Promoting Wellness or Waste? Evidence from Antidepressant Advertising

Bradley Shapiro
March 2018
Promoting Wellness or Waste? Evidence from Antidepressant Advertising

Bradley T. Shapiro*

This Version February 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 in the market for antidepressant drugs. In particular, using individual health insurance claims and human resources data, I estimate the effects of DTCA on outcomes relevant to societal 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. Additionally, I find evidence of no advertising effect on either the prices or co-pays of the drugs prescribed. Next, lagged advertising is associated with higher first refill rates, indicating that the advertising marginal are not more likely to end treatment prematurely due to initial adverse effects. Despite first refill rates being higher for those that are more likely advertising-marginal, concurrent advertising drives slightly lower refill rates overall, particularly among individuals who stand to gain least from treatment. 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 the shapes of the benefit and cost curves for antidepressants with respect to DTCA and where society currently lies on those curves.

*Bradley.Shapiro@chicagobooth.edu. University of Chicago Booth School of Business. I thank Stephen Lamb for excellent research assistance. 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.
In this paper, I evaluate the main proposed social costs associated with advertising marginal prescriptions in the antidepressant category and weigh those against the benefits, if any. 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. Third, I evaluate whether advertising marginal prescriptions are more likely to be discontinued because of worse suitability to treatment or worse adverse effects. The main measured potential benefit evaluated is the wage benefit of increased labor supply. While these costs and benefits are not exhaustive, they provide a first order approximation of the costs and benefits associated with antidepressant use.

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 and has been associated both with large direct costs of medical care as well as large indirect costs of reduced economic activity. Depression has been associated both with large direct costs of medical care as well as large indirect costs of reduced economic activity. Depression has been associated both with large direct costs of medical care as well as large indirect costs of reduced economic activity. Depression has been associated both with large direct costs of medical care as well as large indirect costs of reduced economic activity.

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 that DTCA induces more patients to be prescribed antidepressants with an elasticity of about 0.032, leading to a direct cost of DTCA. Second, I add to the literature on advertising and drug treatment adherence. While that literature finds some mixed effects, I find that advertising reduces adherence, particularly among those less well-suited for treatment. Second month adherence to antidepressant treatment is marginally higher when lagged advertising is higher, suggesting that advertising marginal patients are not worse suited for treatment than other patients. 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. 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 $32.6 million in prescription costs. 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. To my knowledge, this is the first paper linking DTCA to measurable social benefits and costs.

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. 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 direct 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 cost or benefit 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. [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, an average of $7508 per worker per year.

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.

### 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 insure 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. 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:

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

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, \( E[v_{ik}|A = 0] - p_{ig} > E[v_{ij}|A = 0] - p_{ij} \), but \( E[v_{ik}|A = 0] - p_{ig} < E[v_{ij}|A > 0] - p_{ij} \), 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 \( E[v_{ij}|A > 0] > v_{ij} \), 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 could 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 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.

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.[1]

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

---

[1] 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.
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 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. In an average month, 8.12% of the individuals in the data are prescribed an antidepressant. More summary statistics are available in Table 1.

The claims data also contain information on the transacted prices and copayments of prescriptions. Copayments are based on individual formularies that come with insurance contracts. These contracts typically last a year, so there is not variation in copayments of options for a given individual-month. Similarly, prices reflect the copayment 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 copayments 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 $61.50 and the average copayment faced by the individual enrollee for an antidepressant prescription filled is $11.39.

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
eperiment considered is outlined in bold. 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 $GRP_d$ 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

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

where $\beta_1$ and $\beta_2$ capture the causal effects of current and past advertising, respectively, $\alpha_i$ is an individual fixed
effect, $\alpha_{bt}$ is a border area-month fixed effect and $\epsilon_{ibdt}$ is an econometric error term. I consider the outcomes
$Y \in \{ \text{NewRx, RenewalRx, FirstRenewalRx, Price, Copay, DaysAbsent} \}$. 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 effect, residual variation in advertising is
perfectly correlated at the border-DMA-month. Second, from an experimental 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.
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 copayments 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

\footnote{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 $\beta_2$ negatively. $\beta_2$ can alternatively be viewed as a way to control for negative selection into the “new prescription possible” sample for this estimation.}
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.

### 5.2 Features and Limitations

A much 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.

### 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 2 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 column (4), we see a 10% increase in current antidepressant GRPs leads to about a 0.00032 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 a 0.032 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 get an idea of differential effectiveness by sickness, column (5) presents heterogeneous treatment effects by the average number of missed days across the entire sample by an individual, split at the median. That is, if an individual has more than the median number of average missed days of work per month, across the sample, she is categorized as sick. Otherwise, she is categorized as not sick. Column (5) provides some suggestive evidence that advertising is differentially effective on those that are missing more work on average. The estimate on past advertising for those that are high absentee is negative and significant, but as mentioned above, I remain cautious to interpret this too closely, as those that receive high advertising in the recent but do not purchase then are potentially 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, 6, and the average price and copayments of antidepressants, $61.50 and $11.39. Assuming these results apply to all adults in the United States, 230 million, a 10% increase in advertising leads to 530,000 new antidepressant prescriptions to about 88,000 individuals, which yields approximately $33 million in total costs of new prescriptions per year, about $6 million of which is paid in co-pays by the consumer, assuming no changes in prices or copayments. The 95% confidence interval associated with this cost estimate is [$1.77 million, $64.4 million].

6.1.2 Adherence

Previous studies of DTCA focus on adherence [Donohue et al. (2004); Cardon and Showalter (2015); Wosinska (2005)] and tend to find small, inconsistent and sometimes negative effects of DTCA on adherence. Adherence is an important consideration for welfare for a few reasons. First, it is possible that advertising reminds patients to refill their prescriptions appropriately, which would increase the cost of advertising marginal prescriptions, as they would need to be paid for, but also might affect the effectiveness of the eventual treatment. Second, it is possible that advertising helps people currently being treated decide that they are poor matches for treatment and convinces them to discontinue treatment, which would decrease the cost of prescriptions and not worsen welfare outcomes. Third, it is possible that advertising reduces adherence inefficiently due to the list of side effects scaring patients off of treatment.

Renewal prescriptions are also helpful to identify potential indirect harms of advertising marginal prescriptions. 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. If this is the case, we should expect advertising marginal prescriptions to be more likely to be non-adherent in their second month. To test for this possibility, I assess the effect of past advertising on the first refill of antidepressants. That is, if advertising

---

3 As labor supply is only observed for a fraction of the population, for all of those for whom it is not observed, this measure of sickness is imputed from individual year of birth and county fixed effects.

4 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.
causes a prescription in month $t$, I want to know if that prescription is more or less likely to be refilled in period $t+1$ than an average prescription.

Table 3 presents the results of the effects of advertising on adherence. Columns (1)-(5) correspond with Table 2 and column (6) presents results on first month adherence. For conciseness, I will only discuss the preferred specifications in columns (4)-(6). 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 not statistically significant at conventional levels, but most of the mass of the confidence interval is small and negative, consistent with Cardon and Showalter (2015), Donohue et al. (2004) and Wosinska (2005).

It is not clear whether the decreased adherence is good or bad. Column (5) shows that the decreased adherence is exclusively coming from those who tend to miss less work on average. This is consistent with advertising convincing individuals who stand to gain very little from treatment to discontinue use. This would be the case if, for example, they were prescribed an antidepressant for something less serious than depression, saw the ad, and decided the side effects were not worth the potential gain. Column (6) presents estimates on first refill adherence. The coefficient of interest is on past advertising, as we would like to know whether or not advertising marginal prescriptions are more or less likely to be refilled. The point estimate is positive, but not statistically significant. This indicates that if anything, the advertising marginal prescriptions are more likely to be refilled, which is at odds with Cardon and Showalter (2015) and David et al. (2010). That is, the evidence is not consistent with advertising drawing in less appropriate patients. The 95% confidence interval on the elasticity associated with this estimate is [-0.0134, 0.0978], indicating that I can rule out that advertising marginal prescriptions are significantly more likely to be bad matches, but I cannot rule out that the two are identical.

In terms of costs and benefits of adherence, the results indicate that DTCA reduces adherence, particularly among those who have the lowest absenteeism. These individuals stand to gain the least from treatment, so their decreased adherence could be viewed as a social good, but by how much in dollars is unclear. A reasonable lower-bound estimate of that dollar value is the price of one prescription, or about $61.50 per averted prescription that was advertising marginal. 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 $5.58 million, or about 90,750 prescriptions. This is a lower bound because these patients might have seen decreased adverse effects that also have value which is difficult to measure. There could also be costs of decreased adherence in that some who adhered less because of DTCA might have benefited from better labor supply, even though most of the adherence effect is coming from those who stand to gain little. Any such labor supply effects will be accounted for in the labor supply analysis below.

### 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 copayments 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 menu of prices to change. Table 4 shows these effects for copayments while Table 5 shows these effects for prices. The effect of DTCA on copayments affect the private cost of DTCA to those who initiate treatment while the effect of DTCA on transacted

---

5Empirically, 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.
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 copayments, in Table 4, column (4), the effect of advertising on copayments is not statistically significant, and the point estimate is positive and small. It indicates that a 10% increase in advertising leads to a $0.016 increase in the average transacted copayment on an average copayment of $11.62. Taken at face value, this point estimate would imply a small indirect private cost associated with DTCA. A 10% increase in DTCA leads to an additional $84,800 per year to be spent in copayments. Assuming the worst, at the right end of the 95% confidence interval of the estimate, a 10% increase in DTCA leads to an additional $270,300 per year spent on copayments. Assuming the best, at the left end of the 95% confidence interval of the estimate, a 10% increase in DTCA leads to a copayment savings of $95,400 per year. Taken together the results on the effect of DTCA on copayments suggest that the indirect private cost of DTCA are between -1.5% and 4.3% of the direct private cost of DTCA on prescriptions holding copayments fixed. The estimated effects of DTCA on copayments provide evidence against large steering effects.

Next, focusing on transacted prices in Table 5, 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 $0.068 decrease in transacted prices on an average price of $61.15. It implies a small indirect social benefit of DTCA in terms of lowering costs to the insurance company. A 10% increase in DTCA leads to $2.24 million decrease in insurance spending on antidepressants relative to a world where prices remain fixed but the quantity effects continue to hold. Taken together the results on the effect of DTCA on copayments suggest that the indirect social cost of DTCA through prices are between -13.7% and 2.2% of the direct social cost of DTCA on prescriptions holding transacted prices fixed.

While these results on prices might seem counter-intuitive, they are consistent with untreated patients being hesitant to go to the physician to discuss their depression due to concerns about cost, having read about the expense of treatment in the news. Knowing this, the physician might prescribe a generic instead of a brand, even though in practice the copay differences between these drugs are small and the price difference is large. However, I would like to avoid over-interpreting a statistically insignificant result.

### 6.3 Labor Supply

To assess the potential benefits of DTCA, I estimate equation (1) using missed days of work as the dependent variable. 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.

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

---

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.
were effective. Table 6 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) 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 is coming entirely from those who typically miss a lot of work.

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 most 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, 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. If we assume this benefit only applies to those who are advertising marginal to new prescriptions (about 88,000 individuals, as noted above), those individuals gain about 3.78 work days per month and $8,744 per year, or about 17% 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 copayments of $11.39 and 6 months of treatment, the private cost to obtain the $8,744 in benefits is about $68. With the average transaction price of $61.50, the total cost is $369. 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, the decreased adherence among the poorly suited could generate labor supply due to the elimination of adverse effects, individuals could seek non-drug treatment of their depression after seeing ads or individuals could simply be more cognizant of depressive tendencies, leading to increased labor supply.

The 95% confidence interval on the total yearly benefits of a 10% increase in DTCA is [$112.6 million, $1.427 billion]. The 95% confidence interval on the yearly costs of marginal prescriptions from a 10% increase in DTCA was [$1.77

\[7\] 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.

\[8\] 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 88,000 individuals estimated to be advertising marginal to prescriptions in a year.

\[9\] Separation of these mechanisms is difficult due to statistical power limitations, but continuing work on this study is attempting to better separate the mechanisms.
million, $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.

6.4 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.

In terms of costs that I do not measure, while advertising marginal drugs bring with them marginal benefits of labor supply, they may also bring with them marginal costs of adverse effects that do not fully manifest themselves in labor supply reductions. This would be the case if an individual was able to go to work more, but also had more headaches, which could be privately costly as well as make the employee less productive for the employer. Of course, there is free disposal of a prescription. If the benefit of feeling better does not outweigh the pain of adverse effects, then the patient could discontinue use, or rationally not adhere to treatment.

Additionally, 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 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. 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 undertreated due to the stigmatisation, which leaves plenty of opportunity for welfare increasing market expansion. For 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.

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, but 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

---

10 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, which was marinally significant. These numbers suggest the estimated effects on prescriptions are plausible in terms of the firm’s first order condition.
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 considerably 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 $112.6 million.

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 $32.6 million in direct costs of the marginal prescriptions generated. Indirect costs that can be easily assigned a dollar value, increases in copayments and prices, 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 among those who stand to gain little in the form of increased labor supply. In terms of costs and benefits without clear dollar values, I find evidence that DTCA does not produce worse matches and might even produce better matches with treatment, as those with exposed to high past advertising are marginally more likely to adhere to their first refill than are those exposed to lower past advertising.

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.

References


Figures

Figure 1: Antidepressant GRPs
Figure 2: Ohio and DMA Border Example
Figure 3: Antidepressant DTCA GRP variation across Borders
## Tables

### Table 1: Summary Stats

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRP</td>
<td>227.27</td>
<td>148.67</td>
<td>5,428</td>
</tr>
<tr>
<td>Antidepressant Rx</td>
<td>0.0812</td>
<td>0.2731</td>
<td>68,527,405</td>
</tr>
<tr>
<td>New Antidepressant Rx</td>
<td>0.00949</td>
<td>0.0969</td>
<td>68,527,405</td>
</tr>
<tr>
<td>Copayment</td>
<td>11.39</td>
<td>13.25</td>
<td>5,838,225</td>
</tr>
<tr>
<td>Price</td>
<td>61.50</td>
<td>65.51</td>
<td>5,859,963</td>
</tr>
<tr>
<td>Days Absent</td>
<td>2.375</td>
<td>3.155</td>
<td>17,593,438</td>
</tr>
</tbody>
</table>

### Table 2: New Prescriptions

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log(1 + GRP) )</td>
<td>0.00011***</td>
<td>0.00008**</td>
<td>0.00004</td>
<td>0.00032*</td>
<td>0.00011</td>
</tr>
<tr>
<td>( \log(1 + GRP_{past}) )</td>
<td>0.00076***</td>
<td>0.00050***</td>
<td>-0.0001</td>
<td>-0.00019</td>
<td>0.00055</td>
</tr>
<tr>
<td>( x_{HighAbsentee} )</td>
<td>0.00047</td>
<td>-0.00130**</td>
<td>           </td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual FEs</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Month FEs</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Border-Month FEs</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean DV</td>
<td>0.0094</td>
<td>0.0094</td>
<td>0.0094</td>
<td>0.0099</td>
<td>0.0099</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.00003</td>
<td>0.07263</td>
<td>0.07266</td>
<td>0.07446</td>
<td>0.07578</td>
</tr>
<tr>
<td>Observations</td>
<td>66,736,304</td>
<td>66,736,304</td>
<td>66,736,304</td>
<td>12,064,669</td>
<td>12,064,669</td>
</tr>
</tbody>
</table>

*** 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 3: Renewal Prescriptions

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\log(1 + \text{GRP})$</td>
<td>-0.00689***</td>
<td>-0.00358***</td>
<td>0.00009</td>
<td>-0.00411</td>
<td>-0.00874**</td>
<td>-0.00995</td>
</tr>
<tr>
<td></td>
<td>(0.0020)</td>
<td>(0.0005)</td>
<td>(0.0009)</td>
<td>(0.0021)</td>
<td>(0.0030)</td>
<td>(0.0089)</td>
</tr>
<tr>
<td>$x_{\text{HighAbsentee}}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.00198</td>
<td>(0.0032)</td>
</tr>
<tr>
<td>$\log(1 + \text{GRP}_{\text{past}})$</td>
<td>-0.04361*</td>
<td>-0.01091***</td>
<td>-0.0069</td>
<td>-0.00113</td>
<td>-0.00484</td>
<td>0.02534</td>
</tr>
<tr>
<td></td>
<td>(0.0181)</td>
<td>(0.0009)</td>
<td>(0.0015)</td>
<td>(0.0041)</td>
<td>(0.0054)</td>
<td>(0.0167)</td>
</tr>
</tbody>
</table>

Individual FEs: x x x x
Month FEs: x
Border-Month FEs: x x
First Renewal Only: x

Mean DV: 0.84306 0.84946 0.84946 0.84555 0.84561 0.79371
R-squared: 0.00584 0.41108 0.41176 0.4265 0.43095 0.60014
Observations: 5,695,257 5,648,991 5,648,991 1,077,357 1,076,532 91,366

*** 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: Copayments in $

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\log(1 + \text{GRP})$</td>
<td>-0.3269***</td>
<td>-0.4212***</td>
<td>-0.0315</td>
<td>0.1657</td>
</tr>
<tr>
<td></td>
<td>(0.0731)</td>
<td>(0.0235)</td>
<td>(0.0508)</td>
<td>(0.1734)</td>
</tr>
<tr>
<td>Border-DMA FEs: x x x x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Month FEs: x
| Border-Month FEs: x |

Mean DV: $11.54 $11.54 $11.54 $11.00
R-squared: 0.00041 0.04717 0.04938 0.06629
Observations: 5,527,290 5,527,290 5,527,290 1,051,863

*** 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: Transaction Price in $

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log(1 + GRP)</td>
<td>-0.78222***</td>
<td>-1.00527***</td>
<td>0.11412</td>
<td>-0.67818</td>
</tr>
<tr>
<td></td>
<td>(0.2019)</td>
<td>(0.0776)</td>
<td>(0.1277)</td>
<td>(0.3501)</td>
</tr>
<tr>
<td>Border-DMA FEs</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Month FEs</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Border-Month FEs</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean DV</td>
<td>$61.82</td>
<td>$61.82</td>
<td>$61.82</td>
<td>$60.98</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0001</td>
<td>0.01197</td>
<td>0.01303</td>
<td>0.02681</td>
</tr>
<tr>
<td>Observations</td>
<td>5,549,497</td>
<td>5,549,497</td>
<td>5,549,497</td>
<td>1,054,394</td>
</tr>
</tbody>
</table>

*** 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: Labor Supply - Missed Days of Work

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log(1 + GRP)</td>
<td>0.2100*</td>
<td>0.2518**</td>
<td>-0.0650</td>
<td>-0.0339</td>
<td>-0.0206</td>
</tr>
<tr>
<td></td>
<td>(0.0990)</td>
<td>(0.0768)</td>
<td>(0.0904)</td>
<td>(0.0288)</td>
<td>(0.0203)</td>
</tr>
<tr>
<td>xHighAbsentee</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0703</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0509)</td>
</tr>
<tr>
<td>Log(1 + GRP\text{past})</td>
<td>-0.5356</td>
<td>-0.3710*</td>
<td>-0.2757</td>
<td>-0.1382*</td>
<td>0.02199</td>
</tr>
<tr>
<td></td>
<td>(0.2886)</td>
<td>(0.1477)</td>
<td>(0.1830)</td>
<td>(0.0602)</td>
<td>(0.0341)</td>
</tr>
<tr>
<td>xHighAbsentee</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.2652*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.1101)</td>
</tr>
<tr>
<td>Individual FEs</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Month FEs</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Border-Month FEs</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean DV</td>
<td>2.432</td>
<td>2.432</td>
<td>2.432</td>
<td>2.876</td>
<td>2.876</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.00639</td>
<td>0.27876</td>
<td>0.32319</td>
<td>0.35549</td>
<td>0.36187</td>
</tr>
<tr>
<td>Observations</td>
<td>16,310,368</td>
<td>16,303,199</td>
<td>16,303,199</td>
<td>3,363,046</td>
<td>3,362,328</td>
</tr>
</tbody>
</table>

*** 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 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>
<thead>
<tr>
<th>Table A.1: Labor Supply - Log of Absent Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
</tr>
<tr>
<td>Log(1 + GRP)</td>
</tr>
<tr>
<td>0.09429***</td>
</tr>
<tr>
<td>(0.0255)</td>
</tr>
<tr>
<td>xHighAbsentee</td>
</tr>
<tr>
<td>0.0093</td>
</tr>
<tr>
<td>(0.0164)</td>
</tr>
<tr>
<td>Log(1 + GRP_{past})</td>
</tr>
<tr>
<td>-0.16632</td>
</tr>
<tr>
<td>(0.0927)</td>
</tr>
<tr>
<td>xHighAbsentee</td>
</tr>
<tr>
<td>-0.08164**</td>
</tr>
<tr>
<td>(0.0276)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Individual FEs</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month FEs</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Border-Month FEs</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean DV</th>
<th>2.26174</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-squared</td>
<td>0.00331</td>
</tr>
<tr>
<td>Observations</td>
<td>16,310,368</td>
</tr>
</tbody>
</table>

*** 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

<table>
<thead>
<tr>
<th></th>
<th>Days</th>
<th>Log Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log(1 + GRP) )</td>
<td>-0.00004</td>
<td>0.00084</td>
</tr>
<tr>
<td>( \log(1 + GRP_{t-1}) )</td>
<td>-0.03532</td>
<td>0.00148</td>
</tr>
<tr>
<td>( \log(1 + GRP_{t-2}) )</td>
<td>-0.11019</td>
<td>-0.01673</td>
</tr>
<tr>
<td>( \log(1 + GRP_{t-3}) )</td>
<td>-0.00116</td>
<td>-0.01578</td>
</tr>
<tr>
<td>( \log(1 + GRP_{t-4}) )</td>
<td>-0.0046</td>
<td>0.0232</td>
</tr>
<tr>
<td>( \log(1 + GRP_{t-5}) )</td>
<td>-0.01051</td>
<td>-0.00845</td>
</tr>
<tr>
<td>( \log(1 + GRP_{t-6}) )</td>
<td>0.00008</td>
<td>-0.01199</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual FEs</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Border-Month FEs</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Mean DV</td>
<td>2.262</td>
<td>2.505</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.00331</td>
<td>0.3622</td>
</tr>
<tr>
<td>Observations</td>
<td>2,984,534</td>
<td>2,984,534</td>
</tr>
</tbody>
</table>

*** \( p<0.001 \), ** \( p<0.01 \), * \( p<0.05 \)

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 hypothesis that the difference in means is zero. The variables examined are percent of population employed in manufacturing (broadly
Table C.1: Selection into the Border Sample

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

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

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 copays 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 copays. 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.