WORKING PAPER · NO. 2018-14

Promoting Wellness or Waste? Evidence from Antidepressant Advertising

Bradley Shapiro

May 2018

Becker Friedman Institute
FOR ECONOMICS AT UCHICAGO

1126 E. 59th St, Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

Promoting Wellness or Waste? Evidence from Antidepressant Advertising

Bradley T. Shapiro*

This Version May 2018
[Working Draft. Comments Welcome. Check SSRN for new draft before citing or re-circulating.]

Abstract

Direct-to-Consumer Advertising (DTCA) of prescription drugs is controversial and has ambiguous potential welfare effects. In this paper, I estimate costs and benefits of DTCA to patients and payers in the market for antidepressant drugs to assess whether advertising marginal prescriptions are desirable on average. In particular, using individual health insurance claims and human resources data, I estimate the effects of DTCA on outcomes relevant to patient and payer costs: new prescriptions, prices and adherence. Additionally I estimate the effect of DTCA on labor supply, the economic outcome most associated with depression. First, category expansive effects of DTCA found in past literature are replicated, with DTCA particularly causing new prescriptions of antidepressants. Meanwhile, concurrent advertising drives slightly lower refill rates. Additionally, I find evidence of no advertising effect on the prices or co-pays of the drugs prescribed, the generic penetration rate or the rate of adverse effects. Finally, advertising significantly decreases missed days of work, with the effect concentrated on workers who tend to have more absences. Back-of-the-envelope calculations suggest that the wage benefits of the advertising marginal work days are more than an order of magnitude larger than the total cost of the advertising marginal prescriptions.

1 Introduction

Direct-to-Consumer Advertising (DTCA) of prescription drugs is controversial. Much of the controversy stems from ambiguous potential welfare effects. On the positive side, DTCA could provide information that encourages sick people to seek help from their physicians to potentially get better, either through drug treatment or an alternative. Alternatively, DTCA could be socially costly. Since patients tend not to pay the full cost of each prescription with insurance, advertising may inefficiently drive marginal patients to get prescribed when the benefits do not exceed the full cost. Additionally, DTCA may inefficiently induce switches away from inexpensive generic drugs to expensive branded drugs. Finally, DTCA could mislead individuals into believing a drug has value for them when it has little. The net social effect ultimately depends upon to the shapes of the benefit and cost curves for antidepressants with respect to DTCA and where society currently lies on those curves.

In this paper, I evaluate whether or not DTCA generates socially inefficient prescriptions by directly measuring costs and benefits to consumers and insurers. Computing social welfare directly is difficult. First, consumers often value

*Bradley.Shapiro@chicagobooth.edu. University of Chicago Booth School of Business. I thank Stephen Lamb for excellent research assistance. I particularly thank Amanda Starc, Daniel Sacks, Jason Abaluck, Jean-Pierre Dube, Sanjog Misra and Stephen Ryan for detailed comments. Results calculated based on data from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business and from Truven Health Analytics, an IBM company, provided by the Center for Health and the Social Sciences (CHeSS) at the University of Chicago. The conclusions drawn from the Nielsen data and the Truven data are those of the researchers and do not reflect the views of either Nielsen or Truven. Nielsen and Truven are not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. I thank Chris Lyttle for assistance with Truven data. I acknowledge generous financial support from the Beatrice Foods Co. Faculty Research Fund at the University of Chicago Booth School of Business.

attributes of products that are not observable to the researcher. Second, firm profits are part of social welfare, but firm costs are unobservable. As a result, economics researchers often use revealed preference to measure consumer preferences and use that combined with a model of firm conduct to make inferences about social welfare. In the case of DTCA, such a revealed preference approach is not ideal. First, misleading advertising may generate prescriptions that bring patients less value than the price they face, undermining the premise of using revealed preference to compute consumer surplus. Second, advertising may generate prescriptions that are privately more valuable to patients than the copays but are less valuable than the full price the insurers pay, generating externalities onto insurance markets.

As a result, the approach in this paper is to directly measure whether advertising marginal prescriptions bring enough value to justify their costs to consumers and to insurers. There are a number of costs considered. First, increased prescriptions from advertising lead to a direct cost, the price of the drug. Second, it is possible that advertising steers consumers to more expensive drugs, conditional on treatment.[1] Third, I evaluate whether and to what extent advertising leads to increased adverse effects or increased probability of failing to complete a course of treatment. Finally, I measure whether and to what extent advertising increases the rate of adverse drug effects. On the benefit side, I measure the effect of DTCA on labor supply, which is the main observable outcome associated with depression.

This approach has the limitation that only observable benefits and costs can be measured. These potential costs and benefits are not exhaustive, but they are important and provide an guide for thinking about how big any unmeasured costs or benefits must be to swamp those measured here. This approach provides a rather stringent test for thinking about social welfare, as increased firm profits could offset some potential losses to consumers and insurers to the extent that policy makers wish to consider firm profits in social welfare.

Depression is a condition that affects roughly 10% of Americans at any moment in time and is a chemical imbalance in the brain leading to decreased self-worth. In economic terms, it is characterized by the systematic underestimation of one's marginal product of effort (De Quidt and Haushofer (2016)) and has been associated both with large direct costs of medical care as well as large indirect costs of reduced economic activity [Berndt et al. (2000); Currie and Madrian (1999); Greenberg et al. (1993a,b); Stoudemire et al. (1986); Woo et al. (2011); Tomonaga et al. (2013); Stewart et al. (2003); Boyer et al. (1998)]. Survey evidence (Kessler et al. (2003)) suggests that only about half of those who have experienced psychiatric disorders have received any kind of professional treatment.

Total DTCA of prescription drugs, while significant, has decreased from about $3 billion in 2004 to a little over $2 billion in 2012. Meanwhile, antidepressant DTCA makes up an important fraction of total DTCA and has increased from about $200 million in 2004 to a peak of about $400 million in 2011, declining to about $300 million in 2012.

I replicate findings in the literature [Iizuka and Jin (2005, 2007); Shapiro (2018); Sinkinson and Starc (2017); Alpert et al. (2015); Hosken and Wendling (2013)] that DTCA induces more patients to be prescribed antidepressants with an elasticity of about 0.031, leading to a direct cost of DTCA to consumers and insurers. Second, I add to the literature on advertising and drug treatment adherence [Donohue et al. (2004); Cardon and Showalter (2015); Wosinska (2005)]. While that literature finds mixed results, I find that advertising reduces refill prescriptions by a small amount. Next, I find evidence against DTCA having an economically meaningful impact on either the price or the co-pay of the drug, conditional on prescription. I also find evidence against an economically meaningful effect of advertising on the generic penetration rate. Finally, I find that DTCA causes benefits in the form of increased labor supply. The benefits of increased labor supply outweigh the total cost of additional prescriptions by more than an order of magnitude. The preferred estimates imply that 10% increase in DTCA brings $769.5 million in wage benefits while generating $32.4 million in prescription costs. This implies that on average, DTCA is generating prescriptions that are worth more to patients than the cost. In other words, the average DTCA marginal prescription is not a 'mistake.' In addition to the dollar costs and benefits, I find that advertising does not predict increased adverse effects or increased failure to complete a full course of treatment, indicating that advertising marginal prescriptions are no more likely to be poorly

[1] I should note here that if advertising inefficiently steers consumers to expensive drugs, insurers could respond by increasing the co-payment or coinsurance rates for advertised drugs, so steering by itself need not be welfare negative after considering insurer responses. As this study will exploit month-over-month variation in advertising while insurance formularies rarely change within a year, the effect of advertising on transacted prices can be interpreted as the partial effect of steering without insurer response.

tolerated than average prescriptions.

If employers have market power in the labor market and employees are paid less than their marginal product, then employers will also see dollar benefits of the increased labor supply that I do not measure. Additionally, incremental profits to pharmaceutical firms caused by DTCA could further positively tilt the full social welfare. To my knowledge, this is the first paper linking DTCA to measurable benefits and costs to consumers.

This empirical exercise comes with challenges. First, advertising is endogenously chosen by firms in a way that might lead advertising to be spuriously correlated with sales and outcomes. Second, labor supply is determined by many factors other than depression and by extension, antidepressant advertising. This leads to a problem of low statistical power in the estimation of the effect of DTCA on labor supply. Finally, any effects of advertising on labor supply are not expected to materialize immediately, as it takes time for a patient to begin to show improvement from treatment. Antidepressants, in particular, take on average six weeks before they show beneficial effects, but with wide variance (Frazer and Benmansour (2002)). The need to evaluate both current and lagged advertising effects exacerbates power issues.

To overcome the endogeneity of advertising, I take advantage of the panel nature of the data to take into account both individual-specific differences in labor supply and systematic seasonal variation. To control for remaining endogeneity, I make use of random variation in advertising generated by the borders of television markets, as in Shapiro (2018). Despite decreasing the number of observations in estimation, focusing on borders in this case increases statistical power. Seasonal factors that impact labor supply, such as weather and industry type, are highly geographically correlated. By making close geographic comparisons, variation in labor supply driven by factors other than advertising is considerably reduced. The reduction of noise in this case outweighs the reduction in observations that would decrease power.

The contributions of this paper are threefold. First and most importantly, the paper provides the first causal link between DTCA and both social benefits and costs measured in dollars. Previous papers have linked DTCA to non-demand outcomes, but few with clear welfare implications. For example, Kim and KC (2017) links advertising for the erectile dysfunction drug, Viagra, to birth rates, Niederdeppe et al. (2017) links statin advertising both to increased exercise and to increased fast food consumption, David et al. (2010) finds some evidence of advertising increasing adverse effect reporting and Chesnes and Jin (2016) find that advertising drives consumers to search for information about the drugs online. In contrast to these studies, this paper ties advertising to both direct costs to consumers (demand and prices) as well as indirect benefits (increased labor supply). In establishing these costs and benefits, this paper provides the first direct quasi-experimental evidence that DTCA does not steer patients to more expensive drugs. As this is one of the main criticisms of DTCA in the policy community, establishing this fact in the case of antidepressants is independently important. Additionally, this paper provides the first evidence of a link between DTCA and labor market outcomes.

Second, this paper adds to the literature which traces out the benefits of access to medical care in terms of labor market outcomes. While previous papers have found effects of new technologies on stark margins, this is the first paper to show that advertising marginal access to treatment can have meaningful effects. Garthwaite (2012) and Bütikofer and Skira (2016) find that when the Coxx-2 inhibitor, Vioxx, was pulled from the market for fear of adverse effects, there was a substantial decrease in labor supply. Papageorge (2016) links innovation of HIV drugs to patient choices about labor supply and medical treatment. Currie and Madrian (1999) provide an excellent review of the literature linking various types of health and access to treatment through insurance to labor supply, noting a particular link between mental health and labor supply. Deshpande (2016) finds that when low-income youth lose supplemental security income (SSI) benefits and their corresponding eligibility for Medicaid, their economic outcomes become far worse.

Third, this paper adds to the marketing literature thinking about the relationship between advertising and selection. While such a relationship is policy relevant in this particular case, research studying the types of individuals affected by advertising is sparse and is generally important to understanding both whether advertising is good socially and whether it is worthwhile to the firm. In terms of health, Aizawa and Kim (2018) show that if health insurance could select on

health status using advertising, it could have a substantial equilibrium effect on prices, while Shapiro (2017) in that same market finds evidence that advertising provides no such advantageous selection. In the market for mortgages, Grundl and Kim (2017) find that advertising is both targeted at and more effective on people who stand to gain from re-financing. This study shows three pieces of evidence on the relationship between advertising and selection. First, it shows that the balance of the effect of antidepressant DTCA is on those that stand to gain in the form of increased labor supply. Second, it shows that those affected increase their labor supply. Third, it shows that those who are advertising marginal are no more likely to discontinue use than the average patient.

The paper proceeds as follows. Section 2 briefly discusses depression and its economic impacts. Section 3 outlines the possible mechanisms of DTCA being socially beneficial or socially harmful in a simple framework. In section 4, the data used in the study are discussed. Section 5 details the research design, focusing on the borders of television markets. Section 6 presents the results, and section 7 concludes.

2 Depression

Major depressive disorder (MDD) is a chemical imbalance in the brain that leads to numerous direct and indirect costs. It leads to emotional detachment and decreased self worth. De Quidt and Haushofer (2016) provide a nice framework to think about depression from the perspective of economic theory. In particular, it models individuals as unsure of how much to attribute their productivity to luck or to their own efforts. If an individual gets enough repeated unfavorable draws from the luck distribution, he or she will Bayesian update to believe that the low productivity is innate. The belief that the individual has low marginal product leads to lower effort and investment in human capital. This framework provides a theoretical and rational basis for the well documented connection in the medical literature between depression and labor supply.

Providing evidence to posited economic effects, Berndt et al. (2000) finds that early onset depression causes substantial human capital loss. Greenberg et al. (1993a),Stoudemire et al. (1986),Boyer et al. (1998) and Tomonaga et al. (2013) all find that the economic costs of depression in terms of labor supply and productivity are far in excess of the average cost of treatment. Stewart et al. (2003) estimates the productivity cost of depression to be about $31 billion to employers in the US. Greenberg et al. (1993b) similarly estimates the annual costs of depression to be about $44 billion per year in the US. Woo et al. (2011) finds that workers with MDD lose about 30% of their annual salaries to costs associated with missing work or being unproductive at work. Additionally, Greenberg et al. (2015) update previous results and find that the economic burden increased substantially between 2005 and 2010.

In terms of other effects of depression and its treatment, Sobocki et al. (2007) find that depression is associated with significantly lower health related quality of life instrument scores, but people who initiated treatment saw improvement. Stewart et al. (2003) finds that self reported use of antidepressants among people with depression is only around 30% even though reported effectiveness was moderate. Consistently, Bharadwaj et al. (2015) posits that because mental health often carries with it a stigma, it might be expected that society is still in the steep part of the marginal benefit curve with respect to depression treatment. In particular, it finds that survey respondents are likely to lie and say they do not have depression when medical records indicate otherwise. This effect is not the same for less stigmatized health conditions.

Additionally, survey evidence in Kessler et al. (2003) has found that around half of those afflicted with depression have failed to get any kind of treatment. Some of these people may fail to get treated because they are poorly suited to or cannot afford available treatments. Beyond this, it is difficult to know how many people fail to accurately report depression in surveys due to either stigmatization or lack of understanding of the exact nature of the condition. As a result, there may be some margin for improved information, de-stigmatization and treatment of depression, but it is unclear whether or not DTCA is an effective means of targeting those with depression who stand to gain from treatment but who are as of yet untreated.

3 The Welfare Economics of DTCA

The social desirability of DTCA is the subject of considerable controversy, and it is legal in only the United States and New Zealand. A ban on DTCA was part of Hillary Clinton's 2016 platform as a presidential candidate, Senator Al Franken sponsored legislation to end the tax deductibility of DTCA, and recently, the American Medical Association (AMA) and the American Society of Health System Pharmacists (ASHP) came out in favor of a ban on DTCA. The main arguments opposing DTCA are twofold. First, advertising might mislead consumers into believing a drug would benefit them when it would not. This leads them to make unreasonable requests of their physicians which are often honored. Second, advertising steers patients to more expensive brands when less expensive generics are available.

Not all share these views. The position of the Pharmaceutical Research and Manufacturers of America (PhRMA) is that advertising provides information about diseases and treatments that some consumers would otherwise not have. In the absence of that information, these patients would go untreated and miss out on important benefits of treatment. The FDA regulates the content of these ads to ensure that risks are presented and that claims are scientifically justifiable.

DTCA could affect consumers decision in many ways, some of which are good for society and some of which are bad. In this case, I focus on measuring whether advertising marginal prescriptions are inefficient from the perspective of the patient or the payer.[2] To help fix ideas about mechanisms, I present a simple framework for how advertising affects prescription choice. Assume a simple expected utility model whereby each consumer expects utility from antidepressant drugs:

$$E[u_{ij}(A)] = I_{ij}(A) * [E[v_{ij}|A] - p_{ij}]$$

where $I_{ij} \in \{0,1\}$ reflects whether or not consumer i is informed of the existence of product j, $E[v_{ij}]$ is consumer i's expectation of the value received from product j, p_{ij} is the price that consumer i faces for product j and A_j is advertising. Consumer i buys product j if

$$E[u_{ij}(A)] > E[u_{ik}(A)] \, \forall \, k \neq j,$$

where one k is the outside option of getting no antidepressant. Through this simple framework, it is straightforward to highlight arguments for and against DTCA formally. The negative view of DTCA can be translated into this framework as $E[v_{ij}|A > 0] > v_{ij}$. That is, advertising causes consumers to have a more optimistic view of how well a drug will work than reality. If this effect is for the advertised drug, which tends to be branded, it will bias decision making in favor of expensive branded drugs. It could be that for some generic drug g and some branded drug j, $E[v_{ig}|A = 0] - p_{ig} > E[v_{ij}|A = 0] - p_{ij}$, but $E[v_{ig}|A = 0] - p_{ig} < E[v_{ij}|A > 0] - p_{ij}$, leading that consumer to choose the brand when she otherwise would have chosen the generic. If $E[v_{ij}|A > 0] > v_{ij}$, this decision could be a mistake.

As long as $E[v_{ij}|A] > v_{ij}$, the informative effect is welfare ambiguous. It could be that $E[v_{ij}|A] - p_{ij} > v_{ij} - p_{ij} > max_{k \neq j}\{v_{ik} - p_{ik}\}$, so a prescription is still better than no prescription for that individual, despite the biased expectation. Alternatively, it could be that $E[v_{ij}|A] - p_{ij} > max_{k \neq j}\{v_{ik} - p_{ik}\} > v_{ij} - p_{ij}$, making the prescription inefficient– that individual would have been better off with a different choice.

The positive view on DTCA is can be translated into this framework as $|E[v_{ij}|A > 0] - v_{ij}| \leq |E[v_{ij}|A = 0] - v_{ij}|$. That is, advertising serves to give consumers a better idea of their true match value with a product through information. If this is true, any behavioral effect of advertising would improve match value, making consumers at least as well off as if they saw no advertising. In this case, any informative effect of advertising on I_{ij} could only be welfare positive.

A final mechanism through which DTCA could be welfare negative has less to do with advertising in particular and more to do with the nature of health insurance. That is, because the price a consumer pays, p_{ij}, is typically far lower

[2] This is an overly stringent test on total social welfare since the cost of individually inefficient prescriptions could be offset by increases in firm profits. Such offsets might not be viewed as acceptable trade-offs by policy makers.

than the price insurance companies pay on a consumer's behalf, say P_{ij}, the consumer decision problem itself is biased in favor of getting prescribed from the perspective of the other members of the insurance plan. That is, since the end consumer is not bearing the full cost of the prescription, it must be passed through in premiums to other members of the health insurance plan or from the profits of the insurance company itself. With private insurance markets, this behavior would distort the insurance market leading to associated costs of increased premiums on coverage, for example. In this case, there may be some inefficient prescriptions with or without DTCA, and DTCA will amplify the issue.[3]

In practice, each of these mechanisms could be true to varying degrees. Since this framework allows for ex post inefficient purchases from an individual consumer perspective, a standard revealed preference measurement of consumer welfare will not be appropriate. We need additional information beyond purchases to indicate whether the purchases were worthwhile. To that end, this paper will directly measure consumer costs and benefits of DTCA to identify whether purchases are on average worthwhile.

4 Data

4.1 Advertising Data

Advertising data from AC Nielsen's Media database from 2007-2010 is used in this study and are provided by the Kilts Center for Marketing at the University of Chicago Booth School of Business. The database tracks television advertising at the spot-time-DMA level for every product which advertises on television. A DMA, or designated market area, is a collection of counties, defined by the Nielsen company, that all see the same local television stations and affiliates. The top 130 out of 210 DMAs are indicated as "full discovery market" by AC Nielsen, meaning all television advertising occurrences are measured using monitoring devices. In many of the smaller DMAs, only advertising occurrences that match ads in the larger markets are included. This study uses each of these full discovery markets which has a monitoring device on every major network affiliate (ABC, NBC, CBS and FOX), which is 120 DMAs.

In the top 25 DMAs, household impressions are measured from set top viewing information that is recorded in households. In DMAs ranked 26-210, advertising impressions are estimated from quarterly diaries filled out by households.[4] The data also include the total estimated expenditure of the firm on the advertisement; the duration of the advertisement; and very coarse age, race, and gender demographic breakdowns of the impressions data. The data include the parent company of the product advertised, a description of the product being advertised, and a very brief description of the content of the advertising copy.

In addition to local advertising, there is also national advertising. National advertising occurrences are aired in all DMAs. For example, if a firm were to buy a national ad for a product on the CBS evening news, that ad would play in the New York DMA, the Chicago DMA and all other DMAs during that episode of the CBS evening news. As the identification strategy in this paper will exploit variation in local advertising, it is important that there be a significant amount of local advertising.

For an average DMA-month, 7% of the advertising is local advertising, but there is considerable variation in that fraction, with some DMA-months having zero local antidepressant advertising occurrences and some DMA-months having as much as 74%. The standard deviation of the percent of advertising that is local is 13.4%, meaning there is considerable variation both in local advertising and in the share of total advertising that is made up by local advertising in any given DMA-month.

Pairing these data with market size estimates, the total number of Gross Rating Points (GRPs) that each advertisement constituted is computed. A GRP is the typical unit of sale between a firm and a television network for advertising

[3] Of course, copays are typically considerably larger than the marginal costs of production, so these kinds of moral hazard effects will still push total quantity closer to the socially optimal quantity, but there may be externalities in financing this shift.

[4] While impressions are the main advertising measure of interest, there is some concern that the infrequent and self-reported viewing data may be measured with error, all analysis either has been or easily can repeated using ad occurrences as an alternative measure to see if the results are consistent. Please contact the author if you are interested in such analysis.

space: it is calculated as the total number of advertising impressions divided by the population in the DMA, multiplied by 100. As such, a monthly increase of 100 GRP can be interpreted as the average person viewing the ad one additional time over the course of that month.

This study focuses on advertisements for the antidepressant drug category. There are many antidepressants, most of which are now generic. The few products that advertise in the data are branded. The primary brands advertising between 2007 and 2010 are Abilify, Cymbalta, Effexor XR, Pristiq and Seroquel XR. While Cymbalta, Effexor XR and Pristiq are all solely used for the treatment of depression, Abilify and Seroquel XR are used both in the treatment of depression and in the treatment of psychosis. However, the ad copy description indicates that all Abilify and Seroquel XR advertisements over the course of the data are for the depression indication. The average DMA-month has 227.27 GRPs for antidepressants, but with wide dispersion. The standard deviation of DMA-monthly GRPs is 148.67. A histogram of DMA-monthly GRP is provided in Figure 1.

4.2 Claims Data

Insurance claims data come from Truven Health MarketScan® Commercial Database which come from Truven Health Analytics, Inc., an IBM company. The claims are for individuals with employer sponsored insurance in the United States that work for companies that are willing to provide the data. From these claims, I harvest prescription and demographic information on a monthly basis. To make sure I can accurately measure whether a prescription is a new prescription, I focus on those individuals present in the data from the start in January of 2007 and consider their prescription decisions beginning in February 2007. I define a new prescription as a prescription following a month with no prescriptions. The final data set that is cleaned and matched to advertising data in the full discovery markets contains 1,835,265 individuals with employer sponsored insurance that are on average 45 years old. In an average month, 8.1% of the individuals in the data are prescribed an antidepressant, and on average, 59.1% of those antidepressant prescriptions are for generics. More summary statistics are available in Table 1.

The claims data also contain information on the transacted prices and co-payments of prescriptions. Co-payments are based on individual formularies that come with insurance contracts. These contracts typically last a full year, so there is not variation in co-payments of options for a given individual-month. Similarly, prices reflect the co-payment plus the payment by the insurance company, as reported by the insurance company. If there are rebates that the insurance company receives from the drug manufacturer that are not included in the insurance company's reported prices, the reported prices would over-state the true cost of a prescription. The contracts between insurance companies and manufacturers also typically last at least a year. As such, all variation within an insurance-contract year in transacted co-payments or prices would reflect variation in choices of different drugs which carry with them different prices rather than changes in prices for a given drug.

The average price of an antidepressant prescription filled in the data is $62.48 and the average co-payment faced by the individual enrollee for an antidepressant prescription filled is $11.22.

4.3 Labor Supply Data

Information about worker labor supply is provided from Truven Health MarketScan® Health & Productivity Management (HPM) database. A subset of the employers who provide the claims data also provide human resources records on individual enrollee absences from work. Of the individuals in the claims data, there are 518,284 individuals in the labor supply data between 2007 and 2010.

The average number of missed work days in the data is 2.375 with a standard deviation of 3.155 and a median of 1.25. As missed days can be for any reason, there is a huge amount of month-to-month variation in missed work days, even within individuals. The median number of missed days of 1.25 reflects an average of 3 weeks out of the office per

year, which is a reasonable number for a generally healthy person with two weeks of paid vacation as well as some paid sick leave.

5 Research Design

There are two main empirical challenges with identifying the effects of advertising on prescriptions, prices and labor supply in this setting: endogeneity and statistical power. First, as advertising is a firm choice, it is likely targeted at consumers in a non-random way, in particular towards the potential consumers most likely to be responsive to it. Those most likely to be responsive to advertising might also be more likely to get prescribed anyway, eventually receiving any costs or benefits associated with those prescriptions. They could also be more depressed than a randomly selected television viewer, leading the researcher to find worse outcomes associated with advertising even though the causality runs the reverse direction.

A second challenge is statistical power, as advertising is thought to have generally small effects and outcomes are noisy. In particular, workers miss work for many reasons that have nothing to do with depression, advertising or antidepressants. For example, many workers may miss work in a particular area due to a local outbreak of influenza. It would be difficult in the data to know where and when every flu outbreak happens. Even if it happens in a way independent of advertising, it will add considerable noise to any estimates of the advertising effect on labor supply. Similar arguments can be made for vacation days, local labor market conditions or weather conditions, for example. An additional complication that exacerbates power issues in this setting is that treatment of depression takes time to produce individual outcomes. Antidepressants take on average six weeks for improvements to materialize, and there is considerable variability around that amount of time (Frazer and Benmansour (2002)).

To address these challenges, this study exploits random variation in local advertising generated by the borders of DMAs. This design was first used in Shapiro (2018) to study the effects of television advertising on antidepressant demand, but is also used in Tuchman (2016) to study e-cigarette advertising, as well as in Spenkuch and Toniatti (2016) to study political advertising. Consumers who live on different sides of DMA borders face different levels of advertising, due to market factors elsewhere in their DMAs. However, these individuals are otherwise similar, making the cross-border comparison a clean way to identify the effect of the differential advertising. In this way, at the borders, observed advertising is 'out of equilibrium' from what firms would set advertising if they could micro-target very local areas and simulates an experiment.

Capturing this intuition, I estimate the casual effect of advertising on antidepressant prescriptions, prices and labor supply controlling for unobservable geographic characteristics with border-specific brand-time fixed effects. This allows unobservables to be spatially correlated in ways that are consistent with the evolution of the antidepressant market. To control for individual-specific factors that affect antidepressant demand and/or labor supply, individual fixed effects are included. As a number of individual-specific factors and geographic time-specific factors having little to do with depression affect labor supply, these fixed effects will also help to decrease noise in the dependent variables of interest.

The top 120 DMAs contain 209 such borders, 163 of which where the border areas make up no more than 35% of the total DMA population over the course of this time period. Attention will be restricted to these borders that make up a smaller fraction of the whole DMA, as in Shapiro (2017). Each of these brand-border pairs will be considered a separate experiment, with the magnitude of the treatment determined by the advertising in each DMA at a given time, measured in GRPs. Only the individuals residing in counties bordering each other will serve as controls for each other to partial out any local effects that are correlated with outcomes, including any national advertising. The level of an observation is an individual-month.

For an illustrative example, Figure 2 shows the Cleveland and Columbus DMAs in the state of Ohio. The border experiment considered is outlined in bold. People on the Cleveland side of the border might see more antidepressant ads

due to factors affecting the city of Cleveland. Such factors could include things like changes in the rate of depression, home values, employment or crime rates in the city of Cleveland, all of which could be relevant to antidepressant demand. To the extent that those factors also affect the people away from the city of Cleveland at the border of the DMA, they are assumed to also affect the people immediately on the opposite side of the border in the Columbus DMA. I compare how outcomes on the Cleveland side of the border change when when the Cleveland DMA receives a change in advertising GRPs relative to the Columbus DMA.

5.1 Econometric Model

To model the main effects of advertising on demand, let i index individuals, b index borders, and t index time in months. Let Y_{ibdt} the outcome of interest for individual i , in border area b , in DMA d, in month t. Let GRP_{dt} indicate advertising, measured in gross rating points, in DMA d in month t. The effect of an increase in advertising GRP on outcome Y is estimated with regressions of the form

$$Y_{ibdt} = \beta_1 f_1(GRP_{dt}) + \beta_2 f_2(\sum_{\tau=t-t_0}^{t} GRP_\tau) + \alpha_i + \alpha_{bt} + \varepsilon_{ibdt}, \quad (1)$$

where β_1 and β_2 capture the causal effects of current and past advertising, respectively, α_i is an individual fixed effect, α_{bt} is a border area-month fixed effect and ε_{ibdt} is an econometric error term. I consider the outcomes $Y \in \{NewRx, RenewalRx, FirstRenewalRx, Price, Copay, GenericRate, DaysAbsent\}$. All prescription measures are in terms of category-wide rather than brand-specific. I use two-way clustering to account for two forms of correlation between error terms when computing standard errors. First, conditional on the fixed effects, residual variation in advertising is perfectly correlated at the border-DMA-month. Second, from an sampling design standpoint, there are repeated measurements over time in Y at the individual level. As such, I two-way cluster by border-DMA-month and by individual (Abadie et al. (2017)).

For the main results, I will set $f_1(x) = f_2(x) = log(1+x)$. Additionally, I will consider past advertising to be the sum of GRP for the previous six months. In the case of labor market outcomes, this is to account for the fact that it takes time for outcomes to materialize from depression treatment and that there is variance around exactly how much time. In the case of new prescriptions, allowing lagged advertising to have an effect accounts for advertising carry-over.[5] That is, a consumer might watch an ad, but not see the physician for more than a month and at that time remembers last month's ad. In the case of first renewal prescriptions, I interpret the effect of lagged advertising as indicating whether or not those who got a new prescription because of advertising in the previous month are more or less likely to churn. That is, it will signify if advertising induces adverse or advantageous selection in terms of propensity to adhere beyond the first month. In the case of prices and co-payments conditional on a prescription this month, I will set $f_2 = 0$, as there is no clear theoretical link between past advertising and current transaction prices conditional on treatment.

For this approach to be useful in identifying advertising effects, two conditions must hold. First, there must be sufficient variation in advertising across the borders in the data. If all advertising variation were at the national level over time and local stations rarely used their discretion to displace national ads, the border-specific time fixed effects would sweep away all variation in advertising and standard errors would tend to infinity. Second, an individual's location with respect to border side must be quasi-random with respect to changes in preferences for antidepressants and labor supply.

[5] However, this should be treated with caution. One can only get a "new" prescription if one did not receive a prescription in the previous month. Getting a large amount of advertising the previous month and nonetheless not getting a prescription suggests a negative epsilon relative to average, which would bias β_2 negatively. β_2 can alternatively be viewed as a way to control for negative selection into the "new prescription possible" sample for this estimation.

5.2 Features and Limitations

A more detailed analysis accounting of the features and limitations of the approach is available in both Shapiro (2018) and Shapiro (2017). As the identification strategy in particular is not the main contribution of this study, I will focus here only on the most important aspects to validity and interpretation.

Perhaps the largest feature of this approach is that the observed advertising levels at the border are quite different than they would be if firms micro-targeted advertising at individuals or counties rather than DMAs. That is, the variation at the border is driven by the equilibrium supply and demand in other markets. At the border of the Cleveland, OH DMA, viewers see antidepressant ads that were driven by a desire to reach viewers in metro Cleveland, despite the fact that at the border, these viewers can be quite different. If ads were micro-targeted to the county level, these consumers would likely see different ads. Similarly, on the Columbus, OH side of the DMA border, the advertising is largely driven by metro Columbus viewers, which is away from the border, again giving rise to rather different advertising at the Columbus border than if ads could be micro-targeted. If metro Columbus and metro Cleveland are sufficiently different from each other, these very similar consumers right on the border will get very different ads, even though their equilibrium micro-targeted ads would have been very similar. This gives a reasonable amount of variation away from what would be the equilibrium in the micro-targeted world while using the fact that these consumers across the border from one another are very similar to control for unobservable factors driving demand, prices and labor supply.

The border approach has local average treatment effect limitations that are also common to experiments and instrumental variables. In this case, the estimated effect will be local to those consumers who live in border areas. That is, the 'compliers' will be the set of people that live within the border sample, which is a group that can be characterized and compared with the population at large in a straightforward way. An additional potential limitation to this approach is that it relies crucially on variation in local advertising, which is often a remnant of the upfront market and might be systematically different from national network or cable advertising. In this market there is a considerable amount of national advertising, meaning that much of the variation in advertising identifying the effects of interest is away from the zero advertising counterfactual. For an average DMA-month, 7% of the advertising is local advertising, but there is considerable variation in that, with some DMA-months having no local advertising and some DMA-months having as much as 74%. The standard deviation of the percent of advertising that is local is 13.4%, meaning there is considerable variation both in local advertising and in the share of total advertising that is made up by local advertising in any given DMA-month. To get an idea of the amount of variation in GRPs that is helpful for identification using this approach, Figure 3 shows a histogram of GRPs, net of the fixed effects in the border approach, centered at the average GRP level of 227.27 and winsorized at the 0.1st and 99.9th percentiles. The standard deviation of residual GRPs is 34.

In terms of assessing validity, one approach is to show that changes in advertising do not predict changes in observable (non-outcome) control variables at the border in a placebo test. Most demographic or industry of work related variables do not change within individual over time and are thus taken into account with the individual fixed effects. In Table 2, I conduct a placebo analysis as suggested above on variables that do vary at the individual level over time. In this table, the reported variable is used as the dependent variable in the specification from Equation (1). For variables, I use age, whether the worker is paid hourly rather than salary, whether the individual is prescribed in the preceding six months, whether the individual terminates antidepressant treatment conditional on being prescribed in the past six months and whether the individual missed more than the population median number of work days in the preceding six months. Current advertising does not predict any of these variables at the $p<0.05$ level. The only variable that is predicted at the $p<0.1$ level is the fraction of workers who are paid hourly. On one hand, that estimate could be a false positive. On the other hand, if the point estimate on hourly were to be taken seriously, it would imply that a 10% increase in advertising would lead to a 0.007 percentage point (or about 0.02%) decrease in the fraction of workers who are hourly workers, which is economically insignificant.

An additional similar placebo analysis is conducted on county level demographic variables in Appendix D. Advertising changes at the border do not predict county level demographic changes. Additionally, advertising levels at the border do not predict county level demographic levels, lending credibility to the quasi-randomness of advertising at the

borders of DMAs. A final placebo analysis is presented in Appendix E with a placebo treatment, statin advertising, that is not expected to produce the outcome of interest.

6 Results

6.1 The Effect of Advertising on Prescriptions

6.1.1 New Prescriptions

Previous research has shown DTCA to be category expansive in the antidepressant category. Here, I test whether that holds in this data. Table 3 shows the results of estimating equation (1) with new antidepressant prescription as the dependent variable. Column (1) provides estimates from a regression with no controls and no fixed effects, using all of the data both at and away from the borders of DMAs. Column (2) adds individual fixed effects, column (3) additionally adds in month fixed effects and column (4) provides the preferred, border-specification from column (1). Columns (1) and (2) show a small, but statistically significant current advertising effect on new prescriptions and a larger effect of past advertising on new prescriptions. The addition of month fixed effects makes those results disappear. In the preferred specification in column (4), we see that a 10% increase in current antidepressant GRPs leads to about a 0.00031 increase in the probability of a new prescription and past advertising has no effect. As the share of the sample getting a new prescription in a given month is about 0.0099, this amounts to about a 0.031 elasticity and is in line with the category expansive effect of DTCA found in Shapiro (2018). The effect of past advertising on new prescriptions, at first blush, suggests that there is no carry-over effect of advertising. However, that should be interpreted with caution. In order for it to be possible to get a new prescription, one must not have been prescribed in the previous month. Receiving a large treatment of advertising in the past month but not getting prescribed suggests that this treatment, past advertising plus no past prescription, could be negatively selected.

To obtain dollar social and private costs of the new prescriptions generated by advertising, I assume the average number of prescriptions resulting from a new prescription is 6, which is a full course of treatment and roughly the average number of prescriptions resulting from a new prescription in the data. I evaluate the cost of the incremental quantity at the average price and co-payments of antidepressants in the data, $62.48 and $11.22. Assuming these results apply to all adults in the United States, 230 million, a 10% increase in advertising leads to 520,000 new antidepressant prescriptions to about 86,500 individuals, which yields approximately $32.4 million in total costs of new prescriptions per year, about $5.8 million of which is paid in co-pays by the consumer, assuming no changes in prices or co-payments.[6] The 95% confidence interval associated with this cost estimate is [$49,871, $64.4 million].

6.1.2 Renewal Prescriptions

To account for the total cost of advertising marginal prescriptions, it is important also to measure the effect of DTCA on renewal prescriptions. Previous studies of DTCA that focus on adherence [Donohue et al. (2004); Cardon and Showalter (2015); Wosinska (2005)] tend to find small, inconsistent and sometimes negative effects of DTCA on adherence. While the particulars of what causes adherence and non-adherence are very important to welfare, I view the measurement here as contributing only to the cost of advertising marginal prescriptions. To the extent that increased or decreased adherence is important in magnitude and either appropriate or inappropriate, effects of that advertising marginal adherence will be reflected in the measurement of advertising marginal labor supply.[7]

[6] To reach these numbers note that a 10% increase in DTCA corresponds with one tenth the estimated coefficient on log(GRP). The assumed 6 months of treatment following an initial prescription is the average observed in the data. So we take one tenth the coefficient on log(GRP) multiply by 6 months of prescriptions, multiply by the price of a prescription and multiply by twelve months per year.

[7] In particular, if non-adherence is rational and in response to lack of benefit and high adverse effects, decreased adherence would be welfare positive. If non-adherence were due to incorrect expectations about the relative risks and benefits, it would be welfare negative. If non-adherence were due to inattention, it would be welfare negative. However, it would be difficult to argue that advertising marginal non-adherence is due to

Table 4 presents the results of the effects of advertising on renewal prescriptions. Columns (1)-(4) correspond with Table 3. For conciseness, I will only discuss the preferred specifications in column (4). The point estimate on current advertising is negative and small, indicating that it decreases adherence to treatment with an elasticity of about -0.005. The estimate is statistically significant and small in magnitude, consistent with Cardon and Showalter (2015), Donohue et al. (2004) and Wosinska (2005).

In terms of costs and benefits of adherence, the results indicate that DTCA reduces adherence. This implies a reduced cost of advertising associated with averted prescriptions. The dollar value of one advertising marginal prescription averted is about $62.48. Since 8% of adults are taking antidepressants at any given time, the estimate would apply to about 18.4 million individuals, making the total yearly savings from a 10% increase in DTCA using this measure about $6 million, or about 97,152 prescriptions. If some of these marginal prescriptions are averted inappropriately due to over-statement or risk, the cost of that will be reflected in the labor supply analysis below. If some of these marginal prescriptions were averted appropriately due to adverse effects, that will also be reflected in the labor supply analysis if the adverse effects were sufficiently severe.

6.2 Prices

An additional potential cost of DTCA is indirect: it could not only increase total quantity holding average transaction prices fixed, it could also increase the price of the chosen product. This would be the case if advertising caused patients to switch from inexpensive generics to expensive brands or from less expensive brands to more expensive brands. In the literature, Dave and Saffer (2012) have shown a correlation between DTCA and higher prices, though did not leverage any quasi-exogenous variation. I point out here that I am assessing effects on transacted prices rather than list prices. Month-to-month co-payments and prices tend not to vary due to the structure of insurance company-manufacturer bargaining. The effects here reflect whether advertising causes individuals to choose drugs that are already more or less expensive rather than if changes in DTCA cause the profit maximizing menu of prices from the manufacturer to change.[8] However, these two effects are not unrelated. If advertising does not affect the price of the drug chosen, holding prices fixed, then manufacturers would have little incentive to raise retail prices.

I also note that the effect identified here will be the composition of two potential effects. First, advertising could draw more or less price sensitive people into the market. Second, advertising could directly affect the price sensitivity of those currently on antidepressants.[9]

Table 5 shows these effects for co-payments while Table 6 shows these effects for prices. The effect of DTCA on co-payments affect the private cost of DTCA to those who initiate treatment while the effect of DTCA on transacted prices would affect the social cost of DTCA through its externality on the rest of the insurance pool through insurance premiums. For both tables, columns (1) through (4) correspond to the specifications in the previous tables, though this sample is restricted to only those individuals who fill a prescription for an antidepressant in a given month. Rather than individual fixed-effects, these regressions use border-side fixed effects as only using within-individual variation in prices would limit the variation considerably. This fixed effects structure retains the source of variation in advertising as the border-month fixed effects are still employed. Again, I will only discuss the preferred specification in column (4).

First, focusing on co-payments, in Table 5, column (4), the effect of advertising on co-payments is not statistically significant, and the point estimate is positive and small. It indicates that a 10% increase in advertising leads to a $0.015

inattention, since a causal effect of advertising requires attention in this setting. All of these specific mechanisms surrounding adherence are very important but beyond the scope of this study.

[8] Empirically, the menu of prices for drugs are nationally consistent within an insurer-plan type despite considerable differences in DTCA, suggesting the absence of such an effect.

[9] While I do not explicitly measure business stealing effects in this study, it is useful to think about how business stealing affects welfare. First, it reduces category profit. Second, it reallocates quantity between products. As for the first, it would represent a transfer from the drug industry to the advertising industry. As for the second, any such reallocation would only affect consumer welfare if the drug switched to is either a different price, which is measured here, or a better or worse match for the patient, which would show up in the labor supply estimates.

increase in the average transacted co-payment on an average co-payment of $11.36, with a 95% confidence interval of [-$0.017, $0.048]. Taken at face value, this point estimate would imply a small indirect private cost to patients associated with DTCA.

Next, focusing on transacted prices in Table 6, column (4), the effect of advertising on transacted prices is not statistically significant, and the point estimate is negative and small. The point estimate indicates that a 10% increase in advertising leads to a $0.062 *decrease* in transacted prices on an average price of $61.96, with a 95% confidence interval of [-$0.138, $0.0144]. It implies a small indirect social benefit of DTCA in terms of lowering costs to the insurance company.

As prices are measured from insurance claims and might not include information about rebates that are not given on the margin, we might be concerned about measurement error clouding inference on steering. This would be especially concerning if rebates were somehow systematically related to advertising. As a separate measure of steering that does not rely on accurately measuring prices, I use the generic penetration rate. If advertising is steering consumers away from generic drugs to expensive brands, an increase in advertising will drive a lower generic penetration rate. Table 7 provides these estimates. Focusing on the preferred specification in column (4), advertising has a small, statistically insignificant effect on the generic penetration rate. A 10% increase in advertising leads to a 0.03 percentage point (0.05%) decrease in the generic penetration rate with a 95% confidence interval of [-0.08,0.014]. As such, the point estimate implies that a 10% increase in advertising would move the generic penetration rate from 60.13% to 60.10% and at the most pessimistic end of the confidence interval, to 60.05%.

Taken together, these results provide evidence against any economically significant steering effects of antidepressant advertising.

6.3 Labor Supply

To assess the potential benefits of DTCA, I estimate equation (1) using missed days of work as the dependent variable.[10] In this case, any potential effects of DTCA on labor supply would be expected to manifest with a lag, as antidepressants do not work instantaneously. As such, the coefficient of interest is the one attached to lagged DTCA. The coefficient on concurrent advertising is viewed as a placebo. Any significant effect of concurrent advertising on labor supply would presumably indicate targeting on an omitted variable by the firm, such as the level of depression in a given market.

In terms of concerns over endogeneity in a naive correlational analysis, firms might well direct their advertising most at places where individuals have a high incidence of depression. If that is the case, we would find a spurious positive effect of current DTCA on current missed days of work. As many of these individuals might get treated with or without the advertising, we would find a spurious negative effect of past DTCA on current missed days of work if treatment were effective.

Table 8 presents the labor supply results. Column (1) provides the naive regression with no controls and no fixed effects. It indicates that work days missed are significantly increased by current DTCA and significantly decreased by past DTCA. These results are directionally consistent with the expected spurious result. In column (2), individual fixed effects are included. Both effects persist, but the effect of past DTCA is muted to some degree. In column (3), month fixed effects are added, which control for seasonal factors that are correlated with labor supply. For example, many families go on vacation during December or July. With the inclusion of month fixed effects, the estimate on current DTCA changes sign and becomes insignificant. The estimate on past DTCA persists, but is no longer statistically significant. These two forces highlight the two main empirical concerns. First, it seems firms are targeting ads to places where they expect individuals are missing a lot of work, and second, there is a lot of variation in month over month labor supply that has little to do with antidepressant DTCA and that generates significant noise. Column (4)

[10] Using log(absent hours) instead of level of missed days does not change the results significantly quantitatively or qualitatively. Those results will eventually be provided in an appendix.

addresses both of these concerns by implementing the border strategy from equation (1). Despite losing 80% of the observations by focusing only on the borders, standard errors decrease considerably. The border-month fixed effects soak up significant variation in labor supply that has nothing to do with antidepressant DTCA. Current DTCA has no significant effect on days missed, but the point estimate is small and negative. Past DTCA has a significant and negative effect on days missed, suggesting that not only does advertising increase new prescriptions, it eventually leads to individuals missing less work. The point estimate is consistent with an elasticity of labor supply with respect to advertising of about 0.05. Column (5) shows that this effect appears to be coming from those who have missed a lot of work in the previous six months.

I note here that I cannot statistically distinguish the estimates in column (3) from the estimates in column (4). This is partially due to the large amount of noise in the estimates in column (3), but it is also consistent with the panel structure of the data and individual and month fixed effects being sufficient to remove contamination, while the border strategy adds additional statistical power. I will proceed using the estimates in column (4) as the preferred estimates, as they are the both the most conservative with respect to controlling for confounds and exhibit the best statistical power.

The point estimate of column (4) of -0.1382 indicates the average individual in the sample gains about 0.11 hours of monthly time at work from a 10% increase in the past 6 months of DTCA. Assuming this number applies to all working adults (about 145 million) and assuming the national average wage of \$24/hour[11], a sustained 10% increase in DTCA for a year would lead to \$769.5 million in increased wage benefits, or about \$5.31 per working adult.

The 95% confidence interval on the total yearly benefits of a 10% increase in DTCA is [\$99.3 million, \$1.44 billion]. The 95% confidence interval on the yearly costs of marginal prescriptions from a 10% increase in DTCA was [\$49,871, \$64.4 million], indicating that the cost and benefit confidence intervals do not overlap. Assuming the highest cost and lowest benefit still puts the benefits at nearly double the cost. Figure 4 shows the costs to payers coming from new quantity, price and adherence next to the labor supply benefit. Equality of the labor supply benefit and costs to payers is rejected at any conventional level. Additionally, the incremental cost of DTCA to payers is entirely coming through quantity rather than prices.

6.3.1 Mechanisms

If we assume this benefit only applies to those who are advertising marginal to new prescriptions (about 86,500 individuals, as noted above), those individuals gain about 3.86 work days per month and \$8,900 per year, or about 18% of their annual incomes, on average, which is just under half of the effect of depression on annual earnings reported by Woo et al. (2011). With average co-payments of \$11.22 and 6 months of treatment, the private cost to obtain the \$8,744 in benefits is about \$67. With the average transaction price of \$62.48, the total cost is \$375.[12]

Of course it is possible that the effect of DTCA on labor supply acts not only through new prescriptions, but also through other mechanisms. For example, decreased adherence among those poorly suited to treatment could generate labor supply due to the elimination of adverse effects. Individuals could seek non-drug treatment of their depression after seeing ads. Some could simply be more cognizant of depressive tendencies and deal with them through other means, such as exercise, leading to increased labor supply. Individuals viewing ads could recognize the symptoms in their family members and give them additional attention, which could have benefits.[13]

Since many mechanisms are possible, it may make more sense to view the 'per person affected' benefit of DTCA more broadly. Taking as a benchmark the survey evidence in Kessler et al. (2003) that roughly half of people afflicted

[11] I note here that using \$24 as the basis for the social hourly benefit of work assumes that workers are paid their marginal product of labor. If employers have market power in the labor market and pay workers less than their marginal products, this will leave out benefits to employers.

[12] To reach these numbers note that a 10% increase in DTCA corresponds with one tenth the estimated coefficient on log(Past GRP). Since past GRP is defined as 6 months past, the estimate must be divided by 6 to avoid double counting. Then, it is multiplied by 8 hours per day, by \$24 per hour, by 12 months in a year and by 145 million working adults to arrive at the final estimate. To obtain the estimate on only those marginal to prescriptions, I take the \$769.5 million and divide it by the 86,500 individuals estimated to be advertising marginal to prescriptions in a year.

[13] Separation of these mechanisms is difficult due to statistical power limitations, but continuing work on this study is attempting to better separate the mechanisms.

with depression are untreated, then this estimate would apply to roughly 8% of the working population. If the benefit were evenly spread out among these people, it implies a $66.34 per year wage benefit per person, or about 2.75 hours per year.[14] The reality is likely in between this number and the one assuming that all of the effect is coming through advertising marginal prescriptions. While the alternative mechanisms likely drive some of the labor supply effect, it is unlikely that every single untreated depressed person is subject to one of these advertising marginal mechanisms. Broadly speaking, this range of potential per person benefits reflect economically plausible effect sizes, the most extreme of which is still considerably less than the estimated per person costs of untreated depression from Woo et al. (2011).

While it is difficult to separate the mechanisms, doing so would be very useful. If the effects are primarily coming from non-drug pathways, then public service announcements about the disease might be sufficient to generate the positive effects. If the effects are primarily coming from drug pathways, the actual advertising of the drug itself might be important.

6.4 Other costs and benefits

6.4.1 Adverse Effects

One potential cost associated with advertising marginal prescriptions that is difficult to put a dollar value on is adverse effects. In particular, David et al. (2010) and Cardon and Showalter (2015) point out that advertising marginal prescriptions could be worse matches than average prescriptions and result in a greater incidence of adverse effects.[15]

Here, I evaluate the effect of advertising on appropriateness of treatment using six different measures. First, I measure the effect of advertising on the propensity to complete a full course of treatment (six months) conditional on a new prescription. If advertising drives worse matches, then markets with higher advertising should see higher rates of churn in the first six months, prior to completing a full course. Second, the claims data include information on adverse effects reported to physicians. I evaluate these in both a forward-looking and backward looking way, for antidepressant related adverse effects and for adverse effects related to any drug.

The first adverse effect related measure is forward looking adverse effects for any drug (FLAE - Any). In this case, the dependent variable is an indicator for whether or not an adverse effect is reported in the six months following a new prescription. It is important to consider other drug adverse effects rather than just antidepressant related adverse effects, as some adverse effects could come from drug interactions. That is, a bad interaction between an antidepressant and a statin could be caused by antidepressant advertising. The second measure I evaluate is forward looking adverse effects specifically for antidepressants (FLAE - Antidep). These two measures are only comparing individuals who began treatment in the month in question and looking forward to see if those in higher advertising markets were more likely to experience adverse effects.

I also consider backward looking adverse effects (BLAE - Any) and (BLAE - Antidep). In this case, the dependent variable is whether or not an adverse effect was observed in the current period. All observations are included and both concurrent and lagged advertising are of interest. Concurrent advertising could affect concurrent reporting of adverse effects only through those who were already prescribed in previous periods. Past advertising is related to the total number of advertising marginal adverse effects. However, it is a composition of two potential effects. First, past advertising could drive lower current adverse effects by decreasing adherence among those who were experiencing adverse effects before. Second, past advertising could drive higher adverse effects through its effect on new prescriptions and some fraction of new prescriptions will inevitably come with adverse effects. As such, the net effect can be seen as the net effect of advertising on adverse effects through those two channels.

[14] To obtain the estimate assuming advertising impacts all those who are depressed and untreated, I take the $769.5 million and divide it by the (0.08 untreated depressed share)*(145 million)

[15] I note here that any particularly severe adverse effects would likely mute the effect of advertising on labor supply and thus already be accounted for in the analysis above. As such, the effects I measure here are almost by definition second order.

Table 9 presents the results. Column (1) shows that DTCA predicts a reduced chance of an incomplete course of treatment, but is statistically insignificant. This suggests that the advertising marginal are about equally likely to complete a course of treatment as the average new patient, and if anything they are less likely to churn. Columns (2) and (3) show no effect of advertising on the likelihood of observing reported adverse effects (in general or antidepressant related) in the six months following a new prescription. Columns (4) and (5) show no significant effect of concurrent advertising on adverse effect reporting. Column (4) shows no significant effect of past advertising on concurrent general adverse effect reporting. Column (5) shows a negative effect of past advertising on antidepressant adverse effect reporting. This suggests that any effect of advertising on new adverse effects is more than offset by any effect of advertising on adherence for those likely to have adverse effects.

Taken together, these results indicate that the advertising marginal are not substantially different from the average in terms of propensity to experience adverse effects, and as a result, advertising marginal adverse effects will not close the gap between the costs and benefits of advertising marginal prescriptions.

6.4.2 Unmeasured Costs and Benefits

It should be noted that some possible costs and benefits of DTCA are either not measured in this study or do not come with easily calculable dollar values.

In terms of unmeasured benefits, some individuals may have no change in labor supply, but are more productive while they are at work due to being treated. Additionally, some people might simply feel better, and that could have considerable value, but by how much in dollars is unclear. Furthermore, wage dollars represent a strict lower bound on the cost of absenteeism. If employers have market power in the labor market, they are likely to pay their workers less than their marginal product of labor. As such, any measurement using wage dollars alone will understate the social benefit of reduced absenteeism. Finally, measuring cost and benefit at the level of the patient ignores any incremental profits of pharmaceutical companies.

In terms of unmeasured costs, some people may directly dislike watching television ads for antidepressants (see, for example, Wilbur et al. (2013)). If the alternative to antidepressant DTCA were additional television programming, the effect might be significant. However, if we were to remove antidepressant DTCA, it is much more likely the viewer would get a different advertisement, which would presumably imply a smaller welfare effect.

Assuming no unmeasured benefits of DTCA, those costs would have to add up to at least \$737 million per 10% increase in antidepressant DTCA in order to flip this cost-benefit analysis.

How big is a 10% increase in DTCA? As mentioned above, in 2012, the total antidepressant DTCA expenditure was \$300 million, so a 10% sustained increase would be an increase in DTCA spending of \$30 million per year.[16] In the context of private firms spending money on advertising, this \$30 million in expenditure is a transfer from the advertiser to the television network from a welfare perspective. However, if a government or charitable organization wanted to begin a campaign of advertising for depression treatment and thought it could get similar results as found here, \$30 million is quite small in comparison with the \$769.5 million in measured benefit for that money.

6.5 Cautions & Limitations

The reader should take some caution in interpreting these results for policy. First and foremost, they only apply to DTCA as it relates to antidepressant treatment. DTCA for another drug category might have a different interaction with patient selection on potential to gain from treatment. Depression is thought by many physicians to be under-treated due to the stigmatization, which leaves plenty of opportunity for welfare increasing market expansion. For

[16] It is interesting to note that the estimated DTCA-marginal revenue was about \$32.6 million, making a 10% increase in DTCA about 8.7% ROI , not accounting for the decreased adherence. These numbers suggest the estimated effects on prescriptions are plausible in terms of the firm's first order condition.

cholesterol lowering drugs or erectile dysfunction drugs, that might be quite different. However, these estimates do provide evidence that a blanket ban on all DTCA might be a bad idea. Put differently, if a policy maker wanted to do institute a blanket ban on DTCA, he or she should consider how to recover the social losses that would cause in the antidepressant space. Additionally, that advertising draws from a desirable margin in this space might make us more optimistic that a similar mechanism of advertising effectiveness is at play for other products.

Second, the border identification strategy identifies an effect local to the borders of television markets. It is possible that the true effect of advertising away from the borders of TV markets is different from the effect at the border. One reason to potentially worry less about this is that the specification using all of the data, including away from the border with month and individual fixed effects is consistent with the result at the border, simply with more noise. Additionally, there are many borders in this data with many types of individuals. For the non-border counties to flip the cost-benefit in this case would require drastically different advertising effects, both on labor supply and on prescriptions. The effect on prescriptions would need to be much larger and the effect on labor supply much smaller.

Third, all calculations on costs and benefits are computed at the point estimates, but each of these are functions of estimated values with standard errors. Taking a pessimistic view of both the benefits and using the 5% end of the confidence interval for labor supply results and a 95% end of the confidence interval for new prescriptions, the benefits still outweigh the costs, but the gap between them is muted. For a 10% increase in DTCA using these maximally pessimistic estimates, cost of new prescriptions is $64.4 million and the benefit of those prescriptions is $99.3 million. Of course this is a highly pessimistic view, and the joint probability of the most pessimistic estimates on both labor supply and costs of new prescriptions is tiny, and even in that case, the benefits outweigh the costs.

Fourth, it is possible that a substantial amount of the measured benefit of DTCA is operating through channels other than prescriptions. If that is the case, the optimal policy for addressing the remaining untreated depressed population might be through public service announcements rather than through vastly expanded or subsidized branded DTCA.

7 Conclusion

In this paper, I find substantial benefits of antidepressant DTCA on labor supply. The effects are plausible in magnitude at the individual level, with the marginal prescribed individual gaining about 3.78 days of work per month or $8,744 annually from the advertising marginal treatment that comes from a 10% increase in DTCA. The total wage benefits of labor supply from a 10% increase in DTCA, at $769.5 million are substantially larger than the $32.4 million in direct costs of the marginal prescriptions generated. Indirect costs that can be easily assigned a dollar value, increases in co-payments, prices and generic penetration rates, are shown to be small and statistically insignificant, which indicates DTCA does not significantly steer patients to more expensive treatments. Additional savings come from reduced adherence. In terms of costs and benefits without clear dollar values, I find evidence that advertising marginal prescriptions do not produce a higher rate of non-compliance or adverse effects.

These results highlight the importance of understanding which types of consumers are affected by advertising and measuring how much these consumers benefit from marginal treatment when assessing the desirability of DTCA. In the case of antidepressants, the marginal consumers stand to gain and do gain from treatment in a way that far exceeds the social cost. While this result might not be the same across different drug categories, it highlights that a more nuanced approach than a blanket ban on DTCA might be desirable.

The relative magnitude of the different mechanisms behind the advertising marginal labor supply is very interesting and worthy of further study. Whether advertising works mainly through new prescriptions or through alternative strategies for addressing depression is relevant for deciding how to address the remaining untreated depressed population.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Aizawa, N. and Kim, Y. S. (2018). Advertising and risk selection in health insurance markets. *American Economic Review, forthcoming*.

Alpert, A., Lakdawalla, D., and Sood, N. (2015). Prescription drug advertising and drug utilization: the role of medicare part d. Technical report, National Bureau of Economic Research.

Berndt, E. R., Koran, L. M., Finkelstein, S. N., Gelenberg, A. J., Kornstein, S. G., Miller, I. M., Thase, M. E., Trapp, G. A., and Keller, M. B. (2000). Lost human capital from early-onset chronic depression. *American Journal of Psychiatry*, 157(6):940–947.

Bharadwaj, P., Pai, M. M., and Suziedelyte, A. (2015). Mental health stigma. *NBER Working Paper 21240*.

Boyer, P., Danion, J., Bisserbe, J., Hotton, J., and Troy, S. (1998). Clinical and economic comparison of sertraline and fluoxetine in the treatment of depression. *Pharmacoeconomics*, 13(1):157–169.

Bütikofer, A. and Skira, M. M. (2016). Missing work is a pain: The effect of cox-2 inhibitors on sickness absence and disability pension receipt. *Journal of Human Resources*, pages 0215–6958R1.

Cardon, J. H. and Showalter, M. H. (2015). The effects of direct-to-consumer advertising of pharmaceuticals on adherence. *Applied Economics*, 47(50):5432–5444.

Chesnes, M. and Jin, G. Z. (2016). Direct-to-consumer advertising and online search.

Currie, J. and Madrian, B. C. (1999). Health, health insurance and the labor market. *Handbook of labor economics*, 3:3309–3416.

Dave, D. and Saffer, H. (2012). Impact of direct-to-consumer advertising on pharmaceutical prices and demand. *Southern Economic Journal*, 79(1):97–126.

David, G., Markowitz, S., and Richards-Shubik, S. (2010). The effects of pharmaceutical marketing and promotion on adverse drug events and regulation. *American Economic Journal: Economic Policy*, 2(4):1–25.

De Quidt, J. and Haushofer, J. (2016). Depression for economists. Technical report, National Bureau of Economic Research working paper 22973.

Deshpande, M. (2016). Does welfare inhibit success? the long-term effects of removing low-income youth from the disability rolls. *The American Economic Review*, 106(11):3300–3330.

Donohue, J. M., Berndt, E. R., Rosenthal, M., Epstein, A. M., and Frank, R. G. (2004). Effects of pharmaceutical promotion on adherence to the treatment guidelines for depression. *Medical care*, 42(12):1176–1185.

Frazer, A. and Benmansour, S. (2002). Delayed pharmacological effects of antidepressants. *Molecular psychiatry*, 7(S1):S23.

Garthwaite, C. L. (2012). The economic benefits of pharmaceutical innovations: the case of cox-2 inhibitors. *American Economic Journal: Applied Economics*, 4(3):116–137.

Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Pike, C. T., and Kessler, R. C. (2015). The economic burden of adults with major depressive disorder in the united states (2005 and 2010). *The Journal of clinical psychiatry*, 76(2):155–162.

Greenberg, P. E., Stiglin, L. E., Finkelstein, S. N., and Berndt, E. R. (1993a). Depression: a neglected major illness. *The Journal of clinical psychiatry*.

Greenberg, P. E., Stiglin, L. E., Finkelstein, S. N., and Berndt, E. R. (1993b). The economic burden of depression in 1990. *The Journal of clinical psychiatry*.

Grundl, S. and Kim, Y. J. (2017). Consumer mistakes and advertising: The case of mortgage refinancing.

Hosken, D. and Wendling, B. (2013). Informing the uninformed: How drug advertising affects check-up visits. *International Journal of Industrial Organization*, 31(2):181–194.

Iizuka, T. and Jin, G. Z. (2005). The effect of prescription drug advertising on doctor visits. *Journal of Economics & Management Strategy*, 14(3):701–727.

Iizuka, T. and Jin, G. Z. (2007). Direct to consumer advertising and prescription choice. *Journal of Industrial Economics*, 55(4):771–771.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., Rush, A. J., Walters, E. E., and Wang, P. S. (2003). The epidemiology of major depressive disorder: results from the national comorbidity survey replication (ncs-r). *Jama*, 289(23):3095–3105.

Kim, T. and KC, D. S. (2017). Can viagra advertising make more babies?

Niederdeppe, J., Avery, R. J., Kellogg, M. D., and Mathios, A. (2017). Mixed messages, mixed outcomes: Exposure to direct-to-consumer advertising for statin drugs is associated with more frequent visits to fast food restaurants and exercise. *Health Communication*, 32(7):845–856. PMID: 27428179.

Papageorge, N. W. (2016). Why medical innovation is valuable: Health, human capital, and the labor market. *Quantitative Economics*, 7(3):671–725.

Shapiro, B. (2017). Advertising in health insurance markets. *Marketing Science, forthcoming*.

Shapiro, B. T. (2018). Positive spillovers and free riding in advertising of prescription pharmaceuticals: The case of antidepressants. *Journal of Political Economy*, 126(1):381–437.

Sinkinson, M. and Starc, A. (2017). Ask your doctor? direct-to-consumer advertising of pharmaceuticals. *Review of Economic Studies, forthcoming*.

Sobocki, P., Ekman, M., Ågren, H., Krakau, I., Runeson, B., Mårtensson, B., and Jönsson, B. (2007). Health-related quality of life measured with eq-5d in patients treated for depression in primary care. *Value in Health*, 10(2):153–160.

Spenkuch, J. L. and Toniatti, D. (2016). Political advertising and election outcomes.

Stewart, W. F., Ricci, J. A., Chee, E., Hahn, S. R., and Morganstein, D. (2003). Cost of lost productive work time among us workers with depression. *Journal of the American Medical Association*, 289(23):3135–3144.

Stoudemire, A., Frank, R., Hedemark, N., Kamlet, M., and Blazer, D. (1986). The economic burden of depression. *General Hospital Psychiatry*, 8(6):387–394.

Tomonaga, Y., Haettenschwiler, J., Hatzinger, M., Holsboer-Trachsler, E., Rufer, M., Hepp, U., and Szucs, T. D. (2013). The economic burden of depression in switzerland. *Pharmacoeconomics*, 31(3):237–250.

Tuchman, A. E. (2016). Advertising and demand for addictive goods: The effects of e-cigarette advertising. Technical report.

Wilbur, K. C., Xu, L., and Kempe, D. (2013). Correcting audience externalities in television advertising. *Marketing Science*, 32(6):892–912.

Woo, J.-M., Kim, W., Hwang, T.-Y., Frick, K. D., Choi, B. H., Seo, Y.-J., Kang, E.-H., Kim, S. J., Ham, B.-J., Lee, J.-S., et al. (2011). Impact of depression on work productivity and its improvement after outpatient treatment with antidepressants. *Value in Health*, 14(4):475–482.

Wosinska, M. (2005). Direct-to-consumer advertising and drug therapy compliance. *Journal of Marketing Research*, 42(3):323–332.

Figures

Figure 1: Antidepressant GRPs

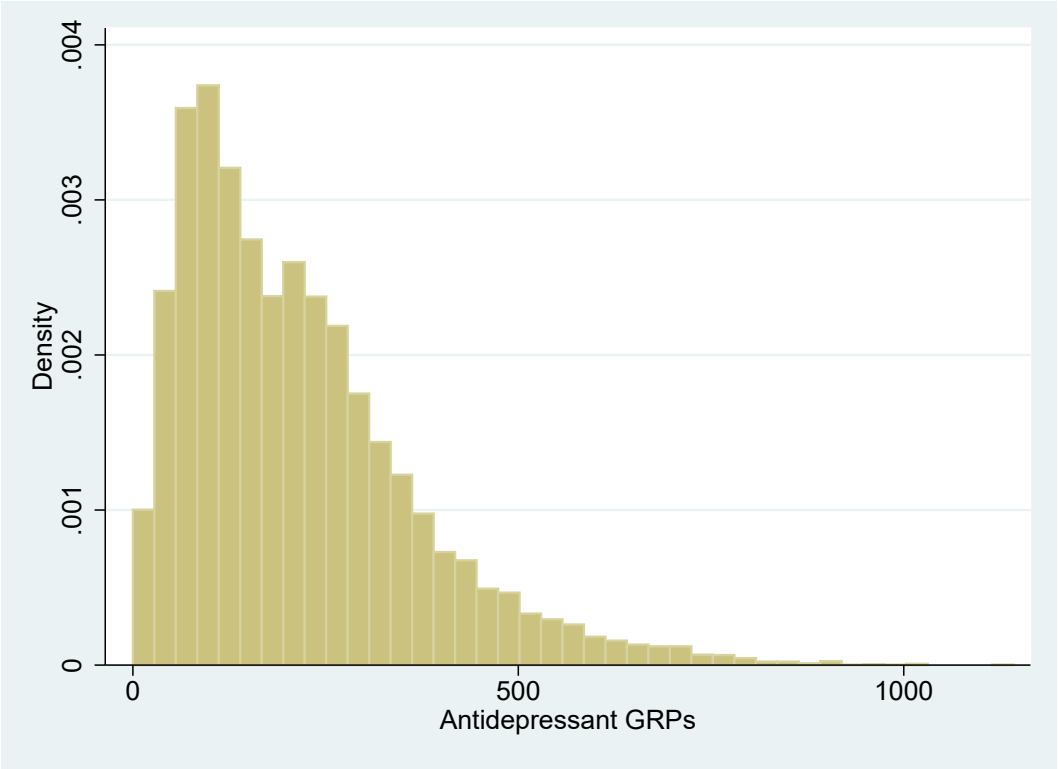


Figure 2: Ohio and DMA Border Example

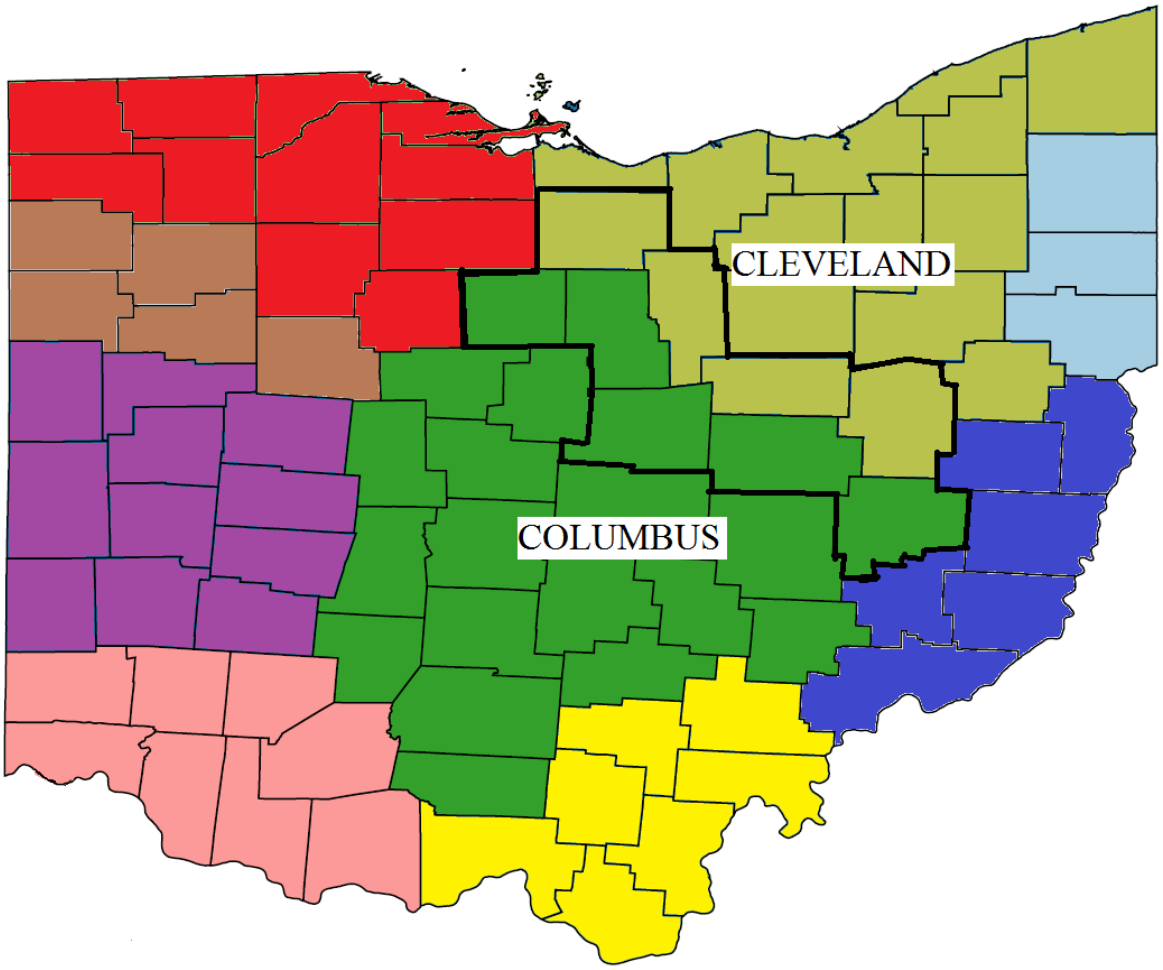


Figure 3: Antidepressant DTCA GRP variation across Borders - SD 34

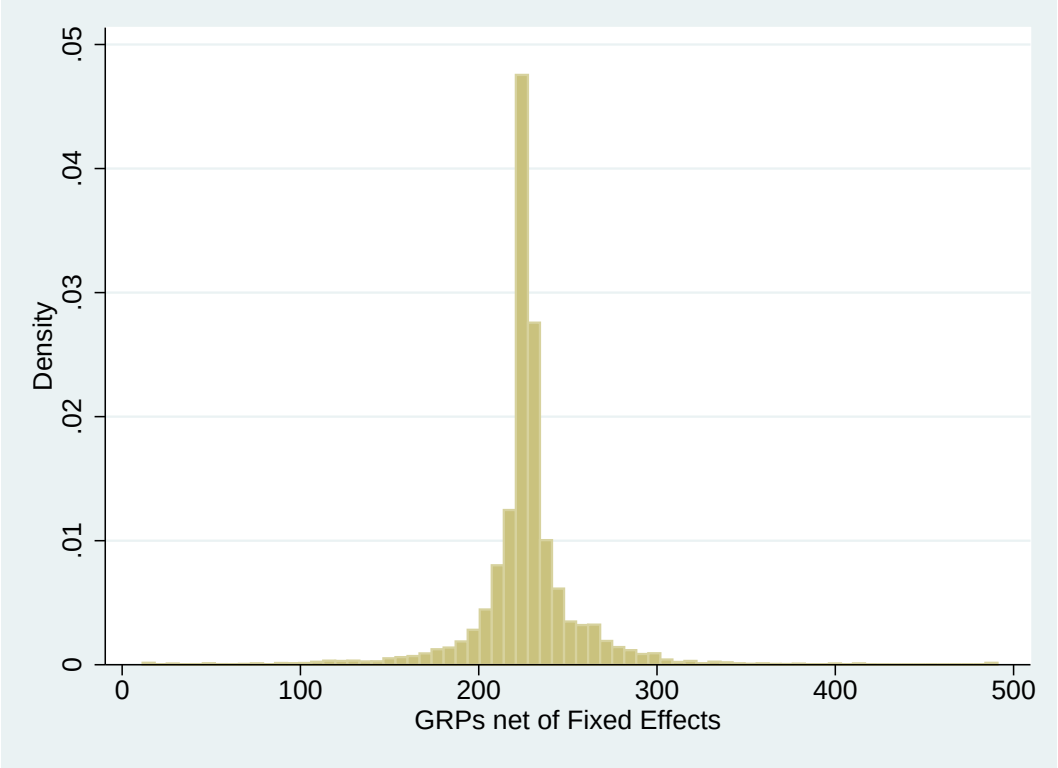
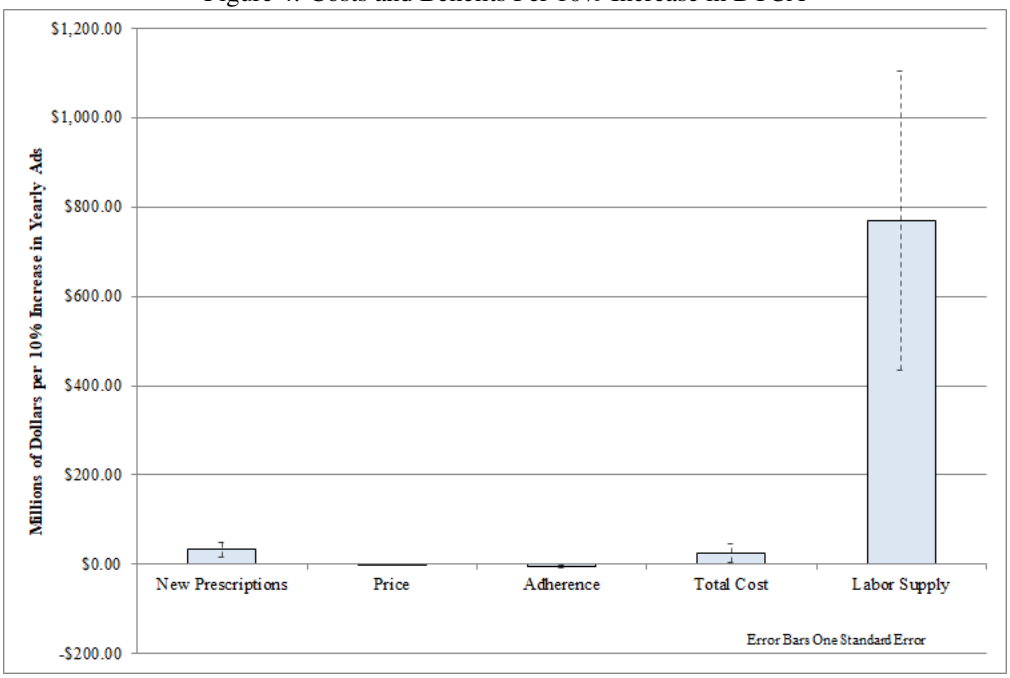


Figure 4: Costs and Benefits Per 10% Increase in DTCA



Tables

Table 1: Summary Stats

	Mean	SD	N
Age	45.39	10.63	68,527,405
GRP	227.27	148.67	5,428
Antidepressant Rx	0.0810	0.2731	68,527,405
New Antidepressant Rx	0.00949	0.0970	68,527,405
Generic Rate	0.5914	0.4916	5,551,058
Copayment	11.22	12.49	5,855,154
Price	62.48	69.58	5,855,154
Days Absent	2.375	3.155	17,593,438

Table 2: Placebo Regressions

	Age	Hourly	Prescribed Past 6	Churn Past 6	High Absentee
$Log(1+GRP)$	-0.0002	-0.00079	0.0005	0.00477	0.00264
	(0.00085)	(0.00045)	(0.00043)	(0.00345)	(0.00429)
Border-DMA FEs	x	x	x	x	x
Border-Month FEs	x	x	x	x	x
Mean DV	45.97566	0.38401	0.11161	0.38258	0.63196
R-squared	0.99923	0.97473	0.74793	0.37623	0.51123
Observations	12,731,985	12,746,153	12,746,153	1,419,929	2,857,781
*** p<0.001, ** p<0.01, * p<0.05					

Each column represents a regression with the given variable as the dependent variable. It shows that changes in these variables are not systematically predicted by changes in antidepressant advertising. Log of past advertising is also included in the regression though not reported, as some of these variables, such as Prescribed Past 6 and Churn Past 6, are potential outcomes of past advertising, though not concurrent advertising.

Table 3: New Prescriptions

	(1)	(2)	(3)	(4)
$Log(1+GRP)$	0.00011***	0.00008**	0.00004	0.00031*
	(0.00003)	(0.00003)	(0.00005)	(0.00015)
$Log(1+GRP_{past})$	0.00076***	0.00050***	-0.0001	-0.00019
	(0.00022)	(0.00005)	(0.00007)	(0.00027)
Individual FEs		x	x	x
Month FEs			x	
Border-Month FEs				x
Mean DV	0.0094	0.0094	0.0094	0.0099
R-squared	0.00003	0.07263	0.07266	0.07446
Observations	66,736,304	66,736,304	66,736,304	12,064,669
*** p<0.001, ** p<0.01, * p<0.05				

Standard errors are two-way clustered. First, they are clustered by (border)x(DMA)x(Month) to account for correlation in the treatment variable. Second, they are clustered by individual to account for the fact that there are repeated observations within individual over the sample.

Table 4: Renewal Prescriptions

	(1)	(2)	(3)	(4)
$Log(1+GRP)$	0.00064**	-0.00358***	0.00007	-0.00441*
	(0.00025)	(0.00052)	(0.00095)	(0.00213)
$Log(1+GRP_{past})$	-0.00052	-0.01105***	-0.00064	-0.00081
	(0.00099)	(0.00089)	(0.00151)	(0.00414)
Individual FEs		x	x	
Month FEs			x	
Border-Month FEs				x
Mean DV	0.88565	0.89202	0.89202	0.89186
R-squared	0.0000	0.1396	0.1406	0.1335
Observations	5,409,591	5,365,812	5,365,812	1,018,908
*** p<0.001, ** p<0.01, * p<0.05				

Standard errors are two-way clustered. First, they are clustered by (border)x(DMA)x(Month) to account for correlation in the treatment variable. Second, they are clustered by individual to account for the fact that there are repeated observations within individual over the sample.

Table 5: Co-payments in $

	(1)	(2)	(3)	(4)
$Log(1+GRP)$	-0.32462***	-0.42855***	-0.03646	0.1499
	(0.07260)	(0.02299)	(0.04888)	(0.16599)
DMA FEs		x	x	
Border-DMA FEs				x
Month FEs			x	
Border-Month FEs				x
Mean DV	$11.36	$11.36	$11.36	$10.84
R-squared	0.0005	0.0512	0.0537	0.0686
Observations	5,544,681	5,544,681	5,544,681	1,053,769
*** p<0.001, ** p<0.01, * p<0.05				

Standard errors are two-way clustered. First, they are clustered by (border)x(DMA)x(Month) to account for correlation in the treatment variable. Second, they are clustered by individual to account for the fact that there are repeated observations within individual over the sample.

Table 6: Transaction Price in $

	(1)	(2)	(3)	(4)
$Log(1+GRP)$	-0.54284**	-0.74045***	0.14356	-0.61605
	(0.20638)	(0.07722)	(0.12612)	(0.38631)
DMA FEs		x	x	
Border-DMA FEs				x
Month FEs			x	
Border-Month FEs				x
Mean DV	$62.78	$62.78	$62.78	$61.96
R-squared	0.00004	0.01033	0.01116	0.02373
Observations	5,544,681	5,544,681	5,544,681	1,053,769
*** p<0.001, ** p<0.01, * p<0.05				

Standard errors are two-way clustered. First, they are clustered by (border)x(DMA)x(Month) to account for correlation in the treatment variable. Second, they are clustered by individual to account for the fact that there are repeated observations within individual over the sample.

Table 7: Generic Penetration Rate

	(1)	(2)	(3)	(4)
$Log(1+GRP)$	0.01983***	0.02454***	-0.00184	-0.00335
	(0.00199)	(0.00099)	(0.00102)	(0.00246)
DMA FEs		x	x	
Border-DMA FEs				x
Month FEs			x	
Border-Month FEs				x
Mean DV	0.59145	0.59145	0.59145	0.59867
R-squared	0.00109	0.01629	0.02045	0.03573
Observations	5,551,058	5,551,058	5,551,058	1,054,561
*** p<0.001, ** p<0.01, * p<0.05				

Standard errors are two-way clustered. First, they are clustered by (border)x(DMA)x(Month) to account for correlation in the treatment variable. Second, they are clustered by individual to account for the fact that there are repeated observations within individual over the sample.

Table 8: Labor Supply - Missed Days of Work

	(1)	(2)	(3)	(4)	(5)
$Log(1+GRP)$	0.2100*	0.2518**	-0.0650	-0.0339	-0.0066
	(0.0990)	(0.0768)	(0.0904)	(0.0288)	(0.0255)
x$HighAbsentee$					-0.0694
					(0.0516)
$Log(1+GRP_{past})$	-0.5356	-0.3710*	-0.2757	-0.1382*	-0.05923
	(0.2886)	(0.1477)	(0.1830)	(0.0602)	(0.0519)
x$HighAbsentee$					-0.2086
					(0.12134)
Individual FEs		x	x	x	x
Month FEs			x		
Border-Month FEs				x	x
Mean DV	2.432	2.432	2.432	2.876	2.876
R-squared	0.00639	0.27876	0.32319	0.35549	0.36187
Observations	16,310,368	16,303,199	16,303,199	3,363,046	3,362,328
*** p<0.001, ** p<0.01, * p<0.05					

Standard errors are two-way clustered. First, they are clustered by (border)x(DMA)x(Month) to account for correlation in the treatment variable. Second, they are clustered by individual to account for the fact that there are repeated observations within individual over the sample.

Table 9: Adverse Effects - Various Measures and Proxies

	Churn	FLAE Any	FLAE Antidep	BLAE Any	BLAE Antidep
$Log(1+GRP)$	-0.00954	-0.0034	0.00013	-0.00004	0.0000005
	(0.01195)	(0.00629)	(0.00064)	(0.00012)	(0.00001)
$Log(1+GRP_{past})$	-0.00597	-0.00372	-0.00115	0.00014	-0.00004*
	(0.02170)	(0.01008)	(0.00128)	(0.00023)	(0.00002)
Individual FEs	x	x	x	x	x
Border-Month FEs	x	x	x	x	x
Mean DV	0.66461	0.03626	0.00042	0.00417	0.00002
R-squared	0.5195	0.54831	0.54514	0.08998	0.0434
Observations	93,995	77,916	77,916	12,438,333	12,438,333

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

Clustered standard errors in parentheses.

Appendix A - Log Hours

In this appendix, I consider the alternative dependent variable of the log of absentee hours per month. As such, these estimates can be read directly as elasticities of labor supply with respect to advertising. The results are presented in Table A.1 are consistent with the main results using levels rather than logs. In particular, in columns (1) and (2), the correlation between current DTCA and labor supply indicates that advertising increases absenteeism, which is a spurious result if firms are targeting advertising at places and during times when individuals are likely to miss a lot of work. In columns (1) and (2), past advertising is negatively correlated with labor supply, but the estimates are not especially precise. In column (3), the positive correlation between absenteeism and current advertising disappears and the magnitude of the point estimate on past advertising decreases and becomes statistically insignificant. In column (4), moving to the borders does not change the point estimates in a meaningful way, but considerable precision is gained, making this column my preferred specification for this table. It indicates that current advertising does not affect absenteeism while the elasticity of labor supply with respect to past advertising is about -0.045, quantitatively consistent with the main results using levels as the dependent variable. Column (5) shows that the labor supply result is driven entirely by those who miss the most days on average. All in all, the results using log of absent hours as the dependent variable do not differ qualitatively or quantitatively from the results using absentee days as the dependent variable.

Table A.1: Labor Supply - Log of Absent Hours

	(1)	(2)	(3)	(4)	(5)
$Log(1+GRP)$	0.09429***	0.11331***	-0.01139	-0.00636	-0.00222
	(0.0255)	(0.0291)	(0.0290)	(0.0124)	(0.0146)
x$HighAbsentee$					-0.01496
					(0.0171)
$Log(1+GRP_{past})$	-0.16632	-0.11521**	-0.04367	-0.04483*	0.01748
	(0.0927)	(0.0434)	(0.0422)	(0.0211)	(0.0308)
x$HighAbsentee$					-0.0783*
					(0.0391)
Individual FEs		x	x	x	x
Month FEs			x		
Border-Month FEs				x	x
Mean DV	2.26174	2.2619	2.2619	2.47743	2.52064
R-squared	0.00331	0.24914	0.3124	0.38071	0.3627
Observations	16,310,368	16,303,199	16,303,199	3,363,046	2,850,985
*** p<0.001, ** p<0.01, * p<0.05					

Standard errors are two-way clustered. First, they are clustered by (border)x(DMA)x(Month) to account for correlation in the treatment variable. Second, they are clustered by individual to account for the fact that there are repeated observations within individual over the sample.

Appendix B - Outcome Timing

As noted above, antidepressants are not expected to work immediately upon prescription. While on average they take about six weeks to work, there is variance around that mean and those who are newly prescribed often take the drugs for a number of months. This is why the conception of past advertising used in the main results is the sum of the past six months of advertising. In this appendix, I relax that parametric assumption and look at each lag of advertising separately. If six weeks is in fact the average amount of time needed for antidepressants to work, the effects should be strongest for advertising that is lagged two months with possible effects continuing in later months minimal effects in current and one month lagged advertising. Given the limited statistical power generally, it is difficult to be say much with precision, however.

The results of this analysis are in Table B.1. In the first column, this analysis is conducted with absentee days as the dependent variable. The largest point estimate is on advertising two months lagged. It is not sufficiently precise to be statistically significant at the p<0.05 level and has p-value of about 0.08. Power is too limited to say much beyond that. In the second column, log of absent hours is the dependent variable. In that column we again face very limited power. However, the effects on two and three month lagged advertising are the largest negative effects, with longer lags also being similarly negative with the exception of five months lagged being positive. In both columns, the effects on the current and one month lagged advertising are the smallest and indicate little or no effect, which is what we should expect from the science of antidepressant effectiveness. While getting the exact nature of the lagged effectiveness proves too challenging in this context due to statistical power, it is reassuring that the effects of lagged advertising on labor supply are not manifesting themselves in the current and one month lagged advertising variable, which would indicate a likely spurious result.

Table B.1: Outcome Timing

	Days	Log Hours
$Log(1+GRP)$	-0.00004	0.00084
	(0.0352)	(0.0144)
$Log(1+GRP_{t-1})$	-0.03532	0.00148
	(0.0577)	(0.0164)
$Log(1+GRP_{t-2})$	-0.11019	-0.01673
	(0.0623)	(0.0168)
$Log(1+GRP_{t-3})$	-0.00116	-0.01578
	(0.0299)	(0.0149)
$Log(1+GRP_{t-4})$	-0.0046	0.0232
	(0.0425)	(0.0145)
$Log(1+GRP_{t-5})$	-0.01051	-0.00845
	(0.0351)	(0.0145)
$Log(1+GRP_{t-6})$	0.00008	-0.01199
	(0.0264)	(0.0133)
Individual FEs	x	x
Border-Month FEs	x	x
Mean DV	2.262	2.505
R-squared	0.3610	0.3622
Observations	2,984,534	2,984,534

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

Standard errors are two-way clustered. First, they are clustered by (border)x(DMA)x(Month) to account for correlation in the treatment variable. Second, they are clustered by individual to account for the fact that there are repeated observations within individual over the sample.

Appendix C - Selection into the Border Sample

In this section, I measure how the border sample used for the preferred specifications is different the non-border sample in terms of observables in the data. I do this at a single point in time, July 2007 and test the null hypothesis that the difference in means is zero. The variables examined are percent of population employed in manufacturing (broadly defined), age, the percent of the population paid hourly (as opposed to salaried), the percent of the population that was prescribed an antidepressant in the previous six months, the percent of the population that was both prescribed in the past six months and failed to adhere, average missed days per month, and antidepressant prices and co-pays conditional on prescription. The results are in Table C.1. Most of the differences are statistically significant given the large sample size, but not all are economically large. Notable economically large differences include the borders having a higher percentage manufacturing jobs and individuals missing more work per month on average . Statistically significant, but smaller differences include the borders being slightly older, slightly more likely to be paid hourly, slightly more likely to be prescribed an antidepressant in the past six months and slightly lower cost antidepressants being chosen conditional on purchase, both in prices and co-pays. There is no detectable difference between the border and non-border counties in propensity to adhere to treatment.

In addition to this, Shapiro (2017) shows that border counties tend to be older, less populous, less wealthy and less racially diverse. None of these differences provide sharp predictions about differences in treatment effects of advertising on sales or outcomes, however. In fact, the simple diff-in-diff specifications of the effect of advertising on labor supply outcomes fail to reject equality with the border approach treatment effects.

Table C.1: Selection into the Border Sample

	Mean Border	Mean Non-Border	Difference	P-Value
% Manufacturing	66.47	54.05	12.41	<0.01
Age	44.62	43.46	1.158	<0.01
% Hourly	37.95	36.08	1.862	<0.01
% Prescribed Past 6	11.47	10.51	0.959	<0.01
% Non-Adherence Past 6	42.38	42.14	0.238	0.444
Avg Absentee Days	2.989	2.359	0.630	<0.01
Price	60.86	62.78	1.92	<0.01
Copay	11.47	12.25	0.78	<0.01

P-value comes from a t-test with null hypothesis of zero difference between means

Appendix D - County Level Placebo Analysis

In this section, I extend the placebo analysis conducted at the individual level in Table 2 to county level covariates collected from the census. As county level demographics change slowly, I assess this analysis both in changes and in levels. While a failed placebo check using levels would not invalidate the research design (only parallel trends are needed given the individual fixed effects), seeing that advertising does not predict demographic levels lends some credence to the intuition that the border counties are not specifically targeted by the advertisers. The demographics considered are population, average income, Hispanic share, Asian share, black share, elderly (over 65 years old) share, and death rates. Each variable, with the exception of population, is normalized to have mean zero and standard deviation one in the full sample for ease of comparison.

Results are presented in Table D.1 and Table D.2. Table D.1 runs the analysis from equation (1), but at the county level with county fixed effects and county-month fixed effects. Changes in antidepressant advertising at the border do not predict changes in any of the demographic variables in a statistically significant way, nor are the point estimates significant in magnitude. Table D.2 drops the county fixed effects to assess whether or not advertising levels predict demographic levels. Levels of antidepressant advertising at the border do not predict levels of any of the demographic variables across borders in a statistically or economically meaningful way.

Table D.1: County Placebo Regressions - Changes

	Pop	AvgIncome	Hispanic	Asian	Black	Elderly	DeathRate
$Log(1+GRP)$	-5.5326	0.0020	0.0008	-0.0004	0.0005	-0.0042	0.0049
	(50.8684)	(0.0033)	(0.0010)	(0.0011)	(0.0010)	(0.0049)	(0.0117)
Mean DV	65,900	-0.26607	-0.21363	-0.14189	0.01517	0.12213	0.26548
R-squared	0.999	0.989	0.997	0.995	0.999	0.975	0.950
Observations	27,445	27,402	27,445	27,445	27,445	27,445	20,229
*** p<0.001, ** p<0.01, * p<0.05							

Each column represents a regression with the given variable as the dependent variable. Each includes county fixed effects and border-month fixed effects and log of past advertising. It shows that changes in these variables are not systematically predicted by changes in antidepressant advertising.

Table D.2: County Placebo Regressions - Levels

	Pop	AvgIncome	Hispanic	Asian	Black	Elderly	DeathRate
$Log(1+GRP)$	-2040	-0.0208	-0.0087	-0.0174	-0.0072	0.0028	0.0276
	(3890)	(0.0190)	(0.0129)	(0.0138)	(0.0261)	(0.0266)	(0.0361)
Mean DV	65,900	-0.26607	-0.21363	-0.14189	0.01517	0.12213	0.26548
R-squared	0.603	0.683	0.678	0.511	0.788	0.502	0.503
Observations	27,445	27,402	27,445	27,445	27,445	27,445	20,229
*** p<0.001, ** p<0.01, * p<0.05							

Each column represents a regression with the given variable as the dependent variable. Each includes border-month fixed effects and log of past advertising, but no county fixed effects. It shows that levels in these variables are not systematically predicted by levels in antidepressant advertising.

Appendix E - Statin Placebo

Another way to validate the approach is to show that an alternative "treatment" at the same level of observation that would not be predicted to show an effect on the outcome of interest does not show such an effect. Here, I use advertising for cholesterol lowering drugs, or statins, as such a treatment. Similar to antidepressant advertising, statin advertising is decided at the DMA level and has discontinuous changes at DMA borders. However, unlike antidepressants, statins are designed to lower cholesterol, which is a proxy for heart disease risk. It is typically prescribed well before heart disease is acute. High cholesterol by itself is not causally linked to reduced labor supply or work functionality, so response to statin advertising by individuals is not predicted to alter labor supply.

Table E.1 is analogous to Table 8, but with statin advertising in place of antidepressant advertising. In columns (1) and (2), concurrent advertising predicts decreased missed days of work. However, in the preferred specification in column (4), neither concurrent advertising nor lagged advertising drives changes in labor supply. The point estimates are very small and not statistically significant. This suggests that the labor supply effects of antidepressant advertising are not simply an artifact of confounds across borders, as we might think if statin advertising produced a similar effect.

Table E.1: Placebo Labor Supply from Statin Ads - Missed Days of Work

	(1)	(2)	(3)	(4)
$Log(1+GRP)$	-0.4630***	-0.4134**	0.1071	0.0111
	(0.0578)	(0.1399)	(0.1955)	(0.0413)
$Log(1+GRP_{past})$	-0.30005	-0.2645	-0.36549	-0.01554
	(0.4030)	(0.3268)	(0.3149)	(0.0922)
Individual FEs		x	x	x
Month FEs			x	
Border-Month FEs				x
Mean DV	2.463	2.463	2.463	2.958
R-squared	0.0086	0.2804	0.3214	0.35549
Observations	14,134,343	14,126,575	14,126,575	2,876,150
*** p<0.001, ** p<0.01, * p<0.05				

Standard errors are two-way clustered. First, they are clustered by (border)x(DMA)x(Month) to account for correlation in the treatment variable. Second, they are clustered by individual to account for the fact that there are repeated observations within individual over the sample.