μ_2	-0.184	0.574	1.410	0.613	0.070	0.269	-1.062	0.330	0.744	0.160
μ_3	-1.463	0.351	-3.944	0.523	0.081	0.205	0.811	0.227	-0.139	0.157
μ_4	-3.512	0.340	-3.560	0.439	0.160	0.275	1.084	0.277	-0.065	0.163

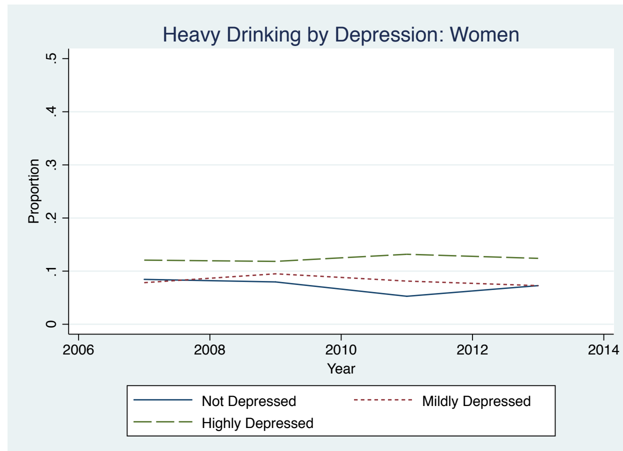
Notes: $n = 17,258$. Selected parameter estimates are from initial condition models. For smoking and drinking, models are estimated on data from exam 2. For depression, data come from the exam 3 CES-D survey.

B Main Figures

Figure 1: Alcohol and Depression: Evidence from NHANES



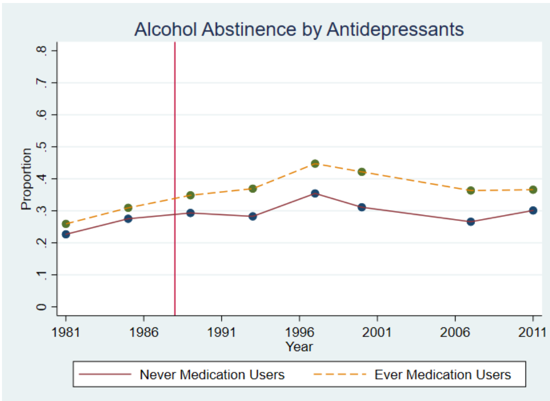
a. Heavy Drinking by Depression, Men



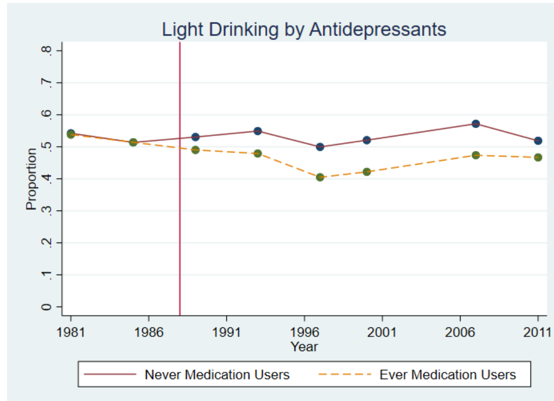
b. Heavy Drinking by Depression, Women

Notes: Author's calculations from NHANES data from 2007-2013. Proportions are weighted by the NHANES full sample 2-year interview weight. Proportions are presented by tertiles of the Patient Health Questionnaire (PHQ-9) Depression Score. $n = 16,940$.

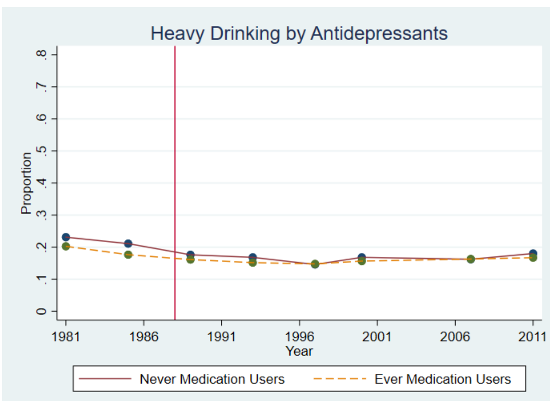
Figure 2: Behavior Over Time by Ever Taking Antidepressants



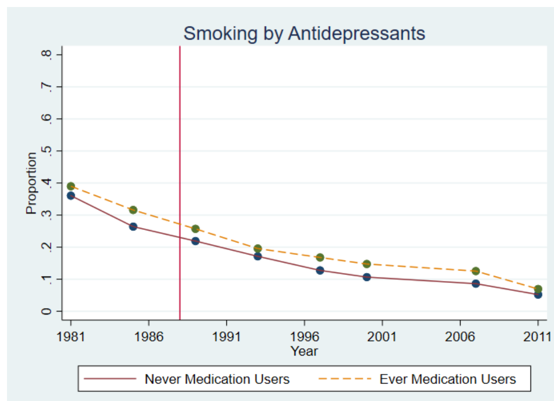
a. No Alcohol



b. Light Drinking



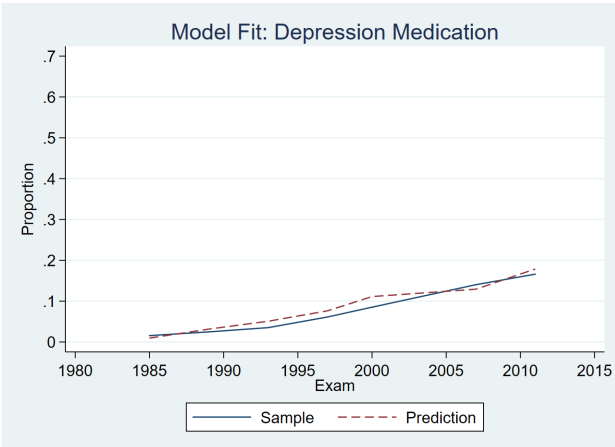
c. Heavy Drinking



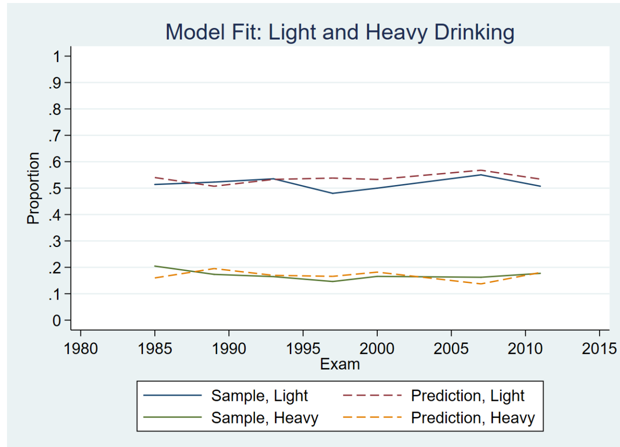
d. Smoking

Notes: The figures represent the time path of each behavior by whether or not an individual is ever observed to take an antidepressant. The vertical line represents 1988, the year of SSRI approval.

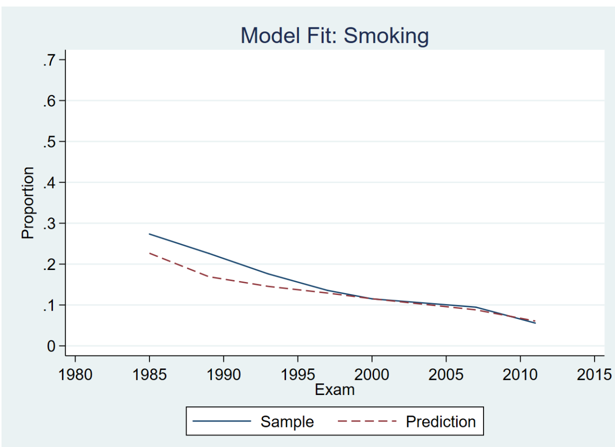
Figure 3: Model Fit



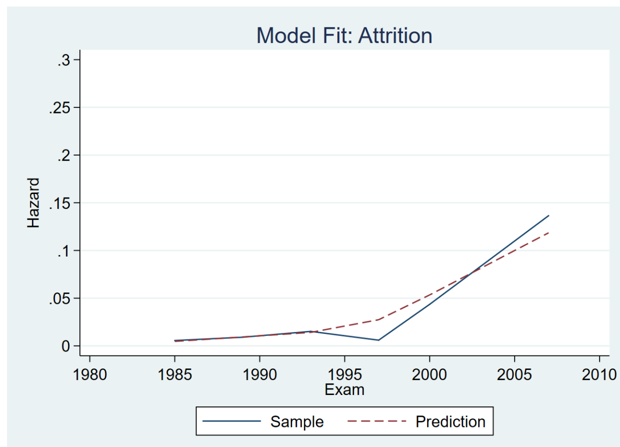
a. Antidepressants



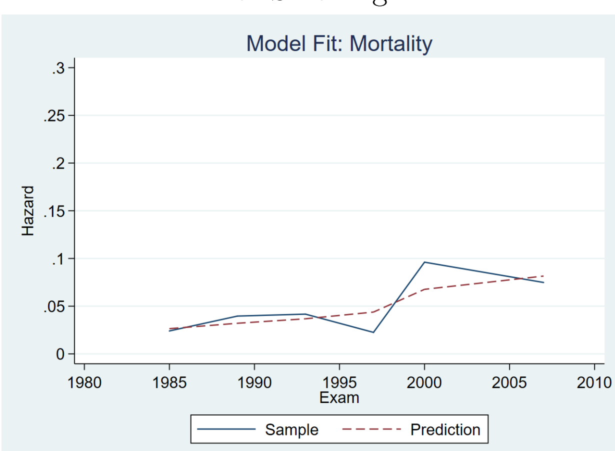
b. Light and Heavy Drinking



c. Smoking



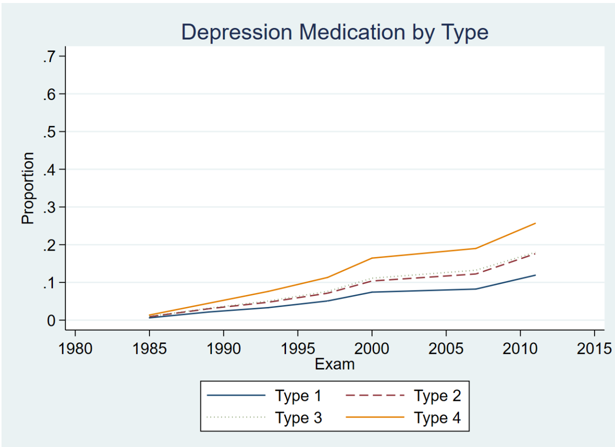
d. Sample Attrition



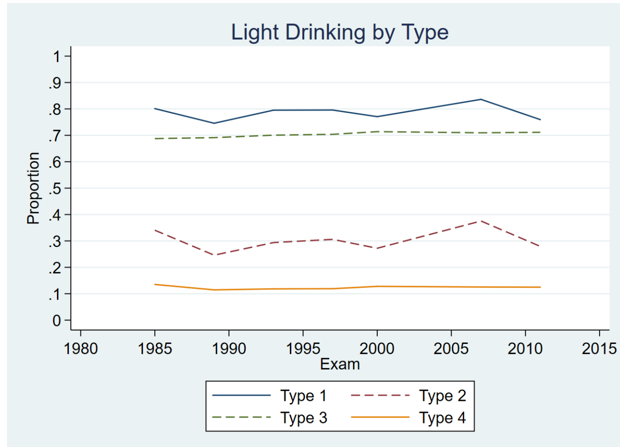
e. Mortality

Notes: Each figure presents results from the baseline simulation of our estimated dynamic model relative to sample data.

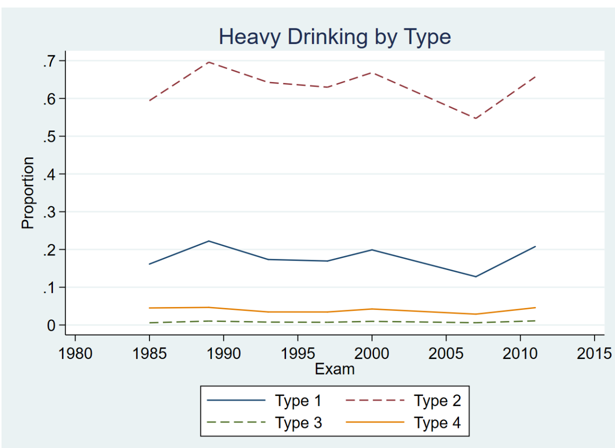
Figure 4: Behaviors and Outcomes by Unobserved Type



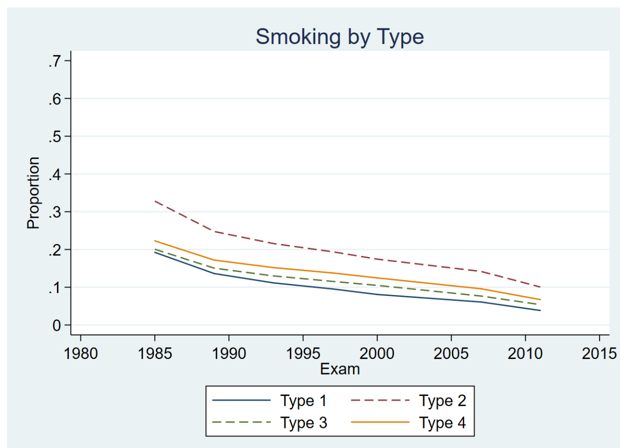
a. Antidepressants



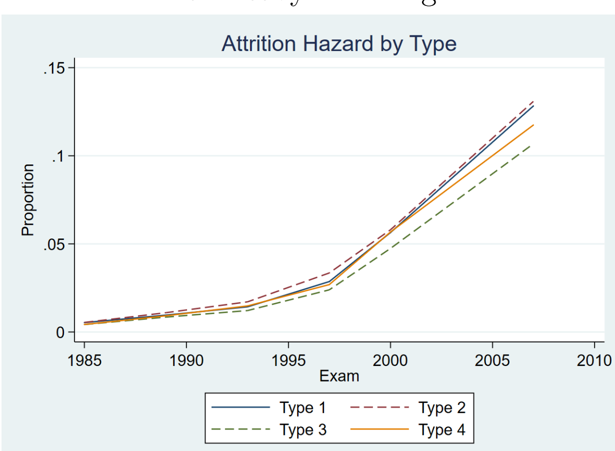
b. Light Drinking



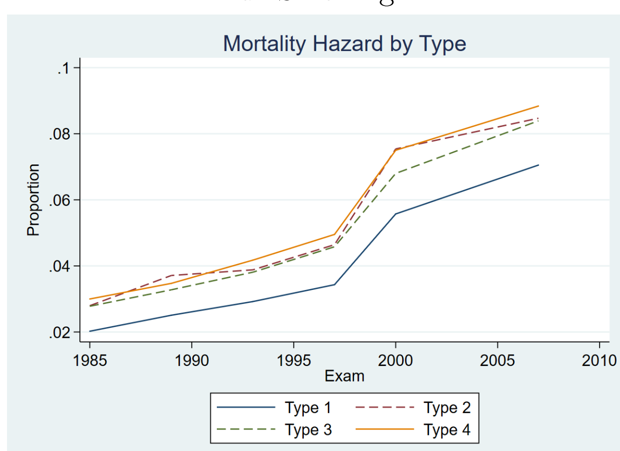
c. Heavy Drinking



d. Smoking



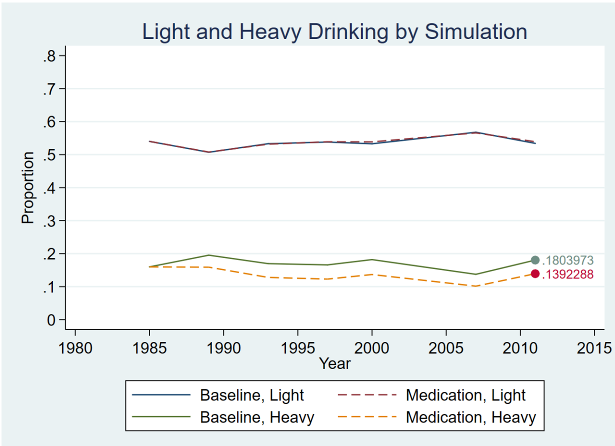
e. Sample Attrition



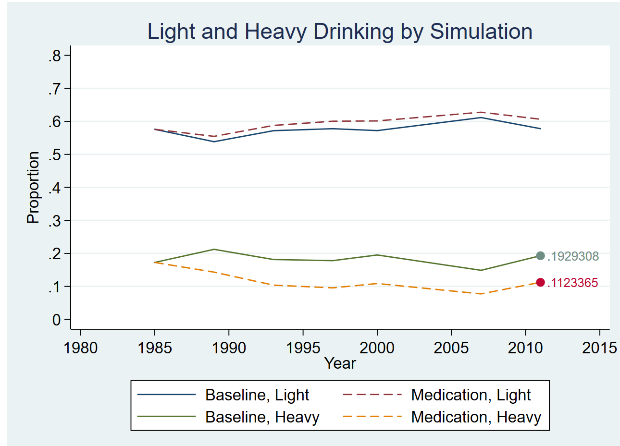
f. Mortality

Notes: Each figure presents results from the baseline simulation of our estimated dynamic model by each of the four unobserved types.

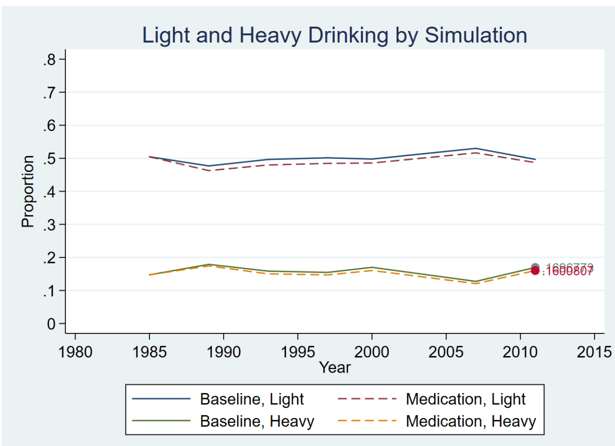
Figure 5: Comprehensive Antidepressants vs. Baseline: Alcohol Consumption



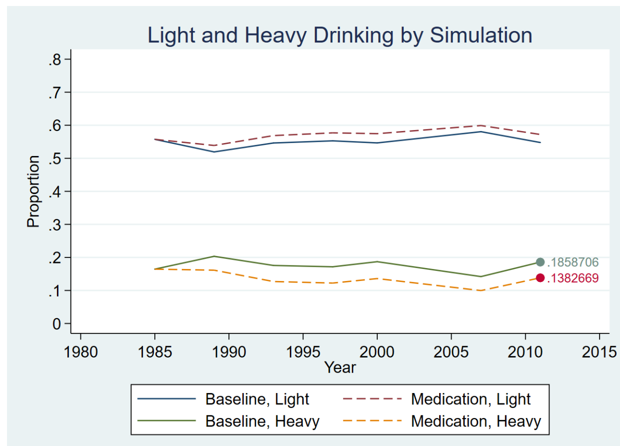
a. Overall



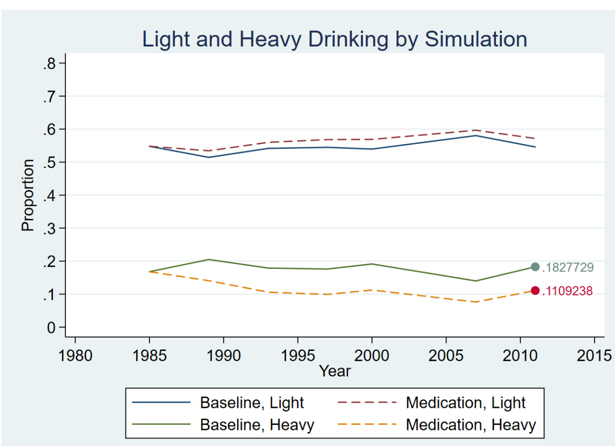
b. Men



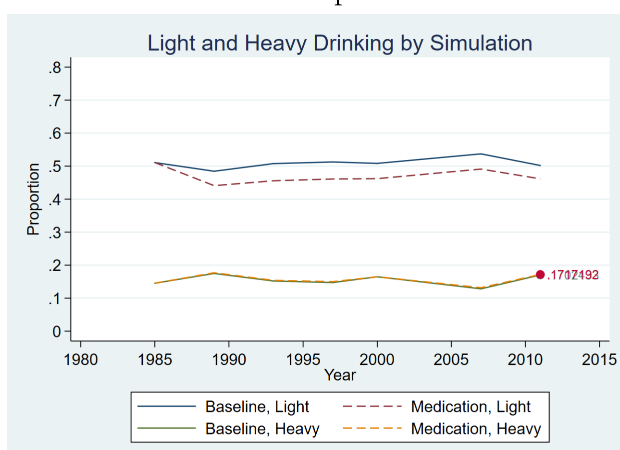
c. Women



d. Low Depression



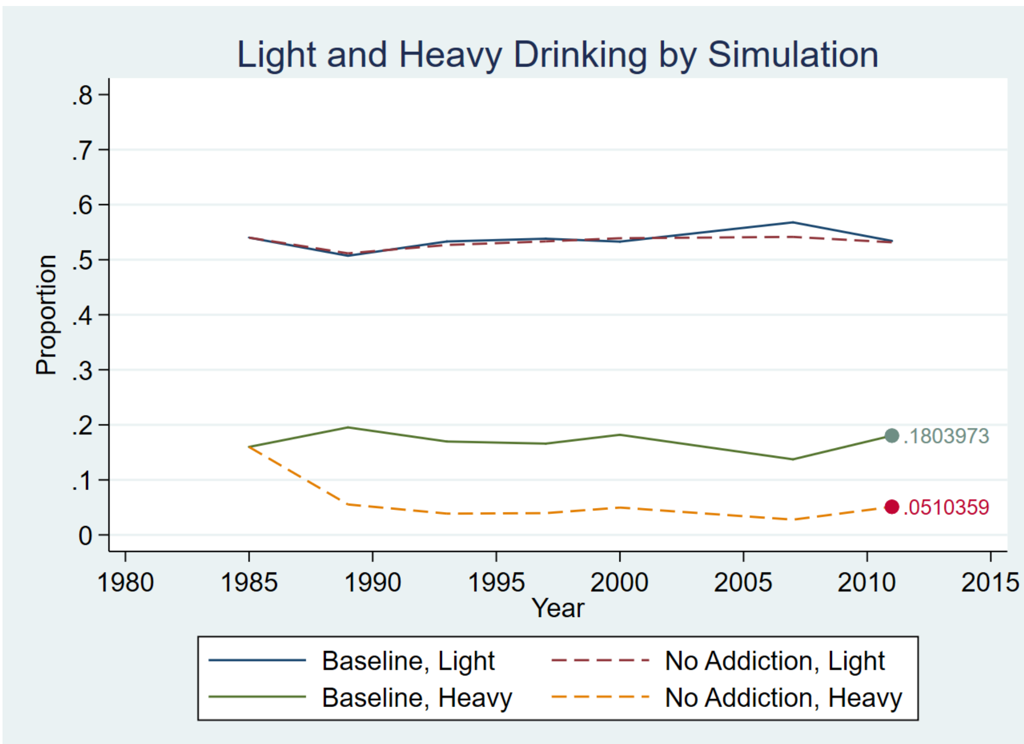
e. Medium Depression



f. High Depression

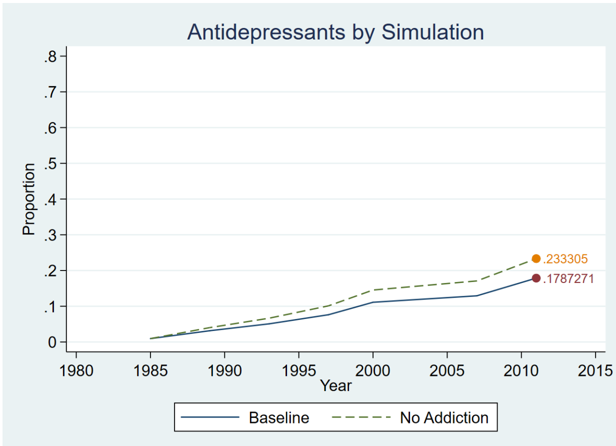
Notes: Each figure presents baseline simulated trends in light and heavy drinking as well as those behaviors when we impose that all individuals take an antidepressant from exam 4 onwards. Figure 5a presents the simulations for the entire sample. Figures 5b and 5c present results separately for men and women. Figures 5d, 5e, 5f present results for those simulated at exam 3 to be in the low, medium, or high tertiles of CES-Depression score.

Figure 6: Alcohol Consumption by Simulation

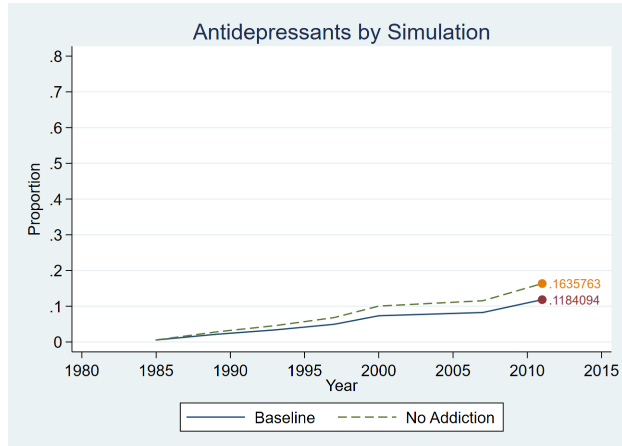


Notes: Figure displays light and heavy smoking under the counterfactual scenario that past alcohol consumption does not factor in any of the contemporaneous period behavioral equations. Results are presented relative to the baseline simulation.

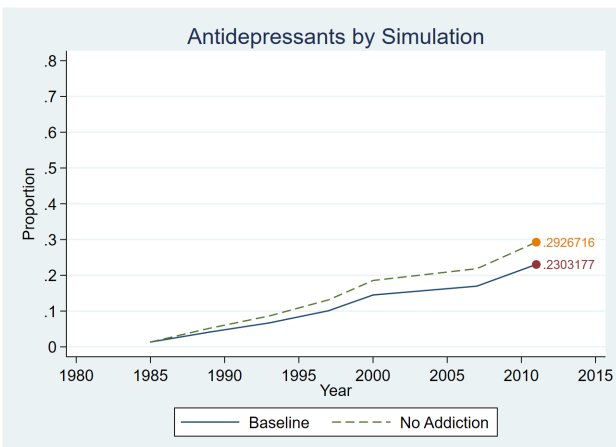
Figure 7: Antidepressant Consumption by Simulation



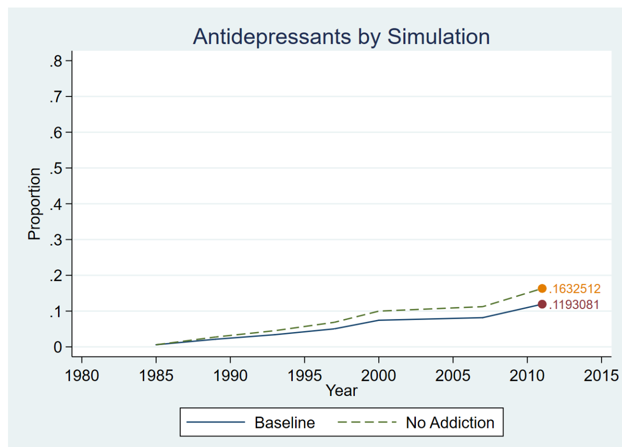
a. Overall



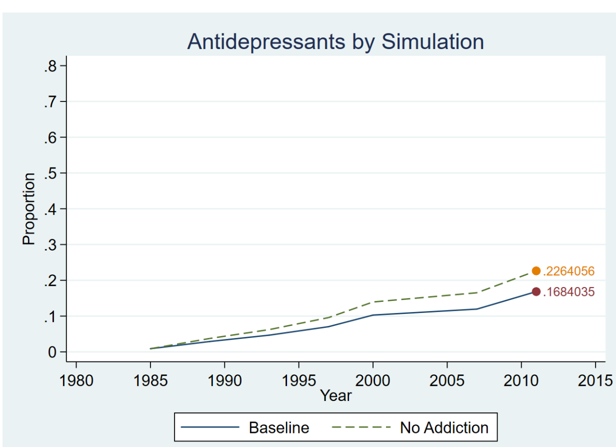
b. Men



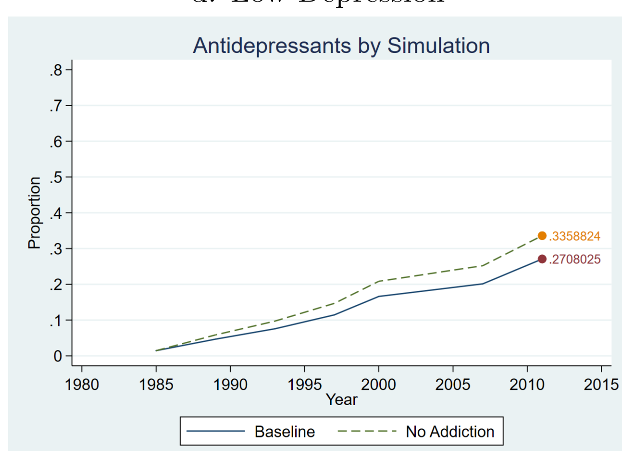
c. Women



d. Low Depression



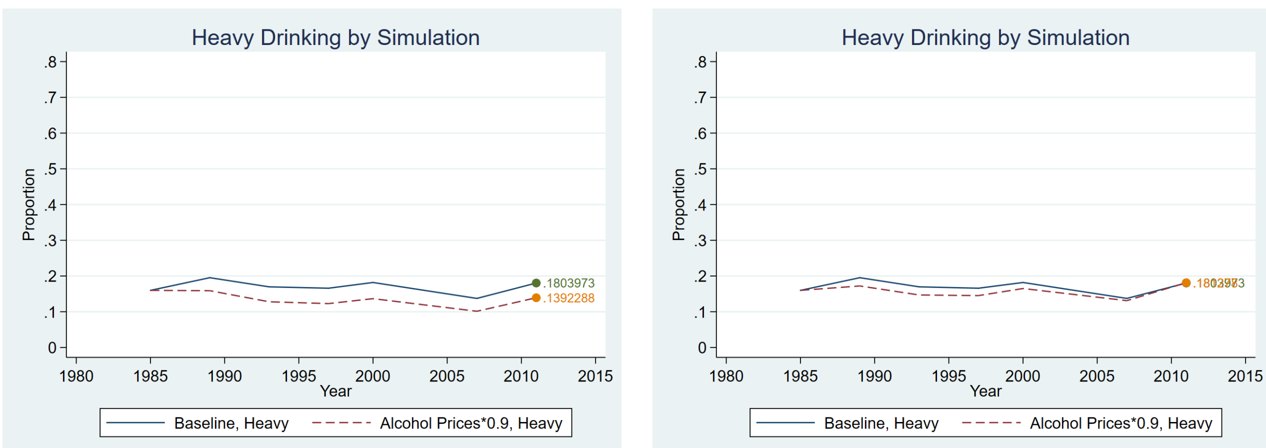
e. Medium Depression



f. High Depression

Notes: Each figure presents simulated trends in antidepressant usage under the baseline scenario as well as under the counterfactual in which we remove the dependence on past alcohol consumption in all behavioral equations. Figure 7a presents the simulations for the entire sample. Figures 7b and 7c present results separately for men and women. Figures 7d, 7e, 7f present results for those simulated at exam 3 to be in the low, medium, or high tertiles of CES-Depression score.

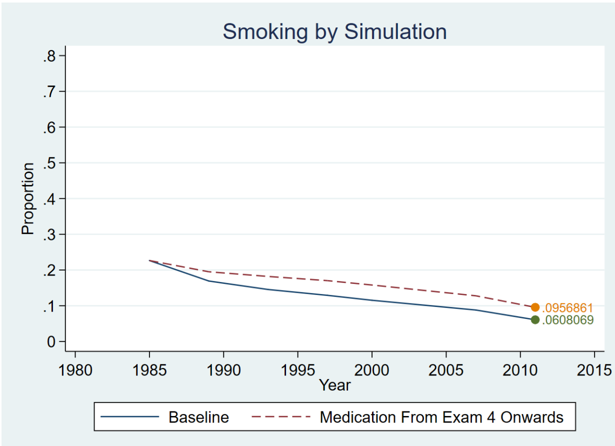
Figure 8: The Role of Alcohol Prices



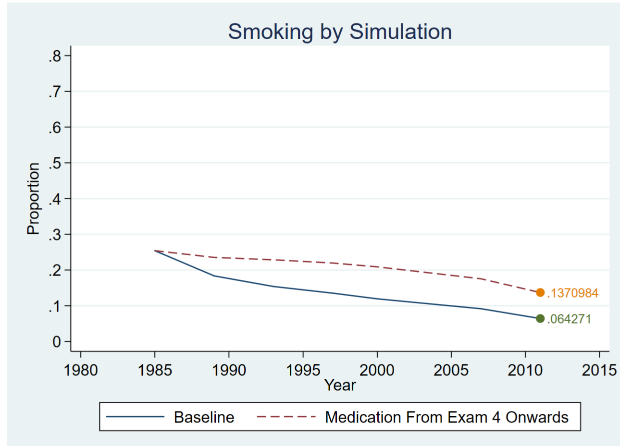
a. Simulation 1 b. Simulation 1 + price effect

Notes: Figure 8a presents simulated trends in heavy alcohol consumption under the baseline scenario as well as under the counterfactual in which we impose antidepressants on all participants at exam 4. This figure is identical to the heavy drinking trend presented in Figure 5. Figure 8b presents the same baseline simulation in heavy drinking along with imposed antidepressants and a decrease in alcohol prices by 10% of baseline levels in all exams after the third.

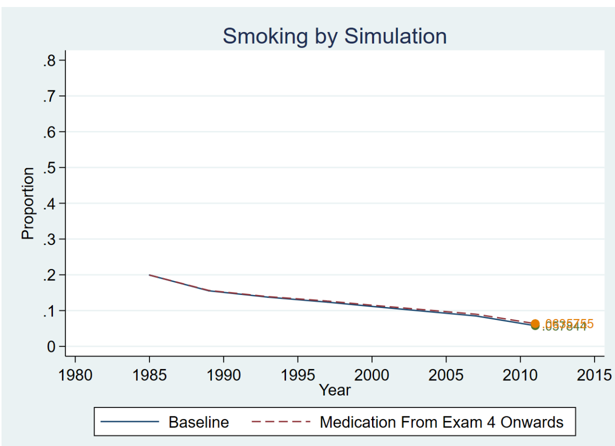
Figure 9: Comprehensive Antidepressants vs. Baseline: Smoking



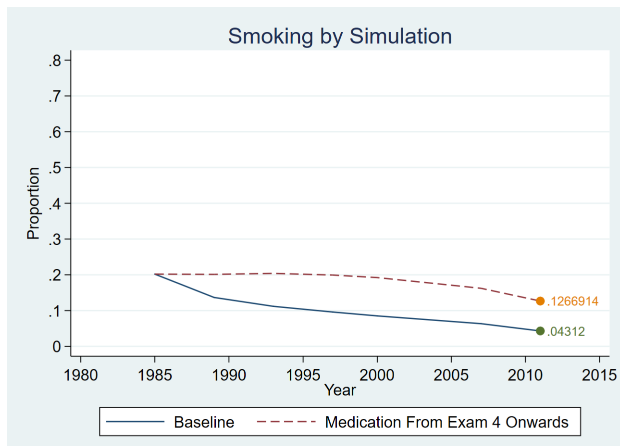
a. Overall



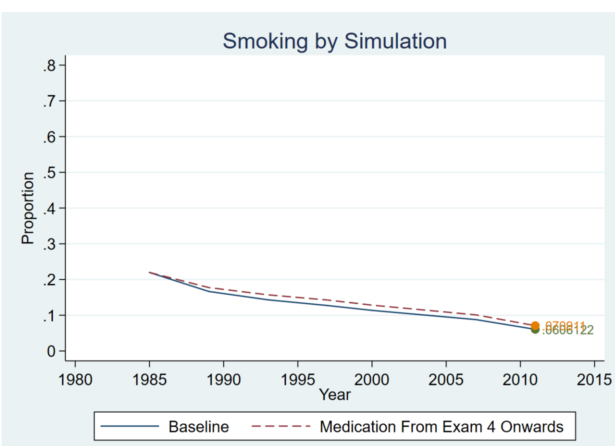
b. Men



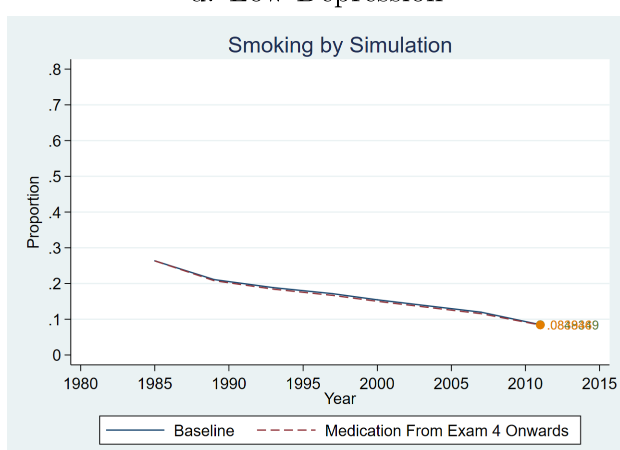
c. Women



d. Low Depression



e. Medium Depression



f. High Depression

Notes: Each figure presents simulated trends in smoking under both the baseline scenario as well as those behaviors when we impose that all individuals take an antidepressant from exam 4 onwards. Figure 9a presents the simulations for the entire sample. Figures 9b and 9c present results separately for men and women. Figures 9d, 9e, 9f present results for those simulated at exam 3 to be in the low, medium, or high tertiles of CES-Depression score.